

南華大學科技學院資訊管理學系

碩士論文

Department of Information Management

College of Science and Technology

Nanhua University

Master Thesis

基於深度學習的中醫虛證辨證

Deep Learning Based Deficiency Syndrome Differentiation
in Traditional Chinese Medicine

劉家佑

Chia-Yu Liu

指導教授：邱宏彬 博士

Advisor: Hung-Pin Chiu, Ph.D.

中華民國 112 年 6 月

June 2023

南 華 大 學
科技學院資訊管理學系
碩 士 學 位 論 文

基於深度學習的中醫虛證辨證

Deep Learning Based Deficiency Syndrome Differentiation in Traditional
Chinese Medicine

研究生：劉家佑

經考試合格特此證明

口試委員：林迺達
葉明憲

邱宏彬

指導教授：邱宏彬

系主任(所長)：陳信良

口試日期：中華民國 112 年 6 月 1 日

論文著作財產權同意書

南華大學資訊管理學系碩士論文著作財產權同意書

立書人： 劉家佑 之碩士畢業論文
中文題目：基於深度學習的中醫虛證辨證

英文題目：Deep Learning Based Deficiency Syndrome Differentiation
in Traditional Chinese Medicine

指導教授：邱宏彬 博士

學生與指導老師就本篇論文內容及資料其著作財產權歸屬如下：

- ☒ 共同享有著作權
☐ 共同享有著作權，學生願「拋棄」著作財產權
☐ 學生獨自享有著作財產權

學 生： 劉家佑 (請親自簽名)

指導老師： 邱宏彬 (請親自簽名)

中華民國 112 年 6 月 20 日

論文指導教授推薦函

南華大學碩士班研究生

論文指導教授推薦函

資訊管理系碩士班 劉家佑君所提之論文

係由本人指導撰述，同意提付審查。

指導教授 邱宏彬

112年6月1日

論文付印申請表

南華大學碩、博士班研究生論文付印申請表

_資訊管理學_系(所)碩、博士班研究生劉家佑已通過論文口試，

其撰寫之論文 基於深度學習的中醫虛證辨證
，經本人確認已依規定完成修改，同意付印。

此致

系主任(所長)

指導教授 邱宏彬 (簽章)
112 年 6 月 23 日

※本表務請填妥繳回系所彙辦，相關文件由系所自行留存。

致謝

首先，必須先感謝指導教授邱老師在這兩年中的指導，從邱老師這邊我也學習到許多做研究該有的觀念，邱老師注重的思維邏輯對我的影響也非常深刻，幫助我了解到思維邏輯的重要性，從碩一開始，每週的報告邱老師都不斷提點我們，不能只是為了報告而讀論文，必須從自身的角度去理解整篇論文，了解作者進行研究時的思維邏輯，也和自己的想法進行比對，試著去找出論文或自身想法的盲點，這樣才能幫助自己進步，讓我在閱讀論文時能夠收穫更多東西。而在進行研究時，並非只是將實驗結果報告出來，需要針對結果進行直觀的解釋，要去思考整體的研究方法與結果是否合乎邏輯，也要找結果能夠證明或是有什麼意義，才能夠確立研究結果的價值。總而言之，非常感謝邱老師這兩年的諄諄教誨，讓我體會到做研究的辛苦與樂趣，也让我學到許多在業界或是課堂上所學不到的知識或觀念，讓我受益良多。

同時，也非常感謝學弟妹與我共同合作，每次都不斷的與邱老師討論論文的方向，也一起解決老師提出來的問題，也感謝珮綺不斷為整個實驗室的付出，讓我們能擁有良好的環境做研究，每次都會一起討論報告或是論文相關的問題。感謝實驗室學弟妹這兩年的陪伴與互相督促，讓我們都變得越來越好，也祝福大家畢業後都能有好工作。

最後，要特別感謝我的親人們一路上的支持與幫助，讓我能夠考上研究所，也付出許多讓我能夠安心讀完碩士，在過程中也不斷關心我，也感謝學長姐、朋友們這兩年來的協助與建議，讓我這一路上走來非常開心也順利，祝福所有的人都能有美好的未來。

摘要

本研究旨在解決中醫辨證中的分類問題，即根據病人的症狀和證候，推斷疾病的原因並選擇相應的治療方法。中醫辨證具有非線性、模糊性和複雜性，因此利用資訊科技進行中醫辨證研究具有挑戰性。

本研究提出一個結合症狀詞向量和深度學習網路的中醫辨證模型。將原始症狀轉換為標準化症狀，以克服中醫用詞差異的問題。使用Word2Vec將標準化症狀映射到中文詞向量空間，以提取字詞間的上下文資訊。通過神經網路模型學習症狀集在證候間的交互作用資訊，從而將輸入的症狀集準確地對應到相應的證候。

研究使用包含18個虛證和50個臨床案例報告的資料集進行實驗。研究採用了四種模型，分別是onehot encoding、embedding layer、Word2Vec和TF-IDF，並針對多標籤問題進行訓練，以評估深度學習網路在中醫辨證方面的效能。研究結果表明四種模型在以下超參數下效能最好，分別為：onehot encoding模型在對齊長度為25和神經元數量為128；embedding layer模型在Vector-Dim為16、對齊長度為25和神經元數量為128；Word2Vec模型在Window_size=4、Vector-Dim為16、對齊長度為25和神經元數量為128；TF-IDF模型在對齊長度為25和神經元數量為128。特別是在四種模型中，Word2Vec預訓練詞向量模型表現最佳，驗證了其在辨證中具有較好的表徵能力。

本研究所提出的中醫辨證模型在中醫研究中有輔助價值。模型可以針對特定虛證進行有效的辨證，同時為深度學習提供臨床病歷中缺乏的證候資訊。這將有助於提升中醫辨證的準確性和效率，為醫生提供更可靠的輔助工具，同時也促進了中醫研究領域與機器學習的結合。

關鍵字：中醫症狀、證候、辨證、詞向量、深度學習

Abstract

The purpose of this study is to address the classification problem in Traditional Chinese Medicine (TCM) pattern identification, which involves inferring the causes of diseases and selecting appropriate treatment methods based on patients' symptoms and patterns. TCM pattern identification is characterized by nonlinearity, fuzziness, and complexity, making it challenging to conduct research using information technology.

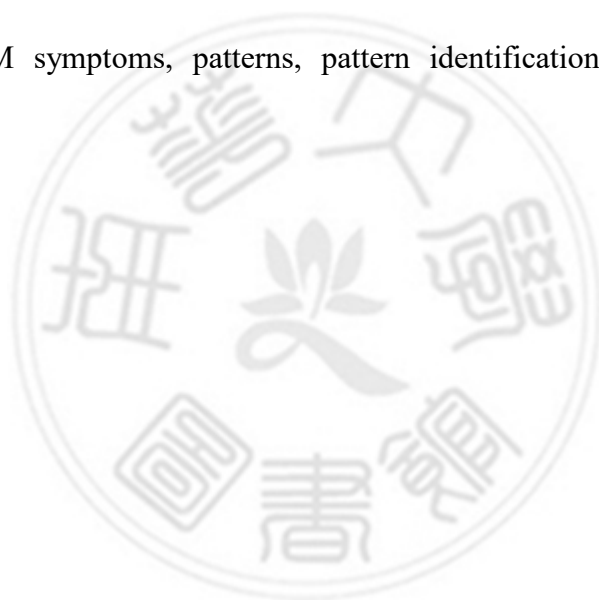
This study proposes a TCM pattern identification model that combines symptom word vectors and deep learning networks. To overcome the issue of terminology variations in TCM, the original symptoms are transformed into standardized symptoms. Word2Vec is used to map standardized symptoms into a Chinese word vector space, extracting contextual information between terms. By training a neural network model, the interaction information among symptom sets in patterns is learned, accurately mapping the input symptom set to the corresponding pattern.

Experiments were conducted using a dataset comprising 18 virtual patterns and 50 clinical case reports. Four models were employed in the study: one-hot encoding, embedding layer, Word2Vec, and TF-IDF. Training was performed for multi-label problems to evaluate the performance of deep learning networks in TCM pattern identification. The research results indicate that the four models performed best under the following hyperparameters: the one-hot encoding model with an alignment length of 25 and 128 neurons; the embedding layer model with a vector dimension of 16, alignment length of 25, and 128 neurons; the Word2Vec model with a window size of 4, vector dimension of 16, alignment length of 25, and 128 neurons; and the TF-IDF model with an alignment length of 25 and 128 neurons. Particularly, among the four models, the Word2Vec pre-trained word vector model demonstrated the best performance, validating

its superior representational capacity in pattern identification.

The proposed TCM pattern identification model in this study holds auxiliary value in TCM research. The model can effectively identify specific virtual patterns while providing pattern information that is lacking in clinical records for deep learning. This will contribute to improving the accuracy and efficiency of TCM pattern identification, providing more reliable support tools for doctors, and promoting the integration of TCM research and machine learning.

Keywords: TCM symptoms, patterns, pattern identification, word vectors, deep learning, TF-IDF



目錄

論文著作財產權同意書	I
論文指導教授推薦函	II
論文付印申請表	III
致謝	IV
摘要	V
ABSTRACT	VI
目錄	VIII
圖目錄	XII
表目錄	XIII
壹、緒論	1
1.1 研究背景	1
1.2 研究動機	2
1.3 研究目的	2
貳、文獻回顧與探討	5
2.1 中醫證候與辨證論治	5
2.1.1 辨證論治	5
2.1.2 辨證	6
2.2 中醫辨證的相關研究	6
2.3 多標籤分類問題相關研究	9
2.4 深度學習網路	10
2.4.1 機器學習 (Machine Learning)	10

2.4.2 神經網路基礎	11
2.4.3 神經網路結構	12
2.4.4 深度學習	13
2.4.5 單一神經元學習方式	14
2.4.6 整個神經網路學習方式	16
2.4.7 多層感知器(Multilayer Perceptron,MLP)	18
2.4.8 卷積神經網路	18
2.5 詞向量	18
2.5.1 詞向量 (Word vector, Word embedding).....	18
2.5.2 Word2vec	19
2.6 TF-IDF	20
2.6.1 TF-IDF 的演變	20
2.6.2 TF-IDF 公式	21
參、 研究方法	24
3.1 資料來源及彙整	24
3.2 資料前處理(Data Preprocessing)	24
3.2.1 建立 Tokenizer 物件	25
3.2.2 症狀詞轉數字(texts_to_sequences)	25
3.2.3 數字序列對齊(pad_sequences)	26
3.2.4 訓練資料及測試資料的產生	26
3.3 模型訓練	30
3.3.1 基於 onehot encoding 模型	30
3.3.2 基於 embedding layer 訓練詞向量的模型	32
3.3.3 基於遷移學習 Word2Vec 詞向量的模型	34
3.3.4 基於 TF-IDFencoding 的模型	35
3.4 多標籤分類評估方法	37
肆、 實驗結果	41
4.1 實驗設計	41
4.2 實驗一、基於 onehot encoding 模型訓練結果	42
4.3 實驗二、基於 TF-IDF encoding 模型訓練結果	45

4.4 實驗三、基於 embedding layer 模型訓練結果.....	47
4.5 實驗四、基於 Word2Vec 模型訓練結果.....	49
4.6 四種模型最佳參數比較.....	53
4.7 實驗五、四種模型在原始資料集的準確率.....	55
4.8 實驗六、原始資料擴充不同筆數對四種模型準確率的影響.....	56
4.9 實驗七、雜訊對四種模型準確率的測試.....	59
伍、 結論與未來發展.....	62
5.1 結論.....	62
5.2 未來發展.....	65
參考文獻.....	67
附錄 1：18 個虛證證候症狀表.....	69
附錄 2：50 筆病歷證候症狀表(1~10 筆).....	70
附錄 3：50 筆病歷證候症狀表(11~20 筆).....	71
附錄 4：50 筆病歷證候症狀表(21~30 筆).....	72
附錄 5：50 筆病歷證候症狀表(31~40 筆).....	73
附錄 6：50 筆病歷證候症狀表(41~50 筆).....	74
附錄 7：TF-IDF 權重證候症狀對照表(症狀 1~15).....	75
附錄 8：TF-IDF 權重證候症狀對照表(症狀 16~30).....	76
附錄 9：TF-IDF 權重證候症狀對照表(症狀 31~45).....	77
附錄 10：TF-IDF 權重證候症狀對照表(症狀 46~60).....	78
附錄 11：TF-IDF 權重證候症狀對照表(症狀 61~75).....	79

附錄 12：TF-IDF 權重證候症狀對照表(症狀 76~90).....	80
附錄 13：TF-IDF 權重證候症狀對照表(症狀 91~105).....	81
附錄 14：TF-IDF 權重證候症狀對照表(症狀 106~116)	82



圖目錄

圖 2-1 人類神經元	11
圖 2-2 人工神經元	12
圖 2-3 神經網路架構圖	13
圖 2-4 單一神經元學習方式	14
圖 2-5 調整權重 W	14
圖 2-6 調整偏值 B	15
圖 2-7 神經網路架構	16
圖 2-8 神經網路訓練流程圖	17
圖 2-9 CBOW 和 SKIP-GRAM	20
圖 3-1 症狀擴充圖	29
圖 3-2 證候症狀重新排列	29
圖 3-3 模型訓練架構圖	30
圖 3-4 基於 ONEHOT ENCODING 模型	32
圖 3-5 基於 EMBEDDING LAYER 模型	33
圖 3-6 基於 WORD2VEC 模型	35
圖 3-7 基於 TF-IDF ENCODING 的模型	37
圖 3-8 單標籤及多標籤	38

表目錄

表 3-1 TF-IDF 權重對照表	36
表 4-1 基於 ONEHOT ENCODING 模型神經元數量對訓練及驗證的影響	43
表 4-2 基於 ONEHOT ENCODING 模型神經元數量對測試集的影響	43
表 4-3 基於 ONEHOT ENCODING 模型對訓練及驗證的準確率	44
表 4-4 基於 ONEHOT ENCODING 模型對測試集的準確率	45
表 4-5 基於 TF-IDF ENCODING 模型對訓練及驗證的準確率	46
表 4-6 基於 TF-IDF ENCODING 模型對測試集的準確率	47
表 4-7 基於 EMBEDDING LAYER 模型對訓練及驗證的準確率	48
表 4-8 基於 EMBEDDING LAYER 模型對測試集的準確率	49
表 4-9 症狀詞相似度	50
表 4-10 基於 WINDOW_SIZE 模型對訓練及驗證的準確率	51
表 4-11 基於 WINDOW_SIZE 模型對測試集的準確率	52
表 4-12 在 15 筆病歷中四種模型最佳參數	53
表 4-13 在 18 筆虛證中四種模型最佳參數	54
表 4-14 原始資料 18 虛證及 50 病歷的準確率	55
表 4-15 不同擴充筆數的四種模型在訓練及驗證的準確率	57
表 4-16 不同擴充筆數的四種模型在測試集的準確率	58
表 4-17 隨機刪減症狀數量 18 虛證及 50 筆病歷的準確率	60

壹、緒論

1.1 研究背景

隨著人們生活水平的提高，對醫療保健的關注度也逐漸提高。然而，醫生在診斷時所記錄的臨床紀錄通常包含豐富的資訊，可用於為患者提供更好的醫療服務，為健康管理者提供決策的價值，甚至可以用來預測疾病，達到早期預防的目的。然而，臨床紀錄通常是非結構化的，而中醫領域的症狀更是難以理解和分析，這使得充分利用臨床紀錄成為一項極具挑戰性的任務。

中醫師在處理疾病時，最核心的方法就是辨證論治[16]。辨證(syndrome differentiation)是對中醫四種主要診斷程序：望、聞、問、切所獲得的臨床訊息的綜合分析，通過對這些臨床訊息所提供之疾病症狀(symptoms)的關聯分析，可以推演出包括病位、病因與病機組合而成的中醫「證候」(TCM syndrome)。證候是綜合分析出的疾病症狀之結果，有了證候，中醫師就能進一步推斷相關的中醫處方及針灸選穴、推拿按摩等治療方法。透過根據症狀辨別證候，推斷疾病發生的原因，再採取相對應的治療方法，顯示出中醫治療疾病的特色：「辨證論治」。中醫辨證論治的重點是辨識出患者的身體狀況，並針對患者的個體特徵進行治療，藉此提高治療效果。

中醫學認為，一種疾病的發展往往經歷多個階段[23]。每個階段都具有獨特的特徵，因此被稱為「證候」。證候可以分為虛證(Deficiency Syndrome)和實證(Excess Syndrome)兩類。虛證是由於臟腑的氣(qi)、血(blood)、陰(yin)、陽(yang)等基本元素不足而引起的病理狀態，而實證則是由外在致病因素或臟腑功能失調所導致的病理變化。症狀是疾病的主觀或客觀指標，用以描述證候的特徵。因此，每一種證候都可以透過一組獨特的症狀來描述[23]。

1.2 研究動機

資訊技術在西醫的疾病診斷中扮演著重要的角色，但是在中醫診斷中卻鮮少被應用。中醫的診斷方法主要是辨證，也就是根據病人的症狀和證候集合進行判斷。這是一個多標籤分類問題(Multi-label classification problem)，需要將病人的症狀與證候進行對比，找出病人的證候。

為了開發一個完善的辨證系統，需要解決三個主要問題。第一個問題是如何判斷病人是否患有某種特定的證候，因為不同的證候與症狀的關聯性不同，需要運用相關的技術來進行準確的判斷。第二個問題是如何確定每個症狀的權重，這將有助於判斷病人是否患有該證候。第三個問題中醫辨證是屬於多標籤的分類問題，在一組症狀詞中須同時判斷多個證候。總之，利用資訊技術來開發中醫辨證系統，可以更好地幫助中醫師提高疾病診斷的準確性和效率，為病人提供更好的治療服務。

1.3 研究目的

多標籤問題和多分類問題是兩種不同的問題形式。在多分類問題中，每個樣本只會被分配到一個類別，而在多標籤問題中，每個樣本可以同時被分配到多個標籤。換句話說，在多分類問題中，每個樣本只有一個正確的答案，而在多標籤問題中，每個樣本可以有多個正確的答案。

例如，在一個多分類問題中，可能希望將一些影像分類為貓、狗、鳥中的一種。在這種情況下，每個影像只能分配到一個類別，例如貓、狗或鳥。然而，在一個多標籤問題中，可能希望將一些文本標註為有關環境、政治和科學中的多個標籤。在這種情況下，每個文本可以同時被標註為有關環境、政治和科學等多個標籤，且每個標籤的存在與否都是獨立的。然而中醫辨證是屬於多標籤的分類問題。

相對於多分類問題，多標籤問題通常更具挑戰性和複雜性，因為它涉及到將樣本分配到多個標籤或類別，同時考慮到這些標籤之間的關聯性。例如，一個圖像可能包含多個對象，每個對象都需要標籤，但這些對象之間可能有某些相關性。在這種情況下，需要考慮到多個標籤之間的相互作用，並進行適當的分類。

與多分類問題相比，多標籤問題的另一個挑戰是樣本不平衡。也就是說，在訓練數據中，某些標籤的分佈可能非常不均衡，這可能會影響模型的預測能力和泛化能力。解決這個問題需要更多的數據、平衡的樣本分佈和特殊的算法技巧，例如參數及權重調整等技巧。

陳信賢[3][23]及其團隊整理了五本中醫診斷學典籍中的虛證證候及其相應的原始症狀，並且進行了標準化處理。隨後，他們利用每個證候的標準化症狀，定義了該證候的模糊歸屬函數，以開發出一個模糊辨證系統，從而提高了系統的辨證精度。然而，這些方法存在一些限制，例如每個證候只能根據其症狀集定義模糊歸屬函數，而忽略症狀之間的語意關係[20]，進而限制了辨證效能。此外，症狀之間可能存在交互作用，包括抑制或激發作用，但這種交互作用通常沒有在辨證區分中被考慮[20]。因此，本研究嘗試運用詞向量和深度學習網路等技術，從中提取這兩種重要資訊，以加強辨證效能。

目前有許多研究致力於建構分散式詞嵌入模型，以將字詞的語法和語義訊息編碼為連續向量。其中，最著名的模型是Word2Vec，它能將語意或脈絡相似的字詞對應到向量空間中相似的位置[20][21]。除此之外，近年來，深度學習作為機器學習的一個分支，通過模擬人腦的運作方式，使機器能夠像人類一樣學習和自主分析。相較於機器學習，深度學習更加強調從原始資料自動提取特徵的能力，而不需要太多領域知識的干預[6][7][8]。因此，透過深度學習網路所學到的特徵權重，可以自動萃取出將症狀集合正確對應到相關證候的交互作用資訊，從而進一步提高辨識效能。

在前述基礎上，本研究提出一個結合症狀詞向量與深度學習網路的中醫辨證模型，以解決直接從臨床病歷中自動推斷和實現中醫的辨證知識非常困難的問題。為了建立該模型，本研究首先收集了自中醫典籍和期刊的證候及其特徵症狀，作為辨證實驗的資料來源。接著，針對18個虛證以及50個病例資料進行了詞向量空間的「預訓練(pre-training)」，再進行遷移(transfer)到深度學習網路來建模。透過分析比較該模型的表現，本研究也將探討模型優劣及其相關原因，以進一步提升中醫辨證的精確度。

本研究結果有助於提高中醫師在對特定證候集進行辨證和診斷時的效率和精準度。未來，本系統可以進一步擴展，包括對其他虛證、實證以及特定疾病的中醫辨證[4][5][19]，從而為臨床診斷提供更有價值的支援，同時也有助於為未來的機器學習提供臨床病歷中缺乏的證候資訊。因此，應用機器學習的方法可以從不斷收集到的臨床醫療記錄中推演出新的知識，自動地不斷升級本系統，並為醫療診斷和治療提供更好的支援和指引。

貳、 文獻回顧與探討

2.1 中醫證候與辨證論治

在中醫學中，治療疾病的核心概念是「辨證求因，審因論治」。這個概念中的「證」是指「證型」或「證候」，即當時身體的整體狀態。每個證型背後都有一套病因和病機，以及相應的一組症狀。證候是病因的外在表現，而病機則是證候的內在本質。由於病機的內涵包括病變的部位、原因、性質和邪正盛衰的變化，因此證候可以用來說明人體的不平衡以及對抗致病因素的能力，揭示了病變的機理和發展趨勢。例如，風寒束表、肝陽上亢、心血虧虛、心脈痹阻等都是常見的證候。在中醫治療中，對證候的準確分析和判斷是十分重要的，因為只有這樣才能制定出針對病因病機的治療方案，從而達到治療疾病的目的[1]。

2.1.1 辨證論治

中醫治療證候的方法不是直接針對症狀進行治療。因為一個證候僅反映了整個疾病發展過程中的一個縱切面，也就是人體這個龐大系統的某一時刻的狀態和模型。一個疾病的發展過程可以看作是許多證候的連續變化和演變。通過辨別不同症狀所表現出的證候，推斷疾病發生的原因，然後根據這些證候採取相應的治療方法，這正是中醫治療疾病的特色：「辨證論治」[1]。

這種中醫辨證論治的方法，對於治療疾病起到了非常重要的作用。因為不同的症狀反映了不同的證候，進而反映出不同的疾病發生原因和機理，因此必須針對具體的證候，對症下藥才能達到良好的治療效果。這種方法與西方醫學的症狀治療不同，它強調的是對整個疾病過程的把握和理解，從而為治療提供更加全面和精準的指導。

2.1.2 辨證

辨證是中醫診斷疾病的重要過程，需要透過四診（望、聞、問、切）所收集的所有資料，包括症狀和體徵，運用中醫學理論進行分析、綜合，從而辨識疾病的原因、部位、性質以及發展的趨向，最終概括、判斷為某種性質之證候。

然而，由於證候只能反映疾病某一個階段和某一個類型的病變本質，故中醫學在辨識證候時，必須同時辨明疾病的病因、病位、病性及病勢，即辨明疾病從發生到轉歸的總體病機[1]。

因此，「辨證論治」是中醫在診斷與治療疾病時，考量所有面向以取得平衡對應的重要觀念與過程，以實現治療效果的最大化。

2.2 中醫辨證的相關研究

近年來，許多學者[3][4][5][15][16][23]開始運用模糊分類方法，以解決傳統中醫的辨證問題。這種方法採用模糊邏輯理論，融合了人類專家知識和現代計算機科學技術，可更精準地辨識和分析疾病證候，提高診斷和治療的準確性。模糊分類方法的優點在於能夠處理中醫辨證時存在的模糊性和不確定性，尤其適合處理多因素影響的複雜疾病。這種方法還可以將中醫的經驗和現代科學結合起來，為中醫學的現代化發展提供了新的思路和方法。

中醫症狀用詞的差異化是一個重要的問題，因為不同的中醫師可能會使用不同的用詞描述相同的症狀，所以中醫診斷具有非線性(nonlinear)、模糊性(fuzzy)和複雜性(complex)，這種差異可能會對中醫辨識系統的性能產生影響。因此，很難通過傳統的數理統計或資料採礦來研究[15]。為了解決這個問題，需要開發一個中醫症狀標準化的方法，將不同的用詞轉換為統一的表達方式。

Yang等人[15]設計基於中醫理論的模糊分類方法來處理中醫診斷中的定性模糊因素(the qualitative Fuzzy factors)。Chen等人[16]定義中醫症狀在疾病的病位、病因、病徵或病況的量化編碼(quantitative code)，並以此記錄的症狀資訊，作為中醫師或系統辨證的依據。然而，這些方法都沒有面對中醫症狀用詞差異化的議題[2][23]。

中醫文獻或臨床病歷中所記載的症狀描述與意義解釋，受到時代背景、傳承派別、地域氣候以及個人習慣等差異的影響，因而出現「多詞一義」與「一詞多義」的問題，使得症狀用詞資訊的一致性產生困難。這種資訊不一致性讓系統難以準確辨別症狀的異同，進而影響到辨證推論的正確性。

在中醫診斷中，常常存在一個症狀描述有著兩種或以上涵義的情況，這種情況被稱為「一詞多義」。舉例來說，原始症狀「二便不利」就可以代表大便和小便皆不順暢這兩種症狀涵義。這樣的情況會對系統辨識造成困難，因為系統很難分辨出這些症狀描述的異同。因此，應該將這些症狀進行標準化，例如將「二便不利」分成「小便不暢」和「大便不暢」這兩個症狀。同樣地，「苔薄白」這個原始症狀也可以被標準化成「舌苔厚度薄」和「舌苔顏色白」兩個症狀 [4]。這樣的標準化能夠幫助系統更準確地識別不同的症狀描述，提高辨識的準確率。

在中醫文獻或臨床病歷中，症狀描述的多樣性往往使得相同的症狀被表達成不同的詞語，這就是所謂的「多詞一義」問題。例如，「大便艱澀」、「大便不通」、「大便不利」等症狀描述實際上都指向同一個意義——大便不順暢。為了解決這個問題，需要對這些描述進行標準化，統一使用「大便不暢」這一症狀描述。類似的，「月經提前」、「月經超前」等症狀描述也都指向「月經先期」這一意義，因此需要統一使用「月經先期」這一症狀描述[5]。

陳南宏[2]在其研究中建構了中醫症狀詞庫以及中醫症狀知識本體，旨在將中醫

症狀術語進行標準化，使其成為一詞一義的統一用語。透過這樣的標準化方式，陳南宏成功地解決了症狀多詞一義與一詞多義的問題，讓中醫文獻和臨床病歷中的症狀描述更加清晰明確。這樣可以保證不同中醫師使用的症狀描述之間的一致性，並且可以提高中醫辨識系統的性能。這項研究成果有助於提高中醫辨識和診斷的精準度，並且可以應用於發展更為智能化的中醫診斷系統。目前，一些研究正在探索如何進行中醫症狀標準化，這些研究可能為中醫辨識系統的發展提供有用的參考。

Yang等人[15]為了解決傳統中醫辨證問題，定義了24個模糊集，對應於24個病位或病因。透過問卷調查收集了21位專家對於22種常見心臟病的症狀權重，利用權重定義歸屬函數，以此計算病人所患症狀集對應的最相似病位模糊集和病因模糊集，並透過這些病位和病因的組合進行間接推斷診斷的證候。在模糊集的論域中， U 表示症狀的字集合，即包含所有可能的結果及其集合的集合。而歸屬函數 $u_d: U \rightarrow [0, 1]$ 則表示每個元素對應到模糊集的度量值，這是定義模糊集的核心要素之一。

在中醫辨證問題的解決中，Chen等研究者[16]應用模糊集理論定義了46個模糊集，分別代表了不同的病位、病因、病徵和病況。透過計算病人症狀集中每個症狀對每個模糊集的歸屬度，綜合判定最相似的病位、病因、病徵和病況，進而推斷出所診斷的證候。其複雜的計算過程涉及到多個變量的組合分析，可謂是模糊集理論在中醫辨證中的成功應用。

陳信賢教授進行了一項研究[3]，從五本中醫診斷學典籍中梳理出18個虛證證候及其特徵症狀。他運用模糊理論，定義模糊集合，採用症狀分群和症狀分權的方式來設計歸屬函數，並開發了一套辨識系統，以計算病人所屬證候的歸屬度。

林柏伸[23]對陳信賢的歸屬函數進行了改進，將虛證症狀分為五個症狀群，並基於疾病的病位、病因和病機分配症狀權重，以提高系統辨識的準確性。這項改進後，後續的研究者將其模糊辨識的方法進一步應用於月經病[5]和呼吸道疾病[4]等相關症狀的辨識上。

Hu等人[22]對於非結構性醫療紀錄中的陽虛和陰虛進行了探究，並使用了卷積神經網路和fastText兩種文本分類方法，以評估辨識的可行性。

Zue等人[20]進行了對濕熱證的分析，並利用Word2Vec和TF-IDF等技術與傳統的機器學習方法相結合，從而進一步提高了分析效果。

2.3 多標籤分類問題相關研究

陳耀男[10]提出一種有效率的標籤空間降維法，稱為條件式主要標籤空間變換(Conditional Principal Label Space Transformation)。相較於現有的標籤空間降維法，此法同時考慮了特徵和標籤的資訊，能夠最小化演算法中常用的漢明距離上限。此法的最小化步驟能夠通過有效率的奇異值分解實現，並且可以擴展到核函數方法中，以使用更複雜的特徵組合進行標籤空間降維。實測結果表明，相較於其他現有方法，此法在多標分類問題中表現更為有效。

林怡如及吳政隆[11]研究運用自然語言處理與機器學習技術，嘗試建構客戶之發問類型辨識模型。其研究目的為建構一個多重標籤之分類任務資料集，其主是露天採用拍賣之問題類型部分作為主要研究資料來源，以及將資料分為粗細分類的兩組資料集。研究嘗試以TFIDF和Word Embeddings 等兩種詞向量演算法進行特徵轉換，以及採用Extra Trees、Logistic Regression、Random Forest、SVM 等四種機器學習分類演算法進行分類模型建立。

傅琦佳[12]使用義大醫院提供的經專業醫生標注的1630筆胸部X光影像，其中1212筆資料是正常無病的，其餘樣本存在「一圖多病」的情況，屬於典型的多標籤分類問題。本研究通過深度學習模型，提取影像特徵，預測疾病標籤；同時，利用美國國立衛生研究院公開的大型胸片影像數據集，採用多重遷移學習的方法減小資料不足的影響，進一步優化分類模型。研究結果顯示，深度學習模型ResNet50和DenseNet121均可實現不錯的準確率，結合遷移學習可以顯著增強模型的預測能力，並節省所需的計算資源。運用疊代式遷移學習有望得到性價比最高的分類模型。

何原野[13]研究使用自動化方法從心理健康網站的心理疾病描述文本中提取情感標籤的問題。使用了一種組合型深度神經網路框架BLSTM_CNN模型來進行多標分類。該模型將BLSTM用於針對每個詞提取整句話的特徵，並使用CNN提取一句話的局部特徵，以避免LSTM和CNN的缺點並結合它們的優點。實驗結果表明，BLSTM_CNN模型優於獨立使用CNN、LSTM和SM模型，可以有效地自動辨識心理疾病標籤，從而提高網上心靈健康服務的效率和效益。

2.4 深度學習網路

2.4.1 機器學習 (Machine Learning)

機器學習 (Machine Learning) 是一門資料科學技術，能讓電腦從過去的資料和經驗中自主學習和分析數據，從中發現隱藏的規律，以便預測未來的結果、行為和趨勢。透過機器學習，電腦可以從大量的數據中學習，並自主提取出有價值的資訊，不斷優化其預測和決策能力，以應對日益複雜的現實挑戰。

2.4.2神經網路基礎

人工智慧技術種類繁多，其中神經網路（Neural Network，簡稱NN）是一種機器學習技術，其模仿了生物神經網路的結構，例如人類大腦，也被稱作類神經網路或人工神經網路（Artificial Neural Network）[8]。

神經網路的最小單位是神經元(neuron)，如圖2- 1所示是人類神經元以及如圖2- 2所示是人工神經元的比較：

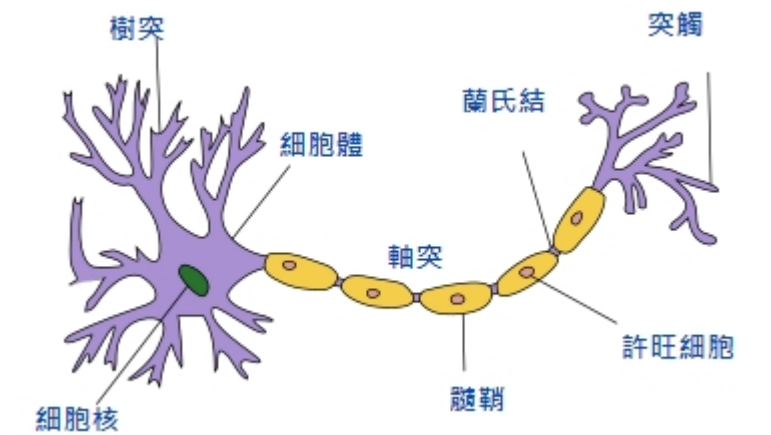


圖 2- 1 人類神經元

資料來源：

<https://zh.wikipedia.org/zh-tw/%E7%A5%9E%E7%B6%93%E5%85%83>

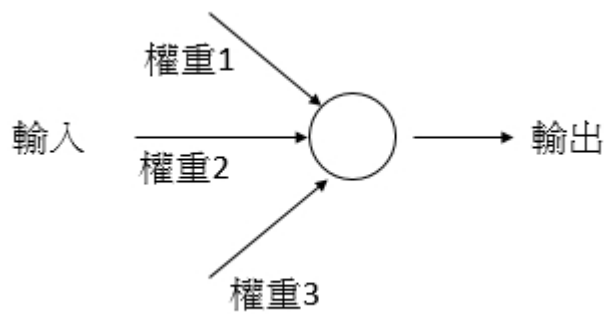


圖 2-2 人工神經元

- a. 如圖2- 1所示，人類神經元是由多個樹突所組成，每一個樹突皆有其獨特的敏感度，接收並處理訊息。經過樹突接收的訊號強度，會被加總並進行必要的轉換後，由軸突傳輸至下一個(或多個)神經元。這樣的神經元結構能夠有效的處理及分析複雜的神經訊息。
- b. 如圖2- 2**錯誤! 找不到參照來源。**所示，如同人類神經元一般，人工神經元也可以接收多個輸入。每個輸入都會經過不同權重的調節，然後被加總計算出輸入的總和，接著進行適當的轉換，最後再輸出到下一個(或多個)神經元。

2.4.3神經網路結構

深度神經網路或深度學習網路是由數百萬個人工神經元隱藏層連結在一起的，這些神經元之間的連接稱為權重。權重值表示節點之間的連線強度，正數表示一個節點激發另一個節點，負數表示一個節點抑制另一個節點。當節點權重值越高時，對其他節點的影響也就越大[8]。

理論上，深度神經網路可以將任何輸入類型映射至任何輸出類型，但相較於其他機器學習方法，它們需要更多的訓練。深度神經網路需要數百萬個訓練資料範例，而一般網路可能只需要數百個或數千個訓練資料。如下圖2- 3所示。

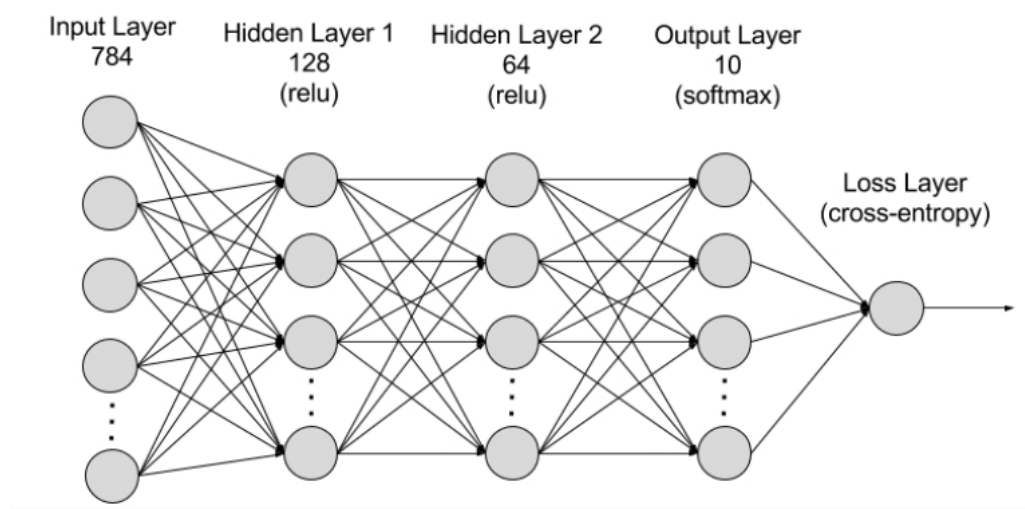


圖 2- 3 神經網路架構圖

資料來源：

<https://aws.amazon.com/tw/what-is/neural-network/>

2.4.4深度學習

深度學習(Deep Learning)是一個從機器學習領域中發展而來的重要技術，通過模擬人類大腦的運作方式，讓機器也能像人類一樣具有自我學習的能力。相較於傳統的機器學習，深度學習不需要人工提取特徵，而是通過自動提取資料集中的特徵，並將這些特徵放入多層神經元的神經網路中進行處理。隱藏層和輸出層的網路架構相互連接，並選擇最適合的網路架構進行深度學習的訓練，從中建立深度學習模型，這些模型被廣泛應用於各個領域[6][7][8]。

2.4.5 單一神經元學習方式

先從只有1個輸入神經元來看，由輸入值 x 到輸出值 y 是一個線性的關係，如圖 2- 4所示。

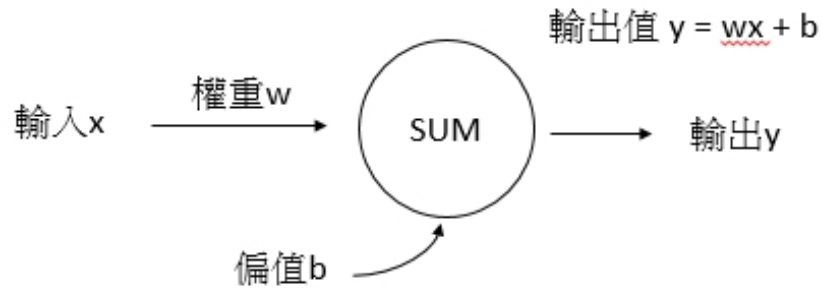


圖 2- 4 單一神經元學習方式

單一神經元是神經網路的基本單元，它通常由一組權重和一個偏值構成。在訓練神經網路時，通常會調整神經元的權重和偏值來使其能夠更好地擬合訓練數據。如果將 $y = w \times x + b$ 畫出來會是一條直線，若是調整權重 w 則可以改變直線的斜率，若調整偏值 b 則可以橫向平移，如圖2- 5、圖2- 6所示：

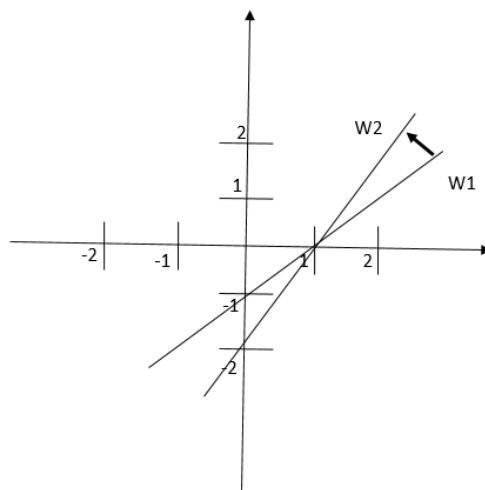


圖 2- 5 調整權重 w

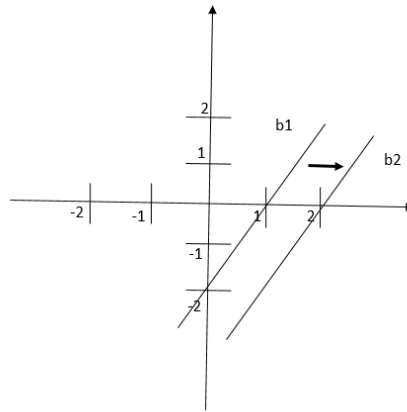


圖 2-6 調整偏值 b

- a. 調整權重的方式是通過計算誤差，然後使用反向傳播算法更新權重。反向傳播算法的基本思想是計算輸出層和隱藏層之間的誤差梯度，然後根據梯度的方向和大小來更新每個權重。具體地，使用梯度下降算法來最小化損失函數，進而更新權重。
- b. 調整偏置的方式與調整權重的方式類似，但只需要計算單個神經元的誤差，並使用梯度下降算法來更新偏置。偏置的調整可以看做是對神經元激活函數的平移，它能夠對神經元的激活狀態產生影響，進而影響神經網路的輸出結果。

2.4.6 整個神經網路學習方式

如圖2-7是一個簡單的3層神經網路架構，每一個神經元的輸出訊息，都會傳遞到下一個神經元，下一層的神經元依照每個輸入的權重來接收上層訊息。如隱藏層有 $3 \times 4 + 4 = 16$ 個權重參數(其中+4是4個偏值bias)[8]。

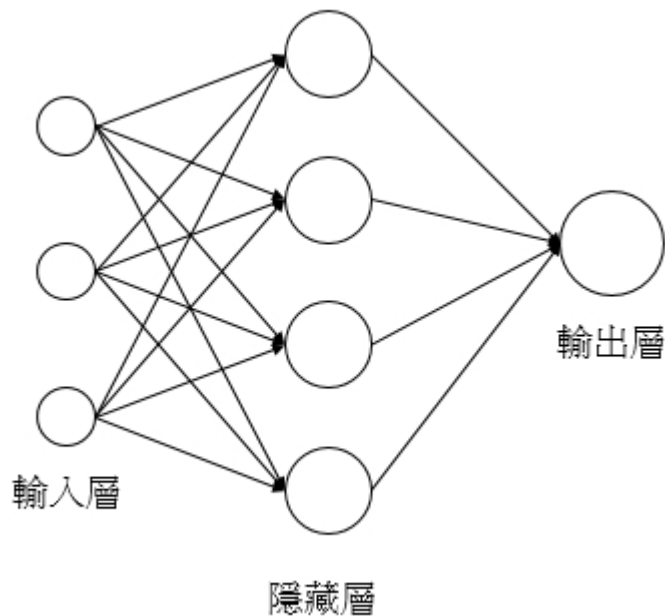


圖 2-7 神經網路架構

- a. 輸入層：是指接收來自外部的數據或資訊的層次，如圖像、聲音、文本等。
 - b. 隱藏層：是指處理這些數據的中間層次，它通常使用多個神經元來進行計算和特徵提取，並將結果傳遞到下一層。
 - c. 輸出層：會根據神經網路的任務，將處理過的數據映射為最終的輸出結果，例如圖像分類中的不同類別、語音識別中的文字、機器翻譯中的翻譯結果等。
- 三層結構的神經網路稱為單隱藏層神經網路，而更多層次的神經網路稱為深度神經網路。

所以神經網路藉由大量的資料來訓練，利用損失函數和優化器找出每個神經元的最佳權重，其流程如圖 2-8 所示：

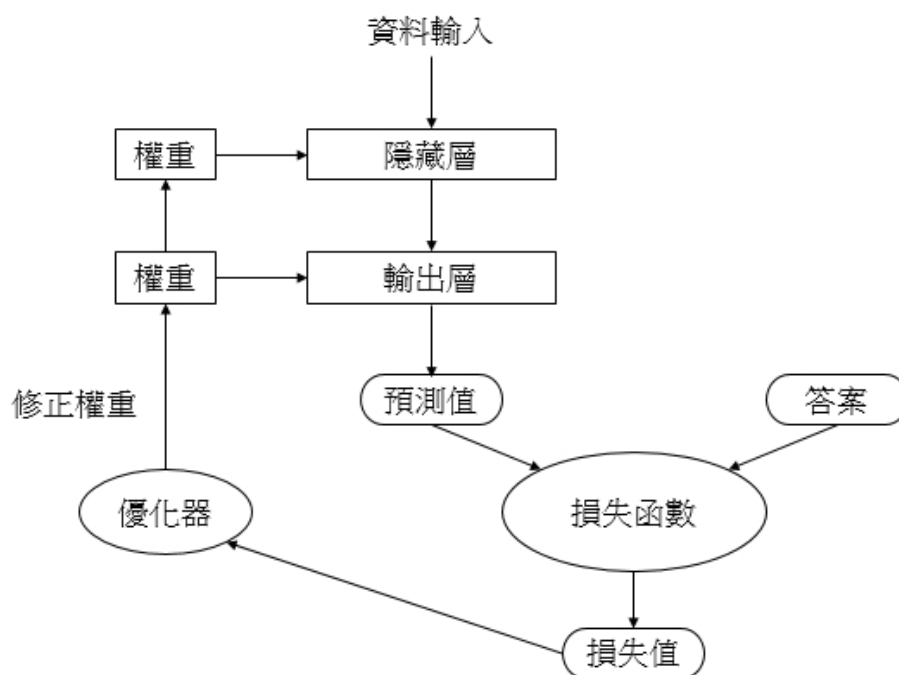


圖 2-8 神經網路訓練流程圖

- a. 損失函數(loss function)：用來度量模型預測值與真實值之間的差距。損失函數越小，表示模型預測越準確。在訓練過程中，通常會最小化損失函數來調整模型參數，使得模型能夠更好地擬合訓練集和泛化到測試集。常見的損失函數包括均方誤差(MSE)、交叉熵(cross-entropy)、對數損失(log loss)等。選擇損失函數時需要考慮模型的任務類型和特點，以及樣本分佈等因素。
- b. 優化器(Optimizer)：用於最小化損失函數的演算法。它是神經網路中的一個重要組件，其目的是根據訓練數據和損失函數來調整神經網路中權重和偏差的值，以使得預測結果更接近於實際值。常見的優化器包括：隨機梯度下降法(Stochastic Gradient Descent, SGD)、動量法(Momentum)等。這些優化器使用不同的策略來調整權重和偏差的值，以求得最小化損失函數的目標。

2.4.7 多層感知器(Multilayer Perceptron,MLP)

多層感知器(Multilayer Perceptron,MLP)是一種常見的神經網路模型，也被稱為深度神經網路(Deep Neural Networks,DNN)。卷積神經網路(Convolutional Neural Network,CNN)是一種專門處理圖像和視頻的神經網路模型，可用於圖像識別和分類等應用。循環神經網路(Recurrent Neural Network, RNN)則適用於處理序列數據，例如語音識別和自然語言處理。這三種模型都有廣泛的應用領域，包括影像辨識、語音識別、文本分類、自動駕駛和人工智慧等領域[6][7][8]。

2.4.8 卷積神經網路

卷積神經網路是深度學習中非常關鍵的一部分，相較於多層感知器，它具有卷積層和池化層的結構。卷積層(Convolution Layer)利用多個濾鏡(filters)對資料進行局部特徵的提取。這些濾鏡的卷積核(Kernel)在資料上滑動並不斷計算，形成了一個特徵圖(Feature map) [6-8]。卷積層所提取的特徵圖可以繼續傳遞到下一層卷積層，以便更加複雜的特徵的提取。池化層(Pooling Layer)則是用來對特徵圖進行壓縮，通常是進行最大化(Max Pooling)或平均化(Average Pooling)的降維計算，以保留最重要的特徵資訊。

2.5 詞向量

2.5.1 詞向量 (Word vector, Word embedding)

為了使電腦能夠理解自然語言中的抽象概念和詞彙，研究人員已開發了向量表示的方法。這種方法將文字轉換為數值形式，並且在轉換過程中保留了原本的語義。通過適當的向量訓練，可以利用這些向量之間的關係來表達詞彙之間的語義關係。例如，將"king"的向量減去"man"的向量再加上"woman"的向量，最終結果最接近於"queen"的向量，這表明向量空間中的算術運算反映了自然語言詞彙之間的語義關係。這種方法在自然語言處理中具有廣泛的應用，例如情感分析、文本分類和自動機器翻譯等。

詞向量技術最初於2000年問世

[24]，當時的研究人員利用神經概率語言模型(neural probabilistic language models)來學習單詞分佈的表示方法[25]。這種方法通過分析大量語言資料，學習單詞之間的相互關係，並將其轉換為向量表示。然而，當時的技術在處理大規模資料時存在效率問題。

後來，研究人員開始使用局部線性嵌入(locally linear embedding)來學習高維資料結構的低維表示方法。這種方法通過維度降低，將複雜的高維數據轉換為更簡單的低維表示，以便更好地進行分析和處理。到了2013年，Mikolov等人[26]則開始使用非監督式學習(unsupervised learning)的Word2vec模型進行詞向量的訓練。這個模型通過分析大量的文本資料，學習單詞之間的相似度和差異性，並將其轉換為向量表示。Word2vec的突破在於其極快的訓練速度和精度，使得詞向量技術得到了極大的發展和應用。現在，詞向量技術已被廣泛應用於自然語言處理領域，如情感分析、文本分類、自動機器翻譯等。

2.5.2 Word2vec

在詞向量訓練領域中，Word2vec[14]是最經典的方法。Word2vec的基本思想是利用一個詞的上下文來學習該詞的向量表示，以便計算詞之間的相似度和關聯度。為此，Word2vec提供了兩種不同的訓練方式：CBOW(Continuous Bag of Words Model)和Skip-gram(ContinuousSkip-gramModel)[14]。

- a. CBOW模型通過一個固定大小的上下文窗口，將上下文詞的向量加總，並以此總和作為預測目標詞的向量表示。這個模型通常用於小資料集，可以生成更快速和更穩定的詞向量。
- b. Skip-gram模型與CBOW相反，通過一個目標詞的向量表示，來預測該詞的上

下文詞。這個模型通常用於大型資料集，可以生成更準確和更細緻的詞向量。

如圖 2- 9 所示，CBOW(Continuous Bag of Words Model) 和 Skip-gram(ContinuousSkip-gramModel)，是Word2vec的兩種訓練模式，兩者都是用神經網路來學習每個單詞的底層單詞表示的架構。

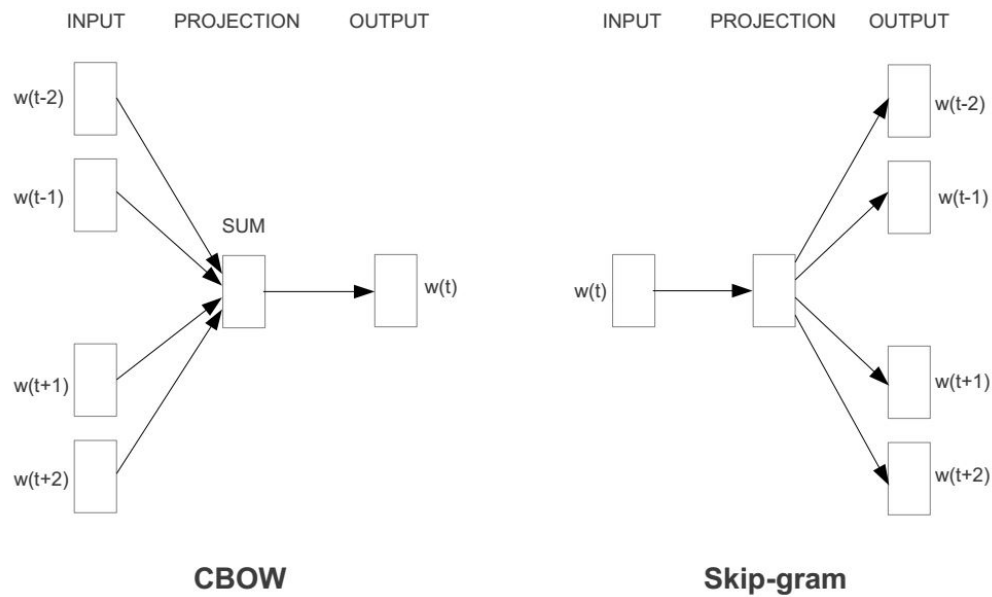


圖 2- 9CBOW 和 Skip-gram

資料來源:

<https://blog.csdn.net/u010665216/article/details/78724856>

2.6TF-IDF

2.6.1 TF-IDF的演變

TF-IDF是一種在自然語言處理中常用的詞語權重計算方法，它是由詞頻(TF)和逆文檔頻率(IDF)兩部分組成。TF指的是詞頻，它衡量的是某個詞在一個文檔中出現的次數，詞頻越高，該詞對應的權重就越大。IDF指的是逆文檔頻率，它衡量的是某個詞在文檔集中出現的次數，文檔集合越大，該詞對應的權重就越小。TF-IDF的計算方法是將詞頻和逆文檔頻率相乘，得到每個詞對應的權重值，該權重值可以用於文本相似度計算、關鍵詞提取等自然語言處理任務。

TF-IDF方法是文本分析中常用的技術。其中，TF（詞頻，term frequency）統計是學界早已公認的技術，不需要進一步的高級技術。而IDF（逆文檔頻率，inverse document frequency）的概念最初是由Sparck Jones於1972年提出的。然而，他並未對IDF的演算法進行理論解釋，也沒有對此進行深入研究。直到Salton等學者多次發表專書和專文，探討TF-IDF技術在資訊檢索中的應用，IDF的概念才得到進一步發展。實際上，IDF的概念是在特定條件下關鍵詞概率分佈的相對熵[9]。

2.6.2 TF-IDF公式

TF-IDF是一種文本特徵提取方法，用於評估詞語對於文本的重要程度。在TF-IDF方法中，詞頻(TF)是指一個詞在文本中出現的頻率，而逆向文件頻率(IDF)則表示包含這個詞的文本數量的倒數，即在整個語料庫中，包含這個詞的文件數越少，IDF值越大，則認為這個詞對於區分文本的能力越強。簡單來說，如果一個詞在某個文本中頻繁出現，但在其他文本中很少出現，則該詞的TF-IDF值就會比較高，進而表明該詞對於該文本的區分能力比較強。TF-IDF方法在文本分類、關鍵詞提取等自然語言處理領域廣泛應用，是一種經典的文本特徵表示方法[9]。

公式上而言，對於在某一特定文件 d_j 裡的詞語 t_i 來說，其詞頻 $tf_{i,j}$ 可表示為：

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2.1)$$

其中 $n_{i,j}$ 是 t_i 有在文件 d_j 中的出現次數，而右式分母 $\sum_k n_{k,j}$ 則是在文件 d_j 中包含的所有字詞(在此設為 k 個)出現次數之加總。

而某一特定詞語的IDF，可以由總文件數目除以包含該詞語之文件的數目，再將

結果取對數而得。其公式為：

$$\text{idf}_i = \log \frac{D}{\{j: t_i \in d_j\}} \quad (2.2)$$

其中D為母體的文件總數， $j: t_i \in d_j$ 則代表包含詞語 t_i 的文件數目如果該詞語不在母體中，就會導致被除數為零，因此一般情況下使用 $1 + \{j: t_i \in d_j\}$ 。

綜上兩式，可得一特定詞語於資料母體之TF-IDF為：

$$\text{tfidf}_{i,j} = \text{tf}_{i,j} \times \text{idf}_i \quad (2.3)$$

TF-IDF技術是一種常用於文本分析的方法，其主要目的是過濾掉常見的詞語，保留重要的詞語。TF-IDF的概念是，如果某個詞或短語在一篇文章中的詞頻（Term Frequency，TF）高，並且在其他文章中很少出現，則認為此詞或者短語具有很好的類別區分能力，適合用來分類。詞頻指的是某一個給定的詞語在該文件中出現的次數。IDF（Inverse Document Frequency，逆向文件頻率）則代表包含該給定詞語的文檔頻率。若包含詞條的文檔越少，IDF越大，則說明詞條具有較好的類別區分能力[9]。

例如，假設蒐集了1000篇新聞作為文本體，其中有一篇新聞提到中油調漲油價的問題。在此篇新聞中，「中油」一詞出現了10次，「油價」出現了1次，「的」出現了50次。以詞頻TF而言，「油」的TF為5，「油價」為4，「的」為50。然而，「的」這個字在本篇文章中並不重要，因此需要以IDF修正之。IDF方面，「中油」這個詞在1000篇新聞中個別出現了10次，「油價」個別出現了1次（即僅出現在一篇），「的」個別出現了1000次（即每篇都出現）。則IDF分別為「中油」 $\log(1000/10)=2$ 、「油價」

$\log(1000/1)=3$ 、「的」 $\log(1000/1000)=0$ 。個別的TFIDF為「中油」 $5*2=10$ 、「油價」 $4*3=12$ 、「的」 $50*0=0$ 。由此可推論，三者中「油價」在本篇新聞中最具重要性，「中油」次之，「的」則最不重要[9]。



參、研究方法

3.1 資料來源及彙整

本研究是由大林慈濟醫院的中醫師和中正大學資工所的中醫資訊化研究團隊聯合參與的。研究的主要目的是使用中正大學資工所從中醫典籍和期刊中收集的證候及其症狀集來作為辨證實驗的資料來源，以期能推動中醫醫療資訊化的發展。

研究所使用的資料來源共分為兩個部分，分別是資料集1和資料集2。資料集1是由五本中醫診斷學典籍所整理而來的，是一個多分類資料集。每個症狀詞組都對應到一個證候，例如「胸悶、自汗、心悸、脈力度弱、情緒神疲」等症狀詞組對應到「心氣虛」證候。此外，資料集1中還包含許多其他常見的證候，共計18種。

資料集2是由50個臨床案例報告整理而來的，是一個多標籤資料集。每個症狀詞組都可以對應到多個證候，例如「心悸、呼吸急促、全身浮腫、情緒憂鬱、畏寒喜溫」等症狀詞組對應到「心陽虛、腎陽虛、脾陽虛」這三種證候。

本研究的重要性在於，它為中醫融入現代科技提供了一個新的思路，同時也有望為中醫醫療資訊化的發展開闢一條新的道路。期待這項研究能為中醫學的現代化發展帶來新的發現。

3.2 資料前處理(Data Preprocessing)

中正大學資訊工程所根據中醫症狀詞庫和症狀知識本體，開發了一個中醫症狀標準化系統，用於收集症狀集和其相對應的一個或多個證候。透過將每筆資料的標準化症狀集轉換成適合神經網路處理的張量格式，進一步轉換成Word2Vec所需的語料庫(corpus)，建立症狀詞向量空間。

3.2.1 建立 Tokenizer 物件

Tokenizer類別負責進行文字序列的各種轉換，包括以下步驟：

- a. 從語料庫編製詞彙對照表，以便將症狀詞的序列轉換成症狀詞數字序列，並確保詞彙表的完整性和準確性。
- b. 建立Tokenizer物件，通常使用預設參數，但也可自訂參數，如num_words、filters、lower、split、char_level和oov_token等。
- c. 使用fit_on_texts()對需要轉換的症狀詞資料做詞彙字典，將全部會用到的詞編制為對照表(字典)，並儲存到Tokenizer物件中，以便後續使用。
- d. 利用已建立好的對照表，對任何文件進行症狀詞與數字的轉換，以利於進行機器學習或深度學習等相關任務。

3.2.2 症狀詞轉數字(texts_to_sequences)

當使用神經網路進行訓練時，資料集的格式必須轉換成可供神經網路使用的數字格式。這時會使用texts_to_sequences函數來將文本中的詞彙映射成對應的數字。為了達到這個目的，會建立一個對照表，將每個症狀詞都對應到一個唯一的數字。例如，對照字典可以包含 {'目眩': 10, '脈力度弱': 4, '心悸': 16, '面色白': 12, '全身無力': 19} 等鍵值對。

然後，可以使用這個對照表來將資料集中的症狀詞轉換成對應的數字。例如，將['心悸 面色白', '脈力度弱 全身無力 心悸 目眩'] 轉換成 [[16, 12], [4, 19, 16, 10]] 的數字格式。

這樣的轉換過程對神經網路的訓練非常重要，因為神經網路只能處理數字形式的資料。透過這種方式，可以確保神經網路可以正確地理解並處理資料，從而提高訓練的效率和準確性。

3.2.3 數字序列對齊(pad_sequences)

神經網路需要一個固定的輸入神經元數量，但是每個證候的症狀數量並不相同，因此需要使用pad_sequences方法來進行"截長補短"。使用這個方法時，可透過參數maxlen指定輸入的最大長度，而padding參數則指定在長度不足時要在序列的前面('pre')或後面('post')進行補0。如果序列的長度超過指定的最大長度，則truncating參數可指定要刪除序列的前面('pre')或後面('post')部分。

舉例來說，如果指定maxlen為3，而原始序列為[[16, 12], [4, 19, 16, 10]]，那麼pad_sequences會在第一個序列的後面補0，使變成[16, 12, 0]，而在第二個序列中，超出的部分會被刪除，使變成[4, 19, 16]，最終將兩個序列對齊為[[16, 12, 0], [4, 19, 16]]。這樣做的目的是為了讓神經網路可以處理不同長度的輸入，並更有效地進行訓練。

在神經網路的訓練過程中，針對不同的證候，症狀的數量是不一定相同的，這可能會對模型的訓練產生影響。因此，為了提高模型的準確性，在訓練過程中使用了不同的序列長度進行實驗，並使用不同的truncating參數進行截斷，以比較不同訓練方式下的準確率表現，從而選擇最適合的訓練方式進行後續的模型訓練和應用。

3.2.4 訓練資料及測試資料的產生

在深度學習中，通常會將資料集切割成三個部分：訓練集(Training Set)、驗證集(Validation Set)和測試集(Test Set)。訓練集是用來訓練模型的，驗證集則是用來評估模型的訓練效果以及調整模型的超參數，而測試集則是用來評估模型最終的泛化能力。

- a. 訓練集主要用於模型的擬合，即通過調整模型參數，使模型能夠更好地擬合訓練數據。
- b. 驗證集則是用來評估模型的訓練效果並根據及結果調整模型的超參數。如果模型在訓練集上表現得很好，但在驗證集上表現得不好，這就意味著模型過度擬合了訓練數據，需要調整超參數或採取其他措施。
- c. 測試集用來評估模型的最終性能，並且用於最終的決策。為了能夠真正評估模型的泛化能力，測試集不能用於訓練或驗證，因為這樣會影響最終評估結果的客觀性和可信度。

當我們需要進行資料分割以建立訓練集、驗證集和測試集時，我們面臨一個多標籤問題，即中醫辨證。為了解決這個問題，我們必須將由50個臨床案例報告整理而來的資料集分割成兩個部分，以供訓練、驗證和測試使用。此外，我們從五本中醫診斷學典籍中整理出的18種證候將作為單標籤資料集直接用於測試。在進行資料分割後，我們將對這三個資料集進行擴充，以增加訓練資料的多樣性，從而使模型能夠學習到更廣泛的特徵和模式。這樣的擴充有助於提高模型的泛化能力，使其能夠更好地應對各種不同的辨證情況。

3.2.4.1 病歷資料分成兩個資料集

病歷資料的切割步驟：

例如有5筆病歷的證候，要從中選取3筆作為訓練資料[A,B,C,D],[A,B],[B,C],[C],[A,B,C]，各個證後出現次數如下：A：3次、B：4次、C：4次、D：1次。

- a. 首先以次數少的證候優先選取，因為D證候最少所以要先選取，所以選擇了[A,B,C,D]這組病歷。選取後剩餘次數：A：2次、B：3次、C：3次、D：0次。
- b. 之後最少的是A證候，所以選取A證候，但因為A證候有兩筆[A,B], [A,B,C]，所以選擇證候數量少的筆數，因此選[A,B]這組病歷。選取後剩餘次數：A：1次、B：2次、C：3次、D：0次。
- c. 當病歷取到剩下的原來一半時就換下一個證候，因此再來就要選C證候，而C證候有三筆[B,C],[C],[A,B,C]，一樣選擇證候數量少的筆數，因此選[C]這組病歷。選取後剩餘次數：A：1次、B：2次、C：2次、D：0次。

依照以上模式選出來的3筆病歷做為訓練資料分別為[A,B,C,D],[A,B], [C]，剩餘2筆作為驗證資料分別為[B,C],[A,B,C]。

3.2.4.2 資料擴充(Data Augmentation)

為了有效地訓練類神經網路，需要豐富的訓練資料。然而，資料的數量常常有限，因此需要進行資料擴充的步驟，以幫助類神經網路實現更深入的學習。通過對現有資料進行一系列的轉換，生成額外的訓練樣本，增加訓練資料的多樣性，使模型能夠學習到更廣泛的特徵和模式。這對於預防過度擬合和提高模型的泛化能力非常重要。

將資料集1的18虛證，資料集2的50筆病例資料切割出來的35筆病歷及15筆病歷。對這三個資料集的資料進行擴充，每筆資料擴充1000次。以下是資料擴充的步驟：

- a. 匯入資料集
- b. 抓取症狀欄位並進行切割，將每個症狀分開，例如，將症狀資料 "自汗，面色白，心悸，胸悶，情緒神疲" 切割成 "自汗"、"面色白"、"心悸"、"胸悶"、

"情緒神疲"。

- c. 將切割後的症狀重新排序組合，產生新的一組資料。例如，將 "自汗"、"面色白"、"心悸"、"胸悶"、"情緒神疲" 重新排序組合成 "心悸，胸悶，情緒神疲，自汗，面色白" 的一組新症狀資料，以達到資料擴充，如圖3-1所示。

證候	症狀
心氣虛、心陰虛	自汗，面色白，心悸...
心氣虛、心陰虛	面色白，心悸，自汗...
心氣虛、心陰虛	心悸，自汗，面色白...
...	...
心氣虛、心陰虛	心悸，自汗，面色白...

圖 3-1 症狀擴充圖

- d. 將擴充完成的資料儲存起來，證候症狀資料重新排列，圖3-2所示。

證候	症狀
心氣虛、心陰虛	自汗，面色白，心悸...
心氣虛、心陰虛	面色白，心悸，自汗...
...	...
心陰虛、心血虛	面色紅，意識健忘，難眠...
心陰虛、心血虛	意識健忘，難眠，面色紅...



證候	症狀
心陰虛、心血虛	意識健忘，難眠，面色紅...
心氣虛、心陰虛	面色白，心悸，自汗...
...	...
心陰虛、心血虛	面色紅，意識健忘，難眠...
心氣虛、心陰虛	自汗，面色白，心悸...

圖 3-2 證候症狀重新排列

3.3 模型訓練

本研究的模型訓練架構如圖3- 3所示。

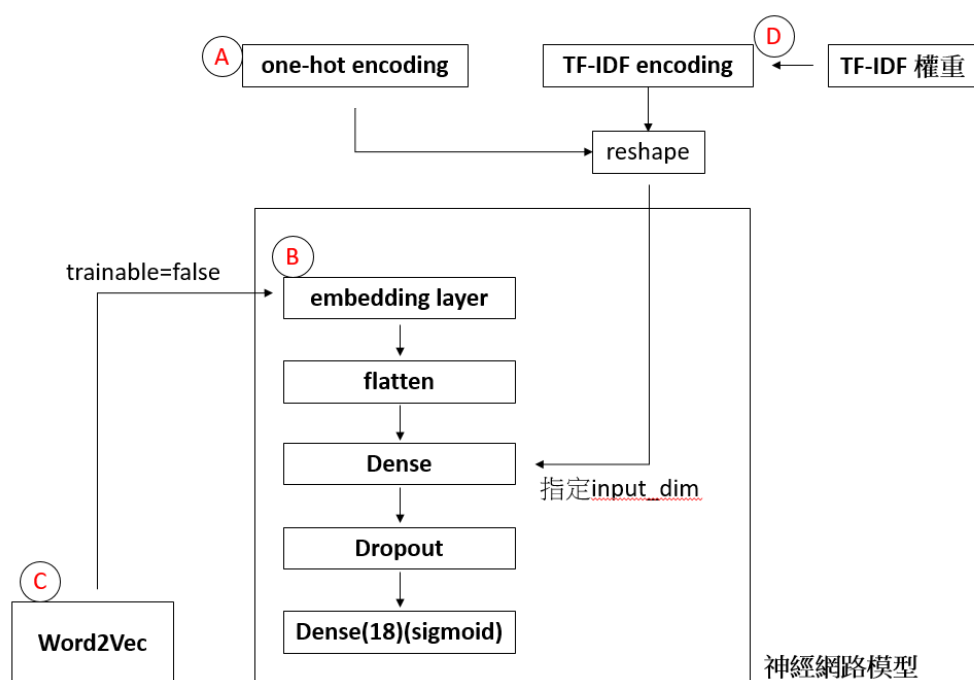


圖 3- 3 模型訓練架構圖

如圖3- 3，是四種模型的架構，分別為onehot encoding、embedding layer、Word2Vec、TF-IDFencoding，四種模型都有著不同的神經網路架構。onehot encoding有以下三層分別為輸入層、丟棄層、輸出層，這種方法將每個單詞表示為一個二進制向量；embedding layer則是有嵌入層作為神經網路的第一層，將資料中的詞彙映射到連續的向量空間，需要設定詞彙量的大小、嵌入向量的維度以及資料的長度；Word2Vec先對資料進行預訓練處理，然後再將其輸入嵌入層，通過學習上下文中單詞的分佈來生成詞向量；TF-IDFencoding則是計算TF-IDF的權重，之後再輸入至神經網路，評估單詞在資料中的重要性。

3.3.1 基於 onehot encoding 模型

在神經網路的訓練過程中，需要先將原始的數字序列轉換成神經網路可接受的onehot編碼格式，再利用reshape()函式將其轉換成1D向量的資料。這是因為神經網路

的輸入層僅能接受1D向量格式的資料，而不同證候的症狀數量不一，因此需要使用 `pad_sequences()` 方法進行序列的截長補短，再將其轉換成onehot編碼格式。

舉例來說，若一筆資料所產生的數字序列為 $[[1, 2], [0, 1]]$ ，經過轉換成onehot編碼後，其shape為 $(2, 2, 3)$ ，其中2表示資料的筆數，3表示2D的onehot編碼格式，類似於一個立體的 $3 \times 3 \times 3$ 魔術方塊的長、寬、高。為了將資料匯入到神經網路模型中，需要將2D向量的資料轉換成1D向量的資料，因此使用`reshape()`函式將其轉換成shape為 $(6,)$ 的1D向量。

值得注意的是，在使用`reshape()`函式之前，需要先確定onehot編碼的長度是否與神經網路輸入層的長度相等，如果不相等，則需要對其進行調整，以避免訓練時產生錯誤。這是因為在神經網路的訓練過程中，任何一個形狀不符合的資料都會導致錯誤的產生，進而影響訓練的準確性。因此，在使用神經網路進行訓練之前，需要對資料進行預處理，以確保資料的形狀符合神經網路的要求，進而提高訓練的效果。

神經網路模型中使用的onehot encoding結構可以分為以下三層：

- a. 第一層為輸入層，其中需要指定神經元個數，本研究分別選擇了32、64、128，多種不同的參數進行測試，同時也需要指定`input_dim`，以設定張量的維度。
- b. 第二層為丟棄層，其主要目的在於減少神經網路的過度適配，以避免過度擬合的問題。在本研究中，將丟棄層的參數設定為0.25，以便於簡化模型。
- c. 第三層為輸出層，神經元個數設定為18，啟動函數則是選擇了sigmoid。透過這樣的模型架構，可以有效地將onehot encoding的資料轉換成神經網路模型可以接受的格式，進而進行後續的訓練和預測。

onehot encoding 的模型訓練架構如圖 3-4 所示。

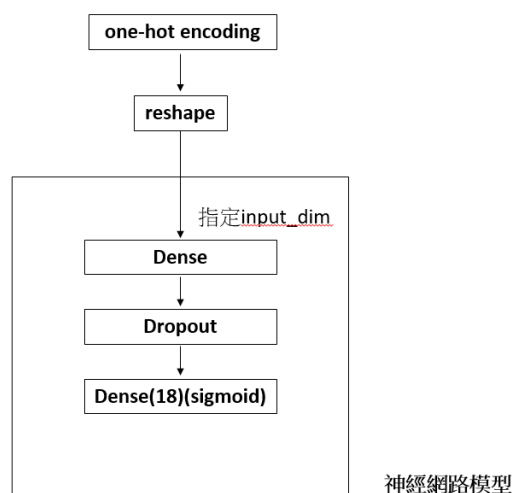


圖 3-4 基於 onehot encoding 模型

3.3.2 基於 embedding layer 訓練詞向量的模型

在本模型中，將使用嵌入層(embedding layer)作為神經網路的第一層，並設定詞彙量大小(症狀詞總數)、維度(詞向量長度)和文字長度(輸入症狀詞個數)。這個嵌入層會將編碼後的症狀詞映射到向量空間中，並轉換為指定維度的詞向量。例如，在的範例中，有兩組症狀詞組，分別為[胸悶，自汗，心悸]和[難眠，喉乾，自汗]。會建立一個詞彙表，包含{胸悶：1，自汗：2，心悸：3，難眠：4，喉乾：5}，然後使用tokenizer物件將這些詞組轉換為數字序列，分別得到[1，2，3]和[4，5，2]。接下來，使用嵌入層(embedding layer)對這些數字序列進行重新分配詞在空間中的位置，並轉換為指定的詞向量。由於嵌入層輸出的詞向量是一個2D陣列，需要使用Flatten()將其展平，以便在之後的神經網路層中使用。

本研究中，的症狀詞數量較多，共228個，每組症狀詞組的長度在10到33之間。將選取每組症狀詞的前10、15、20、25、30個症狀，並將它們轉換為數字序列。接下來，使用嵌入層(embedding layer)將這些數字序列轉換為詞向量，並設定不同的詞向量長度4、8、16和神經元個數分別為32、64、128，多種不同的參數進行測試，來觀

察這些超參數對模型訓練成效的影響。透過這些調整，希望能夠讓神經網路獲取更多的資訊，從而提高模型的效能。

嵌入層(embedding layer)是深度學習模型的一個重要組件，其可將原始文本中的詞彙轉換成向量形式，以便在神經網路中進行處理。嵌入層本身也是神經網路的一個層級，因此可以與深度學習模型一起訓練，以同步更新其權重。

在建立訓練資料用的神經網路模型時，通常會將原始文本中的詞彙傳入嵌入層，以便根據每個詞彙在向量空間中的位置映射成詞向量。接著，可以使用展平層(flatten)來將該詞向量轉換為一維向量，以便進一步進行神經網路的訓練。

此外，可以在嵌入層後方添加一個或多個神經網路層，以進一步提高模型的分類能力。透過這種方式，可以將症狀詞向量傳入神經網路層，讓模型更好地理解這些詞彙之間的關係，並從中學習如何更好地分類證候。因此，嵌入層和神經網路層的組合可以大幅提高深度學習模型的表現和準確率，對於中醫診斷等領域具有非常實用的應用價值。

embedding layer的模型訓練架構如圖3- 5所示。

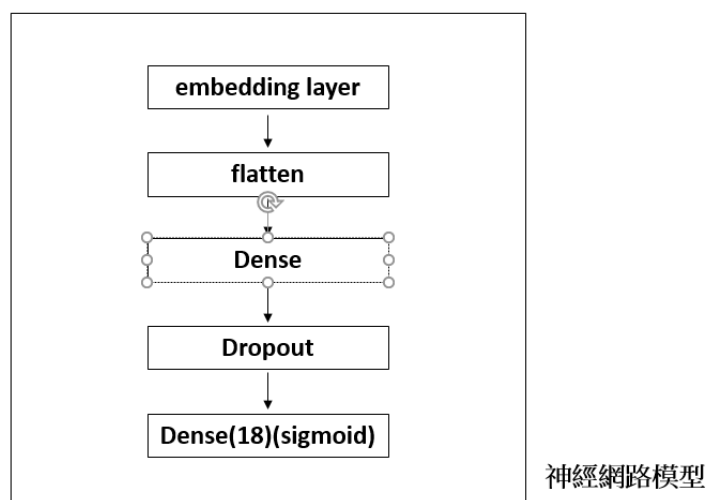


圖 3- 5 基於 embedding layer 模型

3.3.3 基於遷移學習 Word2Vec 詞向量的模型

建立通用性極高的中文詞向量確實是一個非常龐大和困難的任務，特別是對於中醫辨證這樣的領域，因為其中包含大量專業術語和特定背景知識。因此，需要先針對中醫辨證這個特定領域進行「預訓練(pre-training)」，建立一個中文詞向量空間，以將症狀集中的詞語映射到向量空間中的特定位置，進行向量化處理。

一旦建立了這個特定領域的中文詞向量空間，就可以將它遷移(transfer)到這個特定領域想分析的問題來建模。透過這種方式，可以利用先前學到的中文詞向量空間，幫助更有效地分析辨證問題，並且不必從頭開始建立新的中文詞向量空間。透過這樣的方法，可以驗證中文詞向量空間在辨證上的效果，並且嘗試提高分辨能力。這種方法可以提高建立中文詞向量的效率，同時也可以讓更好地理解 and 應用中醫辨證的知識。

Word2Vec詞向量模型訓練步驟如下所示：

- a. 在進行中醫辨證症狀集的詞向量建立時，需要透過Skip-grams()進行採樣以生成訓練資料。這樣做的目的是為了將相似的詞彙轉換成相似的向量，藉此反映詞彙之間的關係，例如顏色詞彙green和pink就應該在向量空間中靠近彼此。透過這樣的語言特性，可以找出相似的詞彙並為建立詞向量打下基礎。
- b. 接下來，需要建立Word2Vec模型，以訓練及檢驗模型。透過先前生成的訓練資料進行模型訓練，然後將Tokenizer物件和Word2Vec模型分別存檔，以便將來能夠將Word2Vec模型的嵌入層遷移到其他網路模型上。透過這樣的方式，可以預先訓練好詞向量，並在建立辨識模型時加以利用。
- c. 在建立辨識模型時，可以將預先訓練好的詞向量嵌入神經網路，再利用先前

建立的建模資料進行訓練，並進行模型效能比較。本研究將使用DNN神經網路結構探討此種模型在中醫辨證上的效能差異，將Word2Vec權重嵌入後，並將該層trainable屬性設為False，以避免損壞已經學習好的詞向量。透過這樣的設計，可以建立一個具有通用性極高的中文詞向量，以進行中醫辨證的研究。

Word2Vec的模型訓練架構如圖3- 6所示。

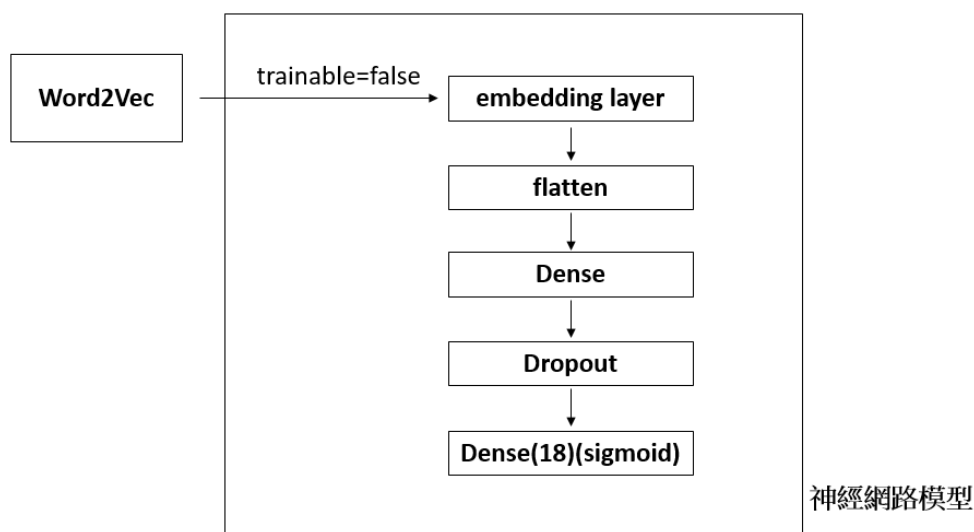


圖 3- 6 基於 Word2Vec 模型

3.3.4 基於 TF-IDFencoding 的模型

TF-IDF 模型的權重是通過將詞頻和逆文檔頻率相乘而得到的，用以表示詞在資料中的重要性。如果某詞頻高且逆文檔頻率低，那麼它將獲得較高的權重，因為這意味著該詞在資料中出現的頻率高，且對區分資料的能力較強。相反，如果某詞頻低且逆文檔頻率高，那麼它將獲得較低的權重，因為該詞對區分資料的能力較弱。透過 TF-IDF 權重，可以選擇最具代表性的詞，進而更好地理解和分析文本，如表3- 1所示。

表 3- 1TF-IDF 權重對照表

證候 \ 症狀	呼吸微弱	多夢	心悸	舌體顏色淡白色
心氣虛	0.391061	0.000000	0.368258	0.317072
心血虛	0.000000	0.426079	0.334780	0.288247
心陰虛	0.000000	0.312458	0.245505	0.000000
心陽虛	0.230036	0.000000	0.216622	0.186513

舉例來說，若一筆資料證後症狀資料序列為[[心氣虛]，[呼吸微弱，多夢]]，經過轉換成TF-IDF的權重編碼後，會轉換成[[心氣虛]，[0.391061，0.000000]]，其數值就是症狀的重要程度，之後再將其TF-IDF的權重編碼的資料匯入神經網路。

值得注意的是，需要先確定TF-IDF的權重編碼的長度是否與神經網路輸入層的長度相等，如果不相等，則需要對其進行調整，以避免訓練時產生錯誤。這是因為在神經網路的訓練過程中，任何一個形狀不符合的資料都會導致錯誤的產生，進而影響訓練的準確性。因此，在使用神經網路進行訓練之前，需要對資料進行預處理，以確保資料的形狀符合神經網路的要求，進而提高訓練的效果。

神經網路模型中使用的TF-IDF encoding結構可以分為以下三層：

- a. 第一層為輸入層，其中需要指定神經元個數，本研究分別選擇了32、64、128，多不同的參數進行測試，同時也需要指定input_dim，以設定張量的維度。
- b. 第二層為丟棄層，其主要目的在於減少神經網路的過度適配，以避免過度擬合的問題。在本研究中，將丟棄層的參數設定為0.25，以便於簡化模型。
- c. 第三層為輸出層，神經元個數設定為18，啟動函數則是選擇了sigmoid。透過這樣的模型架構，可以有效地將TF-IDF的權重編碼的資料轉換成神經網路模

型可以接受的格式，進而進行後續的訓練和預測。

TF-IDF的模型訓練架構如圖3- 7所示。

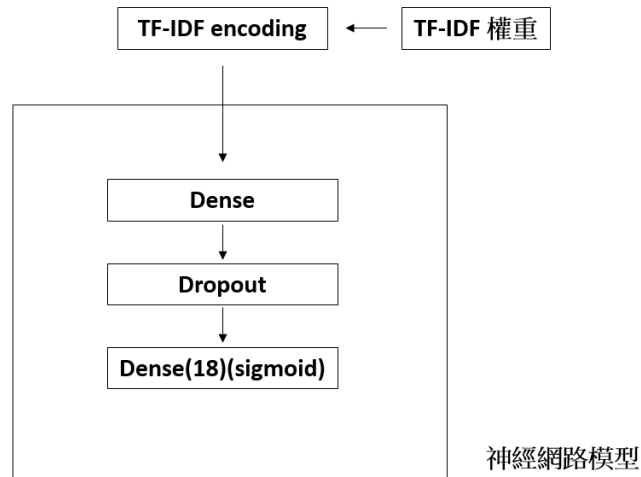


圖 3- 7 基於 TF-IDF encoding 的模型

3.4 多標籤分類評估方法

傳統的分類問題公式中，類別之間是互相排斥的。換言之，每個訓練樣本只能歸屬於一個類別，因此在特徵空間中的類別重疊時，就容易出現分類錯誤。然而，在某些任務中，一個數據點可能會同時屬於多個類別。為了解決這種情況，可以將傳統的分類問題公式轉化為多標籤分類框架。在這個框架中，假設每個標籤都代表不同的分類任務，且是一個伯努利(Bernoulli distribution)隨機變量，如圖3- 8所示。

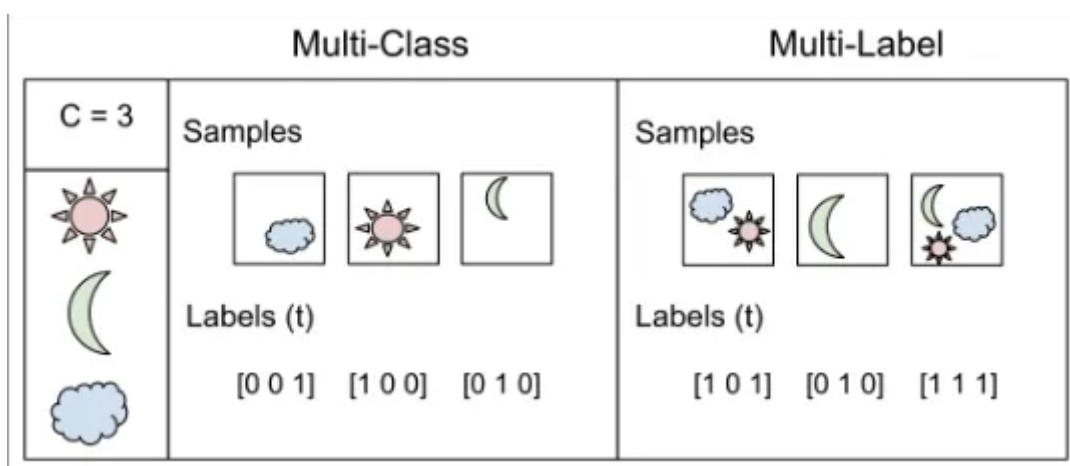


圖 3- 8 單標籤及多標籤

https://gombru.github.io/2018/05/23/cross_entropy_loss/

相對於基於示例(example based)(英文)的度量指標，基於標籤的度量指標是對每個標籤進行獨立評估，然後對所有標籤進行平均。因此，任何適用於二元分類的度量指標都可以用作基於標籤的度量指標。這些指標可以在單個類別標籤上進行計算，然後對所有類別進行平均，這稱為宏平均(macro average)(英文)。或者，可以在所有實例和所有類別標籤上進行全局計算，這稱為微平均(micro average)(英文)。並產生精確率(accuracy)、查準率(precision)、查全率(recall)，三個評價指標，本研究是使用微平均(micro average)[27]分別由以下公式計算如下：

- a. 精確率(accuracy)：在Keras中，準確性指標(accuracy metrics)根據模型的不同而有所不同。本實驗中使用了多標籤模型，accuracy指的是categorical_accuracy準確性指標。該指標主要適用於y_true為onehot標籤，而y_pred為向量的情況。

例如，有4個樣本，其y_true為 $[[0, 0, 1], [0, 1, 0], [0, 1, 0], [1, 0, 0]]$ ，y_pred為 $[[0.1, 0.6, 0.3], [0.2, 0.7, 0.1], [0.3, 0.6, 0.1], [0.9, 0, 0.1]]$ ，根據計算，其accuracy為75%。具體計算方法如下：將y_true轉為非onehot的形式，即y_true_new= $[2, 1, 1, 0]$ ，之後根據y_pred中的每個樣本預測的分數得到y_pred_new= $[1, 1, 1, 0]$ ，最後將y_true_new和y_pred_new對比計算後得到最終的categorical_accuracy=75%。

- b. Binary_accuracy: 是指被「正確預測的二元樣本數目占總二元樣本數的比例」，其中的「二元樣本」指的是正樣本和負樣本，並且Binary_accuracy可以設定門檻值。在二元分類問題中，Binary_accuracy僅計算預測正確的正樣本和預測正確的負樣本分別佔所有正樣本和負樣本的比例。因此，當正負樣本的比例不均時，Binary_accuracy可能會更能反映模型的性能。所以使用Binary_accuracy公式計算如下：

$$\text{Binary_accuracy} = \frac{1}{nk} \sum_{j=1}^k \sum_{i=1}^n [y_j^{(i)} = \hat{y}_j^{(i)}] \quad (3.1)$$

例如有三個樣本y_true為[[1, 0], [1, 1], [0, 1]]，y_pred為[[0.98, 0.2], [0.2, 0.7], [0.2, 0.2]]。其Binary_accuracy預設門檻值為0.5，所以計算出來是0.6666。

- c. 查準率(precision): 是指在所有實際為正例的樣本中，預測為正例的樣本所占的比例，也就是在所有真正的正例樣本中，有多少被預測為正例。查準率高表示模型預測為正例的樣本中，有很少的假正例。

$$\text{precision} = \frac{\sum_{j=1}^k \sum_{i=1}^n [y_j^{(i)} \wedge \hat{y}_j^{(i)}]}{\sum_{j=1}^k \sum_{i=1}^n \hat{y}_j^{(i)}} \quad (3.2)$$

例如有三個樣本y_true為[[1, 0], [1, 1], [0, 1]]，y_pred為[[0.98, 0.2], [0.2, 0.7], [0.2, 0.2]]。precision計算出來是1。

- d. 查全率(recall): 是指在所有實際為正例的樣本中，被預測為正例的樣本所占的比例，也就是在所有真正的正例樣本中，有多少被預測為正例。查全率高

表示模型能夠很好地捕捉到所有真正的正例樣本，但也可能預測出較多的假正例。

$$\text{recall} = \frac{\sum_{j=1}^k \sum_{i=1}^n [y_j^{(i)} \wedge \hat{y}_j^{(i)}]}{\sum_{j=1}^k \sum_{i=1}^n y_j^{(i)}} \quad (3.3)$$

例如有三個樣本 y_true 為 $[[1, 0], [1, 1], [0, 1]]$ ， y_pred 為 $[[0.98, 0.2], [0.2, 0.7], [0.2, 0.2]]$ 。recall 計算出來是 0.5。

本研究使用 accuracy、Binary_accuracy、precision 以 recall 及作為評估指標。



肆、實驗結果

本研究將探究四種不同的模型訓練方法，分別為onehot encoding、embedding layer、Word2Vec以及TF-IDF encoding。每種模型都會受到多種因素影響其網路模型的辨識效能，包括隱藏層層數、對齊長度、詞向量長度、神經網路結構、採樣窗口大小以及網路超參數等。本研究將透過一系列豐富的實驗，探討如何調整這些因素對於辨識效能的影響，進而進行分析比較，以檢驗模型的優劣及其相關原因。相信這些研究結果將對於該領域的進一步研究和實踐具有重要意義。

4.1 實驗設計

每個神經網路模型的辨識能力受到多種因素的影響，例如對齊長度、神經元個數、詞向量長度、神經網路結構、採樣窗口大小等等。因此，在本研究中，進行了一系列豐富的實驗，以探討如何調整這些因素對於辨識能力的影響，進而進行分析比較，以檢驗模型的優劣及其相關原因。

實驗分為以下七個階段：

- a. 第一階段：使用了onehot encoding模型，通過調整Maxlen（對齊長度）和神經元個數，探討這些參數對模型性能的影響。
- b. 第二階段：使用了TF-IDF encoding模型，通過調整Maxlen（對齊長度）和神經元個數，探討這些參數對模型性能的影響。
- c. 第三階段：基於第一步的結果，選取比較好的Maxlen（對齊長度）和神經元個數，然後進行了embedding layer模型實驗，通過調整Vector-Dim參數，探討其對模型性能的影響。
- d. 第四階段：基於前三步的結果，使用了Word2Vec模型進行實驗，通過調整

Window_size參數，探討其對模型性能的影響。

- e. 第五階段：測試原始資料集1（18虛證）及資料集2（50筆病歷），並比較了四種模型的準確率。
- f. 第六階段：原始資料擴充不同筆數，測試擴充50筆、100筆、250筆、500筆，對四種模型的準確度是否有影響。
- g. 第七階段：進行了雜訊的測試，將資料集1（18虛證）及資料集2（50筆病歷）的症狀隨機刪除，探討模型的預測結果。

透過這些豐富的實驗，可以深入了解每個因素對於辨識能力的影響，進而調整模型參數以提高模型的準確性。這些實驗也提供了寶貴的經驗，對於未來深度學習模型的研究和應用都具有重要的參考價值。

4.2 實驗一、基於 onehot encoding 模型訓練結果

當在使用 onehot encoding 模型時，Maxlen（即對齊長度）和神經元個數是非常重要的參數，它們對模型的效能有著關鍵性的影響。為了更深入地研究這些影響，調整了 Maxlen 的值，分別為 5、10、15、20、25 和 30，並且調整了神經元個數 32、64、128 和 256。通過這些調整，希望能夠更全面地觀察模型的變化和表現。

表 4- 1 基於 onehot encoding 模型神經元數量對訓練及驗證的影響

Maxlen (對齊長度)	Neurons_ _number (神經元 個數)	訓練		驗證	
		accuracy (%)	Binary_acc uracy (%)	val_ accuracy (%)	val_ Binary_acc uracy (%)
5	16	0.5599	0.9399	0.5545	0.9320
	32	0.5729	0.9641	0.5651	0.9617
	64	0.6380	0.9744	0.5991	0.9672
	128	0.7144	0.9864	0.6518	0.9812
	256	0.7139	0.9851	0.6516	0.9814

從表4- 1中可以看出，神經元個數為 128 時，訓練及驗證的準確率accuracy=0.7144、Binary_accuracy = 0.9864、val_ accuracy = 0.6518、val_ Binary_accuracy = 0.9812；比起神經元個數為 256，訓練及驗證的準確率accuracy=0.7139，Binary_accuracy = 0.9851、val_ accuracy = 0.6516、val_ Binary_accuracy = 0.9814。可以看出，兩種設定下的性能表現相當良好，但神經元個數為256時有些微下降。這表明神經元個數並非越多越好，存在一定的極限值。

表 4- 2 基於 onehot encoding 模型神經元數量對測試集的影響

Maxlen_ (對齊長 度)	Neurons_number_ (神經元個數)	測試													
		15 筆擴充 1000							18 筆擴充 1000						
		accuracy	Binary_accuracy					Recall	Precision	accuracy	Binary_accuracy				
5	16	0.5300	>0.5	>0.6	>0.7	>0.8	>0.9	0.7269	0.8732	0.5289	>0.5	>0.6	>0.7	>0.8	>0.9
	32	0.5572	0.9614	0.9559	0.9544	0.9489	0.9366	0.7490	0.8974	0.5564	0.9610	0.9551	0.9540	0.9481	0.9353
	64	0.5978	0.9646	0.9616	0.9582	0.9515	0.9417	0.7922	0.9052	0.5971	0.9627	0.9609	0.9563	0.9508	0.9409
	128	0.6505	0.9678	0.9652	0.9643	0.9604	0.9539	0.7959	0.9113	0.6488	0.9662	0.9635	0.9621	0.9587	0.9532
	256	0.6461	0.9665	0.9651	0.9643	0.9615	0.9599	0.7980	0.9130	0.6492	0.9720	0.9745	0.9737	0.9705	0.9691

從表4- 2中可以看出在神經元個數為128時，15筆病歷的準確度為accuracy=0.6505、Binary_accuracy >0.5 = 0.9678、Recall = 0.7959、Precision = 0.9113；18虛證的準確度為accuracy=0.6488、Binary_accuracy >0.5 = 0.9720、Recall = 0.7946、Precision = 0.9096，比起神經元個數為256時準確率來的高。

表 4- 3 基於 onehot encoding 模型對訓練及驗證的準確率

Maxlen (對齊長度)	Neurons_ _number (神經元 個數)	訓練		驗證	
		accuracy (%)	Binary_acc uracy (%)	val_ accuracy (%)	val_ Binary_acc uracy (%)
5	32	0.5729	0.9641	0.5651	0.9617
	64	0.6380	0.9744	0.5991	0.9672
	128	0.7144	0.9864	0.6518	0.9812
10	32	0.6574	0.9918	0.6237	0.9804
	64	0.7504	0.9957	0.7452	0.9812
	128	0.8298	0.9925	0.7908	0.9843
15	32	0.6581	0.9939	0.6338	0.9869
	64	0.7973	0.9963	0.7967	0.9885
	128	0.8829	0.9964	0.8610	0.9915
20	32	0.6665	0.9933	0.6317	0.9909
	64	0.8006	0.9936	0.7661	0.9939
	128	0.9262	0.9975	0.9340	0.9950
25	32	0.6717	0.9886	0.6809	0.9925
	64	0.8242	0.9876	0.7943	0.9934
	128	0.9461	0.9961	0.9654	0.9953

從表4- 3中的研究發現，在調整對齊長度和神經元個數方面，對於訓練和驗證模型的性能有著重大的影響。從對齊長度的角度來看，觀察到隨著神經元個數的增加，模型的訓練和驗證準確度也隨之提高。同樣地，當以神經元個數為基準時，增加對齊長度也會導致更高的準確率。

因此，特別設定了對齊長度為25，並將神經元個數設置為128，這個組合呈現出最佳的性能表現。在這種設置下，觀察到accuracy =0.9461、Binary_accuracy = 0.9961、val_accuracy = 0.9654、val_Binary_accuracy = 0.9951。這些數據表明，這個參數組合在訓練和驗證中展現出最佳的性能和表現。

表 4- 4 基於 onehot encoding 模型對測試及的準確率

Maxlen (對齊長度)	Neurons_number (神經元個數)	測試													
		15 筆病歷 1000							18 虛證 1000						
		accuracy	Binary_accuracy					Recall	Precision	accuracy	Binary_accuracy				
			>0.5	>0.6	>0.7	>0.8	>0.9				>0.5	>0.6	>0.7	>0.8	>0.9
5	32	0.5572	0.9646	0.9616	0.9582	0.9515	0.9417	0.7490	0.8974	0.5564	0.9627	0.9609	0.9563	0.9508	0.9409
	64	0.5978	0.9664	0.9652	0.9643	0.9604	0.9539	0.7922	0.9052	0.5971	0.9662	0.9635	0.9621	0.9587	0.9532
	128	0.6505	0.9678	0.9662	0.9659	0.9632	0.9617	0.7959	0.9113	0.6488	0.9720	0.9745	0.9737	0.9705	0.9601
10	32	0.6124	0.9816	0.9810	0.9799	0.9708	0.9701	0.8702	0.9784	0.6137	0.9830	0.9813	0.9802	0.9711	0.9750
	64	0.7087	0.9855	0.9833	0.9831	0.9830	0.9812	0.8972	0.9889	0.7103	0.9863	0.9843	0.9841	0.9831	0.9818
	128	0.7888	0.9880	0.9874	0.9867	0.9851	0.9827	0.9202	0.9910	0.7890	0.9880	0.9874	0.9874	0.9842	0.9839
15	32	0.6338	0.9945	0.9940	0.9919	0.9815	0.9794	0.9279	0.9887	0.6342	0.9945	0.9934	0.9934	0.9817	0.9805
	64	0.7823	0.9948	0.9946	0.9932	0.9923	0.9919	0.9592	0.9943	0.7820	0.9949	0.9947	0.9943	0.9940	0.9940
	128	0.8599	0.9964	0.9961	0.9956	0.9953	0.9950	0.9728	0.9955	0.8592	0.9964	0.9960	0.9955	0.9955	0.9953
20	32	0.6347	0.9953	0.9946	0.9941	0.9938	0.9935	0.9631	0.9962	0.6348	0.9964	0.9959	0.9949	0.9947	0.9933
	64	0.7637	0.9981	0.9972	0.9967	0.9962	0.9959	0.9855	0.9972	0.7637	0.9982	0.9975	0.9970	0.9965	0.9960
	128	0.9293	0.9993	0.9985	0.9979	0.9975	0.9973	0.9936	0.9976	0.9301	0.9996	0.9994	0.9990	0.9986	0.9975
25	32	0.6675	0.9978	0.9974	0.9968	0.9964	0.9947	0.9774	0.9978	0.6686	0.9975	0.9971	0.9969	0.9969	0.9961
	64	0.7916	0.9981	0.9979	0.9977	0.9974	0.9964	0.9948	0.9989	0.7925	0.9992	0.9990	0.9988	0.9988	0.9981
	128	0.9476	0.9986	0.9983	0.9980	0.9977	0.9971	0.9984	0.9995	0.9483	0.9996	0.9995	0.9994	0.9991	0.9989

從表4- 4中的研究表明，在對齊長度設為25、神經元個數設置為128的情況下，使用了15筆病歷數據進行評估。觀察到該組合呈現出非常優秀的性能表現。accuracy = 0.9476、Binary_accuracy > 0.5 = 0.9986、Recall = 0.9984、Precision = 0.9995；此外，在18個虛證的樣本上，同樣觀察到非常出色的結果，accuracy = 0.9483、Binary_accuracy > 0.5 = 0.9996、Recall = 0.9987、Precision = 0.9995。這些數據表明，在這個特定的參數下，onehot encoding模型在病歷數據的分類上取得了非常高的準確性和性能。

從以上onehot encoding的實驗可以得知，當對齊長度為25和神經元個數為128時，對資料理解的準確度相對也更深入，準確率的效果也更好。

4.3 實驗二、基於 TF-IDF encoding 模型訓練結果

當在使用 TF-IDF encoding 模型時，Maxlen（即對齊長度）和神經元個數是非常重要的參數，它們對模型的效能有著關鍵性的影響。為了更深入地研究這些影響，調整了 Maxlen 的值，分別為 5、10、15、20、25，並且調整了神經元個數 32、64、128。通過這些調整，希望能夠更全面地觀察模型的變化和表現。

表 4- 5 基於 TF-IDF encoding 模型對訓練及驗證的準確率

Maxlen (對齊長度)	Neurons_ _number (神經元 個數)	訓練		驗證	
		accuracy (%)	Binary_acc uracy (%)	val_ accuracy (%)	val_ Binary_acc uracy (%)
5	32	0.5754	0.9657	0.5691	0.9648
	64	0.6428	0.9803	0.6048	0.9692
	128	0.7207	0.9906	0.6588	0.9712
10	32	0.6587	0.9951	0.6257	0.9845
	64	0.7515	0.9992	0.7176	0.9872
	128	0.8318	0.9994	0.7936	0.9883
15	32	0.6593	0.9951	0.6362	0.9912
	64	0.7999	0.9976	0.8003	0.9938
	128	0.8858	0.9992	0.8639	0.9961
20	32	0.6688	0.9960	0.6330	0.9958
	64	0.8063	0.9972	0.7685	0.9974
	128	0.9350	0.9986	0.9380	0.9978
25	32	0.6776	0.9924	0.6857	0.9959
	64	0.8309	0.9930	0.8024	0.9976
	128	0.9475	0.9980	0.9668	0.9983

從表4- 5中研究發現，在調整對齊長度和神經元個數方面，對於訓練和驗證模型的性能有著重大的影響。從對齊長度的角度來看，觀察到隨著神經元個數的增加，模型的訓練和驗證準確度也隨之提高。同樣地，當以神經元個數為基準時，增加對齊長度也會導致更高的準確率。

因此，設定了對齊長度為25，並將神經元個數設置為128，這個組合呈現出最佳的性能表現。在這種設置下，觀察到accuracy = 0.9475、Binary_accuracy = 0.9980、val_accuracy = 0.9668、val_Binary_accuracy = 0.9983。這些數據表明，這個參數組合在訓練和驗證中展現出最佳的性能和表現。

表 4- 6 基於 TF-IDF encoding 模型對測試集的準確率

Maxlen- (對齊長 度)	Neurons_number- (神經元個數)	測試													
		15 筆擴充 1000							18 筆擴充 1000						
		accuracy	Binary_accuracy					Recall	Precision	accuracy	Binary_accuracy				
			>0.5	>0.6	>0.7	>0.8	>0.9				>0.5	>0.6	>0.7	>0.8	>0.9
5	32	0.5575	0.9654	0.9622	0.9590	0.9519	0.9419	0.7494	0.8982	0.5571	0.9631	0.9619	0.9571	0.9515	0.9411
	64	0.5982	0.9669	0.9654	0.9647	0.9607	0.9546	0.7929	0.9062	0.5978	0.9663	0.9641	0.9628	0.9589	0.9537
	128	0.6508	0.9682	0.9666	0.9662	0.9639	0.9620	0.7963	0.9122	0.6489	0.9722	0.9720	0.9715	0.9710	0.9698
10	32	0.6129	0.9825	0.9815	0.9807	0.9715	0.9705	0.8708	0.9785	0.6139	0.9840	0.9822	0.9812	0.9712	0.9698
	64	0.7093	0.9863	0.9839	0.9838	0.9835	0.9817	0.8977	0.9898	0.7106	0.9870	0.9844	0.9840	0.9834	0.9827
	128	0.7897	0.9885	0.9878	0.9869	0.9860	0.9832	0.9208	0.9917	0.7891	0.9884	0.9880	0.9877	0.9846	0.9840
15	32	0.6343	0.9954	0.9948	0.9923	0.9820	0.9804	0.9282	0.9895	0.6349	0.9951	0.9942	0.9940	0.9825	0.9807
	64	0.7826	0.9955	0.9954	0.9938	0.9932	0.9922	0.9593	0.9946	0.7829	0.9956	0.9954	0.9948	0.9945	0.9940
	128	0.8600	0.9969	0.9969	0.9959	0.9956	0.9953	0.9730	0.9957	0.8596	0.9966	0.9964	0.9959	0.9958	0.9955
20	32	0.6348	0.9960	0.9952	0.9945	0.9943	0.9939	0.9633	0.9968	0.6357	0.9972	0.9962	0.9951	0.9950	0.9936
	64	0.7645	0.9984	0.9977	0.9975	0.9963	0.9965	0.9861	0.9980	0.7642	0.9988	0.9980	0.9971	0.9967	0.9965
	128	0.9296	0.9998	0.9990	0.9985	0.9981	0.9980	0.9942	0.9986	0.9309	0.9995	0.9995	0.9993	0.9991	0.9979
25	32	0.6677	0.9980	0.9977	0.9975	0.9974	0.9954	0.9781	0.9987	0.6691	0.9981	0.9975	0.9970	0.9976	0.9966
	64	0.7924	0.9984	0.9986	0.9982	0.9982	0.9973	0.9950	0.9990	0.7932	0.9998	0.9993	0.9990	0.9990	0.9988
	128	0.9486	0.9988	0.9988	0.9986	0.9979	0.9979	0.9992	0.9995	0.9487	0.9999	0.9997	0.9995	0.9995	0.9994

從表4- 6中研究發現，在對齊長度設為25、神經元個數設置為128的情況下，使用了15筆病歷數據進行評估。觀察到該組合呈現出非常優秀的性能表現。accuracy=0.9486、Binary_accuracy >0.5 = 0.9988、Recall = 0.9992、Precision = 0.9995；此外，在18個虛證的資料上，同樣觀察到非常出色的結果，accuracy=0.9487、Binary_accuracy >0.5 = 0.9999、Recall = 0.9995、Precision = 0.9996。這些數據說明，在這個特定的參數下，onehot encoding模型在病歷數據的分類上取得了非常高的準確性和性能。

可以發現，TF-IDF encoding 模型在Maxlen(對齊長度)參數設置為25以及Neurons__number(神經元個數)參數設置為128，達到了最佳的準確率。透過TF-IDF權重，可以選擇出最具代表性的詞語，以更好地理解和分析文本內容。

4.4 實驗三、基於 embedding layer 模型訓練結果

使用 embedding layer模型進行實驗，設定Maxlen（對齊長度）參數設置為 20、25 和神經元個數參數設置為64、128，並調整了Vector-Dim 參數，包括 4、8 和 16。通過這些調整，希望能夠觀察模型在不同的 Vector-Dim 參數下的表現差異。

表 4- 7 基於 embedding layer 模型對訓練及驗證的準確率

Vector-Dim	Maxlen (對齊長度)	Neurons_ _number (神經元 個數)	訓練		驗證	
			accuracy (%)	Binary_acc uracy (%)	val_ accuracy (%)	val_ Binary_acc uracy (%)
4	20	64	0.8143	0.9981	0.7739	0.9980
		128	0.9422	0.9992	0.9456	0.9983
	25	64	0.8391	0.9938	0.8098	0.9985
		128	0.9542	0.9988	0.9730	0.9992
8	20	64	0.8199	0.9952	0.7836	0.9986
		128	0.9506	0.9995	0.9529	0.9993
	25	64	0.8467	0.9947	0.8174	0.9991
		128	0.9620	0.9997	0.9808	0.9997
16	20	64	0.8293	0.9994	0.7888	0.9991
		128	0.9576	1.0000	0.9614	1.0000
	25	64	0.8548	0.9955	0.8431	0.9999
		128	0.9710	1.0000	0.9880	1.0000

從表4- 7中，發現在embedding layer模型中，當Vector-Dim參數設置較大時，模型的訓練和驗證準確率越高。特別是在Vector-Dim參數設置為16，對齊長度參數設置為25，神經元個數設置為128的情況下，觀察到非常出色的性能。具體而言，在這個參數組合下，觀察到accuracy =0.9710、Binary_accuracy = 1.0000、val_accuracy = 0.9880、val_Binary_accuracy = 1.000。這些數據結果顯示，這個參數組合在訓練和驗證中展現出最佳的性能和表現。這表明著當Vector-Dim參數設置較大時，模型能夠更好地捕捉數據中的特徵並進行準確的分類。這些結果對於改進模型的準確性和性能，以及在相關應用領域中的潛在優勢具有重要意義。

表 4- 8 基於 embedding layer 模型對測試集的準確率

vector-Dim	Maxlen (對齊長度)	Neurons_number (神經元個數)	測試													
			15 筆擴充 1000							18 虛證擴充 1000						
			accuracy	Binary_accuracy					Recall	Precision	accuracy	Binary_accuracy				
				>0.5	>0.6	>0.7	>0.8	>0.9				>0.5	>0.6	>0.7	>0.8	>0.9
4	20	64	0.7705	0.9985	0.9980	0.9979	0.9965	0.9963	0.9867	0.9989	0.7732	0.9993	0.9985	0.9972	0.9972	0.9968
		128	0.9371	0.9987	0.9985	0.9983	0.9980	0.9975	0.9950	0.9991	0.9373	0.9996	0.9996	0.9995	0.9994	0.9980
	25	64	0.7991	0.9986	0.9986	0.9986	0.9983	0.9976	0.9957	0.9997	0.7993	0.9994	0.9994	0.9994	0.9993	0.9992
		128	0.9565	0.9990	0.9990	0.9988	0.9984	0.9983	0.9992	0.9992	0.9577	0.9997	0.9996	0.9995	0.9994	0.9997
8	20	64	0.7786	0.9988	0.9981	0.9980	0.9970	0.9969	0.9876	0.9998	0.7799	0.9997	0.9987	0.9975	0.9973	0.9971
		128	0.9458	0.9992	0.9990	0.9990	0.9985	0.9981	0.9958	0.9997	0.9469	0.9995	0.9995	0.9995	0.9994	0.9982
	25	64	0.8077	0.9990	0.9990	0.9988	0.9985	0.9979	0.9962	0.9995	0.8055	0.9995	0.9995	0.9995	0.9993	0.9993
		128	0.9647	0.9993	0.9992	0.9992	0.9988	0.9985	0.9993	0.9996	0.9671	0.9996	0.9996	0.9995	0.9993	0.9993
16	20	64	0.7856	0.9990	0.9985	0.9984	0.9974	0.9971	0.9886	0.9996	0.7862	0.9997	0.9989	0.9985	0.9983	0.9980
		128	0.9557	0.9994	0.9993	0.9990	0.9988	0.9986	0.9964	0.9996	0.9524	0.9998	0.9997	0.9997	0.9996	0.9989
	25	64	0.8168	0.9994	0.9993	0.9990	0.9990	0.9981	0.9970	0.9996	0.8153	0.9998	0.9998	0.9998	0.9997	0.9997
		128	0.9731	0.9995	0.9995	0.9994	0.9991	0.9989	0.9993	0.9996	0.9726	1.0000	1.0000	1.0000	0.9999	0.9998

從表4- 8中研究發現，在Vector-Dim參數設置為16、對齊長度設置為25且神經元個數設置為128的情況下，使用了15筆病歷數據進行評估。觀察到該組合呈現出非常優秀的性能表現。accuracy=0.9731、Binary_accuracy>0.5 = 0.9995、Recall = 0.9993、Precision = 0.9996；此外，在18個虛證的樣本上，同樣觀察到非常出色的結果，accuracy=0.9726、Binary_accuracy>0.5 = 1.000、Recall = 0.9995、Precision = 0.9998。這些數據說明，在這個特定的參數下，embedding layer模型在病歷數據的分類上取得了非常高的準確性和性能。

從實驗結果得知，在使用 embedding layer 模型時，神經網路的 Vector-Dim 參數越大，可以獲得更好的效果。這是因為較大的Vector-Dim 參數可以增加詞向量的維度，使模型能夠更好地理解單詞之間的語義關係，進而提高準確率。因此，在使用 embedding layer 模型時，調整 Vector-Dim 參數可以是一種有效的提升模型性能的方法。

4.5 實驗四、基於 Word2Vec 模型訓練結果

依據 Word2Vec 模型的原理，詞彙在向量空間的位置能夠反映詞彙之間的語意相似性。因此，本研究使用了 get_simword() 方法來找出與目標症狀詞意思相近的其他症狀詞。透過傳遞目標症狀詞作為參數，該方法得出了與目標症狀詞相似的其他症狀詞。經實驗得知，利用 Word2Vec 模型所找出的相似症狀詞，其相似度高達 85% 以上，表4- 9也呈現了相似詞的清單及其相似度分數，具有高度的可信度。

表 4-9 症狀詞相似度

Word2Vec	傳入目標症狀詞	得出相似症狀詞	相似度
2	目眩	頭暈	94%
	心悸	胸悶	91%
	四肢麻木	四肢拘急	98%
	舌體顏色淡白色	舌苔顏色白	87%
3	目眩	頭暈	89%
	心悸	心痛	94%
	四肢麻木	四肢拘急	95%
	舌體顏色淡白色	舌苔顏色白	86%
4	目眩	頭暈	98%
	心悸	胸悶	94%
	四肢麻木	四肢拘急	98%
	舌體顏色淡白色	舌苔顏色白	91%

使用 Word2Vec 模型進行實驗，設定了 Maxlen（對齊長度）參數設置為 20、25；神經元個數參數設置為 64、128；W2V_DIM 參數設置為 4、8、16，並調整了 Window_size 參數，包括 2、3、4。透過這些調整，希望觀察模型在不同 Window_size 參數下的表現差異。

表 4- 10 基於 Window_size 模型對訓練及驗證的準確率

window-size ^o	vector-Dim ^o	Maxlen ^o (對齊長度) ^o	Neurons_number ^o (神經元個數) ^o	訓練 ^o		驗證 ^o	
				accuracy (%) ^o	Binary_accuracy (%) ^o	val_accuracy (%) ^o	val_Binary_accuracy (%) ^o
2 ^o	4 ^o	20 ^o	64 ^o	0.8231 ^o	0.9982 ^o	0.7838 ^o	0.9980 ^o
			128 ^o	0.9512 ^o	0.9993 ^o	0.9539 ^o	0.9984 ^o
		25 ^o	64 ^o	0.8459 ^o	0.9940 ^o	0.8170 ^o	0.9986 ^o
			128 ^o	0.9619 ^o	0.9989 ^o	0.9823 ^o	0.9993 ^o
	8 ^o	20 ^o	64 ^o	0.8278 ^o	0.9952 ^o	0.7921 ^o	0.9988 ^o
			128 ^o	0.9565 ^o	0.9996 ^o	0.9611 ^o	0.9995 ^o
		25 ^o	64 ^o	0.8543 ^o	0.9949 ^o	0.8250 ^o	0.9993 ^o
			128 ^o	0.9717 ^o	0.9999 ^o	0.9887 ^o	0.9998 ^o
	16 ^o	20 ^o	64 ^o	0.8350 ^o	0.9996 ^o	0.7983 ^o	0.9991 ^o
			128 ^o	0.9627 ^o	0.9997 ^o	0.9695 ^o	0.9997 ^o
		25 ^o	64 ^o	0.8608 ^o	0.9956 ^o	0.8503 ^o	0.9990 ^o
			128 ^o	0.9769 ^o	0.9999 ^o	0.9972 ^o	0.9998 ^o
3 ^o	4 ^o	20 ^o	64 ^o	0.8289 ^o	0.9982 ^o	0.7913 ^o	0.9980 ^o
			128 ^o	0.9594 ^o	0.9992 ^o	0.9608 ^o	0.9984 ^o
		25 ^o	64 ^o	0.8545 ^o	0.9940 ^o	0.8241 ^o	0.9986 ^o
			128 ^o	0.9693 ^o	0.9991 ^o	0.9856 ^o	0.9993 ^o
	8 ^o	20 ^o	64 ^o	0.8331 ^o	0.9953 ^o	0.7968 ^o	0.9988 ^o
			128 ^o	0.9673 ^o	0.9997 ^o	0.9652 ^o	0.9994 ^o
		25 ^o	64 ^o	0.8580 ^o	0.9949 ^o	0.8343 ^o	0.9992 ^o
			128 ^o	0.9765 ^o	0.9998 ^o	0.9941 ^o	0.9999 ^o
	16 ^o	20 ^o	64 ^o	0.8463 ^o	0.9996 ^o	0.8049 ^o	0.9993 ^o
			128 ^o	0.9706 ^o	0.9998 ^o	0.9803 ^o	0.9998 ^o
		25 ^o	64 ^o	0.8720 ^o	0.9957 ^o	0.8619 ^o	0.9957 ^o
			128 ^o	0.9856 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o
4 ^o	4 ^o	20 ^o	64 ^o	0.8381 ^o	0.9984 ^o	0.8005 ^o	0.9981 ^o
			128 ^o	0.9648 ^o	0.9992 ^o	0.9675 ^o	0.9985 ^o
		25 ^o	64 ^o	0.8612 ^o	0.9942 ^o	0.8302 ^o	0.9988 ^o
			128 ^o	0.9780 ^o	0.9992 ^o	0.9934 ^o	0.9995 ^o
	8 ^o	20 ^o	64 ^o	0.8395 ^o	0.9954 ^o	0.8049 ^o	0.9988 ^o
			128 ^o	0.9724 ^o	0.9998 ^o	0.9728 ^o	0.9995 ^o
		25 ^o	64 ^o	0.8666 ^o	0.9951 ^o	0.8421 ^o	0.9992 ^o
			128 ^o	0.9848 ^o	0.9999 ^o	0.9995 ^o	0.9999 ^o
	16 ^o	20 ^o	64 ^o	0.8559 ^o	0.9997 ^o	0.8102 ^o	0.9994 ^o
			128 ^o	0.9769 ^o	0.9999 ^o	0.9854 ^o	0.9999 ^o
		25 ^o	64 ^o	0.8772 ^o	0.9958 ^o	0.8671 ^o	0.9996 ^o
			128 ^o	0.9921 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o

通過觀察表4- 10 中，發現在Window_size中，當神經網路的 Window_size參數設置為 4 時，模型的訓練和驗證準確率越高。特別是在Window_size參數設置為 4、Vector-Dim參數設置為16、對齊長度參數設置為25、神經元個數設置為128的情況下，觀察到非常出色的性能。具體而言，在這個參數組合下，觀察到accuracy =0.9921、Binary_accuracy = 0.9999、val_accuracy = 0.9999、val_Binary_accuracy = 0.9999。

表 4- 11 基於 Window_size 模型對測試集的準確率

window-size	vector-Dim	Maxlen (對齊長度)	Neurons_number (神經元個數)	測試 ^o															
				15 筆擴充 1000 ^o							18 虛徵擴充 1000 ^o								
				accuracy	Binary_accuracy					Recall	Precision	accuracy	Binary_accuracy					Recall	Precision
			>0.5 ^o	>0.6 ^o	>0.7 ^o	>0.8 ^o	>0.9 ^o				>0.5 ^o	>0.6 ^o	>0.7 ^o	>0.8 ^o	>0.9 ^o				
2 ^o	4 ^o	20 ^o	64 ^o	0.7788 ^o	0.9990 ^o	0.9984 ^o	0.9981 ^o	0.9968 ^o	0.9965 ^o	0.9876 ^o	0.9995 ^o	0.7786 ^o	0.9995 ^o	0.9989 ^o	0.9987 ^o	0.9985 ^o	0.9984 ^o	0.9886 ^o	0.9996 ^o
			128 ^o	0.9452 ^o	0.9991 ^o	0.9987 ^o	0.9985 ^o	0.9985 ^o	0.9978 ^o	0.9956 ^o	0.9995 ^o	0.9443 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9995 ^o	0.9994 ^o	0.9970 ^o	0.9996 ^o
		25 ^o	64 ^o	0.8072 ^o	0.9991 ^o	0.9991 ^o	0.9988 ^o	0.9984 ^o	0.9980 ^o	0.9966 ^o	0.9995 ^o	0.8075 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9996 ^o	0.9995 ^o	0.9976 ^o	0.9996 ^o
			128 ^o	0.9644 ^o	0.9994 ^o	0.9992 ^o	0.9989 ^o	0.9988 ^o	0.9984 ^o	0.9997 ^o	0.9996 ^o	0.9652 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9996 ^o	0.9996 ^o	0.9986 ^o	0.9996 ^o
	8 ^o	20 ^o	64 ^o	0.7844 ^o	0.9990 ^o	0.9985 ^o	0.9983 ^o	0.9975 ^o	0.9973 ^o	0.9884 ^o	0.9995 ^o	0.7861 ^o	0.9999 ^o	0.9998 ^o	0.9995 ^o	0.9995 ^o	0.9994 ^o	0.9891 ^o	0.9996 ^o
			128 ^o	0.9554 ^o	0.9995 ^o	0.9995 ^o	0.9993 ^o	0.9988 ^o	0.9984 ^o	0.9966 ^o	0.9996 ^o	0.9558 ^o	0.9999 ^o	0.9998 ^o	0.9997 ^o	0.9997 ^o	0.9995 ^o	0.9986 ^o	0.9996 ^o
		25 ^o	64 ^o	0.8151 ^o	0.9994 ^o	0.9992 ^o	0.9991 ^o	0.9989 ^o	0.9983 ^o	0.9967 ^o	0.9995 ^o	0.8135 ^o	0.9999 ^o	0.9998 ^o	0.9997 ^o	0.9997 ^o	0.9995 ^o	0.9985 ^o	0.9996 ^o
			128 ^o	0.9735 ^o	0.9997 ^o	0.9995 ^o	0.9994 ^o	0.9991 ^o	0.9988 ^o	0.9997 ^o	0.9996 ^o	0.9764 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9996 ^o	0.9996 ^o	0.9996 ^o
	16 ^o	20 ^o	64 ^o	0.7909 ^o	0.9991 ^o	0.9988 ^o	0.9986 ^o	0.9977 ^o	0.9975 ^o	0.9891 ^o	0.9995 ^o	0.7923 ^o	0.9999 ^o	0.9998 ^o	0.9996 ^o	0.9995 ^o	0.9995 ^o	0.9898 ^o	0.9996 ^o
			128 ^o	0.9638 ^o	0.9998 ^o	0.9996 ^o	0.9992 ^o	0.9991 ^o	0.9989 ^o	0.9971 ^o	0.9996 ^o	0.9590 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9996 ^o	0.9986 ^o	0.9997 ^o
		25 ^o	64 ^o	0.8244 ^o	0.9997 ^o	0.9997 ^o	0.9992 ^o	0.9992 ^o	0.9984 ^o	0.9979 ^o	0.9995 ^o	0.8224 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9997 ^o	0.9988 ^o	0.9996 ^o
			128 ^o	0.9796 ^o	0.9998 ^o	0.9996 ^o	0.9996 ^o	0.9995 ^o	0.9993 ^o	0.9997 ^o	0.9996 ^o	0.9811 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9997 ^o	0.9998 ^o	0.9997 ^o
4 ^o	20 ^o	64 ^o	0.7871 ^o	0.9994 ^o	0.9989 ^o	0.9984 ^o	0.9971 ^o	0.9968 ^o	0.9884 ^o	0.9997 ^o	0.7841 ^o	0.9998 ^o	0.9992 ^o	0.9990 ^o	0.9990 ^o	0.9988 ^o	0.9894 ^o	0.9997 ^o	
		128 ^o	0.9533 ^o	0.9994 ^o	0.9989 ^o	0.9988 ^o	0.9988 ^o	0.9981 ^o	0.9962 ^o	0.9997 ^o	0.9513 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9995 ^o	0.9977 ^o	0.9997 ^o	
	25 ^o	64 ^o	0.8154 ^o	0.9995 ^o	0.9995 ^o	0.9989 ^o	0.9985 ^o	0.9983 ^o	0.9974 ^o	0.9997 ^o	0.8157 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9996 ^o	0.9996 ^o	0.9984 ^o	0.9997 ^o	
		128 ^o	0.9722 ^o	0.9998 ^o	0.9995 ^o	0.9991 ^o	0.9991 ^o	0.9986 ^o	0.9997 ^o	0.9998 ^o	0.9727 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9997 ^o	0.9993 ^o	0.9998 ^o	
3 ^o	8 ^o	20 ^o	64 ^o	0.7901 ^o	0.9991 ^o	0.9989 ^o	0.9985 ^o	0.9979 ^o	0.9977 ^o	0.9891 ^o	0.9997 ^o	0.7924 ^o	0.9998 ^o	0.9997 ^o	0.9996 ^o	0.9996 ^o	0.9992 ^o	0.9899 ^o	0.9998 ^o
			128 ^o	0.9650 ^o	0.9998 ^o	0.9998 ^o	0.9996 ^o	0.9992 ^o	0.9988 ^o	0.9975 ^o	0.9998 ^o	0.9648 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9993 ^o	0.9998 ^o
		25 ^o	64 ^o	0.8224 ^o	0.9998 ^o	0.9996 ^o	0.9994 ^o	0.9994 ^o	0.9987 ^o	0.9972 ^o	0.9997 ^o	0.8216 ^o	0.9999 ^o	0.9999 ^o	0.9997 ^o	0.9997 ^o	0.9996 ^o	0.9995 ^o	0.9998 ^o
			128 ^o	0.9822 ^o	0.9999 ^o	0.9998 ^o	0.9996 ^o	0.9993 ^o	0.9990 ^o	0.9998 ^o	0.9998 ^o	0.9857 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9997 ^o	0.9998 ^o
	16 ^o	20 ^o	64 ^o	0.7962 ^o	0.9993 ^o	0.9991 ^o	0.9988 ^o	0.9979 ^o	0.9978 ^o	0.9896 ^o	0.9997 ^o	0.7984 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9997 ^o	0.9998 ^o	0.9905 ^o	0.9998 ^o
			128 ^o	0.9720 ^o	0.9998 ^o	0.9998 ^o	0.9994 ^o	0.9993 ^o	0.9992 ^o	0.9978 ^o	0.9998 ^o	0.9657 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9993 ^o	0.9998 ^o
		25 ^o	64 ^o	0.8320 ^o	0.9999 ^o	0.9999 ^o	0.9995 ^o	0.9995 ^o	0.9988 ^o	0.9988 ^o	0.9997 ^o	0.8294 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9997 ^o	0.9998 ^o	0.9995 ^o	0.9998 ^o
			128 ^o	0.9861 ^o	0.9999 ^o	0.9997 ^o	0.9996 ^o	0.9996 ^o	0.9995 ^o	0.9998 ^o	0.9998 ^o	0.9896 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9998 ^o
	4 ^o	20 ^o	64 ^o	0.7954 ^o	0.9999 ^o	0.9993 ^o	0.9986 ^o	0.9975 ^o	0.9970 ^o	0.9893 ^o	0.9998 ^o	0.7895 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9902 ^o	0.9998 ^o
			128 ^o	0.9614 ^o	0.9998 ^o	0.9991 ^o	0.9990 ^o	0.9990 ^o	0.9984 ^o	0.9968 ^o	0.9998 ^o	0.9583 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9985 ^o	0.9998 ^o
		25 ^o	64 ^o	0.8235 ^o	0.9999 ^o	0.9999 ^o	0.9995 ^o	0.9986 ^o	0.9987 ^o	0.9983 ^o	0.9999 ^o	0.8239 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9997 ^o	0.9992 ^o	0.9998 ^o
			128 ^o	0.9801 ^o	0.9999 ^o	0.9997 ^o	0.9992 ^o	0.9991 ^o	0.9990 ^o	0.9993 ^o	0.9999 ^o	0.9802 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9999 ^o	0.9999 ^o
8 ^o	20 ^o	64 ^o	0.7959 ^o	0.9993 ^o	0.9993 ^o	0.9988 ^o	0.9984 ^o	0.9981 ^o	0.9899 ^o	0.9999 ^o	0.7986 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9997 ^o	0.9906 ^o	0.9998 ^o	
		128 ^o	0.9746 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9995 ^o	0.9991 ^o	0.9983 ^o	0.9999 ^o	0.9737 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9999 ^o	0.9999 ^o	
	25 ^o	64 ^o	0.8298 ^o	0.9999 ^o	0.9995 ^o	0.9995 ^o	0.9995 ^o	0.9993 ^o	0.9977 ^o	0.9999 ^o	0.8236 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9999 ^o	0.9998 ^o	
		128 ^o	0.9910 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9996 ^o	0.9993 ^o	0.9999 ^o	0.9999 ^o	0.9949 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9999 ^o	0.9999 ^o	
16 ^o	20 ^o	64 ^o	0.8014 ^o	0.9994 ^o	0.9994 ^o	0.9990 ^o	0.9989 ^o	0.9987 ^o	0.9901 ^o	0.9999 ^o	0.8045 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9911 ^o	0.9998 ^o	
		128 ^o	0.9801 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9996 ^o	0.9995 ^o	0.9986 ^o	0.9999 ^o	0.9723 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	
	25 ^o	64 ^o	0.8396 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9993 ^o	0.9996 ^o	0.9999 ^o	0.8365 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	
		128 ^o	0.9925 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9998 ^o	0.9998 ^o	0.9999 ^o	0.9999 ^o	0.9900 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	

從表4- 11中，發現15筆病例在Window_size參數設置為4、Vector-Dim參數設置為4、對齊長度參數設置為25、神經元個數設置為128時，accuracy、Binary_accuracy>0.5、Recall、Precision就已經達到0.9999的準確率了；18虛證在Window_size參數設置為4、Vector-Dim參數設置為4、對齊長度參數設置為25、神經元個數設置為128時，accuracy、Binary_accuracy>0.5、Recall、Precision就已經達到0.9999的準確率了，但是Binary_accuracy>0.5在Window_size參數設置為2、Vector-Dim參數設置為8、對齊長度參數設置為20、神經元個數設置為64就已經達到0.9999的準確率了。

從實驗結果得知，在使用 Word2Vec 模型時，隨著window_size採樣窗口大小的增加，模型的效果會逐漸提升。這意味著先前Word2Vec的詞向量產生了意義相近的詞之間的關聯性，並成功地發揮了其效用。

4.6 四種模型最佳參數比較

根據前面四種不同模型的實驗，分別為onehot、TF-IDF、embedding layer、Word2Vec四種模型，從中個別尋找最佳的準確率，來分別比較模型差異。

表 4- 12 在 15 筆病歷中四種模型最佳參數

模型 ^o	window-size ^o	vector-Dim ^o	Maxlen ^o (對齊長度) ^o	Neurons_number ^o (神經元個數) ^o	測試 ^o				Total params ^o	Trainable params ^o	Non-trainable params ^o
					15 筆擴充 1000 ^o						
					accuracy ^o	Binary_accuracy ^o >0.5 ^o	Recall ^o	Precision ^o			
One hot ^o	→ ^o	→ ^o	25 ^o	128 ^o	0.9476 ^o	0.9986 ^o	0.9984 ^o	0.9995 ^o	674,220 ^o	674,220 ^o	0 ^o
TF-IDF ^o	→ ^o	→ ^o	25 ^o	128 ^o	0.9486 ^o	0.9988 ^o	0.9992 ^o	0.9995 ^o	3,328 ^o	3,328 ^o	0 ^o
embedding layer ^o	→ ^o	16 ^o	25 ^o	128 ^o	0.9731 ^o	0.9995 ^o	0.9993 ^o	0.9996 ^o	58,450 ^o	58,450 ^o	0 ^o
Word2Vec ^o	4 ^o	4 ^o	25 ^o	128 ^o	0.9801 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	14,257 ^o	10,929 ^o	3,328 ^o
	4 ^o	8 ^o	25 ^o	128 ^o	0.9910 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	28,257 ^o	24,929 ^o	3,328 ^o
	4 ^o	16 ^o	25 ^o	128 ^o	0.9925 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	56,257 ^o	52,929 ^o	3,328 ^o

從表4- 12中，發現onehot模型在對齊長度設置為 25、神經元個數設置為128時準確率是最好的；TF-IDF模型在對齊長度設置為 25、神經元個數設置為128時準確率是最好的；embedding layer模型在vector-Dim設置為 16、對齊長度設置為 25、神經元個數設置為128時準確率是最好的；Word2Vec模型分別在三種不同參數下準確率都達到最好的，分別為第一種window-size設置為 4 、vector-Dim設置為 4、對齊長度設置為 25、神經元個數設置為 128；第二種window-size設置為 4 、vector-Dim設置為 8、對齊長度設置為 25、神經元個數設置為 128；第三種window-size設置為 4 、vector-Dim設置為 16、對齊長度設置為 25、神經元個數設置為 128。

表 4- 13 在 18 筆虛證中四種模型最佳參數

模型 ^o	window-size ^o	vector-Dim ^o	Maxlen ^o (對齊長度) ^o	Neurons_number ^o (神經元個數) ^o	測試 ^o				Total params ^o	Trainable params ^o	Non-trainable params ^o
					18 虛證擴充 1000 ^o						
					accuracy ^o	Binary_accuracy ^o >0.5 ^o	Recall ^o	Precision ^o			
One hot ^o	-- ^o	-- ^o	25 ^o	128 ^o	0.9483 ^o	0.9996 ^o	0.9987 ^o	0.9995 ^o	674,220 ^o	674,220 ^o	0 ^o
TF-IDF ^o	-- ^o	-- ^o	25 ^o	128 ^o	0.9487 ^o	0.9999 ^o	0.9995 ^o	0.9996 ^o	455,798 ^o	455,798 ^o	0 ^o
embedding layer ^o	-- ^o	16 ^o	25 ^o	128 ^o	0.9726 ^o	1.0000 ^o	0.9995 ^o	0.9998 ^o	3,328 ^o	3,328 ^o	0 ^o
Word2Vec ^o	4 ^o	4 ^o	25 ^o	128 ^o	0.9802 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	14,257 ^o	10,929 ^o	3,328 ^o
	4 ^o	8 ^o	20 ^o	128 ^o	0.9737 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	23,137 ^o	20,449 ^o	2,688 ^o
	4 ^o	8 ^o	25 ^o	128 ^o	0.9949 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	28,257 ^o	24,929 ^o	3,328 ^o
	4 ^o	16 ^o	20 ^o	128 ^o	0.9723 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	46,017 ^o	43,329 ^o	2,688 ^o
	4 ^o	16 ^o	25 ^o	64 ^o	0.8365 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	30,529 ^o	28,865A ^o	1,664 ^o
	4 ^o	16 ^o	25 ^o	128 ^o	0.9980 ^o	0.9999 ^o	0.9999 ^o	0.9999 ^o	56,257 ^o	52,929 ^o	3,328 ^o

從表4- 13中，發現onehot模型在對齊長度設置為 25、神經元個數設置為 128時準確率是最好的；TF-IDF模型在對齊長度設置為 25、神經元個數設置為 128時準確率是最好的；embedding layer模型在vector-Dim設置為 16、對齊長度設置為 25、神經元個數設置為 128時準確率是最好的；Word2Vec模型分別在五種不同參數下準確率都達到最好的，分別為第一種window-size設置為 4、vector-Dim設置為 4、對齊長度設置為 25、神經元個數設置為 128；第二種window-size設置為 4、vector-Dim設置為 8、對齊長度設置為 20、神經元個數設置為 128；第三種window-size設置為 4、vector-Dim設置為 16、對齊長度設置為 20、神經元個數設置為 128；第四種window-size設置為 4、vector-Dim設置為 16、對齊長度設置為 25、神經元個數設置為 64；第五種window-size設置為 4、vector-Dim設置為 16、對齊長度設置為 25、神經元個數設置為 128。

從這兩個不同的資料集的實驗結果來看。首先是18筆虛證資料集，這是一個單標籤的資料集，表示每個樣本僅屬於一個類別。發現在使用Word2Vec模型在window-size設置為 4、vector-Dim設置為 4、對齊長度設置為 25、神經元個數設置為 128的情況下，該模型已經能夠取得很好的預測結果。

相比之下，15筆病例資料集則是一個多標籤的資料集，表示每個樣本可能同時屬於多個類別。發現，在這種情況下，Word2Vec模型需要更多的調整才能獲得良好的

準確率。需要將window-size設置為 4 、vector-Dim設置為 4、對齊長度設置為 25、神經元個數設置為 128，才能獲得相對較好的預測結果。

這些結果暗示了單標籤和多標籤問題在使用Word2Vec模型時的差異。單標籤問題可能更容易解決，並在較小的參數設置下達到良好的預測結果。而多標籤問題則需要更多的參數調整才能獲得相似的性能。這些發現為後續的研究和模型優化提供了有價值的指引。

4.7 實驗五、四種模型在原始資料集的準確率

進行了四種模型的實驗，以探討原始資料在準確率方面的表現。在這些實驗中，固定了以下參數：Maxlen（對齊長度）參數設置為 25、神經元個數參數設置為128、W2V_DIM參數設置為16，Window_size參數設置為4。透過這些實驗，的目標是評估這四種模型在處理原始資料時的準確率。通過固定這些參數，能夠在相同條件下比較不同模型的性能。這些實驗結果將提供有關這四種模型在處理原始資料時的準確率的資訊。這有助於了解模型的效能，並選擇最適合的模型來處理原始資料。同時，這些結果也為後續研究和應用提供了有價值的參考。

表 4- 14 原始資料 18 虛證及 50 病歷的準確率

模型	測試							
	18 虛證				50 筆病例			
	accuracy	Binary accuracy	Recall	Precision	accuracy	Binary accuracy	Recall	Precision
		>0.5				>0.5		
One hot	0.9420	0.9936	0.9901	0.9931	0.9388	0.9893	0.9886	0.9926
TF-IDF	0.9389	0.9917	0.9921	0.9929	0.9415	0.9928	0.9921	0.9924
embedding layer	0.9667	0.9950	0.9917	0.9940	0.9680	0.9926	0.9937	0.9930
Word2Ve	0.9910	0.9919	0.9903	0.9930	0.9854	0.9911	0.9947	0.9905

根據表4- 14的結果中，根據對原始資料中18筆虛證和50筆病例的四種模型（onehot、Embedding layer、Word2Vec、TF-IDF）的評估結果，可以得出以下結論：

在18筆虛證資料中，所有模型在各項指標（accuracy、Binary_accuracy、Recall和Precision）上表現優異。這表明這些模型的訓練已經達到相當高的水準。這些指標的優秀表現意味著模型能夠準確識別出虛證資料，並具有出色的全面性和精確性。這些結果顯示出模型已成功地學習並捕捉到資料中的關鍵特徵，並在預測虛證方面具有可靠性和有效性。

對於50筆病例資料，四種模型在所有指標上也呈現出優秀的表現。這再次暗示模型訓練已經達到了相當高的水準。這些指標的優秀表現意味著模型能夠準確識別出50筆病例資料，並具有出色的全面性和精確性。這些結果同樣顯示出模型已成功地學習並捕捉到資料中的關鍵特徵，並在預測病例方面具有可靠性和有效性。

總結而言，根據這些評估結果，可以得出結論：所使用的四種模型在虛證和病例預測上都呈現出出色的表現，證明了這些模型的訓練已經非常成功。這些模型能夠準確地辨識並預測資料中的虛證和病例，並具有可靠性、全面性和精確性。這對於該領域的應用具有重要的意義，並為進一步的研究和應用提供了堅實的基礎。

4.8 實驗六、原始資料擴充不同筆數對四種模型準確率的影響

基於先前的實驗結果，當資料筆數擴充為1000筆時，使用參數window-size設置為4、vector-Dim設置為16、對齊長度設置為25、神經元個數設置為128時，四種模型均取得了極高的準確率。現在資料筆數擴充對這些模型的準確率是否有影響，將分別測試資料擴充為50、100、250、500筆數，並觀察對模型的影響。這個實驗設定以評估資料筆數對於模型性能的影響。通過比較不同筆數擴充的實驗結果，可以了解模型在不同資料量下的表現差異。

將使用相同的參數設置，包括window-size為4、vector-Dim為16、對齊長度為25

和神經元個數為128，並進行四種模型的測試。通過觀察模型在50、100、250、500筆數擴充下的準確率變化，探討資料量對於模型性能的影響。

表 4- 15 不同擴充筆數的四種模型在訓練及驗證的準確率

擴充筆數	模型	window-size	vector-Dim	Maxlen (對齊長度)	Neurons_number (神經元個數)	訓練		驗證	
						accuracy (%)	Binary_accuracy (%)	val_accuracy (%)	val_Binary_accuracy (%)
50	One hot	-	-	25	128	0.3873	0.7457	0.4013	0.7775
	TF-IDF	-	-	25	128	0.4476	0.7426	0.4543	0.7712
	embedding layer	-	16	25	128	0.6994	0.8143	0.6838	0.8350
	Word2Vec	4	16	25	128	0.7756	0.8509	0.8131	0.8431
100	One hot	-	-	25	128	0.4690	0.8809	0.5539	0.8684
	TF-IDF	-	-	25	128	0.5437	0.8818	0.5841	0.8780
	embedding layer	-	16	25	128	0.7382	0.8846	0.7512	0.9062
	Word2Vec	4	16	25	128	0.8297	0.8881	0.8598	0.9173
250	One hot	-	-	25	128	0.5508	0.9161	0.5865	0.8994
	TF-IDF	-	-	25	128	0.6397	0.9210	0.6540	0.9048
	embedding layer	-	16	25	128	0.7771	0.9349	0.7785	0.9375
	Word2Vec	4	16	25	128	0.8839	0.9454	0.9065	0.9515
500	One hot	-	-	25	128	0.6926	0.9512	0.6992	0.9303
	TF-IDF	-	-	25	128	0.7358	0.9602	0.7238	0.9515
	embedding layer	-	16	25	128	0.8159	0.9652	0.7958	0.9687
	Word2Vec	4	16	25	128	0.9380	0.9726	0.9532	0.9797

從表4- 15中，發現訓練及驗證在只有擴充50筆資料時，onehot模型：accuracy=0.3873、Binary_accuracy=0.7457、val_accuracy=0.4013、val_Binary_accuracy=0.7775；TF-IDF模型：accuracy=0.4476、Binary_accuracy=0.7426、val_accuracy=0.4543、val_Binary_accuracy=0.7712；embedding layer模型：accuracy=0.6994、Binary_accuracy=0.8143、val_accuracy=0.6838、val_Binary_accuracy=0.8350；Word2Vec模型：accuracy=0.7756、Binary_accuracy=0.8509、val_accuracy=0.8131、val_Binary_accuracy=0.8431。可以觀察到，在資料擴充50筆時，所有四個模型的準確率都有所降低。然而，Word2Vec模型在這種情況下仍然保持相對較高的準確率。這可能是因為Word2Vec模型能夠更好地捕捉資料的語義和關聯性，並從中提取有用的特徵。儘管準確率有所降低，但Word2Vec模型仍然具有一定的準確性。

表 4- 16 不同擴充筆數的四種模在型測試集的準確率

擴充筆數	模型	測試							
		15 筆病歷				18 虛證			
		accuracy	Binary accuracy >0.5	Recall	Precision	accuracy	Binary accuracy >0.5	Recall	Precision
50	One hot	0.3195	0.9366	0.5732	0.6223	0.3566	0.9318	0.4853	0.6647
	TF-IDF	0.5203	0.9608	0.5929	0.6722	0.5009	0.9473	0.6468	0.7415
	embedding layer	0.5672	0.9736	0.7985	0.7892	0.5224	0.9684	0.7924	0.7301
	Word2Vec	0.6833	0.9759	0.8348	0.8506	0.6586	0.9770	0.8260	0.8178
100	One hot	0.3790	0.9436	0.6172	0.6911	0.4065	0.9395	0.5507	0.7227
	TF-IDF	0.5675	0.9672	0.6691	0.7516	0.5533	0.9572	0.7099	0.8040
	embedding layer	0.6274	0.9800	0.8480	0.8418	0.5932	0.9637	0.8437	0.7975
	Word2Vec	0.7575	0.9819	0.8758	0.8879	0.7370	0.9752	0.8545	0.8633
250	One hot	0.4384	0.9506	0.6611	0.7599	0.4565	0.9472	0.6162	0.7807
	TF-IDF	0.6148	0.9736	0.7453	0.8310	0.6058	0.9672	0.7731	0.8564
	embedding layer	0.6875	0.9863	0.8974	0.8944	0.6639	0.9889	0.8950	0.8648
	Word2Vec	0.8317	0.9879	0.9167	0.8952	0.8155	0.9834	0.9030	0.9089
500	One hot	0.4978	0.9576	0.7051	0.8286	0.5064	0.9550	0.6816	0.8387
	TF-IDF	0.6620	0.9799	0.8215	0.9104	0.6582	0.9771	0.8362	0.9288
	embedding layer	0.7476	0.9927	0.9468	0.9469	0.7347	0.9942	0.9463	0.9322
	Word2Vec	0.9059	0.9939	0.9577	0.9226	0.8939	0.9917	0.9414	0.9544

從表4- 16中，發現測試集在只有擴充50筆資料時，15筆病歷資料：onehot模型：accuracy = 0.3195、Binary_accuracy>0.5 = 0.9366、Recall = 0.5732、Precision = 0.6223；TF-IDF模型：accuracy = 0.5203、Binary_accuracy>0.5 = 0.9608、Recall = 0.5929、Precision = 0.6722；embedding layer 模型：accuracy = 0.5672、Binary_accuracy>0.5 = 0.9736、Recall = 0.7985、Precision = 0.7892；Word2Vec模型：accuracy = 0.6833、Binary_accuracy>0.5 = 0.9759、Recall = 0.8348、Precision = 0.8506；18筆虛證資料中onehot模型：accuracy = 0.3566、Binary_accuracy>0.5 = 0.9318、Recall = 0.4853、Precision = 0.6647；TF-IDF模型：accuracy = 0.5009、Binary_accuracy>0.5 = 0.9473、Recall = 0.6468、Precision = 0.7415；embedding layer 模型：accuracy = 0.5224、Binary_accuracy>0.5 = 0.9684、Recall = 0.7924、Precision = 0.7301；Word2Vec模型：accuracy = 0.6586、Binary_accuracy>0.5 = 0.9770、Recall = 0.8260、Precision = 0.8178。從這些結果可以觀察到，在資料擴充數量較少的情況下，模型的準確率確實較低。不同模型在預測15筆病歷資料和18筆虛證資料時都呈現出相對較低的準確率、Recall、Precision值。這表明資料擴充的數量對模型的性能有一定的影響。因此，在選擇擴充

數量時，需要找到一個適合的平衡點，使模型能夠在擴充後的資料上取得更好的準確率。

這些結果也再次強調了資料量對於模型性能的重要性。當資料量有限時，模型難以充分學習資料中的模式和特徵，因此準確率下降。適當的資料擴充能夠增加資料量，有助於提高模型的準確性和性能。進一步增加資料擴充的數量可能有助於改善模型的準確率和性能。

4.9 實驗七、雜訊對四種模型準確率的測試

進行了四種模型的實驗，以探討在資料存在雜訊的情況下，這些模型在準確率方面的魯棒性。在這些實驗中，固定了以下參數：Maxlen（對齊長度）設置為25、神經元個數設置為12、W2V_DIM設置為16，Window_size設置為4。這些參數的固定使能夠比較四種不同模型在相同條件下的性能表現。同時，的目標是觀察模型在面對有雜訊的資料時的魯棒性，即其在噪聲存在的情況下是否能夠保持良好的準確率。透過這些實驗，可以獲得關於不同模型在處理有雜訊資料時的性能比較，進而評估它們的魯棒性。這將有助於選擇最適合處理有雜訊資料的模型，並為未來相關應用提供有價值的參考。以下是雜訊處理的步驟：

- a. 抓取症狀欄位並進行切割，將每個症狀分開，例如，一筆證候症狀資料["心氣虛"， "自汗，面色白，心悸，胸悶，情緒神疲"]將症狀取出來切割成 "自汗"、"面色白"、"心悸"、"胸悶"、"情緒神疲"。
- b. 將切割後的症狀隨機刪除N個，產生新的一組資料。例如，將 "自汗"、"面色白"、"心悸"、"胸悶"、"情緒神疲" 隨機刪除1個組合成 "自汗，心悸，胸悶，情緒神疲" 的一組新症狀資料，再將證候組合["心氣虛"， "自汗，心悸，胸悶，情緒神疲"]回去，以達到雜訊的效果。

表 4- 17 隨機刪減症狀數量 18 虛證及 50 筆病歷的準確率

模型	隨機砍症狀數	18 虛證				50 筆病歷			
		accuracy	Binary_accuracy>0.5	recall	precision	accuracy	Binary_accuracy>0.5	recall	precision
One hot	1	0.9336	0.9928	0.9855	0.9831	0.9191	0.9788	0.9698	0.9694
	3	0.9212	0.9804	0.9707	0.9699	0.9062	0.9644	0.9555	0.9579
	5	0.7790	0.8370	0.8326	0.8313	0.7568	0.8159	0.8078	0.8114
	7	0.5576	0.8168	0.6107	0.6092	0.5312	0.7906	0.5803	0.5827
TF-IDF	1	0.9371	0.9938	0.9851	0.9836	0.9201	0.9755	0.9701	0.9730
	3	0.9253	0.9809	0.9735	0.9741	0.9081	0.9645	0.9602	0.9609
	5	0.7915	0.8482	0.8446	0.8428	0.7707	0.8298	0.8209	0.8210
	7	0.5742	0.8316	0.6265	0.6245	0.5461	0.8056	0.5968	0.5992
Embedding layer	1	0.9460	0.9927	0.9853	0.9893	0.9320	0.9807	0.9756	0.9743
	3	0.9362	0.9816	0.9738	0.9774	0.9196	0.9657	0.9601	0.9598
	5	0.7933	0.9431	0.8364	0.8347	0.7718	0.8195	0.8166	0.8130
	7	0.5966	0.9109	0.6365	0.6366	0.5682	0.8144	0.6108	0.6067
Word2Vec	1	0.9573	0.9949	0.9876	0.9898	0.9466	0.9804	0.9781	0.9767
	3	0.9503	0.9816	0.9805	0.9768	0.9361	0.9698	0.9661	0.9644
	5	0.8355	0.9712	0.8671	0.8638	0.8178	0.8522	0.8475	0.8483
	7	0.7035	0.9374	0.7335	0.7333	0.6833	0.8190	0.7123	0.7127

根據表4- 17的結果，可以觀察到在使用四種不同的編碼方式（onehot、Embedding layer、Word2Vec、TF-IDF）的模型時，隨著症狀數量的逐漸減少，模型在一層隱藏層中的準確率也逐漸下降。具體而言，當症狀數量減少到7個時，可以看到不同模型的準確率，以Word2Vec模型在18虛證中的accuracy = 0.7035、Binary_accuracy >0.5 = 0.9374、recall = 0.7335、precision = 0.7333；50筆病歷中的accuracy = 0.6833、Binary_accuracy >0.5 = 0.8190、recall = 0.7123、precision = 0.7127，表現最好。

Word2Vec模型在這些編碼方式中表現出色，其準確率能夠保持在0.7，即使在減少到7個症狀時仍然如此。這一結果強調了Word2Vec模型在處理自然語言處理任務中的優勢。Word2Vec模型通過預先訓練詞向量，能夠有效地捕捉單詞之間的語義和關聯性，從而提高模型對症狀的特徵表示能力；相比之下，onehot編碼方式的模型在症狀數量減少時表現最差，其準確率只剩下0.55。這是因為onehot編碼僅考慮單詞是否存在，而忽略了單詞之間的語義相似性和上下文資訊；Embeddinglayer和TF-IDF模型的準確率介於onehot模型和Word2Vec模型之間，表明它們在處理特徵表示方面相對中等。

綜上所述，編碼方式對於模型在症狀預測任務中的準確率有著明顯的影響。Word2Vec模型通過詞向量的預訓練，在症狀數量減少時仍能保持較高的準確率，這為基於自然語言處理的疾病診斷和預測提供了有價值的參考和指導。



伍、 結論與未來發展

5.1 結論

本研究旨在探討症狀詞向量與深度學習網路在提升中醫辨證效能方面的可行性，並收集了由五本中醫診斷學典籍所整理而來的18虛證，以及從期刊中整理出來的50個臨床案例報告，這些資料來自中醫典籍和期刊。

建立了一個語料庫，其中包含了將症狀詞轉換成數字以及訓練詞向量所需的詞組。使用擴充後的資料集，訓練和驗證了四種不同架構的神經網路模型。實驗結果表明，對齊長度的長短、輸入症狀詞的數量、詞向量長度、神經元數量和詞向量模型的採樣窗口大小等因素都對辨證效能有影響。

實驗結果如下：

實驗一是基於 one-hot encoding 模型的訓練結果。我們發現將對齊長度設置為25並將神經元個數設置為128時，模型達到了最佳的性能表現。在這種設置下，我們觀察到 $\text{accuracy} = 0.9461$ 、 $\text{Binary_accuracy} = 0.9961$ 、 $\text{val_accuracy} = 0.9654$ 、 $\text{val_Binary_accuracy} = 0.9951$ 。這些數據表明，這組參數在訓練和驗證過程中展現出最佳的性能和表現。

實驗二則是基於 TF-IDF encoding 模型的訓練結果。我們將對齊長度設置為25並將神經元個數設置為128，這組參數也呈現出最佳的性能表現。在這種設置下，我們觀察到 $\text{accuracy} = 0.9475$ 、 $\text{Binary_accuracy} = 0.9980$ 、 $\text{val_accuracy} = 0.9668$ 、 $\text{val_Binary_accuracy} = 0.9983$ 。這些數據表明，這組參數在訓練和驗證過程中展現出最佳的性能和表現。

根據實驗三的結果，當使用 embedding layer 模型時，增加神經網路的 Vector-Dim 參數可以獲得更好的效果。較大的 Vector-Dim 參數增加了詞向量的維

度，使模型能夠更好地理解單詞之間的語義關係，進而提高準確率。因此，在使用 embedding layer 模型時，調整 Vector-Dim 參數可以是一種有效的提升模型性能的方法。

實驗四則是基於 Word2Vec 模型的訓練結果。根據實驗結果，我們發現隨著 window_size 採樣窗口大小的增加，模型的效果逐漸提升。這意味著先前使用 Word2Vec 所產生的詞向量成功捕捉到了詞彙之間的關聯性，並有效地應用於模型中。

根據實驗五的結果顯示，四種模型在原始資料集上都表現出色，對於虛證和病例的預測都具有可靠、全面和精確的能力。這些模型能夠準確地識別和預測資料中的虛證和病例，為該領域的應用提供了重要的支持，同時為進一步的研究和應用奠定了堅實的基礎。

實驗六針對資料擴充不同筆數對四種模型準確率的影響進行了測試。結果再次強調了資料量對於模型性能的重要性。當資料量有限時，模型難以充分學習資料中的模式和特徵，從而降低了準確率。適當的資料擴充可以增加資料量，有助於提高模型的準確性和性能。進一步增加資料擴充的數量可能有助於進一步改善模型的準確率和性能。

實驗七是針對雜訊對四種模型準確率的測試。結果顯示，在這些編碼方式中，Word2Vec模型表現出色，其準確率能夠保持在0.7，即使在減少到7個症狀時仍然如此。這強調了Word2Vec模型在處理自然語言處理任務中的優勢。Word2Vec模型通過預先訓練詞向量，能夠有效地捕捉單詞之間的語義和關聯性，從而提高模型對症狀的特徵表示能力。相比之下，one-hot編碼方式的模型在症狀數量減少時表現最差，其準確率只剩下0.55。這是因為one-hot編碼僅考慮單詞是否存在，而忽略了單詞之間的語義相似性和上下文資訊。Embedding layer和TF-IDF模型的準確率介於one-hot模型和

Word2Vec模型之間，表明它們在特徵表示方面相對中等。

總結而言，這些實驗結果提供了有關不同編碼方式和模型參數對於模型性能的重要性。我們可以根據這些結果來選擇最佳的設置，以提高模型的準確率並應用於相關領域中。基於實驗結果，建議在使用自然語言處理模型時考慮以下因素：

編碼方式的選擇：根據實驗結果，Word2Vec模型在處理文本特徵表示方面表現出色，能夠有效地捕捉單詞之間的語義關聯性。然而，Embedding layer和TF-IDF模型也具有一定的性能，在特定情況下可能是更好的選擇。因此，在應用中，應該根據具體任務需求和資料特徵來選擇最合適的編碼方式。

參數調整：根據實驗結果，神經網路的參數調整對於模型性能具有重要影響。例如，在使用embedding layer模型時，調整Vector-Dim參數可以提高模型對於單詞語義關聯性的理解能力。因此，在訓練模型時，應該進行參數調優以獲得最佳的性能。

資料擴充：實驗結果顯示，資料量對於模型性能至關重要。在資料量有限的情況下，進行適當的資料擴充可以增加資料量，從而提高模型的準確性和性能。因此，應該考慮使用資料擴充技術來增加訓練資料的多樣性。

綜上所述，這些實驗結果為我們選擇和設計自然語言處理模型提供了有價值的指導。透過適當的編碼方式、參數調整和資料擴充，我們可以提高模型的準確率和性能，並應用於各種文本相關的應用領域。

這些發現強調了深度學習和詞向量模型在提升中醫辨證效能方面的可行性。研究為中醫醫學領域的自動化辨證提供了有價值的參考和指導，也為未來進一步的研究和應用奠定了基礎。未來的研究可以探索更多模型架構和技術，以提高中醫辨證的準確性和可靠性，並推動中醫醫學的現代化和智能化發展。

本研究致力於解決中醫辨證中的多標籤分類問題，探討症狀詞與證候之間的關聯性，以提高辨證的準確性。實驗結果顯示，通過調整模型參數可以明顯提升準確率。在各種不同的實驗中，相對於採用onehot encoding、embedding layer訓練詞向量以及TF-IDF權重模型，遷移學習中的Word2Vec詞向量模型展現出最佳的驗證準確率效能。這結果表明預訓練模型在表徵能力上優越，並適用於不同自然語言處理的情境，保留了字詞間的交互關係。

本研究透過設計一系列實驗，比較不同實驗的成果，旨在找出提升中醫辨證效果和可行性的最佳方法。這些實驗有助於深入了解模型參數對辨證效果的影響。同時，透過遷移學習的應用，利用Word2Vec詞向量模型的預訓練知識，獲得了更優秀的結果。這些研究結果對於中醫辨證的自動化具有重要意義，為中醫醫學領域提供了有價值的指導和啟示。同時，本研究也為機器學習在中醫辨證上的應用開拓了新的可能性。未來的研究可以進一步探索更多的模型和技術，以提高中醫辨證的效果和準確性，推動中醫醫學的現代化和智能化發展。

5.2 未來發展

本研究透過不斷調整超參數，訓練深度學習神經網路模型，旨在讓模型學習識別與證候相關的特徵症狀。訓練完成後，可以利用此模型對新的資料集進行證候預測，並從預測結果評估模型的訓練效果。目前，該模型僅在中醫的18個虛證上應用，但未來的研究計劃將逐步加入更多的實證資料集，以擴展神經網路模型的訓練資料。

然而，中醫的資料集並不容易取得，需要與中醫師討論實證的證候及其對應的症狀組合，以避免遺漏任何重要的症狀。藉由擁有更完整的資料集，神經網路模型可以與實證證候一同進行預測，進一步擴大中醫的診斷範圍。

期望未來本研究的預測結果能夠應用於中醫證候預測系統，同時將人工智慧的觀點應用於中醫辨證，以減輕醫生在中醫辨證方面的負擔和不確定性。這將有助於提升中醫診斷的準確性和效率，為醫生提供更可靠的輔助工具，同時推動中醫與機器學習的結合，促進醫學領域的進步。



參考文獻

- [1]張慈文(2017)。中醫證治學。國防醫學院。
- [2]陳南宏(2015)。中醫症狀知識本體的建構，碩士論文，國立中正大學資訊工程學系暨研究所，台灣。
- [3] 陳信賢(2016)。基於知識本體的中醫虛證辨證系統，碩士論文，國立中正大學資訊工程學系，台灣。
- [4]蔡昀庭(2021)。呼吸道疾病的中醫模糊辨證，碩士論文，國立中正大學資訊工程學系，台灣。
- [5]張純華(2021)。月經病模糊中醫辨證，碩士論文，國立中正大學資訊工程學系，台灣。
- [6] 陳允傑(2019)。《TensorFlow + Keras : Python 深度學習應用實務》。出版社：旗標。
- [7] Jon Krohn , Grant Beyleveld , Aglae Bassens 。黃駿，哈雷譯(2021)。《深度學習的 16 堂課》。出版社：旗標。
- [8] 施威銘研究室(2020)。《tf.Keras：深度學習攻略手冊》。出版社：旗標。
- [9] 黃柏叡(2013)。第八屆立法委員質詢行為與臉書經營關係之研究，碩士論文，國立臺灣師範大學政治學研究所，台灣。
- [10]陳耀男(2012)。多標籤分類問題中考量特徵的標籤空間降維法，碩士論文，國立臺灣大學，台灣。
- [11]林怡如，吳政隆(2020)。建構拍賣網提問類型之多重標籤辨識模型，碩士論文，東吳大學巨量資料管理學院，台灣。
- [12]傅琦佳(2020)。胸部 X 射線之多標籤深度學習分類—常見胸部疾病預測，碩士論文，國立交通大學統計學研究所，台灣。
- [13]何原野(2017)。應用深度學習于文本多標籤分類之研究，碩士論文，元智大學資訊工程學系，台灣。
- [14] 毛楚琳(2020)。結合知識圖譜與文本來訓練專業領域詞彙詞向量，碩士論文，國立交通大學數據科學與工程研究所，台灣。
- [15] T. Yang, C. Wu, Z. Xu, Y. Ding, “The syndrome differentiation model and program of traditional Chinese medicine based on the fuzzy recognition,” Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, Shanghai, China, pp. 285-287, 2013.
- [16] J. Chen, D. Yang, Y. Cao, Y. Ma, C. Wen, X. Huang, J. Guo, “Syndrome differentiation and treatment algorithm model in traditional Chinese medicine based on disease cause, location, characteristics and conditions,” IEEE Access, vol. 6, pp. 71801-71813, 2018.
- [17] M. Jiang, C. Lu, C. Zhang, J. Yang, Y. Tan, A. Lu, K. Chan, “Syndrome differentiation in modern research of traditional Chinese medicine,” Journal of Ethnopharmacology, vol. 140, no. 3, pp. 634-642, April 2012.

- [18] Y. Wang, Y. Yu, Y. Jiang, Y. Liu, L. Chen, Y. Liu, “A framework and its empirical study of automatic diagnosis of traditional Chinese medicine utilizing raw free-text clinical records,” *Journal of Biomedical Informatics*, vol. 45, no. 2, pp. 210-223, April 2012.
- [19] Yufeng Zhao, Liyun He, Qi Xie, Guozheng Li, Baoyan Liu, Jian Wang, Xiaoping Zhang, Xiang Zhang, Lin Luo, Kun Li, and Xianghong Jing, “A Novel Classification Method for Syndrome Differentiation of Patients with AIDS,” *Evidence-Based Complementary and Alternative Medicine*, vol. 2015, Article ID 936290, 8 pages.
- [20] W. Zhu, W. Zhang, G. Z. Li, C. He, L. Zhang, “A study of damp-heat syndrome classification using Word2vec and TF-IDF,” *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine*, Shenzhen, China, pp. 1415-1420, 2016.
- [21] L. Yao, Y. Zhang, B. Wei, Z. Li, X. Huang, “Traditional Chinese medicine clinical records classification using knowledge-powered document embedding,” *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine*, Shenzhen, China, pp. 1926-1928, 2016.
- [22] Hu, Q.; Yu, T.; Li, J.; Yu, Q.; Zhu, L.; Gu, Y. End-to-End syndrome differentiation of Yin deficiency and Yang deficiency in traditional Chinese medicine. *Computer Methods and Programs in Biomedicine*, vol. 174, pp. 9-15, 2019.
- [23] P.-S. Lin, N.-W. Lin, M.-H. Yeh, C.-C. Yeh, H.-P. Chiu, “Improving fuzzy syndrome differentiation for deficiency syndromes in traditional Chinese medicine,” *Proceedings of International Conference on Consumer Electronics-TW*, Taoyuan, Taiwan, 2020.
- [24] Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research*, 3(Feb), 1137-1155.
- [25] Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500), 2323-2326.
- [26] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv : 1301.3781*.
- [27] Pritish Jadhav (2021) ◦ Evaluation Metrics for Multi-Label Classification ◦
url: <https://medium.datadriveninvestor.com/a-survey-of-evaluation-metrics-for-multi-label-classification-bb16e8cd41cd>

附錄

附錄 1：18 個虛證證候症狀表

證候	症狀
心氣虛	自汗，面色白，心悸，胸悶，情緒神疲，全身無力，呼吸微弱，舌體顏色淡白色，舌苔顏色白，脈力度弱
心血虛	意識健忘，難眠，唇色白，面色白，面色暗，心悸，頭暈，目眩，多夢，舌體顏色淡白色，脈寬度細
心陰虛	意識健忘，難眠，面色紅，心悸，頭暈，目眩，多夢，手足心熱，潮熱，盜汗，口乾，喉乾，舌體顏色紅色，脈寬度細，脈速度快
心陽虛	自汗，面色白，胸痛，心痛，面色暗，心悸，胸悶，情緒神疲，呼吸微弱，四肢冷，畏寒喜溫，舌體顏色淡白色，舌苔顏色白，舌苔濕度滑，舌形狀胖大，脈力度弱，脈部位沉
心陽暴脫	面色白，面色暗，心痛，意識昏迷，意識恍惚，冷汗，唇色紫，心悸，胸悶，呼吸微弱，四肢厥冷，舌體顏色紫色，脈力度弱，脈寬度細
肝血虛	四肢拘急，四肢麻木，面色暗，月經閉經，月經量少，月經色淡紅，目暗，耳鳴，頭暈，目眩，多夢，舌體顏色淡白色，脈寬度細
肝陰虛	情緒憤怒，脅痛，目乾，面色紅，目暗，耳鳴，頭暈，目眩，手足心熱，潮熱，盜汗，口乾，喉乾，舌體顏色紅色，脈寬度細，脈速度快
肺氣虛	自汗，面色白，畏風，痰色清，痰質稀，情緒聲弱，咳嗽，呼吸急促，情緒神疲，全身無力，呼吸微弱，舌體顏色淡白色，舌苔顏色白，脈力度弱
肺陰虛	咳時無痰，痰難咳出，聲啞，涎量少，痰量少，痰挾血，咳血，面色紅，咳嗽，手足心熱，潮熱，盜汗，口乾，喉乾，舌體顏色紅色，舌苔顏色白，脈力度弱，脈寬度細，脈速度快
脾氣虛	意識懶言，全身浮腫，全身消瘦，面色黃，面色暗，食欲低落，腹脹，大便質稀，情緒神疲，全身無力，呼吸微弱，舌苔顏色白，舌體顏色淡白色，脈力度弱
脾陽虛	全身浮腫，小便量少，面色白，白帶量多，白帶質稀，食欲低落，腹脹，大便質稀，腹痛，情緒神疲，全身無力，四肢冷，畏寒喜溫，舌苔顏色白，舌體顏色淡白色，舌苔濕度滑，舌形狀胖大，脈力度弱，脈部位沉
脾氣陷	意識懶言，面色黃，面色暗，頭暈，目眩，小便挾乳糜，大便次多，食欲低落，腹脹，大便質稀，情緒神疲，全身無力，脫肛，子宮脫垂，舌苔顏色白，舌體顏色淡白色，脈力度弱
脾不統血	意識懶言，面色暗，面色白，食欲低落，腹脹，大便質稀，情緒神疲，全身無力，呼吸微弱，大便挾血，小便挾血，瘀血，鼻衄，牙齦出血，月經量多，舌體顏色淡白色，脈力度弱，脈寬度細
腎精不足	意識健忘，意識恍惚，意識癡呆，全身消瘦，鬚髮脫落，牙齒動搖，耳聾，耳鳴，幽門遲閉，腰無力，腰酸，膝酸，下肢無力，下肢酸，精子量少，月經閉經，妊娠不孕，舌體顏色淡白色，舌苔顏色白，脈寬度細
腎不納氣	情緒聲弱，自汗，四肢冷，腰無力，腰酸，膝酸，下肢無力，下肢酸，情緒神疲，呼吸急促，咳嗽，舌體顏色淡白色，舌苔顏色白，脈寬度細，脈部位沉
腎氣不固	小便不盡，小便量多，小便色清，遺精，早洩，白帶色清，白帶質稀，腰無力，腰酸，膝酸，情緒神疲，小便失禁，遺尿，小便次多，夜時小便次多，舌體顏色淡白色，舌苔顏色白，脈寬度細，脈力度弱
腎陰虛	全身消瘦，鬚髮脫落，耳鳴，難眠，小便色黃，小便量少，涎量少，骨蒸，腰痛，膝痛，面色紅，遺精，早洩，月經閉經，月經量少，腰無力，腰酸，膝酸，下肢無力，下肢酸，頭暈，目眩，多夢，手足心熱，潮熱，盜汗，口乾，喉乾，舌體顏色紅色，舌苔顏色白，舌苔厚度薄，脈寬度細，脈速度快
腎陽虛	小便量多，小便色清，遺精，早洩，大便質稀，大便次多，目眩，全身浮腫，性欲低落，面色黧黑，腰冷，頭暈，面色白，陽痿，妊娠不孕，腰無力，腰酸，膝酸，下肢無力，下肢酸，情緒神疲，小便次多，夜時小便次多，四肢冷，畏寒喜溫，舌體顏色淡白色，舌苔顏色白，舌苔濕度滑，舌形狀胖大，脈力度弱，脈部位沉

附錄 2：50 筆病歷證候症狀表(1~10 筆)

證候	症狀
肺氣虛。肺陰虛。 心氣虛。脾氣虛	情緒神疲，面乾，面色黃，面色暗，情緒煩躁，情緒憤怒，全身消瘦，痰色黃，腹脹，呼吸急促，夜時呼吸急促，全身無力，惡寒，發熱，目乾，口乾，舌苔濕度乾，咳嗽，痰量多，痰質濃，胸悶，心悸，惡熱，食欲低落，大便不暢，大便質稀，易醒，多夢，脈部位沉，脈力度弱，痰難咳出，痰量少，舌體顏色紅色，涎量少，自汗，面色白，心悸，胸悶，情緒神疲，全身無力，呼吸微弱，舌體顏色淡白色，舌苔顏色白，脈力度弱
心陽虛。腎陽虛。 脾陽虛	心悸，呼吸急促，全身浮腫，情緒憂鬱，畏寒喜溫，畏風，頭暈，胸悶，腹痛，大便質稀，大便次多，情緒神疲，難眠，四肢冷，腹脹，舌體顏色淡紅色，舌苔顏色白，脈部位沉，脈力度滑，脈力度弱，意識健忘，難眠，唇色白，面色白，面色暗，心悸，頭暈，目眩，多夢，舌體顏色淡白色，脈寬度細
肺氣虛。脾氣虛	大便質硬，大便次少，難眠，易醒，口乾，食欲低落，畏寒喜溫，四肢冷，舌體顏色紅色，舌苔厚度薄，舌苔顏色黃，脈緊度緊，脈長度長，脈力度滑，意識健忘，難眠，面色紅，心悸，頭暈，目眩，多夢，手足心熱，潮熱，盜汗，口乾，喉乾，舌體顏色紅色，脈寬度細，脈速度快
心氣虛。肺氣虛。 腎氣不固	情緒神疲，面色白，舌體顏色淡紅色，舌苔顏色白，舌苔厚度薄，咳嗽，痰色黃，難眠，易醒，口乾，大便質稀，小便次多，小便量多，脈力度弱，自汗，面色白，胸痛，心痛，面色暗，心悸，胸悶，情緒神疲，呼吸微弱，四肢冷，畏寒喜溫，舌體顏色淡白色，舌苔顏色白，舌苔濕度滑，舌形狀肥大，脈力度弱，脈部位沉
心陽虛。腎陽虛	頭暈，目眩，畏寒喜溫，四肢冷，嘔吐，腰酸，腰冷，腰痛，脈部位沉，脈力度弱，舌體顏色淡白色，舌形狀肥大，舌苔顏色白，舌苔濕度濕，面色白，面色暗，心痛，意識昏迷，意識恍惚，冷汗，唇色紫，心悸，胸悶，呼吸微弱，四肢厥冷，舌體顏色紫色，脈力度弱，脈寬度細
脾陽虛。腎陽虛	腰酸，腰痛，腰無力，背痛，情緒神疲，全身無力，畏寒喜溫，四肢冷，難眠，易醒，面色暗，面色黃，面色白，小便挾乳糜，小便色白，夜時小便次多，大便質硬，大便次少，舌體顏色淡紅色，舌苔厚度薄，舌苔顏色黃，脈寬度細，脈力度弱，四肢拘急，四肢麻木，面色暗，月經閉經，月經量少，月經色淡紅，目暗，耳鳴，頭暈，目眩，多夢，舌體顏色淡白色，脈寬度細
脾氣虛。脾氣陷。 心氣虛。腎陽虛	面色黃，面色暗，全身消瘦，情緒煩躁，情緒恐懼，情緒心慌，唇乾，情緒聲弱，頭暈，目眩，胸悶，心悸，喉乾，情緒神疲，易醒，多夢，四肢汗出，胃脹，噯氣，大便質稀，大便次多，腰酸，舌體顏色紅色，舌苔顏色白，舌苔厚度薄，脈部位沉，脈緊度緊，脈長度長，脈力度弱，情緒憤怒，脅痛，目乾，面色紅，目暗，耳鳴，頭暈，目眩，手足心熱，潮熱，盜汗，口乾，喉乾，舌體顏色紅色，脈寬度細，脈速度快
心血虛。心陰虛。 腎陰虛	唇色紫，難眠，易醒，多夢，大便質稀，喉乾，咳嗽，痰色黃，痰質濃，鼻流涕，胸悶，脅脹，自汗，舌體顏色紅色，舌體形體裂紋，舌苔厚度薄，脈寬度細，脈部位沉，脈緊度緊，脈長度長，自汗，面色白，畏風，痰色清，痰質稀，情緒聲弱，咳嗽，呼吸急促，情緒神疲，全身無力，呼吸微弱，舌體顏色淡白色，舌苔顏色白，脈力度弱
肝陰虛。腎陰虛	面色白，目散，難眠，情緒煩躁，易醒，多夢，盜汗，潮熱，喉乾，咳嗽，情緒神疲，頭暈，腰酸，腰痛，腹脹，大便質硬，舌體顏色紅色，舌形狀肥大，舌體形體齒痕，舌苔顏色白，舌苔厚度薄，脈部位沉，脈寬度細，脈速度快，脈緊度緊，脈長度長，咳時無痰，痰難咳出，聲啞，涎量少，痰量少，痰挾血，咳血，面色紅，咳嗽，手足心熱，潮熱，盜汗，口乾，喉乾，舌體顏色紅色，舌苔顏色白，脈力度弱，脈寬度細，脈速度快

附錄 3：50 筆病歷證候症狀表(11~20 筆)

證候	症狀
心氣虛。心血虛。 心陰虛	心悸，難眠，多夢，頭暈，頭重，意識恍惚，情緒神疲，食欲低落，面色暗，情緒心慌，情緒恐懼，舌體顏色淡白色，舌苔顏色白，舌苔質地膩，脈寬度細，脈力度弱，全身浮腫，小便量少，面色白，白帶量多，白帶質稀，食欲低落，腹脹，大便質稀，腹痛，情緒神疲，全身無力，四肢冷，畏寒喜溫，舌苔顏色白，舌體顏色淡白色，舌苔濕度滑，舌形狀胖大，脈力度弱，脈部位沉
脾氣虛。心血虛	面色黃，面色暗，情緒聲弱，呼吸微弱，意識懶言，頸項僵，頸項痛，乳房結塊，情緒神疲，食欲低落，自汗，心悸，難眠，舌體顏色淡白色，舌苔形體部分剝落，舌苔厚度薄，舌苔顏色白，脈部位沉，脈寬度細，脈力度弱，意識懶言，面色黃，面色暗，頭暈，目眩，小便挾乳糜，大便次多，食欲低落，腹脹，大便質稀，情緒神疲，全身無力，脫肛，子宮脫垂，舌苔顏色白，舌體顏色淡白色，脈力度弱
脾氣虛。腎陽虛	全身消瘦，情緒神疲，面色暗，唇色白，情緒聲弱，畏寒喜溫，偏頭痛，頭昏，頸項僵，頸項痛，鼻塞，鼻流涕，涕色清，耳鳴，胸悶，少腹痛，腰酸，背痛，膝酸，腰無力，四肢冷，大便質稀，大便次多，舌體顏色淡紅色，舌苔厚度薄，舌苔顏色黃，脈部位沉，脈寬度細，妊娠不孕，意識懶言，面色暗，面色白，食欲低落，腹脹，大便質稀，情緒神疲，全身無力，呼吸微弱，大便挾血，小便挾血，瘀血，鼻衄，牙齦出血，月經量多，舌體顏色淡白色，脈力度弱，脈寬度細
心血虛	心悸，情緒心慌，頭暈，意識恍惚，情緒神疲，全身無力，面色暗，舌體顏色淡白色，舌苔顏色白，舌苔厚度薄，脈部位沉，脈寬度細，脈速度慢，意識健忘，意識恍惚，意識癡呆，全身消瘦，鬚髮脫落，牙齒動搖，耳聾，耳鳴，凶門遲閉，腰無力，腰酸，膝酸，下肢無力，下肢酸，精子量少，月經閉經，妊娠不孕，舌體顏色淡白色，舌苔顏色白，脈寬度細
心氣虛	心悸，呼吸微弱，胸悶，呼吸不暢，難眠，易醒，情緒神疲，全身無力，面色白，舌苔顏色白，舌苔厚度薄，舌苔濕度滑，脈部位沉，脈寬度細，脈力度弱，情緒聲弱，自汗，四肢冷，腰無力，腰酸，膝酸，下肢無力，下肢酸，情緒神疲，呼吸急促，咳喘，舌體顏色淡白色，舌苔顏色白，脈寬度細，脈部位沉
脾氣虛	腹脹，大便不暢，面色暗，面色黃，情緒神疲，胸悶，口渴，大便質稀，大便次多，大便挾血，舌體顏色淡紅色，舌苔顏色白，脈力度弱，小便不盡，小便量多，小便色清，遺精，早洩，白帶色清，白帶質稀，腰無力，腰酸，膝酸，情緒神疲，小便失禁，遺尿，小便次多，夜時小便次多，舌體顏色淡白色，舌苔顏色白，脈寬度細，脈力度弱
脾氣虛。脾陽虛	畏寒喜溫，全身消瘦，面色白，唇色紅，目乾，大便質稀，大便次多，夜時小便次多，腹脹，腹痛，四肢冷，舌體顏色淡紅色，舌苔顏色白，舌苔厚度厚，脈寬度細，脈速度快
心陽虛	心悸，頭暈，情緒神疲，全身無力，四肢冷，難眠，腹鳴，舌苔厚度薄，舌苔顏色白，舌體顏色淡紅色，脈部位沉，脈速度慢
肺氣虛	胸悶，咳嗽，呼吸急促，夜時呼吸急促，心悸，全身無力，呼吸微弱，情緒聲弱
肺氣虛。心氣虛。 心血虛	面色白，唇暗，情緒聲弱，情緒神疲，全身痛，呼吸急促，畏寒喜溫，少汗，皮膚乾，皮膚癢，頭暈，胸悶，四肢浮腫，大便不暢，難眠，舌體顏色淡紅色，舌苔厚度薄，舌苔顏色白，脈部位沉，脈寬度細，脈力度弱，脈利度澀

附錄 4：50 筆病歷證候症狀表(21~30 筆)

證候	症狀
脾氣虛	胸悶，胸痛，呼吸微弱，全身無力，腹脹，食欲低落，大便質稀，舌體顏色淡白色，舌形狀胖大，舌體形體齒痕，舌苔顏色白，脈寬度細，脈力度弱，脈部位浮
肺陰虛	面色紅，皮膚色暗，咳嗽，咳時無痰，咳聲小，夜咳，口渴，大便質硬，大便不暢，舌體顏色紅色，舌苔厚度薄，舌苔濕度乾，脈寬度細，脈速度快，脈力度弱
心氣虛。脾氣虛	口目歪斜，張目，難眠，易醒，情緒神疲，目乾，胸悶，腹脹，腰酸，舌體顏色淡紅色，舌苔顏色白，舌苔厚度薄，舌苔濕度濕，脈力度弱，脈利度滑
肺陰虛。心氣虛	情緒神疲，全身浮腫，皮膚乾，咳嗽，惡熱，指甲色白，小便色黃，情緒聲弱，全身無力，意識懶言，呼吸急促，咳鳴，小便味臭，自汗，情緒煩躁，頭暈，頭重，鼻塞，口乾，喉乾，痰色白，痰質濃，痰難咳出，心悸，食欲低落，大便質稀，大便味臭，舌體顏色淡紅色，舌形狀胖大，舌體形體裂紋，舌下脈形體曲張，舌苔厚度薄，舌苔質地黏膩，脈緊度散，脈長度短，脈速度快，脈止率亂，脈力度弱
心血虛。脾陽虛。腎陽虛	面色暗，鬚髮脫落，情緒聲弱，唇色紅，畏寒喜溫，自汗，目眩，頭暈，心悸，胸悶，呼吸微弱，胃痛，腹痛，腰無力，下肢無力，四肢冷，大便質稀，夜時小便次多，性欲低落，舌體顏色淡紅色，舌體形體胖厚，舌苔顏色白，舌苔厚度薄，舌苔濕度滑，脈寬度細，脈部位沉，脈緊度緊，脈長度長
腎氣不固。腎陽虛	畏寒喜溫，畏風，情緒神疲，情緒心慌，情緒憂鬱，易醒，多夢，頭暈，耳鳴，口乾，腰酸，背痛，小便次多，夜時小便次多，小便不暢，小便不盡，四肢冷，舌體顏色淡白色，舌體形體胖厚，舌體形體齒痕，舌苔厚度薄，舌苔顏色白，脈部位沉
脾氣虛。脾陽虛。腎氣不固。腎陽虛	面色白，面色暗，皮膚乾，全身浮腫，畏寒喜溫，情緒神疲，目暗，腹脹，腰酸，腰無力，膝酸，下肢無力，四肢冷，大便質稀，夜時小便次多，小便色清，舌體形體胖厚，舌苔厚度薄，舌苔顏色黃，舌下脈顏色紫色，脈部位沉，脈利度澀，脈力度弱
心氣虛。脾氣虛。心血虛	頭暈，目眩，情緒神疲，全身無力，情緒聲弱，面色白，面色暗，難眠，易醒，多夢，口乾，月經量少，月經色淡紅，食欲低落，大便質稀，大便次多，舌體顏色淡紅色，舌苔厚度薄，舌苔顏色白，舌苔濕度乾，脈寬度細，脈力度弱，脈緊度緊，脈長度長
脾陽虛。腎陽虛	月經量少，月經先後不定期，月經痛，腰酸，腹脹，白帶量多，皮膚癢，大便質硬，大便次少，情緒神疲，難眠，面色白，全身浮腫，食欲低落，易醒，大便不暢，舌體顏色淡紅色，舌苔顏色白，舌苔厚度薄，脈部位沉，脈部位浮，脈力度弱，脈寬度細
脾陽虛。腎陽虛	面色暗，面色白，全身消瘦，情緒聲弱，月經量少，月經痛，月經挾塊，腰酸，腰無力，膝酸，下肢無力，下肢酸，頭昏，情緒神疲，腹痛，大便質稀，大便次多，小便量多，小便次多，畏風，畏寒喜溫，四肢冷，食欲低落，情緒心慌，舌體顏色淡白色，舌苔顏色白，舌苔厚度薄，脈部位沉，脈寬度細，脈力度弱

附錄 5：50 筆病歷證候症狀表(31~40 筆)

證候	症狀
脾陽虛。腎陽虛	面色暗，指甲色白，唇色白，畏寒喜溫，四肢冷，頭暈，大便質稀，小便次多，小便量多，噴嚏，鼻塞，鼻流涕，涕色清，月經痛，白帶量多，白帶質稀，白帶色清，舌體顏色淡白色，舌體質地細膩，舌體形體齒痕，舌苔顏色白，脈部位沉，脈速度慢
心陰虛	全身消瘦，唇乾，盜汗，情緒神疲，頭暈，口乾，口苦，喉乾，心悸，胸悶，胸痛，小便次多，小便量少，小便不暢，小便色棕，情緒憂鬱，難眠，易醒，月經後期，月經痛，月經挾塊，月經量少，月經色深紅，舌體顏色深紅色，舌苔形體部分剝落，脈寬度細，脈速度快
脾氣虛。心氣虛。心血虛	面色黃，面色暗，情緒神疲，畏寒喜溫，自汗，盜汗，情緒煩躁，情緒心慌，情緒憤怒，意識健忘，情緒心驚，磨牙，多夢，腹脹，大便質稀，四肢冷，下肢浮腫，月經量多，月經經期過長，月經後期，舌體顏色淡白色，舌苔厚度薄，舌苔顏色白，脈寬度細，脈力度弱
脾氣虛	頭昏，頭暈，目眩，耳鳴，噁心，嘔吐，盜汗，全身無力，全身消瘦，面色白，呼吸微弱，意識懶言，食欲低落，腹脹，月經量少，月經色淡紅，白帶質稀，舌體顏色淡白色，舌苔形狀鏡面，脈部位沉，脈力度弱，脈速度慢
肺氣虛。肺陰虛	情緒煩躁，情緒心慌，面色紅，咳嗽，畏寒喜溫，情緒神疲，頭暈，口乾，舌苔濕度乾，呼吸急促，易醒，脈部位浮，脈緊度緊，脈速度快，四肢暖
心氣虛。脾氣虛	情緒神疲，舌體顏色淡白色，舌苔厚度薄，舌苔顏色白，畏寒喜溫，自汗，盜汗，情緒煩躁，情緒心慌，情緒憤怒，意識健忘，情緒心驚，磨牙，多夢，腹脹，月經量多，脈寬度細，脈力度弱
脾氣虛	面色暗，面色黃，舌體顏色淡紅色，舌苔顏色白，舌苔濕度乾，腹脹，食欲低落，情緒神疲，全身無力，口乾，頭暈，四肢冷，小便量少，脈緊度緊，脈長度長，脈利度澀，脈幅異常，脈速度快
脾陽虛。腎陽虛	舌苔顏色白，舌苔濕度濕，畏寒喜溫，腹痛，矢氣量多，多夢，腰酸，腰痛，脈緊度緊，脈長度長
脾陽虛。腎陽虛	情緒神疲，皮膚乾，舌體顏色深紅色，涎量少，情緒聲弱，畏寒喜溫，四肢無力，易醒，目花，大便質硬，小便量少，腰酸，腰無力，膝酸，下肢無力，下肢酸，步態不穩，脈部位沉，脈部位浮，脈力度弱，脈寬度細，脈利度滑，脈速度慢
腎陽虛	面色白，唇色紅，舌體顏色淡紅色，舌苔顏色白，舌苔厚度薄，腰酸，脈部位沉

附錄 6：50 筆病歷證候症狀表(41~50 筆)

證候	症狀
肝陰虛。腎陰虛	大汗，難眠，易醒，目乾，口乾，口苦，耳鳴，咽阻，咳嗽，痰色白，胸悶，呼吸微弱，小便次多，腰酸，腰痛，下肢無力，右手麻，脈緊度緊，脈長度長，脈寬度細
肺氣虛。脾氣虛。 胃陰虛。腎陰虛	面色白，小便色黃，情緒聲弱，食欲低落，大便不暢，情緒神疲，呼吸急促，口乾，腰痛，脈部位沉，脈力度弱，脈寬度細，脈緊度散，脈緊度緊，脈長度長
肝血虛。心血虛。 心陰虛。肝陰虛。 腎陰虛	面色暗，唇暗，唇乾，唇色紫，舌體顏色紅，舌苔顏色黃，舌苔濕度黏，鼻塞，難眠，易醒，多夢，意識健忘，心悸，情緒神疲，情緒憤怒，目乾，四肢拘急，耳鳴，腰痠，畏風，四肢重，口乾，頭暈，潮熱，脈寬度細，脈力度滑
腎氣不固。腎陽 虛。脾陽虛	舌體顏色淡紅色，舌苔顏色白，腰酸，下肢浮腫，口乾，夜時小便次多，小便量多，小便次多，小便色清，小便不盡，脈部位沉，脈力度弱，遺尿，四肢冷，畏寒喜溫
腎精不足。腎陽虛	全身消瘦，四肢冷，難眠，精子量少，脈部位沉，脈速度慢，舌體顏色淡紅色
心陽暴脫。心陰 虛。心陽虛。心血 虛	面色白，唇色白，舌體顏色淡白，舌體顏色紫色，舌苔厚度薄，舌苔顏色白，情緒聲弱，胸痛，胸悶，情緒驚恐，心悸，難眠，頭暈，頸項僵，冷汗，四肢厥冷，目眩，大便質稀，小便不暢，口乾，畏寒喜溫，腰酸，背痛，食欲低落，脈部位沉，脈寬度細，脈力度弱，脈速度快
大腸陽虛。脾氣虛	舌體顏色淡紅色，舌形狀胖大，舌苔濕度滑，舌體質地細緻，舌尖顏色紅，舌苔顏色白，腹痛，大便質稀，脈緊度緊，脈長度長，脈寬度細
脾不統血。脾氣虛	腹痛，大便挾血，大便次多，食欲低落，情緒神疲，全身消瘦，大便失禁，舌體顏色淡白色，舌苔顏色白，脈部位浮，脈力度弱，脈寬度細，脈速度慢
腎不納氣	痰量多，四肢冷，小便次多，小便量多，小便色清，呼吸急促，脈部位沉，脈速度慢，舌苔濕度滑，舌苔顏色白
脾不統血。脾氣虛	瘀血，牙齦出血，面色暗，面色黃，情緒神疲，大便質稀，舌苔顏色白，舌苔厚度薄，脈寬度細，脈力度弱

附錄 7：TF-IDF 權重證候症狀對照表(症狀 1~15)

證候/症狀	下肢無力	下肢酸	全身浮腫	全身消瘦	全身無力	冷汗	口乾	呼吸微弱	呼吸急促	咳喘	咳嗽	咳時無痰	咳血	唇色白
心氣虛	0	0	0	0	0.582017	0	0	0.560308	0	0	0	0	0	0
心血虛	0	0	0	0	0	0	0	0	0	0	0	0	0	1
心陰虛	0	0	0	0	0	0	0.607593	0	0	0	0	0	0	0
心陽虛	0	0	0	0	0	0	0	0.329593	0	0	0	0	0	0
心陽暴脫	0	0	0	0	0	1	0	0.40022	0	0	0	0	0	0
肝血虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
肝陰虛	0	0	0	0	0	0	0.569618	0	0	0	0	0	0	0
肺氣虛	0	0	0	0	0.415727	0	0	0.40022	0.731055	0	0.805056	0	0	0
肺陰虛	0	0	0	0	0	0	0.479679	0	0	0	0.593199	1	1	0
脾氣虛	0	0	0.756601	0.773827	0.415727	0	0	0.40022	0	0	0	0	0	0
脾陽虛	0	0	0.557496	0	0.306325	0	0	0	0	0	0	0	0	0
脾氣陷	0	0	0	0	0.342363	0	0	0	0	0	0	0	0	0
脾不統血	0	0	0	0	0.323343	0	0	0.311282	0	0	0	0	0	0
腎精不足	0.529901	0.529901	0	0.541679	0	0	0	0	0	0	0	0	0	0
腎不納氣	0.706534	0.706534	0	0	0	0	0	0	0.682318	1	0	0	0	0
腎氣不固	0	0	0	0	0	0	0	0	0	0	0	0	0	0
腎陰虛	0.321152	0.321152	0	0.32829	0	0	0.276179	0	0	0	0	0	0	0
腎陽虛	0.341871	0.341871	0.341691	0	0	0	0	0	0	0	0	0	0	0

附錄 8：TF-IDF 權重證候症狀對照表(症狀 16~30)

證候/症狀	唇色紫	喉乾	四肢冷	四肢厥冷	四肢拘急	四肢麻木	胸門遲閉	多夢	夜時小便次多	大便挾血	大便次多	大便質稀	妊娠不孕	子宮脫垂
心氣虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
心血虛	0	0	0	0	0	0	0	0.650275	0	0	0	0	0	0
心陰虛	0	0.607593	0	0	0	0	0	0.476868	0	0	0	0	0	0
心陽虛	0	0	0.543468	0	0	0	0	0	0	0	0	0	0	0
心陽暴脫	1	0	0	1	0	0	0	0	0	0	0	0	0	0
肝血虛	0	0	0	0	1	1	0	0.550232	0	0	0	0	0	0
肝陰虛	0	0.569618	0	0	0	0	0	0	0	0	0	0	0	0
肺氣虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
肺陰虛	0	0.479679	0	0	0	0	0	0	0	0	0	0	0	0
脾氣虛	0	0	0	0	0	0	0	0	0	0	0	0.574482	0	0
脾陽虛	0	0	0.486261	0	0	0	0	0	0	0	0	0.423303	0	0
脾氣陷	0	0	0	0	0	0	0	0	0	0	0.876812	0.473103	0	1
脾不統血	0	0	0	0	0	0	0	0	0	1	0	0.446819	0	0
腎精不足	0	0	0	0	0	0	1	0	0	0	0	0	0.840297	0
腎不納氣	0	0	0.61593	0	0	0	0	0	0	0	0	0	0	0
腎氣不固	0	0	0	0	0	0	0	0	0.852601	0	0	0	0	0
腎陰虛	0	0.276179	0	0	0	0	0	0.216758	0	0	0	0	0	0
腎陽虛	0	0	0.298031	0	0	0	0	0	0.522562	0	0.480833	0.259444	0.542127	0

附錄 9：TF-IDF 權重證候症狀對照表(症狀 31~45)

證候/症狀	小便不盡	小便失禁	小便挾乳糜	小便挾血	小便次多	小便色清	小便色黃	小便量多	小便量少	心悸	心痛	性欲低落	情緒憤怒	情緒神疲
心氣虛	0	0	0	0	0	0	0	0	0	0.565494	0	0	0	0.492449
心血虛	0	0	0	0	0	0	0	0	0	0.514085	0	0	0	0
心陰虛	0	0	0	0	0	0	0	0	0	0.376996	0	0	0	0
心陽虛	0	0	0	0	0	0	0	0	0	0.332644	0.635707	0	0	0.289676
心陽暴脫	0	0	0	0	0	0	0	0	0	0.403924	0.77193	0	0	0
肝血虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
肝陰虛	0	0	0	0	0	0	0	0	0	0	0	0	1	0
肺氣虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0.351749
肺陰虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
脾氣虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0.351749
脾陽虛	0	0	0	0	0	0	0	0	0.866622	0	0	0	0	0.259184
脾氣陷	0	0	1	0	0	0	0	0	0	0	0	0	0	0.289676
脾不統血	0	0	0	1	0	0	0	0	0	0	0	0	0	0.273583
腎精不足	0	0	0	0	0	0	0	0	0	0	0	0	0	0
腎不納氣	0	0	0	0	0	0	0	0	0	0	0	0	0	0.328299
腎氣不固	1	1	0	0	0.852601	0.852601	0	0.852601	0	0	0	0	0	0.259184
腎陰虛	0	0	0	0	0	0	1	0	0.498964	0	0	0	0	0
腎陽虛	0	0	0	0	0.522562	0.522562	0	0.522562	0	0	0	1	0	0.158854

附錄 10：TF-IDF 權重證候症狀對照表(症狀 46~60)

證候/症狀	情緒聲弱	意識健忘	意識恍惚	意識懶言	意識昏迷	意識癡呆	手足心熱	早洩	月經色淡紅	月經量多	月經量少	月經閉經	涎量少	潮熱
心氣虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
心血虛	0	0.737154	0	0	0	0	0	0	0	0	0	0	0	0
心陰虛	0	0.54058	0	0	0	0	0.607593	0	0	0	0	0	0	0.607593
心陽虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
心陽暴脫	0	0	0.819232	0	1	0	0	0	0	0	0	0	0	0
肝血虛	0	0	0	0	0	0	0	0	1	0	0.930408	0.79614	0	0
肝陰虛	0	0	0	0	0	0	0.569618	0	0	0	0	0	0	0.569618
肺氣虛	0.731055	0	0	0	0	0	0	0	0	0	0	0	0	0
肺陰虛	0	0	0	0	0	0	0.479679	0	0	0	0	0	0.866622	0.479679
脾氣虛	0	0	0	0.661811	0	0	0	0	0	0	0	0	0	0
脾陽虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
脾氣陷	0	0	0	0.545021	0	0	0	0	0	0	0	0	0	0
脾不統血	0	0	0	0.514742	0	0	0	0	0	1	0	0	0	0
腎精不足	0	0.405435	0.573462	0	0	1	0	0	0	0	0	0.517491	0	0
腎不納氣	0.682318	0	0	0	0	0	0	0	0	0	0	0	0	0
腎氣不固	0	0	0	0	0	0	0	0.765358	0	0	0	0	0	0
腎陰虛	0	0	0	0	0	0	0.276179	0.440661	0	0	0.366525	0.313631	0.498964	0.276179
腎陽虛	0	0	0	0	0	0	0	0.46909	0	0	0	0	0	0

附錄 11：TF-IDF 權重證候症狀對照表(症狀 61~75)

證候/症狀	牙齒動搖	牙齦出血	畏寒喜溫	畏風	痰挾血	痰色清	痰質稀	痰量少	痰難咳出	瘀血	白帶色清	白帶質稀	白帶量多	盜汗
心氣虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
心血虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
心陰虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0.607593
心陽虛	0	0	0.689855	0	0	0	0	0	0	0	0	0	0	0
心陽暴脫	0	0	0	0	0	0	0	0	0	0	0	0	0	0
肝血虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
肝陰虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0.569618
肺氣虛	0	0	0	1	0	1	1	0	0	0	0	0	0	0
肺陰虛	0	0	0	0	1	0	0	1	1	0	0	0	0	0.479679
脾氣虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
脾陽虛	0	0	0.617239	0	0	0	0	0	0	0	0	0.707107	1	0
脾氣陷	0	0	0	0	0	0	0	0	0	0	0	0	0	0
脾不統血	0	1	0	0	0	0	0	0	0	1	0	0	0	0
腎精不足	1	0	0	0	0	0	0	0	0	0	0	0	0	0
腎不納氣	0	0	0	0	0	0	0	0	0	0	0	0	0	0
腎氣不固	0	0	0	0	0	0	0	0	0	0	1	0.707107	0	0
腎陰虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0.276179
腎陽虛	0	0	0.378308	0	0	0	0	0	0	0	0	0	0	0

附錄 12：TF-IDF 權重證候症狀對照表(症狀 76~90)

證候/症狀	目乾	目暗	目眩	精子量少	耳聾	耳鳴	聲啞	胸悶	胸痛	脅痛	脈力度弱	脈寬度細	脈速度快	脈部位沉
心氣虛	0	0	0	0	0	0	0	0.733981	0	0	0.473175	0	0	0
心血虛	0	0	0.543758	0	0	0	0	0	0	0	0	0.432784	0	0
心陰虛	0	0	0.398756	0	0	0	0	0	0	0	0	0.317375	0.607593	0
心陽虛	0	0	0	0	0	0	0	0.431753	1	0	0.278339	0	0	0.543468
心陽暴脫	0	0	0	0	0	0	0	0.524272	0	0	0.337982	0.340045	0	0
肝血虛	0	0.776114	0.460103	0	0	0.668475	0	0	0	0	0	0.366202	0	0
肝陰虛	1	0.630593	0.373834	0	0	0.543136	0	0	0	1	0	0.297539	0.569618	0
肺氣虛	0	0	0	0	0	0	0	0	0	0	0.337982	0	0	0
肺陰虛	0	0	0	0	0	0	1	0	0	0	0.24904	0.250559	0.479679	0
脾氣虛	0	0	0	0	0	0	0	0	0	0	0.337982	0	0	0
脾陽虛	0	0	0	0	0	0	0	0	0	0	0.24904	0	0	0.486261
脾氣陷	0	0	0.351843	0	0	0	0	0	0	0	0.278339	0	0	0
脾不統血	0	0	0	0	0	0	0	0	0	0	0.262875	0.264479	0	0
腎精不足	0	0	0	1	1	0.434509	0	0	0	0	0	0.238031	0	0
腎不納氣	0	0	0	0	0	0	0	0	0	0	0	0.317375	0	0.61593
腎氣不固	0	0	0	0	0	0	0	0	0	0	0.24904	0.250559	0	0
腎陰虛	0	0	0.181253	0	0	0.263339	0	0	0	0	0	0.144261	0.276179	0
腎陽虛	0	0	0.192946	0	0	0	0	0	0	0	0.152637	0	0	0.298031

附錄 13：TF-IDF 權重證候症狀對照表(症狀 91~105)

證候/症狀	脫肛	腰冷	腰無力	腰痛	腰酸	腹痛	腹脹	膝痛	膝酸	自汗	舌形狀胖大	舌苔厚度薄	舌苔濕度滑	舌苔顏色白
心氣虛	0	0	0	0	0	0	0	0	0	0.659285	0	0	0	0.47491
心血虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
心陰虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
心陽虛	0	0	0	0	0	0	0	0	0	0.387814	0.689855	0	0.689855	0.279359
心陽暴脫	0	0	0	0	0	0	0	0	0	0	0	0	0	0
肝血虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
肝陰虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0
肺氣虛	0	0	0	0	0	0	0	0	0	0.470918	0	0	0	0.339222
肺陰虛	0	0	0	0	0	0	0	0	0	0	0	0	0	0.249953
脾氣虛	0	0	0	0	0	0	0.594851	0	0	0	0	0	0	0.339222
脾陽虛	0	0	0	0	0	1	0.438311	0	0	0	0.617239	0	0.617239	0.249953
脾氣陷	1	0	0	0	0	0	0.489877	0	0	0	0	0	0	0.279359
脾不統血	0	0	0	0	0	0	0.462662	0	0	0	0	0	0	0
腎精不足	0	0	0.462777	0	0.462777	0	0	0	0.462777	0	0	0	0	0.237455
腎不納氣	0	0	0.617036	0	0.617036	0	0	0	0.617036	0.439523	0	0	0	0.316607
腎氣不固	0	0	0.487134	0	0.487134	0	0	0	0.487134	0	0	0	0	0.249953
腎陰虛	0	0	0.280471	1	0.280471	0	0	1	0.280471	0	0	1	0	0.143912
腎陽虛	0	1	0.298566	0	0.298566	0	0	0	0.298566	0	0.378308	0	0.378308	0.153197

附錄 14：TF-IDF 權重證候症狀對照表(症狀 106~116)

證候/症狀	舌體顏色淡白色	舌體顏色紅色	舌體顏色紫色	遺尿	遺精	陽痿	難眠	面色暗	面色白	面色紅	面色黃	面色黧黑	頭暈	食欲低落	骨蒸	鬚髮脫落	鼻衄
心氣虛	0.415522	0	0	0	0	0	0	0	0.507504	0	0	0	0	0	0	0	0
心血虛	0.377747	0	0	0	0	0	0.778761	0.490202	0.461367	0	0	0	0.543758	0	0	0	0
心陰虛	0	0.607593	0	0	0	0	0.571092	0	0	0.607593	0	0	0.398756	0	0	0	0
心陽虛	0.244425	0	0	0	0	0	0	0.31719	0.298532	0	0	0	0	0	0	0	0
心陽暴脫	0	0	1	0	0	0	0	0.385159	0.362503	0	0	0	0	0	0	0	0
肝血虛	0.319632	0	0	0	0	0	0	0.414786	0	0	0	0	0.460103	0	0	0	0
肝陰虛	0	0.569618	0	0	0	0	0	0	0	0.569618	0	0	0.373834	0	0	0	0
肺氣虛	0.296802	0	0	0	0	0	0	0	0.362503	0	0	0	0	0	0	0	0
肺陰虛	0	0.479679	0	0	0	0	0	0	0	0.479679	0	0	0	0	0	0	0
脾氣虛	0.296802	0	0	0	0	0	0	0.385159	0	0	0.77193	0	0	0.594851	0	0	0
脾陽虛	0.218696	0	0	0	0	0	0	0	0.267107	0	0	0	0	0.438311	0	0	0
脾氣陷	0.244425	0	0	0	0	0	0	0.31719	0	0	0.635707	0	0.351843	0.489877	0	0	0
脾不統血	0.230846	0	0	0	0	0	0	0.299568	0.281947	0	0	0	0	0.462662	0	0	1
腎精不足	0.207761	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.855198	0
腎不納氣	0.277015	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
腎氣不固	0.218696	0	0	1	0.765358	0	0	0	0	0	0	0	0	0	0	0	0
腎陰虛	0	0.276179	0	0	0.440661	0	0.259587	0	0	0.276179	0	0	0.181253	0	1	0.518302	0
腎陽虛	0.134039	0	0	0	0.46909	1	0	0	0.163711	0	0	1	0.192946	0	0	0	0