

國立中正大學

醫療資訊管理研究所

碩士論文

自動化護理記錄產生-輔助記錄檢索與生成

Automated Nursing Record Generation-
Supporting Record Retrieval and Generation

研究生：黃曼瑄 撰

指導教授：吳 帆 博士

中華民國 114 年 6 月

致謝

本研究得以順利完成，首先要特別感謝我的指導教授吳帆博士。在研究期間，老師無論在系統設計、研究方向，皆給予專業且清晰的建議與實質的協助，使我能順利完成此篇論文。

其次，感謝參與專家評估的洗腎室與內科加護病房的護理師們，感謝您們提供寶貴的臨床意見與回饋，讓本研究更加貼近實務需求。

同時，衷心感謝研究所的同學們，謝謝你們在課業討論、專題協作中的相互幫助與一路陪伴，彼此鼓勵，讓這段研究歷程更加充實與順利。

此外，也特別感謝實驗室的學姊與學弟妹們，謝謝大家日常的互助合作與經驗分享，無論在技術或心理層面，都讓我獲益良多。

最後，我要深深感謝家人一直以來的理解與鼓勵，因為你們的支持，我才能安心且專注地完成這篇論文。

摘要

在臨床照護現場中，護理記錄不僅是病人安全與照護品質的重要依據，更是醫療團隊進行溝通與決策的關鍵，然而護理人力資源長期面臨嚴重短缺，尤其在醫學中心與地區醫院，日夜班護病比常高達 1:10 甚至 1:15，使得護理人員必須同時照顧多位病患，增加工作負荷與紀錄壓力。

為因應此一挑戰，本研究提出一套「自動化護理記錄產生系統」，結合自然語言處理（Natural Language Processing [NLP]）與全文檢索技術所發展出的語意向量檢索方法，協助護理人員快速從既有資料中擷取語意相近之歷史記錄，並依照標準化 DART（Data、Action、Response、Teaching）格式提供可參考與修改的建議內容，以提升紀錄效率並減輕第一線人員負擔。

本系統整合 Flask 框架、PostgreSQL 關聯式資料庫及 Elasticsearch 向量檢索引擎，並搭配 Sentence-BERT（all-MiniLM-L6-v2）語句嵌入模型，將使用者輸入的病人症狀轉換為語意向量，透過 cosine 相似度查詢方式，即時擷取語意最相近的歷史記錄。系統亦提供異常生理數據偵測功能與前端互動介面，使用者可針對建議內容進行挑選與修改後直接儲存。

本研究透過臨床專家問卷方式進行系統效能評估，邀請具有豐富護理經驗之專家實際評估系統回傳的推薦記錄。研究結果顯示，系統在資料選取階段與專家判斷間具高度一致性（Cohen's Kappa = 0.73~1.00），在記錄排序方面亦呈現中度一致性（Kendall's τ = 0.60~0.73），證實系統具備良好的臨床適用性與輔助決策之潛力。

關鍵字：自動化護理記錄、語意向量檢索、Sentence-BERT、Elasticsearch

Abstract

In clinical settings, nursing records are crucial for patient safety, care quality, and effective medical team communication. However, severe nurse shortages, especially in hospitals where nurse-to-patient ratios often reach 1:10 or even 1:15, substantially increase workloads and documentation pressure on nursing staff.

To address this, we developed an Automated Nursing Record Generation System that integrates Natural Language Processing (NLP) and semantic vector retrieval techniques. The system rapidly retrieves semantically similar historical records, offering editable recommendations structured by the standardized DART format (Data, Action, Response, Teaching) to streamline documentation and ease frontline workloads.

Our system combines Flask, PostgreSQL, and Elasticsearch with a Sentence-BERT embedding model (all-MiniLM-L6-v2) to convert user-entered patient symptoms into semantic vectors. It uses cosine similarity for real-time retrieval of the most relevant historical records and includes anomaly detection for physiological data and an interactive interface for easy modification and storage of records.

Clinical experts evaluated the system's performance, showing high agreement in record selection (Cohen's Kappa = 0.73–1.00) and moderate agreement in record ranking (Kendall's τ = 0.60 -0.73). These results demonstrate the system's strong clinical applicability and potential as a decision-support tool.

Keywords: automated nursing record, semantic vector retrieval, Sentence-BERT, Elasticsearch

目錄

致謝.....	i
摘要.....	ii
Abstract	iii
目錄.....	iv
圖目錄.....	vi
表目錄.....	vii
第一章 緒論.....	1
1.1 研究背景	1
1.2 研究動機	1
1.3 研究目的與問題	3
1.4 研究結果與貢獻	3
第二章 文獻探討.....	4
2.1 護理記錄系統與資訊系統概述	4
2.1.1 護理記錄法	5
2.2 自然語言處理應用於護理	6
2.3 全文檢索引擎(Full-Text Search Engine).....	9
第三章 系統演算法與建置.....	12
3.1 資料前處理與資料庫	12
3.2 系統演算法邏輯	12
3.2.1 檢索流程演算法	12
3.2.2 偵測異常生理數據邏輯	15

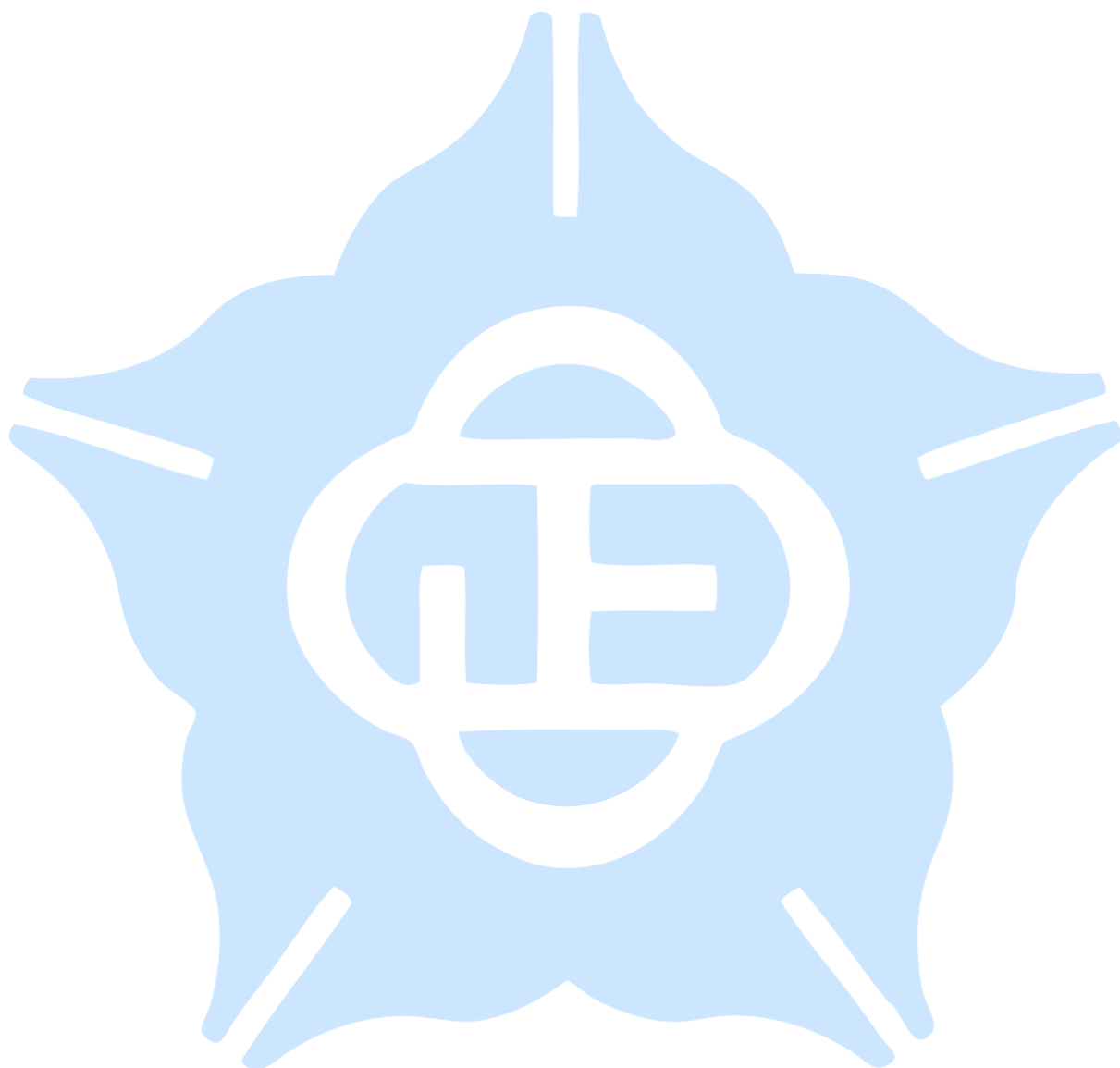
3.3 Flask 架構實作.....	16
3.4 系統流程	18
3.5 環境建置	20
3.6 護理記錄呈現介面	22
第四章 實驗結果與分析.....	23
4.1 實驗結果	23
4.2 系統與專家一致性分析	26
4.2.1 資料選取一致性分析	28
4.2.2 排序一致性分析	30
第五章 結論.....	35
5.1 結論	35
5.2 研究限制	35
5.3 未來發展與建議	36
參考文獻.....	37

圖目錄

圖 2.1 SBERT 訓練階段模型架構	8
圖 2.2 SBERT 應用階段模型架構	8
圖 2.3 Elasticsearch 文字資料處理流程.....	11
圖 3.1 Elasticsearch 索引設計與向量欄位設定.....	15
圖 3.2 研究架構圖	18
圖 3.3 系統活動圖（Activity Diagram）	19
圖 3.4 系統容器化服務設定.....	21
圖 3.5 病患基本資料呈現介面.....	22
圖 3.6 病患護理記錄呈現介面.....	22
圖 4.1 病患資料與症狀輸入畫面.....	24
圖 4.2 語意檢索結果.....	25
圖 4.3 洗腎室情境之專家問卷 1 結果.....	27
圖 4.4 內科加護病房情境之專家問卷 2 結果.....	27
圖 4.5 洗腎室情境之 Cohen's Kappa 一致性分析長條圖	29
圖 4.6 內科加護病房情境之 Cohen's Kappa 一致性分析長條圖	29
圖 4.7 洗腎室情境之 Kendall's τ 排序一致性散點圖	32
圖 4.8 洗腎室情境之 Kendall's τ 排序一致性散點圖	32
圖 4.9 內科加護病房情境之 Kendall's τ 排序一致性散點圖	33
圖 4.10 內科加護病房情境之 Kendall's τ 排序一致性散點圖	33

表目錄

表 3.1 檢索流程虛擬碼.....	14
表 3.2 生理數據異常值判斷虛擬碼.....	16



第一章 緒論

1.1 研究背景

隨著全球人口老化加劇和慢性患者數量的增加，醫療機構的護理工作量日益增大，護理人員的工作壓力隨之提升。在此背景下，智能技術的應用，特別是在護理記錄（Nursing Records）生成和病歷檢索方面，被視為提升護理效率和準確性（Accuracy）的可行途徑（World Health Organization, 2021）。這些技術結合了自然語言處理（Natural Language Processing [NLP]）和機器學習（Machine Learning [ML]），不僅能自動生成基於患者具體情況的護理記錄（Nursing Records），還可以有效協助護理人員檢索出相似病歷紀錄，輔助護理人員的決策（Topol, 2019）。

護理記錄對於病患的健康管理和醫療決策至關重要，但傳統的護理記錄撰寫過程耗時且容易受到人為疏忽的影響導致錯誤或不一致性。智能護理系統的引入，能通過自動化記錄生成來確保紀錄的一致性與準確性，減少人為錯誤，並且提高紀錄效率（Krittanawong et al., 2017）。此外，通過相似病歷檢索來提供個性化的照護建議，為護理人員減少工作負擔並提升照護品質（Dilsizian & Siegel, 2014）。

目前已有研究顯示，智能系統能夠透過加權相似度（Weighted Similarity）匹配技術，從大量病歷中檢索出相似病歷，並生成相應的護理建議（Esteva et al., 2017）。例如，對於一位年齡 95 歲、服用安眠藥的患者，系統能夠根據患者的年齡、藥物使用情況等條件從數萬筆資料中篩選出類似案例，並列出最相近的十筆供護理人員參考。

1.2 研究動機

為了可以更有效率並且減緩護理人員的工作負擔，智能的護理記錄系統可以減少輸入的成本，將護理人員的精力更多地投入到病人的直接照護中。傳統的護理記錄依賴護理人員手動輸入，這在繁忙的醫療環境中耗時耗力。護理人員往往需要花費大量時間填寫紀錄，容易

因疲勞和時間壓力而導致紀錄不完整或錯誤，進而影響護理品質和病患安全（Aiken, Clarke, Sloane, Sochalski, & Silber, 2014）。自動化護理記錄生成系統運用自然語言處理（Natural Language Processing [NLP]）和人工智慧（Artificial Intelligence [AI]）技術，可以根據護理人員的口述或選擇項目自動生成標準化紀錄，大幅提高了紀錄的效率和準確性（McGonigle & Mastrian, 2017）。例如，智齡科技的 Jubo 系統便是一種先進的自動化護理記錄工具，內建詞彙智慧語辭資料庫和多種記錄模組，使護理人員能通過模組引導完成紀錄。研究顯示，這種系統能夠節省護理人員約 30% 的紀錄時間，並降低因人工抄寫而產生的錯誤（智齡科技, 2024）。此外，這類系統的標準化功能確保了跨部門協作時紀錄的一致性，有助於醫療團隊更好地理解病患狀況，減少訊息差的風險，尤其在病情複雜的患者處理中顯得尤為重要。

另一個例子是中國醫藥大學附設醫院的「智海系統」，該系統通過語音辨別（Speech Recognition）和生成式 AI（Generative artificial intelligence [Generative AI]）技術，讓護理人員可口述病歷資訊，並由 AI 自動生成護理記錄。這一過程不僅減少了手動輸入的時間，還在提高照護效率的同時降低了人為錯誤的可能性，提升了整體照護品質（中國醫藥大學附設醫院, 2024）。與傳統記錄相比，這種系統使護理人員能更專注於病患照護，而非繁瑣的紀錄工作，從而實現了時間資源的最佳配置。

自動化護理記錄生成系統還具有輔助決策的潛力。Google Cloud 的 Medical Language Model（MedLM）模型引入了視覺分析功能，使護理人員能夠快速檢索歷史病歷並獲取相關建議，這對於需要快速決策的急診照護特別有用。相比之下，傳統記錄系統僅能提供單一病患的當前紀錄訊息，護理人員需要花費大量時間來查找歷史病歷，不利於快速做出決策。這類智能系統的引入，有助於護理人員在面對複雜病歷時得到更多支持資訊，提升決策的精確性和科學性。

綜上所述，現行醫療體系中的自動化護理記錄生成系統，提升紀錄效率、確保紀錄一致性、輔助護理決策，展現出相較於傳統護理記錄的多方面優勢。

本研究旨在開發一個具備自動生成紀錄、標準化護理記錄的智能護理系統，透過相似病歷檢索和記錄生成，有效減少護理人員的紀錄負擔，提升病人安全（Patient Safety）。

1.3 研究目的與問題

本研究的目的旨在設計一個智能護理記錄生成系統，結合自動化記錄生成、相似病歷檢索和標準化紀錄功能，以減少護理人員的紀錄負擔、提升病人安全，並提供精準的護理照護支持。這三項策略是開發此系統的有效方法，並將在文獻探討中分別闡述其優缺點及採用原因；文中亦將透過具體案例說明其運作方式。具體目標如下：

1. 減少護理人員的紀錄負擔：透過自動化生成紀錄，護理人員可顯著節省手動輸入的時間，將更多精力專注於病人照護上（Poissant, Pereira, Tamblyn, & Kawasumi, 2005）。
2. 提升病人安全：利用相似病歷檢索和記錄生成功能，為護理人員提供即時建議，協助其在短時間內做出護理決策，提升病人安全（Menachemi & Collum, 2011）。
3. 標準化護理記錄：自動生成的紀錄能按照標準格式，確保紀錄的一致性和精確性，減少人為錯誤，並支持後續數據分析。
4. 輔助護理決策：系統通過歷史數據的檢索，提供基於相似病人的建議，幫助護理人員在面對複雜或罕見情況下做出更具依據的決策，提升整體護理品質（Esteva et al., 2019）。

1.4 研究結果與貢獻

本研究結果減少護理人員的手動紀錄負擔，使他們能夠專注於病患照護，從而提升整體品質和效率。技術貢獻方面，本研究展示自動化護理記錄產生器的應用潛力，為智慧護理的推廣提供具體的實踐參考。同時，在病人安全性和決策支持方面，提供自動化記錄生成和相似病歷檢索功能將提供有力的決策支持，使護理工作更加高效、安全。

第二章 文獻探討

2.1 護理記錄系統與資訊系統概述

護理資訊系統（Nursing Information System [NIS]）是現代醫療機構的重要基礎設施，旨在提升效率與資料管理能力。作為一個綜合性的醫療資訊平台，NIS 包括了護理記錄系統（Nursing Record System [NRS]），並整合了其他模組，例如醫囑管理、檢驗數據與影像報告，實現全方位的醫療資訊管理能力（馮容莊 & 葉雨婷, 2014）。

護理資訊系統（Nursing Information System [NIS]）具備多項功能，能夠大幅提升醫療機構的運作效率與數據管理能力。以下是其主要特點的詳細說明：

1. 數據整合：與其他醫療系統連結，促進多部門的數據共享與協作（Chiang & Lee, 2018；楊漢淙 & 劉立, 2014）。
2. 輔助分析：透過數據分析工具，幫助提升護理品質與效率（Chiang & Lee, 2018；楊漢淙 & 劉立, 2014）。
3. 提升病人安全：提供如藥物過敏提醒的警示功能，減少潛在醫療錯誤（Chiang & Lee, 2018；楊漢淙 & 劉立, 2014）。

作為 NIS 的核心組成部分，護理記錄系統（Nursing Record System [NRS]）更專注於護理過程中的數據記錄，負責病患健康資訊、護理計畫與執行情況的詳細記錄。其核心特色包括：

1. 提升效率：幫助護理人員快速輸入與查詢數據，減少手寫紀錄所需的時間與人為錯誤風險（林育如等, 2019）。
2. 統一標準：確保紀錄的一致性與完整性，減少因人工輸入風格差異導致的數據不一致問題（林育如等, 2019）。
3. 支援決策：集中管理護理數據，為醫療團隊提供護理記錄以輔助診斷與決策（Zhai et al., 2022）。

護理記錄系統（NRS）負責病患健康資訊的紀錄與管理。護理記錄的結構與品質會影響 NRS 的應用成效，不同紀錄法在資料呈現的層次性、標準化程度及系統整合能力上各具差異，對於後續資料檢索、分析應用及介面設計均具有實質上的影響。以下為幾種常見護理記錄法。

2.1.1 護理記錄法

根據美國健康照護機構鑑定聯合委員會（Joint Commission on Accreditation of Healthcare Organizations [JCAHO]）於 1991 年實施的護理服務標準，護理紀錄應包括照護計畫與執行過程的完整記載，以確保照護品質與連續性。目前臨床常見之護理記錄方式主要包括問題導向護理記錄法（Problem-Oriented Nursing Record [PONR]）與焦點護理記錄法（Focus Charting）。

問題導向護理記錄法（Problem-Oriented Nursing Record [PONR]）為一種以病人健康問題為核心所發展之紀錄模式，常使用 S.O.A.P.（Subjective, Objective, Assessment, Plan）或 S.O.A.P.I.E.（Subjective, Objective, Assessment, Plan, Implementation, Evaluation）作為書寫結構，本法強調資料收集與邏輯判斷的連貫性，紀錄方式呈現問題與照護對應措施的因果關係，特別適合用於需長期追蹤或進行深入分析的臨床情境。不過，其紀錄過程相對繁瑣，因此對臨床人員而言在時間與書寫上有不小的負擔（黃芳子，2003）。

焦點護理記錄法（Focus Charting）則是由 Lampe 在 1985 提出，並發展為結構化的 DART 格式，分別紀錄病人的資料（Data）、護理措施（Action）、病人反應（Response）與衛教內容（Teaching）。此方法以病人當下照護重點為焦點，紀錄方式簡潔明瞭，有助於提升書寫效率並強化臨床紀錄的一致性，其格式固定特別適用於需快速回顧與標準化記錄的臨床環境，亦能作為資訊系統進行資料結構化與分析的良好依據（Lampe, 1985）。

不同記錄法各具有其特色與應用場景。本研究系統採用結構性良好、易標準化的 DART 紀錄法作為主要資料格式，並輔以向量語意分析技術，以達成病歷檢索與紀錄輔助生成之目標。

2.2 自然語言處理應用於護理

自然語言處理（Natural Language Processing [NLP]）技術作為人工智慧（Artificial Intelligence [AI]）的一個重要分支，憑藉著卓越的語言理解與生成能力，逐步改變各行各業的運作方式。NLP 技術不僅在處理結構化和非結構化數據提供解決方案，也為實現跨領域的智能應用上奠定基礎。

在醫療領域中，NLP 技術因其卓越的語言處理能力而備受關注。隨著深度學習（Deep Learning [DL]）和機器學習（Machine Learning [ML]）演算法的不斷進步，NLP 在醫療應用中展現了強大的潛力，特別是在電子健康記錄（Electronic Health Record [EHR]）的處理、臨床決策支援以及患者數據分析等方面。NLP 系統在語義檢索（Semantic Retrieval）和關鍵字匹配（Keyword Matching）上表現優異，能夠快速從大量文獻中檢索出相關資料，為各領域的應用提供有力支持（Liu & Lapata, 2019）。例如，基於深度學習模型，NLP 技術可以從海量 EHR 中高效提取相關病例，並生成簡潔明瞭的內容概要，幫助臨床醫生快速掌握關鍵資訊（Rajkomar, Dean, & Kohane, 2019；Li et al., 2021）。

例如，花蓮慈濟醫院與台灣人工智慧實驗室合作開發的「i 菩提」護理 AI 助手，是 NLP 技術在護理記錄生成中的成功應用實例。該系統結合了語音辨別和生成式人工智慧（Generative artificial intelligence [Generative AI]）技術自動生成標準化的護理紀錄，有效降低人工輸入造成的錯誤風險（花蓮慈濟醫院, 2024）。

本研究導入語句層級的語意向量技術，藉由結合自然語言處理與語句嵌入模型，提升相似護理記錄檢索的準確性與效率。

語句嵌入（Sentence Embedding）是將自然語言中的句子轉換為固定長度向量的重要技術，使得能夠計算出不同句子之間意思的相似程度，其中，Sentence-BERT（SBERT）為一種具代表性的嵌入模型。在傳統的 BERT 模型中，雖然具備強大的語言表示能力，但其主要設計是作為跨句編碼器（Cross-Encoder），需要成對輸入句子以進行相似性計算，因此在處理大量句子對比任務時效率較低。為了解決效率較低的問題，Reimers 和 Gurevych (2019) 提出了 Sentence-BERT（SBERT）模型，採用 Siamese 和 Triplet 網路架構，透過微調（fine-tune）

方式訓練，將 BERT 在語意理解上的優勢與向量表示方法結合起來，以此來提高句子相似度的計算效能。

SBERT 模型主要利用 Siamese network 的架構，對每個句子獨立地進行處理並產生固定長度的向量表示，如圖 2.1 所示，此模型在訓練階段使用了共享權重 (tied weights) 的雙 BERT 模型，將待處理的 Sentence A 和 Sentence B 兩個句子分別透過 BERT 編碼為初始向量序列，並透過池化 (Pooling) 層產生固定長度的句子向量 u 和向量 v 。之後將這兩個句子向量連同其絕對值的差 $|u - v|$ 串接後輸入到 softmax 分類器，以判斷句子之間的語意關係，並進行模型參數的微調訓練。透過這種方式，SBERT 可以學習到語意相似的句子向量空間表徵，使得語義相似的句子在向量空間中更加靠近，有助於後續語意檢索及分類任務 (Reimers & Gurevych, 2019)。

在實際應用與推論階段，SBERT 模型的架構與操作方式則如圖 2.2 所示。此階段中，模型不再需要進行參數更新，而是直接使用已訓練完成的模型來計算句子間的語意相似性。將任意兩個句子如圖中的 Sentence A 和 Sentence B，將它們輸入至共享權重的雙 BERT 模型進行編碼後，再經由池化層產生各自的句子向量 u 和向量 v 。最後透過計算兩個向量間的餘弦相似度 (cosine similarity)，快速得出這兩個句子之間的語意相似程度。這樣的架構在計算效率上具有明顯的優勢，透過簡單且高效的向量運算即可完成大量的語義比對工作，適合用於本研究所涉及的護理記錄檢索 (Reimers & Gurevych, 2019)。

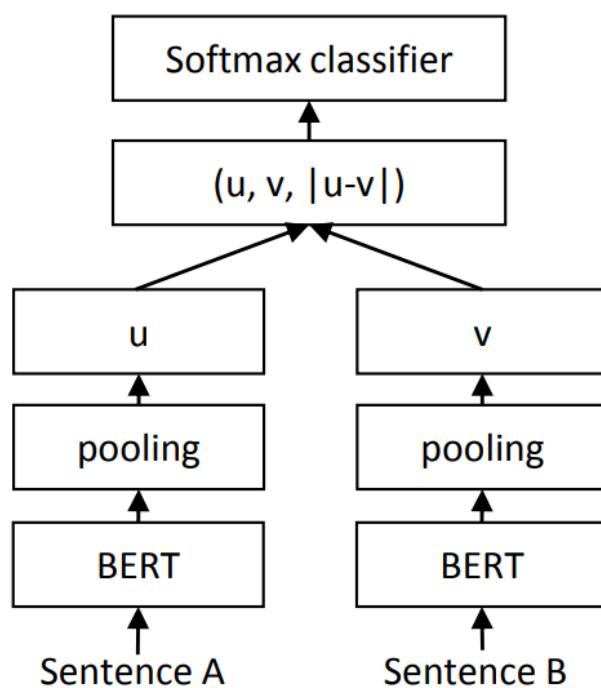


圖 2.1

SBERT 訓練階段模型架構

參考文獻：Reimers & Gurevych (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks.

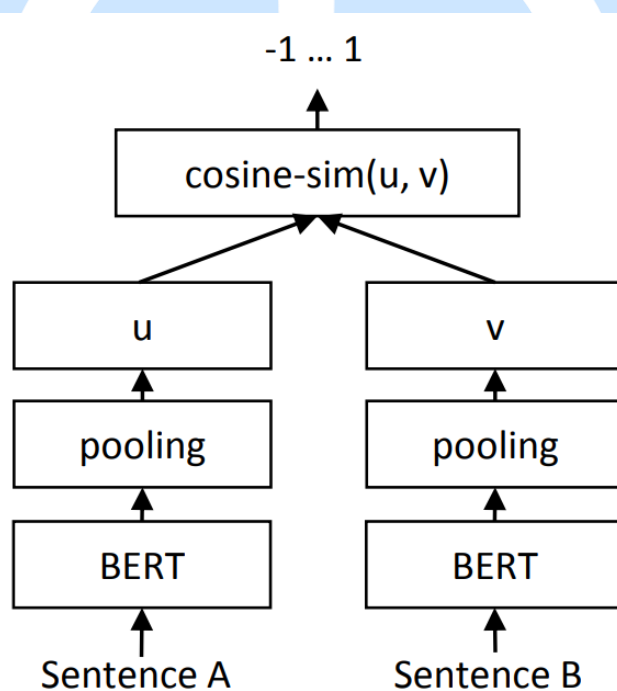


圖 2.2

SBERT 應用階段模型架構

參考文獻：Reimers & Gurevych (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks.

本研究所採用的模型「sentence-transformers/all-MiniLM-L6-v2」即為 SBERT 技術架構的延伸版本，使用更輕量級的 MiniLM 模型作為基礎，特別適用於實務場景中大量語句資料的語意檢索與比對。具體而言，本系統利用該模型將護理紀錄中「Data、Action、Response、Teaching (DART)」等欄位文字資料進行萃取，並將其轉化為 384 維度的固定長度向量，384 維度為該模型之預設輸出大小，以便後續透過 Elasticsearch 進行高效語意向量檢索。

2.3 全文檢索引擎(Full-Text Search Engine)

全文檢索引擎 (Full-Text Search Engine) 是一種高效率的資訊檢索工具，能夠快速地從大量的文字資料中檢索出使用者所需的內容，且不僅限於結構化資料，還包含非結構化及半結構化的文本資料。全文檢索引擎常見的索引方法有倒排索引 (Inverted Index) 及後綴索引 (Suffix Index)，並且透過智慧搜尋及個人化推薦技術，能有效提高使用者資訊存取的精確性及效率 (Wang, 2021)。倒排索引是一種基於關鍵字的索引方式，會將關鍵字與其在文檔中的位置資訊建立映射關係，因此能快速找到相關文件，但缺點是無法有效處理難以清晰分詞的資料，如生物資訊資料、日誌檔或感測資料等。為解決此問題，後綴索引則以文檔的所有後綴排列建立索引，能實現真正意義上的全文檢索，且保證檢索召回率達到 100%，特別適合於非結構化或異質性資料的處理需求 (Xu et al., 2021)。

在醫學領域中也有全文檢索引擎應用的例子，Garcelon 等人 (2017) 提出了一種改良式全文檢索方法，透過自然語言處理技術 (Natural Language Processing [NLP]) 辨識敘述性醫療報告中的否定詞 (negation) 和家族病史 (family history) 資訊，並將其有效地整合至臨床資料倉儲 (Clinical Data Warehouse [CDW])。此研究針對狼瘡伴隨腹瀉、克隆氏症伴隨糖尿病和 NPHP1 基因疾病這三種罕見疾病案例進行了完整的實驗驗證，結果顯示該方法能有效提升檢索的準確度和召回率，在實際應用上具有高度實用性，可協助研究人員迅速從龐大的電子病歷資料庫中識別出具有特定症狀與病史的患者，進而提高罕見疾病研究與診斷的效率 (Garcelon et al., 2017)。本研究在系統設計上導入 Elasticsearch 作為系統的全文檢索引擎，以提升非結構化護理紀錄資料的檢索效率與準確性。

Elasticsearch 是一款強大的分散式全文檢索與分析引擎，基於 Apache Lucene 所建構，廣泛應用於結構化資料、非結構化資料、地理位置、圖形及數值資料等多樣化數據處理。Elasticsearch 具有高度的即時性與可擴展性，能快速地查詢大量文件，提供接近即時的搜尋體驗，尤其適合處理大規模與異質性的數據 (Kathare et al., 2022) (Akca et al., 2016)。

如圖 2.3 所示，Elasticsearch 採用文件導向 (document-oriented) 的儲存方式，資料以 JSON 檔案的形式存放，並透過索引 (Indexing) 和映射 (Mapping) 等機制進行管理。索引的概念似於傳統資料庫中的資料庫概念，而檔案則相當於資料庫中的表格資料列。每份檔案包含多個欄位，每個欄位都會被建立成倒排索引 (Inverted Index)，以便在全文檢索時提供快速與高效的搜尋能力 (Kathare et al., 2022)。

在索引與搜尋過程中，Elasticsearch 使用分析器 (Analyzer) 對文字資料進行處理。分析器主要包含字元過濾器 (Character filter)、標記器 (Tokenizer) 與標記過濾器 (Token filter) 三個重要元件，如圖 2.3 所示。首先 Character filter 對文字資料進行初步的清理及轉換；接著 Tokenizer 負責將文字分割為獨立的詞彙單位 (Token)；最後，Token filter 會進一步處理這些詞元，例如將字母轉換為小寫或移除常見但意義不大的停用詞等，以提高搜尋的準確度與效率 (Kathare et al., 2022)。

此外，Elasticsearch 具有優秀的分散式架構能力。透過分片 (Shard) 與副本 (Replica) 的機制，能夠輕易將資料儲存分散於多個節點 (Node) 上，實現數據的水平擴展，並提高系統可用性與容錯能力。當系統中某一節點發生故障時，Elasticsearch 可自動啟用副本分片，確保數據完整性與服務不中斷。

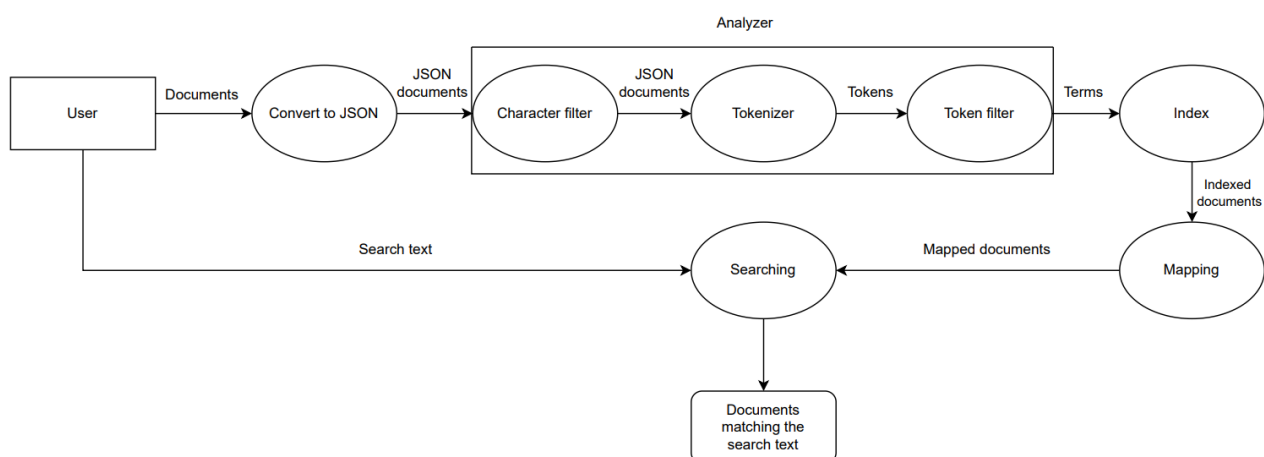


圖 2.3

Elasticsearch 文字資料處理流程

參考文獻：Kathare, N., Reddy, O. V., & Prabhu, V. (2022). A comprehensive study of Elastic Search. *Journal of Research in Science and Engineering (JRSE)*, 4(11), 34–38
[https://doi.org/10.53469/jrse.2022.04\(11\).07](https://doi.org/10.53469/jrse.2022.04(11).07)

第三章 系統演算法與建置

3.1 資料前處理與資料庫

本系統的核心功能分為兩大部分：基於相似病歷的檢索功能和病人資料管理功能。因此，系統採用了兩套專門設計的資料庫架構以支援不同功能需求：一個專注於檢索的資料庫和一個專注於病人資料管理的資料庫。

檢索資料庫主要存放用於相似記錄檢索的歷史記錄，資料以 JSON 格式存取於 Elasticsearch 檢索引擎中，為了專注於內容檢索，建置初期需進行資料前處理，以支援後續的語意檢索功能，主要目的是將過去累積的焦點護理紀錄資料轉換為可供語意向量比對的格式。原始資料內容包含病人識別編碼、所屬病房(ward)，以及依據 DART(Data、Action、Response、Teaching) 格式記錄的四段護理敘述。系統會先擷取每筆紀錄的四個欄位內容進行整合處理，接著使用預訓練語句嵌入模型將該筆文字資訊轉換為 384 維的語意向量，該維度為該模型之預設輸出大小。為了保留原始語意內容與支援語意檢索，系統會將原始紀錄內容與對應語意向量一併整合為一筆結構化的 JSON 物件，並寫入檢索資料庫中，檢索資料庫使用 Elasticsearch 為檢索引擎，其主要儲存經過語意向量轉換後的歷史焦點護理紀錄。

病人資料庫以 PostgreSQL 作為後端儲存平台，資料表結構涵蓋病人基本資訊、生理數據以及對應之護理焦點計畫，並透過欄位間主鍵與外鍵的關聯設計，以支援跨表查詢並確保資料一致性。

3.2 系統演算法邏輯

3.2.1 檢索流程演算法

本系統之核心功能為根據使用者輸入之症狀敘述，於檢索資料庫中找出語意最相近的歷史護理記錄。整體運作流程可分為三個主要步驟，如表 3.1 的語意檢索流程虛擬碼所示。

首先，系統會將使用者輸入的文字套入預先設計好的 DART 結構中，作為查詢模板。這個模板的「Data」欄位填入使用者的敘述，其餘欄位則先留空。

接著，系統呼叫語句嵌入模組 `generate_vector()`，將此樣板轉換為語意向量 `vec`。成功產生語意向量後，使用系統構建的 Elasticsearch 查詢語法 `es_body`，此查詢語法採用 `script_score` 配對機制，它結合 `cosineSimilarity` 函數計算查詢向量與索引中每筆記錄向量的相似度。索引設計方面如圖 3.1 所示，將向量欄位的 `index` 屬性設定為 `false`，意思是不使用向量索引技術（Approximate Nearest Neighbor [ANN]）而是採取暴力全量比對（Brute-Force Vector Search）的方法，其時間複雜度為 $O(N)$ 。由於處理的資料規模屬於中小型資料量，因此即使 $O(N)$ 查詢仍能在合理時間內完成，且暴力比對能確保不會遺漏最相關向量，結果精確無近似誤差。而相較之下，ANN 可將查詢複雜度降至次線性（約 $O(\log N)$ ），但其原理是先篩出若干候選向量再計算精確距離，因此召回率通常在 95 %~99 % 之間，仍有極小機會遺漏真正最相近的向量。因此本研究在現行資料規模下以暴力比對為優先選擇；若日後資料量倍增，再評估引入 ANN 加速。

最後，系統會根據語意相似程度找出最接近的前五筆歷史護理記錄。這些記錄會依相似度由高至低排序，並回傳至使用者介面供參考與選擇。若查詢過程中發生錯誤，系統會自動返回空結果，避免影響整體操作流程。

透過這樣的語意比對與自動查詢機制是為了確保即使面對使用者輸入形式不一的自然語言描述時，系統仍能有效辨識其意圖並提供具參考價值的歷史記錄來有效支援護理人員的工作。

表 3.1

 檢索流程虛擬碼

Input:

keyword_txt

 $k \leftarrow 5$

Output:

results_list // Up to k records, from ES_source

```

1.  PROC DART_SEMANTIC_RETRIEVAL(keyword_txt, k):
2.    query_record ← {
3.      "DART": {"Data" : keyword_txt, "Action": "", "Response": "", "Teaching": ""}
4.    }
5.    TRY:
6.      vec ← generate_vector(query_record)
7.    EXCEPT Exception:
8.      RETURN [ ]
9.    END TRY
10.   es_body ← {
11.     "size": k,
12.     "query": {
13.       "script_score": {
14.         "query": {"match_all": {}},
15.         "script": {
16.           "source": "cosineSimilarity(params.qv,'vector') + 1.0",
17.           "params": {"qv": vec.tolist()}
18.         }
19.       }
20.     }
21.   }
22.   TRY:
23.     es_res ← es_client.search(index="internal_medicine", body=es_body)
24.   EXCEPT Exception:
25.     RETURN [ ]
26.   END TRY
27.   hits ← es_res["hits"]["hits"]
28.   results_list ← [ hit["_source"] for hit in hits ]
29.   RETURN results_list
30. END PROC

```

參考文獻：本研究自行整理


```

"mappings": {
  "properties": {
    "vector": {
      "type": "dense_vector",
      "dims": 384,
      "index": false,
      "similarity": "cosine"
    },
    "patientID": {"type": "keyword"},
    "ward": {"type": "text"},
    "DART": {
      "properties": {
        "Data": {"type": "text"},
        "Action": {"type": "text"},
        "Response": {"type": "text"},
        "Teaching": {"type": "text"}
      }
    }
  }
}

```

圖 3.1

Elasticsearch 索引設計與向量欄位設定

參考文獻：本研究自行整理

3.2.2 偵測異常生理數據邏輯

本研究設計一套即時生理數據異常判斷邏輯，並用紅字標示異常項目，輔助使用者快速辨識潛在風險，並自動將異常結果標註於記錄中以供檢索，作為後續語意檢索之依據。異常判斷包括體溫（Temperature）、脈搏（Heart Rate）、呼吸速率（Respiration Rate）、血壓（Blood Pressure），等四項指標，其邏輯條件以虛擬碼（pseudocode）形式說明，詳如表 3.1 所示。

本邏輯模組應用於系統前端之「新增病人資料」介面中，當使用者輸入病人生理數據後，系統會自動比對判斷條件並進行異常值偵測，若有異常即即時以紅字提示於畫面中呈現，並同步自動填入 DART 記錄中的 Data 欄位。此設計除可協助使用者即時察覺潛在風險外，亦能有效提升紀錄撰寫效率

表 3.2

生理數據異常值判斷虛擬碼

```
1. //Temperature
2. if temperature  $\geq$  38.0 then
3.   Abnormalities. add ("fever")
4. else if temperature  $<$  35.0 then
5.   Abnormalities. add ("hypothermia")
6. end if
7. //Heart Rate
8. if pulse  $<$  60 then
9.   Abnormalities. add ("bradycardia ")
10. else if pulse  $>$  100 then
11.   Abnormalities. add ("tachycardia ")
12. end if
13. // Respiration rate
14. if respiration  $<$  12 then
15.   Abnormalities. add ("bradypnea ")
16. else if respiration  $>$  20 then
17.   Abnormalities. add ("tachypnea")
18. end if
19. //Blood Pressure
20. if systolic  $\neq$  0 and diastolic  $\neq$  0 then
21.   if systolic  $<$  90 or diastolic  $<$  60 then
22.     Abnormalities. add ("hypotension")
23.   else if systolic  $>$  140 or diastolic  $>$  90 then
24.     Abnormalities. add ("hypertension")
25.   end if
```

參考文獻：本研究整理

3.3 Flask 架構實作

本研究系統後端以 Python Flask 框架進行開發，整體設計強調模組化、分層化，將系統各項核心功能比如說資料儲存、語意向量處理與檢索等整合至單一伺服器，並透過 RESTful API 對外提供標準化的存取與操作服務，以利與前端介面進行即時的資料互動。

系統架構設計如圖 3.2 所示，共分為兩層架構。第一層為 Flask 主應用程式，主要功能為啟動伺服器、設定系統基礎環境，以及載入並整合所有路由模組；第二層則由路由模組組成，負責處理所有 HTTP 請求與響應，作為系統邏輯的核心樞紐。當使用者透過前端介面發出請求後，路由模組會依據使用者所請求的 URL 路徑與 HTTP 協定中的 GET 與 POST 功能呼叫對應之功能模組，完成資料處理與回傳結果。整體系統設計的主要功能區塊如下：

1. 病人資料管理：透過前端表單輸入資料後，系統將接收並解析請求，接著路由模組會呼叫 PostgreSQL 資料庫介面，將使用者所輸入的病人資訊（如基本資料、生理數據與護理紀錄內容）進行新增或更新。查詢操作則透過 HTTP 的 GET 方法從資料庫取得相應資料，並以 JSON 格式回傳至前端頁面呈現。
2. 生理數據異常偵測：當使用者在前端頁面輸入病人之生理數據後，系統即時將這些數據提交至路由模組，並依據預設的異常判斷邏輯，分析輸入數據是否存在異常狀態。一旦偵測到異常，系統會自動在前端即時顯示紅色提示字樣，並同步將異常狀態標記自動填入 DART 記錄的 Data 欄位內，以利後續檢索使用。
3. 護理記錄檢索（DART 檢索）：使用者在系統前端輸入病人症狀或其他敘述內容後，系統將該文字透過路由模組傳遞至向量處理模組，透過 Sentence-BERT 語句嵌入模型轉換為固定長度之語意向量，隨後送至 Elasticsearch 全文檢索引擎，以 script_score 搭配 cosine similarity 方式，逐筆計算該向量與資料庫中歷史護理紀錄向量的相似性，最後返回語意相似度最高的前五筆紀錄供使用者參考。
4. 護理記錄儲存：當使用者於前端介面確認或修改了系統推薦的紀錄後，路由模組將接收處理後的紀錄內容，並將其儲存至 PostgreSQL 病人資料庫，形成病人的最新護理紀錄歷程，以便後續查詢或分析。

系統前端介面則採用 HTML 配合 Flask 所使用的 Jinja2 模板語法進行開發，支援動態資料渲染與條件式顯示，可根據後端回傳的資料動態調整網頁內容，達到即時互動與資料同步的效果。

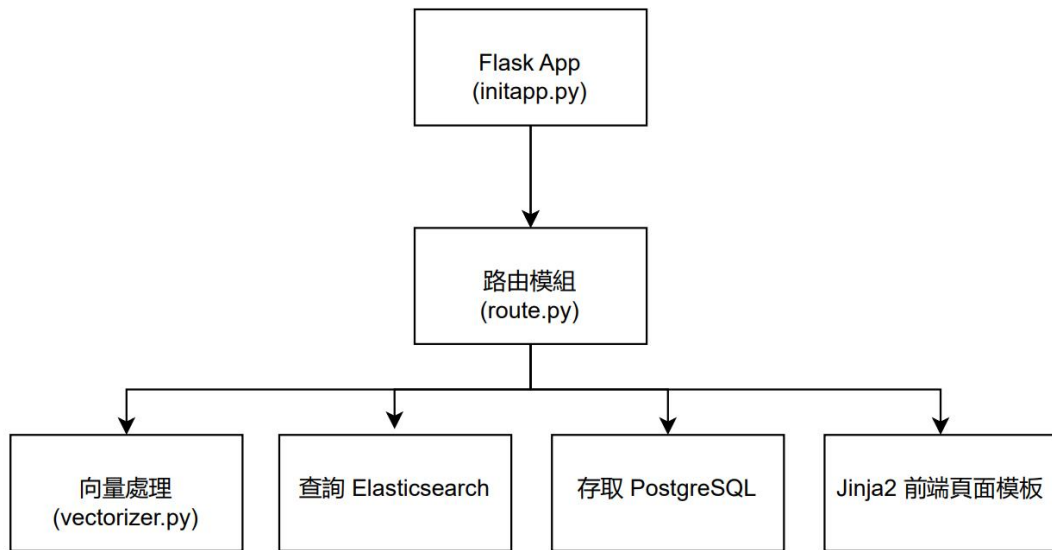


圖 3.2

研究架構圖

參考文獻：本研究自行繪製

3.4 系統流程

本研究透過建立一個自動化的護理記錄檢索生成系統，以智能化技術解決傳統護理記錄中手動操作帶來的高負擔與易錯問題。整體運行流程緊密圍繞「自動化生成護理記錄」與「相似記錄檢索」兩大核心功能設計。系統的運作機制如圖 3.1 所示，以活動圖 (Activity Diagram) 呈現系統的運作方式與使用流程。

首先，護理人員需在系統介面上輸入患者的基本資料與相關症狀描述，例如「65 歲男性，持續咳嗽三個禮拜了，曾使用過抗生素」。系統使用 Sentence-BERT 語句嵌入模型將這些文字資料轉換為語意向量，並結合 Elasticsearch 全文檢索引擎進行語意相似度比對。透過這種方式，系統能夠從資料庫中快速檢索出與輸入內容語意最相近的歷史護理記錄。

完成資料的處理後，系統會啟動相似病歷的檢索功能。基於患者輸入的條件，系統使用分散式檢索引擎 Elasticsearch，從大量護理記錄資料中找到與患者條件相符的記錄。檢索過程中，系統採用多層次的匹配技術，包括關鍵詞比對與語意相似度計算，確保返回的記錄符合

表面條件。

檢索結果按相似度排序，將與患者條件最為相近的十條護理記錄以列表形式呈現。生成格式採用護理記錄中常用的 DART 記錄方法，以結構化方式生成護理記錄。

最後，生成的護理記錄會通過介面呈現，護理人員可根據他的需求來選擇或修改記錄內容，並將最終版本儲存到資料庫中。

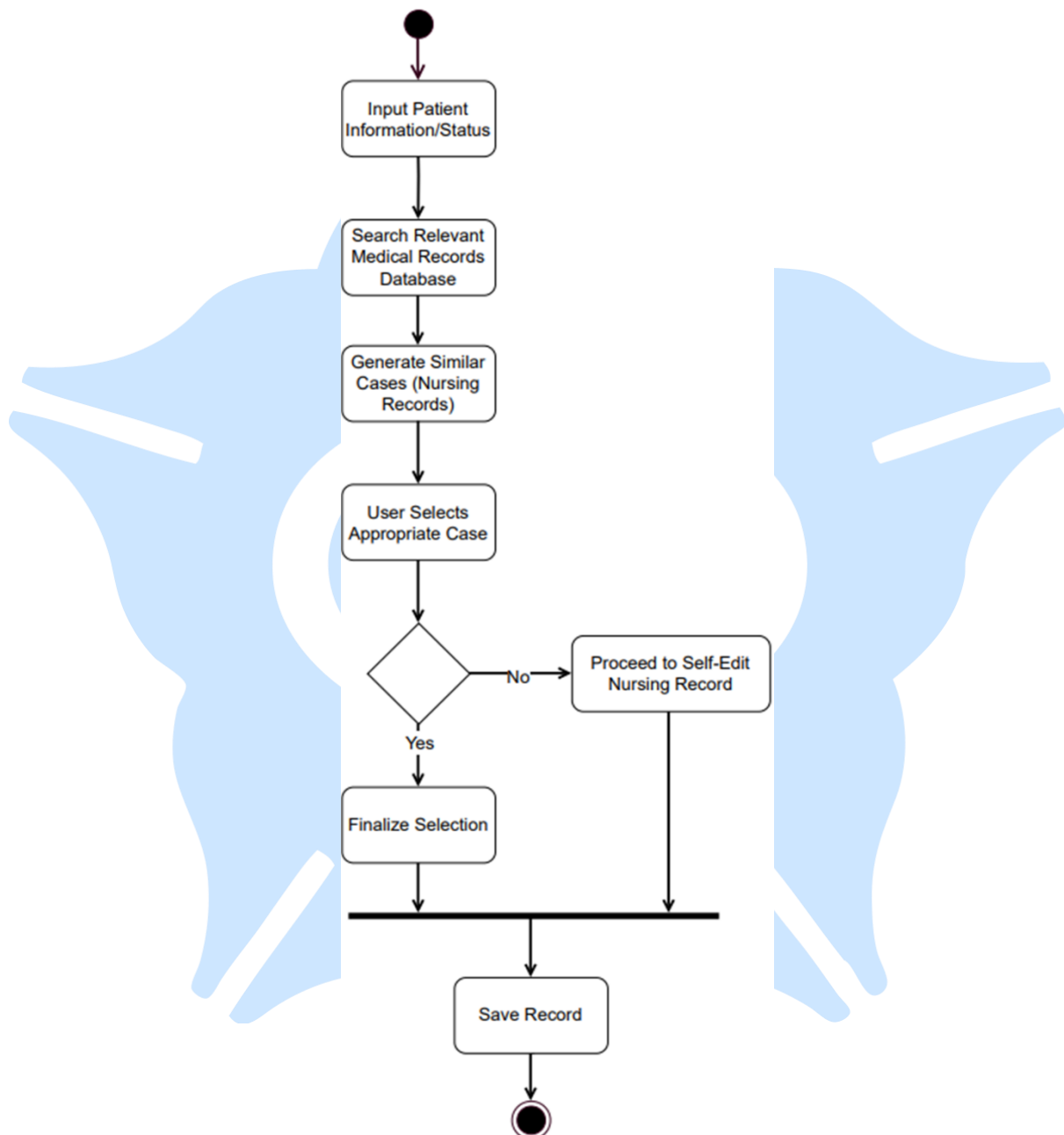


圖 3.3

系統活動圖 (Activity Diagram)

參考文獻：本研究自行繪製

3.5 環境建置

為確保本系統具備穩定且便於部署的執行環境，本研究採用容器化技術搭配輕量級伺服器框架進行後端建置。透過 Docker Compose 的多容器協同部署功能，統一管理包括資料庫與全文檢索引擎等核心服務，不僅簡化了系統部署流程，亦提升了整體系統的可移植性。各項服務可藉由 `docker-compose.yml` 這個設定檔案來進行配置與建立，進一步確保服務之間的相依性與正確執行順序。如圖 3.3 所示在本系統中 Docker Compose 共定義以下三個主要服務：

1. **Elasticsearch**：作為全文檢索與向量相似度查詢引擎，支援護理紀錄的語意檢索功能。服務使用映像（image）版本 8.5.1，容器開放 9200 埠進行 HTTP 通訊、9300 埠進行叢集通訊，設置單節點模式（`discovery.type=single-node`）。將 `bootstrap.memory_lock` 為設為 `true` 及設定虛擬機器記憶體參數（`ES_JAVA_OPTS=-Xms512m -Xmx512m`），以確保在執行語意向量查詢時具有穩定的效能。此外，Elasticsearch 資料目錄掛載至主機的 `esdata` Volume 以實現資料持久化。
2. **PostgreSQL**：作為主要資料庫系統，用於儲存病人基本資料、生理數據及護理紀錄。服務使用映像（image）版本 14，資料庫服務預設開放 5432 埠。而在環境變數設定部分，將容器名稱設為 `my_postgres`，並設定使用者名稱、密碼跟初始資料庫名稱。為確保資料持久化，資料庫資料目錄掛載於主機的 `./postgres_data` 目錄，對應至容器內部的 `/var/lib/postgresql/data`。
3. **pgAdmin**：提供 Web 介面的 PostgreSQL 資料庫管理工具，協助開發人員可視化操作資料庫。服務綁定在容器內部的 5050 埠，並設定初始登入信箱及密碼。

```

45  >Run Service
46  elasticsearch:
47    image: docker.elastic.co/elasticsearch/elasticsearch:8.5.1
48    environment:
49      - discovery.type=single-node
50      - xpack.security.enabled=false
51      - xpack.security.http.ssl.enabled=false
52      - xpack.security.transport.ssl.enabled=false
53      - bootstrap.memory_lock=true
54      - "ES_JAVA_OPTS=-Xms512m -Xmx512m"
55    ulimits:
56      memlock:
57        soft: -1
58        hard: -1
59    ports:
60      - "9200:9200"
61    volumes:
62      - esdata:/usr/share/elasticsearch/data
63    healthcheck:
64      test: ["CMD-SHELL", "curl -f http://localhost:9200 || exit 1"]
65      interval: 30s
66      timeout: 10s
67      retries: 5
68
69  >Run Service
70  postgres:
71    image: postgres:14
72    container_name: my_postgres
73    restart: always
74    environment:
75      POSTGRES_USER: mythesis
76      POSTGRES_PASSWORD: thesis1234
77      POSTGRES_DB: mydatabase
78    ports:
79      - "5432:5432" # 開放 PostgreSQL 端口
80    volumes:
81      - postgres_data:/var/lib/postgresql/data
82
83  >Run Service
84  pgadmin:
85    image: dpage/pgadmin4
86    container_name: my_pgadmin
87    restart: always
88    environment:
89      PGADMIN_DEFAULT_EMAIL: thesis@example.com
90      PGADMIN_DEFAULT_PASSWORD: thesis1234
91    ports:
92      - "5050:80" # pgAdmin 界面使用 5050 端口來訪問

```

圖 3.4

系統容器化服務設定

參考資料：本研究自行整理

3.6 護理記錄呈現介面

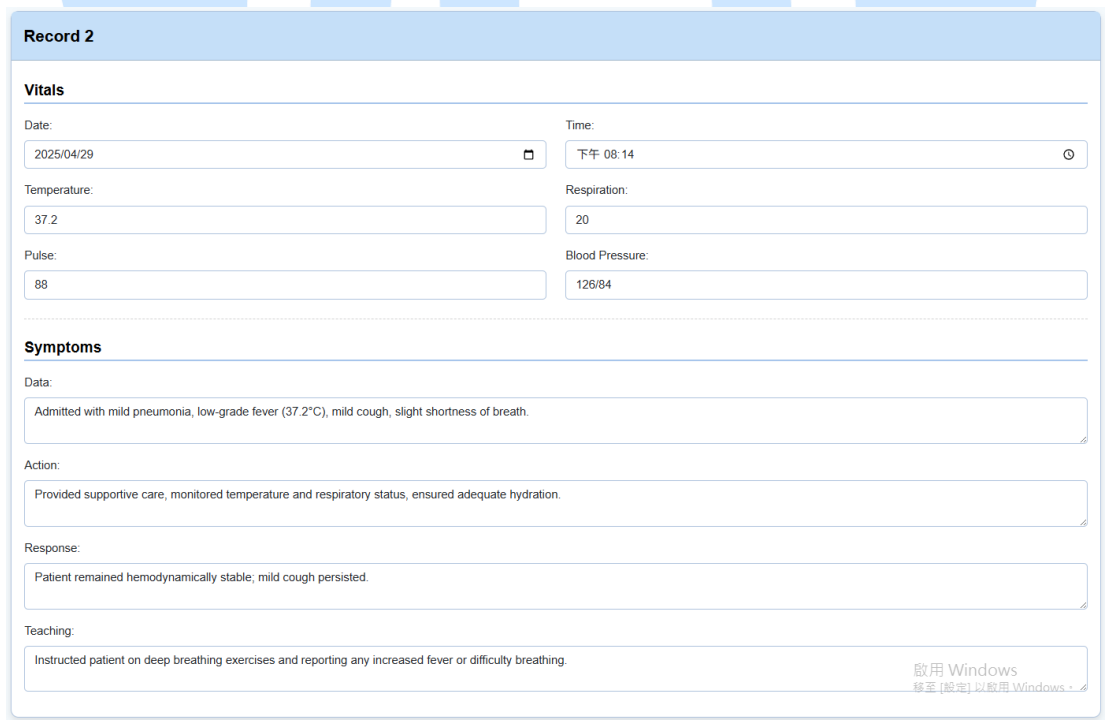
本研究設計之護理紀錄呈現介面如圖 3.5 與圖 3.6 所示，整體頁面可區分為兩大區塊，分別為個人資訊 (Personal Information) 與紀錄內容 (Record)。使用者可於介面中檢視並修改病患之生理資訊與語意檢索結果，系統亦支援內容編輯與即時更新，以協助護理人員依實際臨床狀況調整紀錄內容。



Personal Information	
Patient ID:	Patient Name:
<input type="text" value="A123456789"/>	<input type="text" value="王曉明"/>
Gender:	Age:
<input type="radio"/> 女 <input checked="" type="radio"/> 男	<input type="text" value="66"/>
Ward Name:	
<input type="text" value="Pulmonary001"/>	

圖 3.5
病患基本資料呈現介面

參考文獻：本研究自行整理



Record 2	
Vitals	
Date:	Time:
<input type="text" value="2025/04/29"/>	<input type="text" value="下午 08:14"/>
Temperature:	Respiration:
<input type="text" value="37.2"/>	<input type="text" value="20"/>
Pulse:	Blood Pressure:
<input type="text" value="88"/>	<input type="text" value="126/84"/>
Symptoms	
Data:	
<input type="text" value="Admitted with mild pneumonia, low-grade fever (37.2°C), mild cough, slight shortness of breath."/>	
Action:	
<input type="text" value="Provided supportive care, monitored temperature and respiratory status, ensured adequate hydration."/>	
Response:	
<input type="text" value="Patient remained hemodynamically stable; mild cough persisted."/>	
Teaching:	
<input type="text" value="Instructed patient on deep breathing exercises and reporting any increased fever or difficulty breathing."/>	

圖 3.6
病患護理記錄呈現介面

參考文獻：本研究自行整理

第四章 實驗結果與分析

4.1 實驗結果

本研究語意檢索模組於實際操作流程中的回傳結果樣態如圖 4.1、4.2 所示，說明系統如何依據使用者輸入之症狀敘述，自動比對既有記錄並依語意相似度排序顯示。

輸入之病患資料如圖 4.1 所示，患者為 77 歲男性，主訴咳嗽與輕微呼吸困難，系統偵測到體溫達 38.0°C，自動標示為異常，並加入「fever」至症狀敘述中。完成輸入後，系統立即執行語句向量化與語意比對，並於數秒內回傳相似紀錄。

檢索後回傳的前三筆記錄如圖 4.2 所示，整體語意內容與本例輸入有高度重疊，尤以第一筆紀錄最為貼近，包括輕微肺炎、低燒與輕度呼吸症狀等敘述，符合度甚高。第二、三筆紀錄亦屬肺炎個案，但症狀表現程度有所差異，例如黃痰、倦怠或體溫更高，顯示系統仍能保有語意延展的彈性。

初步觀察檢索結果可見，系統回傳之紀錄與原始症狀敘述在語意上呈現一定程度的重疊與延伸，排序順序亦與臨床直覺有所呼應。然而其排序合理性與臨床適切性仍有待進一步專家評估，後續將於下一節進行一致性分析，以深入探討其應用價值。

Personal Information

Patient ID : D123456789

Gender : ☐女 ☒男

Patient Name : 陳大銘

Age : 77

Date : 2025/06/24

Time : 上午 12:03

Ward : Pulmonary666

Temperature : 38 °C

Respiration : 20 bpm

Pulse : 88 beat/min

Blood Pressure : 126/84 mmHg

Generate Symptoms

Symptom Details

Patient has been coughing a bit and feels a little short of breath. Slight fever observed, might be pneumonia. (fever)

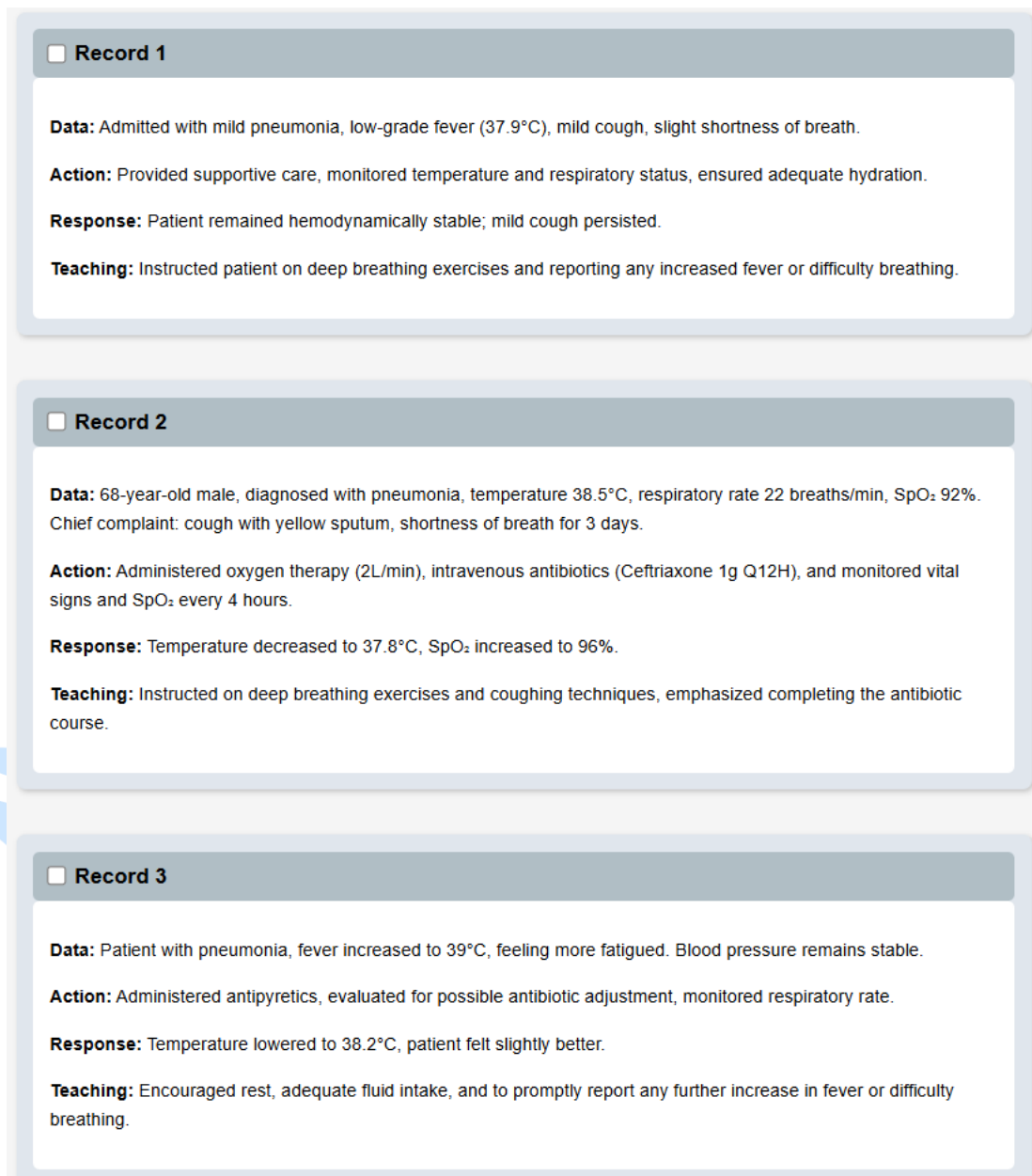
Retrieval

啟用 W
移至 [設定]

圖 4.1

病患資料與症狀輸入畫面

參考文獻：本研究自行整理



☐ **Record 1**

Data: Admitted with mild pneumonia, low-grade fever (37.9°C), mild cough, slight shortness of breath.

Action: Provided supportive care, monitored temperature and respiratory status, ensured adequate hydration.

Response: Patient remained hemodynamically stable; mild cough persisted.

Teaching: Instructed patient on deep breathing exercises and reporting any increased fever or difficulty breathing.

☐ **Record 2**

Data: 68-year-old male, diagnosed with pneumonia, temperature 38.5°C, respiratory rate 22 breaths/min, SpO₂ 92%. Chief complaint: cough with yellow sputum, shortness of breath for 3 days.

Action: Administered oxygen therapy (2L/min), intravenous antibiotics (Ceftriaxone 1g Q12H), and monitored vital signs and SpO₂ every 4 hours.

Response: Temperature decreased to 37.8°C, SpO₂ increased to 96%.

Teaching: Instructed on deep breathing exercises and coughing techniques, emphasized completing the antibiotic course.

☐ **Record 3**

Data: Patient with pneumonia, fever increased to 39°C, feeling more fatigued. Blood pressure remains stable.

Action: Administered antipyretics, evaluated for possible antibiotic adjustment, monitored respiratory rate.

Response: Temperature lowered to 38.2°C, patient felt slightly better.

Teaching: Encouraged rest, adequate fluid intake, and to promptly report any further increase in fever or difficulty breathing.

圖 4.2

語意檢索結果

參考文獻：本研究自行整理

4.2 系統與專家一致性分析

為了評估本研究所建構之護理記錄生成系統在不同臨床情境中的泛用性與適用性，本研究設計了兩份專家問卷，分別邀請具洗腎室與內科加護病房之臨床經驗的護理專家參與系統結果的回饋與評估。洗腎室專家由新竹某醫院之洗腎室護理師擔任，兩位護理師分別有 10、20 年以上的工作經驗，內科加護病房則邀請兩位具加護病房工作背景之新竹某醫院護理師進行評估。兩份問卷操作情境皆為：系統接受使用者輸入之病人主訴，其中包含括號內容，括號內容是系統根據生理數據自動偵測並補充之資訊，系統將整體輸入語句進行語意分析，根據語意相似度從資料庫中檢索出前五筆建議之 DART 記錄，供臨床使用者修改與使用，以提升記錄撰寫之效率與準確性。

每位專家在問卷中需進行兩項任務，第一項為系統提供的 20 筆候選記錄中，選出最符合該病人狀況的 5 筆記錄；再將這 5 筆記錄依照其臨床重要性排序。兩份專家問卷結果分別如圖 4.3 和圖 4.4 所示，圖 4.3 為洗腎室之情境問卷結果，圖 4.4 為內科加護病房之問卷結果，橫軸為記錄編號，縱軸為獲得票數；不同顏色代表資料來源，數字則為該來源對該筆記錄的排序名次（數字愈小表示排序愈前面）。為了客觀評估本系統與臨床專家判斷之間的一致性，故將專家所選取與排序之紀錄與系統自動生成的建議結果進行比對分析，並分別針對「選取哪些記錄」與「排序先後順序」兩個層面進行探討。

在資料選取部分，專家與系統皆須從 20 筆候選記錄中挑選出最具臨床參考價值的前五筆資料。為檢驗雙方在選取項目上的判斷一致性，採用 Cohen's Kappa 一致性係數（科恩一致性係數）進行分析。Cohen's Kappa 適用於二元分類問題，可分析系統與專家在「是否選取某筆記錄」上的整體一致程度。

接著，針對雙方所共同選取的紀錄，進一步分析其排序的一致性，亦即判斷系統在排序上是否能擬仿專家對資料重要性的評估邏輯。採用 Kendall's τ （Kendall rank correlation coefficient）作為排序一致性評估指標，此指標可衡量兩組排名之間的秩序相關性，越接近 1 代表排序越一致，越接近 -1 則代表排序方向相反。

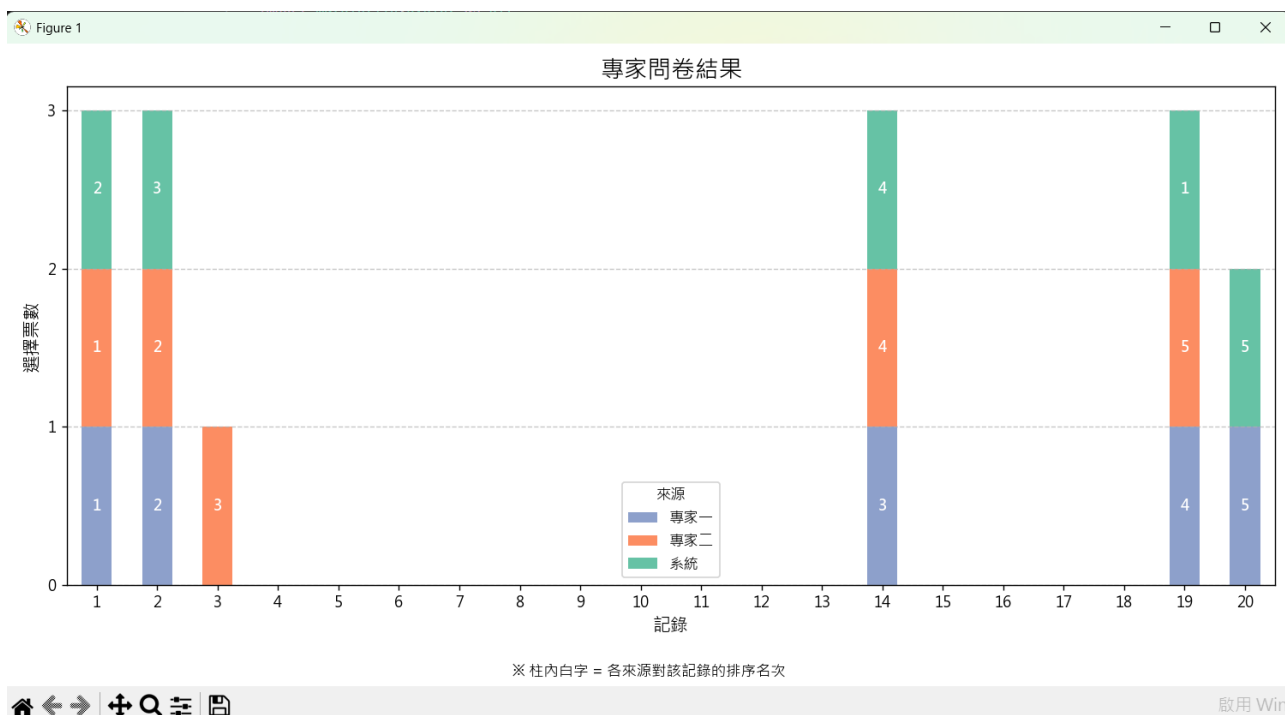


圖 4.3

洗腎室情境之專家問卷 1 結果

參考文獻：本研究自行整理

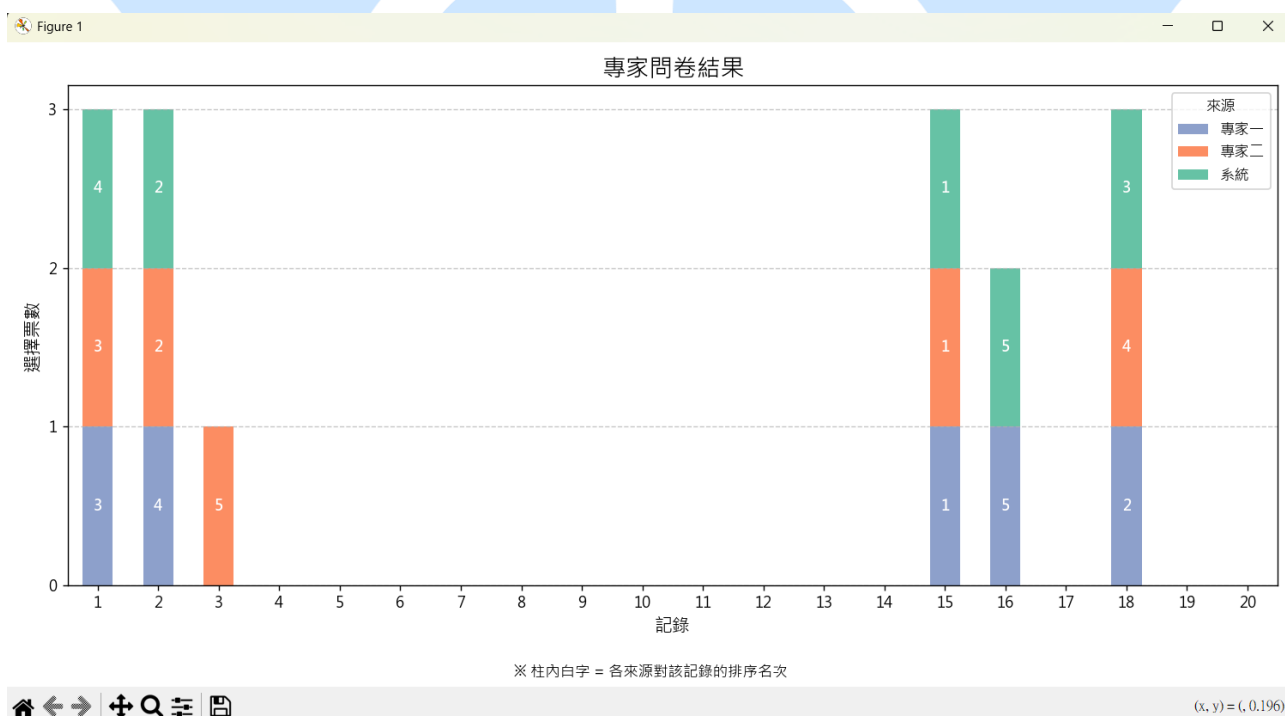


圖 4.4

內科加護病房情境之專家問卷 2 結果

參考文獻：本研究自行整理

4.2.1 資料選取一致性分析

Cohen's Kappa 一致性係數是一項常用於檢驗觀察者間判斷結果一致性的統計工具，尤其在醫療人工智慧系統評估中廣受應用，其計算公式如下：

$$k = \frac{p_0 - p_e}{1 - p_e}$$

其中， p_0 為觀察到的實際一致比例， p_e 為依機率推估之預期一致比例。Kappa 值越接近 1，表示一致性越高；若小於 0，則可能代表一致性低於隨機水準。

根據 Kromrey 等人 (2024) 發表於 Healthcare 期刊的研究，他們採用 Cohen's Kappa 作為衡量指標，以評估機器學習輔助之 AI 系統在臨床胸腔 X 光判讀上與放射科醫師之診斷結果間的一致性，並此來判斷該系統在實務情境中的應用潛力 (Kromrey et al., 2024)。本研究參考此一驗證邏輯，選擇 Cohen's Kappa 作為系統與臨床專家在資料選取階段的一致性分析工具，以評估系統所挑選之 DART 記錄是否與專家判斷一致，進而佐證系統在輔助臨床決策上的可行性與可信度。

根據本研究之洗腎室情境設計，由圖 4.3 可知專家 1 選擇了記錄 1、2、14、19、20；專家 2 選擇了記錄 1、2、3、14、19；系統則選擇了記錄 1、2、14、19、20，可見系統與專家 1 完全重疊，與專家 2 則有 4 筆資料相符。再來，根據內科加護病房情境設計，由圖 4.4 可知專家 1 選擇記錄 1、2、15、16、18；專家 2 則選擇記錄 1、2、3、15、18；系統則選擇了記錄 1、2、15、16、18，由此亦可得知系統與專家 1 完全重疊，與專家 2 則有 4 筆資料相符。

透過 Cohen's Kappa 統計方法進一步分析系統與專家在選項判斷之間的一致性。如圖 4.5 所示，在洗腎室情境中，系統與專家 1 的 Kappa 值為 1.00，達到完全一致的水準；系統與專家 2 的 Kappa 值則為 0.73，屬於高度一致的程度。再根據內科加護病房情境的 Cohen's Kappa 分析，如圖 4.6 可知系統與專家 1 之間的 Kappa 值同樣為 1.00，達到完全一致；而與專家 2 的 Kappa 值亦為 0.73，同屬高度一致的範疇。

以上兩個不同病房情境之分析結果，皆顯示本研究之系統不僅在資料選擇上能與臨床專家達到高重疊性，並且能夠在不同臨床情境下穩定維持極佳的判斷一致性，證明系統具備良好的語意分析準確性及跨場域之臨床應用潛力，適合作為輔助臨床決策的工具。

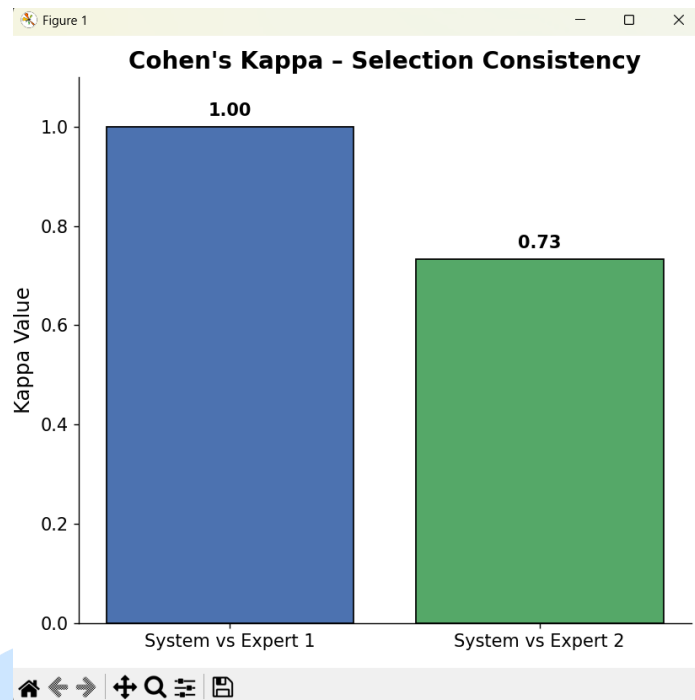


圖 4.5
洗腎室情境之 *Cohen's Kappa* 一致性分析長條圖

參考文獻：本研究自行整理

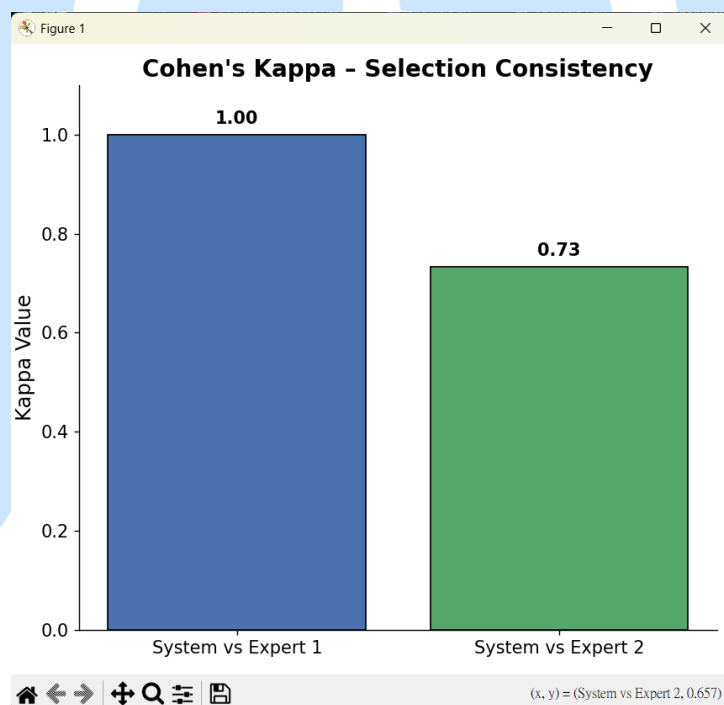


圖 4.6
內科加護病房情境之 *Cohen's Kappa* 一致性分析長條圖

參考文獻：本研究自行整理

4.2.2 排序一致性分析

Kendall's τ (Kendall rank correlation coefficient) 是一種常用於評估兩組排名之間秩序一致性的非參數統計方法，特別適合用於樣本數較小、排序資料為次序型、類別距離不明確的資料，其計算公式如下：

$$\tau = \frac{(C - D)}{\frac{1}{2}n(n - 1)}$$

其中 C 為符合相同排序方向的 concordant 對數， D 為排序相反的 discordant 對數， n 為樣本數量， τ 越接近越接近 1 代表排序越一致，越接近 -1 則代表排序方向相反

在醫療人工智慧領域，Hoebel 等人 (2023) 於其針對深度學習腦瘤分割演算法品質之評估研究中，即採用 Kendall's τ 作為分析工具，用以探討臨床專家對影像分割品質之主觀評分與 Dice score、Hausdorff distance 等常用定量指標之間的排序相關性。研究結果顯示，即便上述指標廣為使用，其與臨床專家評分之間的 Kendall's τ 相關係數仍偏低，突顯秩序相關性於臨床評估中具備獨立且不可忽視的參考價值 (Hoebel et al., 2023)。本研究此一驗證邏輯，選擇使用 Kendall's τ 作為評估工具，進一步檢驗系統所提供的五筆建議記錄之排序結果，是否與專家實際排序具有一致性。

根據本研究之情境設計，由圖 4.3 可見，在洗腎室情境中專家 1 的排序為 1、2、14、19、20；專家 2 的排序為 1、2、3、14、19；而系統的排序則為 19、1、2、14、20，由於記錄 3 不被系統選取，故視為排序第 6 名。再來由圖 4.4 可見，在內科加護病房情境中，專家 1 的排序為 15、18、1、2、16；專家 2 的排序為 15、2、1、18、3；系統的排序則為 15、2、18、1、16。記錄 3 在此同樣未被系統選取，故亦視為排序第 6 名。

在系統與專家於洗腎室情境使用 Kendall's τ 分析結果如圖 4.7、圖 4.8 所示，系統與專家 1 的 Kendall's τ 為 0.60，表示雙方在排序趨勢上具有中度一致性，系統所推薦之排序與專家 1 的臨床判斷呈現相似的邏輯方向。雖然對應的 p 值為 0.136，未達傳統統計顯著性標準，但在樣本數僅為五筆的情境下，仍可視為具有一定程度的臨床參考意義。另一方面，系統與專家 2 的 Kendall's τ 為 0.20，屬於低度相關，顯示排序趨勢不一致，系統排序邏輯與專家 2 的臨床評估出現落差。此一結果意味著系統在不同專家的主觀排序邏輯之間，可能存在適應性

差異，亦提示本系統在處理不同臨床評估邏輯時，尚需進一步優化其適應性與泛用性。

在系統與專家於內科加護病房情境使用 Kendall's τ 分析結果如圖 4.9、圖 4.10 所示，系統與專家 1 的 Kendall's τ 為 0.73，顯示雙方排序趨勢具有高度一致性，代表系統的推薦排序邏輯與專家 1 的臨床判斷呈現相當相似的趨勢；而系統與專家 2 的 Kendall's τ 同樣為 0.73，亦屬於高度一致性。儘管兩組分析對應的 p 值皆為 0.056，僅略高於傳統統計顯著性門檻 ($p < 0.05$)，但在樣本數有限的情況下，仍充分反映出系統的排序邏輯與專家們之臨床經驗具有良好的共識性。



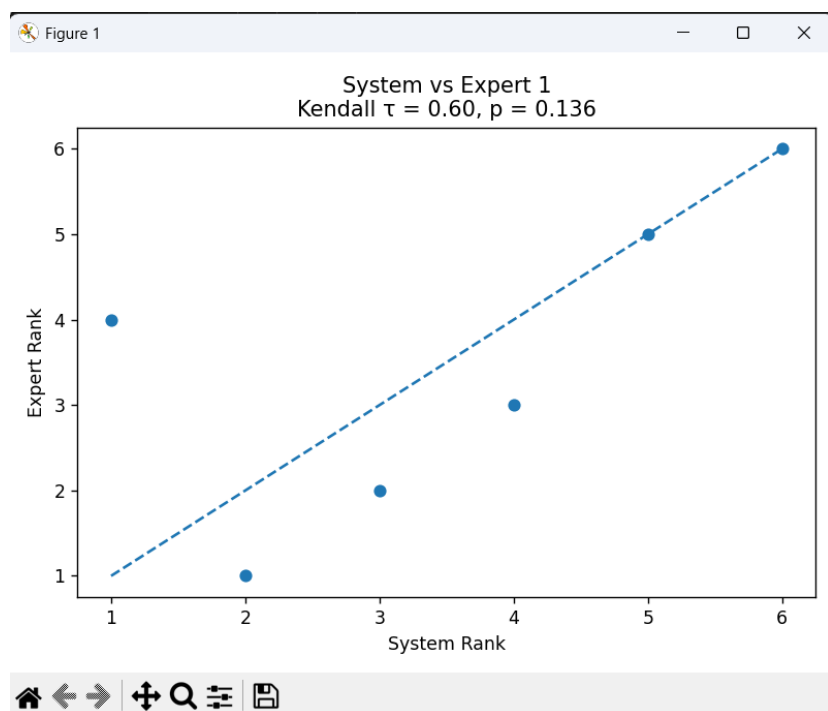


圖 4.7

洗腎室情境之 Kendall's τ 排序一致性散點圖

參考文獻：本研究自行整理

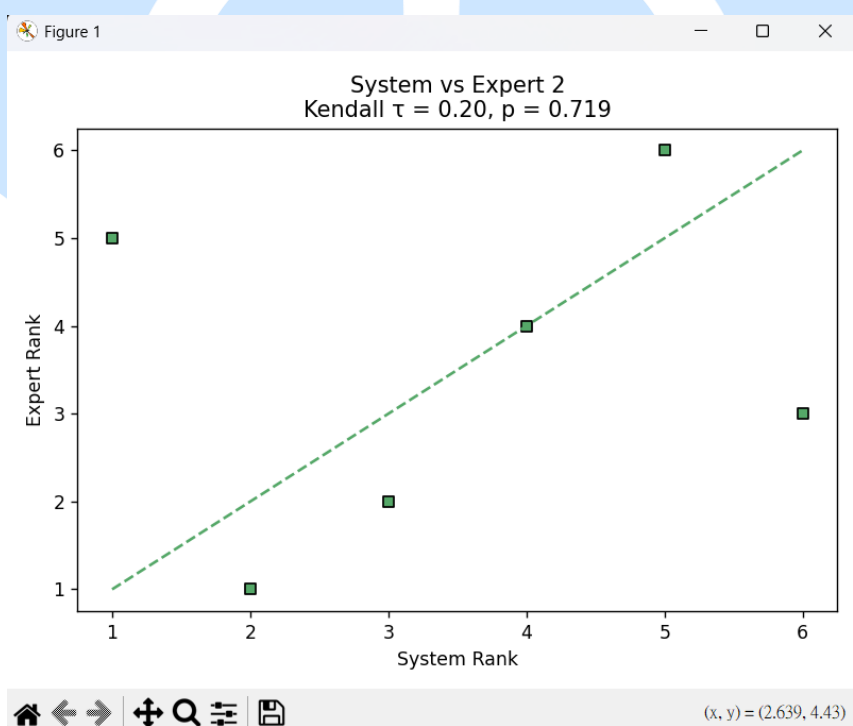


圖 4.8

洗腎室情境之 Kendall's τ 排序一致性散點圖

參考文獻：本研究自行整理

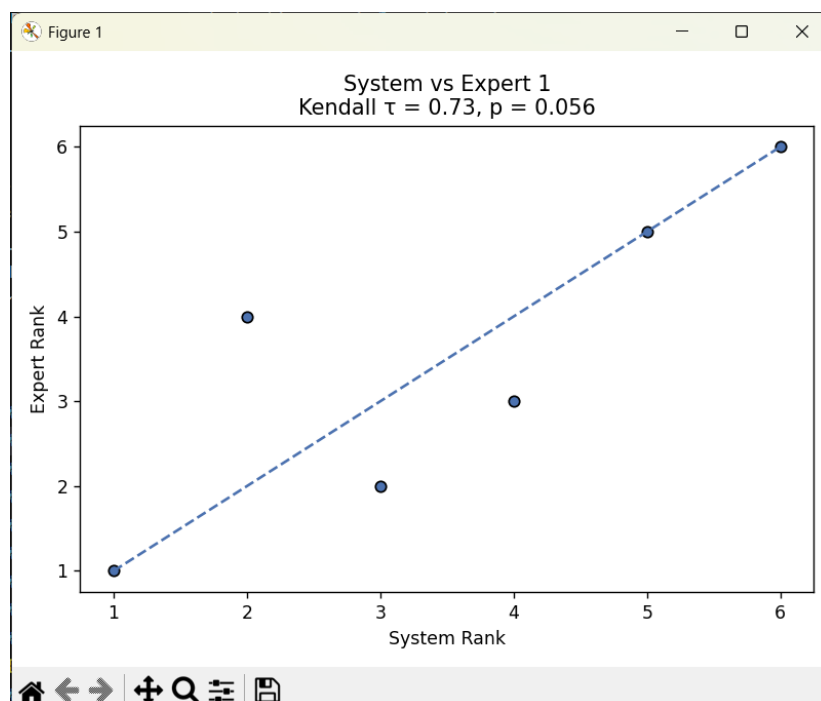


圖 4.9

內科加護病房情境之 Kendall's τ 排序一致性散點圖

參考文獻：本研究自行整理

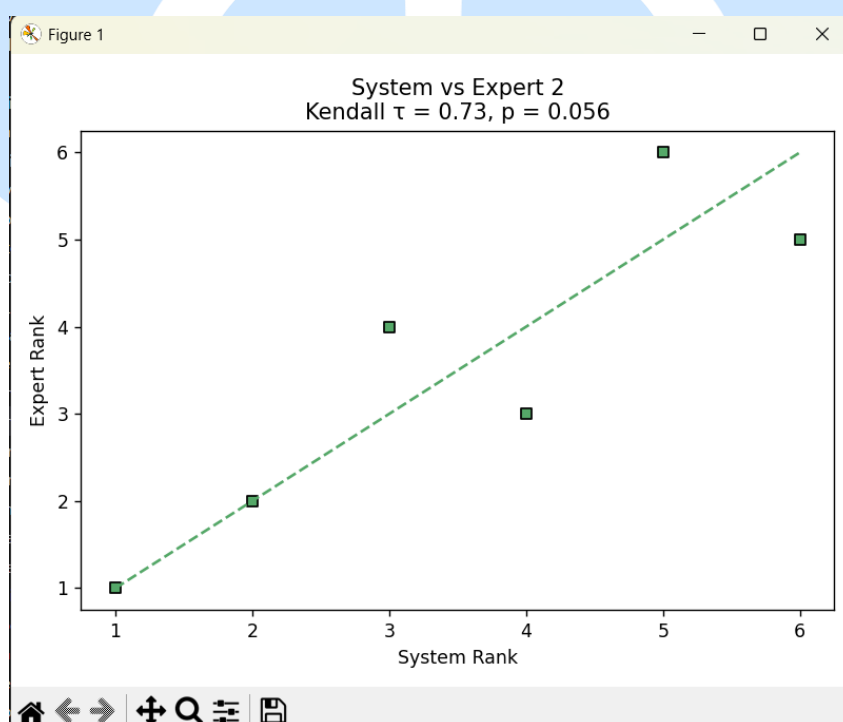


圖 4.10

內科加護病房情境之 Kendall's τ 排序一致性散點圖

參考文獻：本研究自行整理

綜合上述結果可知，本研究所建構之護理記錄生成系統，在「資料選取」與「排序建議」兩個層面，皆展現出具備臨床一致性的潛力。在資料選取方面，無論是在洗腎室或內科加護病房情境下，系統與其中一位專家的選項皆達成完全一致，根據 Kappa 值達 0.73 的結果可得知與另一位專家也呈現高度一致。此一致的結果顯示本系統在不同臨床情境中，具備穩定而可靠的模擬臨床專業判斷的能力。

然而，在排序建議方面，本系統在兩種不同情境之表現則有差異。在洗腎室情境下，系統與專家 1 的排序呈現中度正相關 (Kendall's $\tau = 0.60$)，表示系統的排序邏輯可在一定程度上反映臨床專家的思維，但與專家 2 之相關性較低 (Kendall's $\tau = 0.20$)，顯示系統於此情境下的排序策略尚需調整，以提高對不同專家邏輯的適應性。另一方面，在內科加護病房情境中，系統與兩位專家之排序一致性皆達到高度正相關 (Kendall's $\tau = 0.73$)，充分展現系統能在特定臨床情境下更穩定且有效地貼近專家排序邏輯。

因此，雖然本系統整體已展現出高度的臨床參考價值與應用潛力，但在面對不同科別或專家個體判斷之情境時，其排序邏輯仍存在進一步優化與調整的空間，以提升其泛用性及實務價值。

第五章 結論

5.1 結論

隨著護理臨床工作日趨繁重，如何透過智慧化工具協助護理人員進行記錄撰寫，已成為醫療照護系統發展中的重要課題。因此，本研究以「結合語意檢索記錄來輔助自動化護理記錄系統」為研究目標，旨在解決護理紀錄撰寫繁瑣、耗時等問題，希望可以透過智能技術的應用，來提升護理師的工作效率及減少人為上的疏失。

本研究建構之系統整合了 Flask 架構、PostgreSQL 資料庫、Elasticsearch 檢索引擎與語句嵌入模型 Sentence-BERT (all-MiniLM-L6-v2)，實現從病人資料輸入、生理數據異常判讀、護理紀錄語意向量化、相似紀錄回傳至勾選儲存等完整流程。使用者輸入任意症狀描述後，系統將其轉為 384 維語意向量，並透過 Elasticsearch 的 script_score + cosineSimilarity 機制，計算其與索引資料庫中記錄的語意相似度，即時回傳最相關的五筆 DART 記錄，供使用者作為參考與修改。

最後透過實驗操作與專家評估，本系統初步展現其臨床適用性與實作潛力，證實語意檢索技術結合自動化推薦邏輯，確實能提升護理記錄的個別化程度與生成效率，並具備作為決策輔助工具的應用價值。

5.2 研究限制

儘管本研究系統已涵蓋從資料輸入到記錄生成的完整流程，但在實務應用中仍存在若干限制。首先，系統回傳的紀錄依賴語句向量相似度判斷，當輸入為複雜、多重症狀的敘述時，系統可能無法準確捕捉重點語意，導致回傳內容不夠精準或臨床適配度不足。

其次，語句嵌入模型雖具語意理解能力，然目前並未針對特定病科進行微調(fine-tuning)，在處理專科用語或罕見敘述時仍可能產生語意偏差。此外，本系統使用暴力全量比對 (Brute-

Force Vector Search) 策略以保留最大檢索彈性，雖能提升語意比對準確度，但對於資料筆數大量成長的情境下，可能會面臨效能瓶頸。

最後，專家評估樣本數仍屬有限，雖可作為初步可行性驗證，但尚不足以全面反映各類臨床場域或人員的多元需求，後續仍需擴大實測規模。

5.3 未來發展與建議

為進一步提升本系統的實務應用效能與彈性，未來可從多個面向持續優化與拓展。首先，在模型層級方面，建議針對不同病房類型與專科語境進行語句嵌入模型的微調 (fine-tuning)，以提升系統對特定醫療語彙與臨床情境的理解與適配能力，強化語意檢索的精準度與回應力。

此外，系統的查詢效率亦可透過技術強化來優化，例如導入近似近鄰搜尋 (Approximate Nearest Neighbor [ANN]) 技術中的 HNSW (Hierarchical Navigable Small World) 結構，取代目前使用之暴力全量比對策略，以有效降低查詢時間並維持相似度判斷的準確性，使系統能更順暢地應對大型資料庫的擴展。

進一步地，系統也可考慮導入 Retrieval-Augmented Generation (RAG) 架構，結合語意檢索與語言生成模型，實現自動生成建議紀錄的能力。此舉將不僅限於比對既有資料，更能針對使用者輸入的臨床敘述生成個別化、語意連貫且符合邏輯的護理建議內容，進一步提升自動化層級與臨床實用性。

綜合而言，若能在模型精度、查詢效率與生成能力等多層面持續優化，並配合實務現場需求調整應用邏輯與介面設計，本系統將具備更高的實作價值與發展潛力，為智慧護理決策支援系統的應用落地提供有力支撐。

參考文獻

- 中醫大附醫與台灣微軟強強聯手共創未來智慧醫院 引領創新「智海系統-生成式語音智慧醫療」 提高醫護效率、拯救更多病人生命 | 最新消息—中國醫藥大學附設醫院. (2024). 讀取於 2025年6月25日, 從 <http://www.cmuh.cmu.edu.tw/NewsInfo/NewsArticle?no=8505>
- 比較護理診斷為主的問題導向記錄法與焦點護理紀錄法之成效差異—臺灣博碩士論文知識加值系統. (2003). 讀取於 2025年6月25日, 從 <https://ndltd.ncl.edu.tw/cgi-bin/gs32/gswweb.cgi?o=dnclcdr&s=id=%22091TMC00563008%22.&searchmode=basic>
- 林育如, 林雯萱, 王敏芳, 葉淑敏, 徐紫娟, & 張玉婷. (2019). 提升透析電子護理記錄完整性之專案. 護理雜誌, 66(1), 93–100. [https://doi.org/10.6224/JN.201902_66\(1\).11](https://doi.org/10.6224/JN.201902_66(1).11)
- 花蓮慈院開發「i菩提」護理AI助手記錄、衛教全自動. (2024). 讀取於 2025年6月25日, 從 https://hlm.tzuchi.com.tw/home/index.php/news/reports/i-ai_3354
- 姜瑤娟, & 李亭亭. (2018). 護理人員使用跨團隊照護資訊系統的成效探討. 護理雜誌, 65(4), 49–59. [https://doi.org/10.6224/JN.201808_65\(4\).08](https://doi.org/10.6224/JN.201808_65(4).08)
- 首頁—照護科技 | Jubo Health 智齡科技 | 台灣最懂你的照護系統. (2024). 讀取於 2025年6月25日, 從 <https://jubo-health.com/>
- 馮容莊, & 葉雨婷. (2014). 護理新視界—護理資訊的演變與發展. 護理雜誌, 61(4), 78–84. <https://doi.org/10.6224/JN.61.4S.78>
- Akca, M. A., Aydoğan, T., & İlkuçar, M. (2016). An analysis on the comparison of the performance and configuration features of big data tools Solr and Elasticsearch. International

Journal of Intelligent Systems and Applications in Engineering, 4(Special Issue), 8–12.

<https://doi.org/10.18201/ijisae.271328-252047>

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., & Silber, J. H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *JAMA*, 288(16), 1987–1993.

<https://doi.org/10.1001/jama.288.16.1987>

Digital health. (2021). 讀取於 2025年6月25日, 從 <https://www.who.int/health-topics/digital-health>

Dilsizian, S. E., & Siegel, E. L. (2014). Artificial intelligence in medicine and cardiac imaging: Harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Current Cardiology Reports*, 16(1), 441. <https://doi.org/10.1007/s11886-013-0441-8>

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118. <https://doi.org/10.1038/nature21056>

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>

Garcelon, N., Neuraz, A., Benoit, V., Salomon, R., & Burgun, A. (2017). Improving a full-text search engine: The importance of negation detection and family history context to identify cases in a biomedical data warehouse. *Journal of the American Medical Informatics Association: JAMIA*, 24(3), 607–613. <https://doi.org/10.1093/jamia/ocw144>

Hoebel, K. V., Bridge, C. P., Ahmed, S., Akintola, O., Chung, C., Huang, R. Y., Johnson, J. M., Kim, A., Ly, K. I., Chang, K., Patel, J., Pinho, M., Batchelor, T. T., Rosen, B. R., Gerstner, E. R., & Kalpathy-Cramer, J. (2024). Expert-centered Evaluation of Deep Learning Algorithms for Brain Tumor Segmentation. *Radiology. Artificial Intelligence*, 6(1), e220231.

<https://doi.org/10.1148/ryai.220231>

Kathare, N., Reddy, O. V., & Prabhu, V. (2022). A comprehensive study of Elastic Search. *Journal of Research in Science and Engineering (JRSE)*, 4(11), 34–38
[https://doi.org/10.53469/jrse.2022.04\(11\).07](https://doi.org/10.53469/jrse.2022.04(11).07)

Kromrey, M.-L., Steiner, L., Schön, F., Gamain, J., Roller, C., & Malsch, C. (2024). Navigating the spectrum: Assessing the concordance of ML-based AI findings with radiology in chest X-rays in clinical settings. *Healthcare*, 12(2225). <https://doi.org/10.3390/healthcare12222225>

Krittawong, C., Rogers, A. J., Aydar, M., Choi, E., Johnson, K. W., Wang, Z., & Narayan, S. M. (2020). Integrating blockchain technology with artificial intelligence for cardiovascular medicine. *Nature Reviews. Cardiology*, 17(1), 1–3. <https://doi.org/10.1038/s41569-019-0294-y>

Lampe, S. S. (1985). Focus charting: Streamlining documentation. *Nursing Management*, 16(7), 43–46.

Li, I., Pan, J., Goldwasser, J., Verma, N., Wong, W. P., Nuzumlalı, M. Y., Rosand, B., Li, Y., Zhang, M., Chang, D., Taylor, R. A., Krumholz, H. M., & Radev, D. (2021). *Neural Natural Language Processing for Unstructured Data in Electronic Health Records: A Review* (No. arXiv:2107.02975). arXiv. <https://doi.org/10.48550/arXiv.2107.02975>

Liu, Y., & Lapata, M. (2019). Text Summarization with Pretrained Encoders. *Proceedings of*

the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (頁 3730–3740). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1387>

Menachemi, N., & Collum, T. H. (2011). Benefits and drawbacks of electronic health record systems. *Risk Management and Healthcare Policy*, 4, 47–55.
<https://doi.org/10.2147/RMHP.S12985>

Poissant, L., Pereira, J., Tamblyn, R., & Kawasumi, Y. (2005). The Impact of Electronic Health Records on Time Efficiency of Physicians and Nurses: A Systematic Review. *Journal of the American Medical Informatics Association : JAMIA*, 12(5), 505–516.
<https://doi.org/10.1197/jamia.M1700>

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *The New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMra1814259>

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* (No. arXiv:1908.10084). arXiv. <https://doi.org/10.48550/arXiv.1908.10084>

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>

Wang D., (2021). Digital archive management based on Lucene full-text search engine. *Journal of Physics: Conference Series*, 2074(1), 012001.
<https://doi.org/10.1088/1742-6596/2074/1/012001>

Xu, W., Chen, H., Huan, Y., Hu, X., & Nong, G. (2021). Full-text search engine with suffix

index for massive heterogeneous data. *Information Systems*, 104, 101893.

<https://doi.org/10.1016/j.is.2021.101893>

Zhai, Y., Yu, Z., Zhang, Q., Qin, W., Yang, C., & Zhang, Y. (2022). Transition to a new nursing information system embedded with clinical decision support: A mixed-method study using the HOT-fit framework. *BMC Medical Informatics and Decision Making*, 22(1), 310.

<https://doi.org/10.1186/s12911-022-02041-y>

