

國立中興大學資訊管理學系

碩士學位論文

利用遷移學習進行跨疾病病程預測的模型建立

A Transfer Learning Framework for Modeling

Prediction of Cross Disease Course

National Chung Hsing University

指導教授： 英家慶 教授 Jia - Ching Ying

研 究 生： 張雁婷 Yen - Ting Chang

中華民國 一一〇年 七月

國立中興大學 資訊管理學系

碩士學位論文

題目：利用遷移學習進行跨疾病病程預測的模型建立

姓名：張雁婷

學號：7108029131

經 口 試 通 過 特 此 證 明

論文指導教授

吳宗慶

論文考試委員

陳宏華

吳宗慶

林冠成

中華民國 110 年 07 月 02 日

授權書編號：nchu-110-7108029131-1

## 中興大學博碩士論文授權書

本授權書所授權之論文為立書人在 中興大學管理學院資訊管理學系所，109學年度第2學期取得 碩士學位之論文。

論文題目：利用遷移學習進行跨疾病病程預測的模型建立

指導教授：英家慶

授權事項：

- 一、立書人 張雁婷 同意無償授權 中興大學將上列論文全文資料之以微縮、數位化或其他方式進行重製作為典藏及網際網路公開傳輸之用。中興大學在上述範圍內得再授權第三者進行重製或網際網路公開傳輸等其他利用。
- 二、立書人 張雁婷 同意無償授權教育部指定送繳之圖書館及 中興大學將上列論文全文資料之紙本為學術研究之目的予以重製。同意全本重製，惟每人以一份為限。
- 三、立書人 張雁婷 同意中興大學以有償方式再授權資料庫廠商 將前條典藏之資料收錄於廠商之資料庫，並以電子形式透過單機、網際網路、無線網路或其他公開傳輸方式授權用戶進行檢索、瀏覽、下載、傳輸、列印等。立書人可選擇之權利金方案如下列二種擇一：  
※由立書人收取有償授權之權利金，其權利金額度由中興大學與再授權之廠商議定之。  
※由立書人將有償授權之權利金捐贈中興大學校務基金。
- 四、中興大學得將第三條之權利再授權予其他第三人進行網際網路公開傳輸等其他加值利用。
- 五、前四條授權均為非專屬授權，立書人仍擁有上述授權著作之著作權。立書人擔保本著作為立書人所創作之著作，有權依本授權書內容進行各項授權，且未侵害任何第三人之智慧財產權。如有侵害他人權益及觸犯法律之情事，立書人願自行負責一切法律責任，被授權人一概無涉。

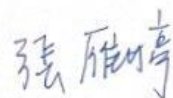
論文紙本於中興大學圖書館內公開陳列上架時間：立即公開

有償授權條件：享有權利金的回饋，權利金捐贈校務發展基金。

論文全文上載網路公開時間：立即公開

立書人：張雁婷

簽名：



中華民國 110 年 7 月 13 日

## 摘要

近年來，人工智慧的發展及應用都是人們所關注的議題，在醫療領域上如何利用現有的醫療紀錄並從中提取有用的資訊，提供更精確、更有幫助的診斷建議是醫療科技(Medical Technology, MedTech)發展的一個重要方向。因此本篇論文提出能夠從容易發現病徵的疾病病程發展來預測醫學上證實相關但不同的其他缺乏數據之疾病病程發展，目的為協助評估不易發現病徵之疾病的病程發展，並提升醫療資料的利用率；首先利用時間序列模型捕捉不同時間間隔內容易發現病徵之疾病的連續表現狀況，接著藉由遷移學習搭配注意力機制將捕捉到的通用特徵運用到另一數據量不足且不易發現病徵的疾病病程發展預測模型上。



關鍵字： 深度學習、時間序列模型、轉移學習、電子健康紀錄

# Abstract

In recent years, the development and application of artificial intelligence has been an issue of concern. In the medical field, how to use existing medical records and extract useful information from them to provide more accurate and helpful diagnostic recommendations is an important direction in the development of Medical Technology (MedTech). Therefore, this paper proposes a framework for predicting the course of diseases that lack data from other course of related diseases that easily detectable symptoms. The purpose is to assist in the assessment of the course of diseases that do not have easily detectable symptoms and to improve the utilization of medical data.

First, using the time series model to capture the successive performance of diseases course with easily detectable symptoms at different time intervals, and then the generic features captured are applied to another model for predicting the course of diseases with insufficient data and less easily detectable symptoms by using transfer learning and attention mechanism.

Key words: Deep Learning, Time Series Model, Transfer Learning, Electronic Health Record(s)

# 目錄

摘要.....	i
Abstract .....	ii
目錄.....	iii
表目錄.....	v
圖目錄.....	vi
第一章 緒論 .....	1
1.1. 研究動機與背景 .....	1
1.2. 研究目的 .....	3
1.3. 論文組織與架構.....	3
第二章 文獻探討 .....	5
2.1. 電子健康紀錄利用深度學習模型進行分析.....	5
2.1.1. 電子健康紀錄.....	5
2.1.2. 利用深度學習模型於電子健康紀錄的相關文獻 .....	6
2.2. 遷移學習應用於電子健康紀錄 .....	7
2.2.1. 遷移學習 .....	7
2.2.2. 遷移學習應用於電子健康紀錄 .....	10
2.3. 遞迴神經網路(Recurrent Neural Network, RNN).....	10
2.3.1. LSTM .....	11
2.3.2. GRU .....	11
2.4. 注意力機制(Attention Mechanism) .....	12
第三章 研究設計與實施 .....	14
3.1. 研究流程 .....	14
3.2. 資料集 .....	15
3.2.1. 來源域資料.....	15
3.2.2. 目標域資料.....	16

3.3. 資料預處理 .....	16
3.3.1. 來源域資料預處理方法 .....	17
3.3.2. 目標域資料預處理方法 .....	19
3.3.3. 病程發展表示矩陣 .....	21
3.4. TL-CDC .....	22
3.4.1 來源疾病病程發展預測模型 .....	22
3.4.2 目標疾病病程發展預測模型 .....	25
第四章 實驗結果與分析 .....	28
4.1. 實驗數據設置 .....	28
4.2. 實驗結果評估函數 .....	28
4.3. 實驗結果 .....	29
4.3.1. TL-CDC 實驗結果 .....	29
4.3.2. TL-CDC 與其他時間序列模型比較結果 .....	33
4.3.3. 注意力機制之解釋力呈現 .....	34
第五章 結論與未來研究方向 .....	37
5.1. 結論 .....	37
5.2. 未來研究 .....	37
參考文獻 .....	39

## 表目錄

表 1	牙周病病程表示矩陣計算方法 .....	18
表 2	ID 為 8F14 的病患之 1999-2010 年醫療紀錄 .....	19
表 3	ID 為 8F14 的病患之病程表示矩陣(一個月為一筆的時序資料).....	19
表 4	ID 為 791F 的病患之 1999-2010 年醫療紀錄 .....	21
表 5	ID 為 791F 的病患之病程表示矩陣(一個月為一筆的時序資料).....	21
表 6	ID 為 791F 的病患之病程表示矩陣( $K=3$ ) .....	22
表 7	參數凍結與微調方法對目標疾病病程預測效果比較表 .....	30
表 8	參數凍結與微調方法對目標疾病病程預測效果比較表 .....	31
表 9	注意力機制對目標疾病病程發展預測效果比較表 .....	32
表 10	TL-CDC 與其他時間序列模型對目標疾病病程發展預測效果比較表 .....	33
表 11	模型訓練時間比較表 .....	34
表 12	PATIENT_1 的病程發展表示矩陣( $K=3$ ).....	36
表 13	PATIENT_2 的病程發展表示矩陣( $K=3$ ).....	36
表 14	PATIENT_3 的病程發展表示矩陣( $K=3$ ).....	36



## 圖目錄

圖 1	PUBMED 上利用人工智慧進行健康資訊學研究的已發表論文分布 .....	6
圖 2	傳統機器學習與遷移學習訓練過程示意圖 .....	8
圖 3	遷移學習的凍結與微調方法訓練過程示意圖 .....	9
圖 4	研究流程圖 .....	14
圖 5	就醫紀錄資料表示圖 .....	15
圖 6	資料前處理流程圖 .....	17
圖 7	來源疾病病程發展預測模型 .....	24
圖 8	目標疾病病程發展預測模型 .....	27
圖 9	注意力機制個案分析 .....	35



# 第一章 緒論

本章節將以研究的動機及背景說明本研究之重要性，後以研究的主要目的與任務說明本研究之貢獻，最後針對論文的整體組織與架構作介紹。

## 1.1. 研究動機與背景

根據 2020 年世界衛生組織(World Health Organization, WHO)發表的 2019 年全球健康估計[1]，統計了 2000 至 2019 年期間的數據，證實 2019 年人們平均壽命較 2000 年增長 6 年，全球平均壽命從 2000 年的 67 歲至 2019 年超過 73 歲，但壽命延長並非全為健康生命年，其中心臟病、糖尿病、心血管疾病及慢性呼吸道疾病等慢性疾病皆嚴重縮短人們的健康生命年，而因非傳染性疾病導致死亡更是佔世界十大死因中的七項。

慢性疾病是指人的身體狀況或疾病是持續且發病期長，常見的慢性疾病除了以上提及之外，還包括關節炎、哮喘及癌症等，而患者在接受藥物治療，得到週期性的短暫緩解，隨後又再次復發，在反覆的症狀復發與控制當中存活，使得患者飽受疾病摧殘。由以上的統計報告與慢性疾病的症狀發展介紹，可以發現慢性疾病的預防與治療更是需要加強的醫療項目，而慢性疾病的預防與治療最直接的方法，就是從疾病前期尚未有明顯症狀時進行治療，以免在疾病後期症狀嚴重至影響生活時，已難以控制病情發展。

現今醫療科技的蓬勃發展，醫療產業進行了各種不同應用的拓展與進化，其業務涵蓋了掛號、問診、病例、手術、門診流程、醫療設備和照護…等，而在這些應用中的病患資料也透過儲存與同步技術的進步，能將代表病患身體狀態的歷史就醫經驗得以有系統性的留存下來，成為人們所熟知的電子健康紀錄(Electronic Health Records, EHRs)。透過對 EHRs 數據進行分析、分類和整合，開發與疾病相關的預測機制，提供準確的病程發展預測，包含了診斷、治療及推薦

藥物…等；然而在 EHRs 資料分析的過程中仍有某些疾病資料不足的問題，例如類風濕性關節炎(Rheumatoid Arthritis, RA)，此疾病不會直接造成病患死亡，而是漸進式的傷害關節，因此較不易被大眾重視，導致病患未能在初期就醫治療，錯失治療的黃金時期，也造成未能在疾病初期進行治療並記錄患者的關節炎病徵發展歷程，使得能提供完整關節炎病程發展的資料樣本數少。

近年來，深度學習方法已經能實現端到端學習，並且對於醫學預測任務有著很好的表現[2,3]，Yang 等人[4]提取住院資訊的臨床數據形成臨床序列，並使用雙向長短期記憶模型搭配注意力機制進行未來的罹病機率預測，He 等人[5]透過結合深度殘差網絡與雙向長短期記憶模型建立自動進行心律失常分類的模型，Edward 等人[6]則是利用醫學本體固有的層次知識補充 EHRs 的資訊，建立一個基於圖(Graph)的注意力模型(GRAM)對 EHRs 進行分析，預測下次就診的診斷與心臟衰竭的機率；從以上醫學任務預測的文獻回顧中可以發現大多數的研究都是應用深度學習於單一疾病的預測任務，鮮少研究將深度學習應用於跨疾病的分析預測。

而在慢性疾病的相關研究中發現牙周病(Periodontal Disease, PD)的治療可以有效減緩 RA 的症狀[7,8]，由此可知，這兩種疾病之間有相似的病理特徵分布，因此，我的這項研究，主要是利用深度學習中的遷移學習(Transfer Learning, TL)分析 EHRs 中 PD 與 RA 的資料，透過時間序列模型進行特徵提取，建立以 PD 就診紀錄資料實現的來源疾病病程發展預測模型，了解病患醫療紀錄所呈現出的健康狀態，再藉由來源疾病病程發展預測模型訓練出的權重遷移至預測目標疾病病程發展的目標預測模型中，而此目標預測模型將以 RA 就診紀錄資料實現，再由目標疾病的少量數據集對目標預測模型進行微調，以達到準確預測目標疾病的病程發展。

## 1.2. 研究目的

這份研究主要任務為利用遷移學習框架搭配深度學習方法，建置跨疾病病程發展預測模型，歸納現有慢性疾病患者資料的病程發展模式，目的是協助醫生進行慢性疾病病程發展之推估，及時給予準確的治療方案推薦；鑑於慢性疾病病程發展特色所造成之訓練資料不足的問題，因此本研究希望能夠在台灣健保資料中，透過深度學習模型提取出各牙周病患者病程發展特徵並進行病程發展預測，結果用於類風濕性關節炎患者病程發展模型特徵提取之訓練，最終希望能夠透過慢性疾病患者的就醫資料所建構之跨疾病病程發展預測模型，使數據量少的疾病資料能夠得到好的病程發展預測效果，使患者能提早發現疾病，免於疾病造成之不適的症狀，進而提升人民的健康生命年。利用深度學習的特徵提取技術為基礎，建構適合進行病程發展分析的預測模型並訓練，通過遷移學習搭配時間序列模型處理時間序列以及樣本數量少的資料之優勢，對資料進行細部特徵提取，建立可以達成下列目的之模型框架以幫助在慢性疾病病程發展的分析評估：

1. 提出用於相關但不同的疾病間之跨疾病的病程發展預測模型。
2. 使用遷移學習解決在特定疾病資料量少的情況下出現預測效果較不準確的問題。
3. 應用注意力機制於遷移學習中，提升對目標疾病病程發展預測的效果。
4. 利用注意力機制使預測模型擁有解釋力。

## 1.3. 論文組織與架構

本研究利用深度學習之遷移學習進行跨疾病病程發展預測模型建置總共分為五大章節，第二章節為針對本研究資料與方法相關應用之研究文獻的比較討論；第三章節詳細介紹本研究使用之全民健保資料庫、資料前處理流程及研究方法設計的遷移學習搭配時間序列模型架構；第四章節進行模型實驗評估的結果呈現以

及對注意力機制的結果解釋探討；第五章節則是總結本研究所得出來的結論以及未來研究方向。



## 第二章 文獻探討

本章節中，將針對研究使用的方法做介紹，鑒於本研究利用深度學習中的遷移學習作為研究方法框架，將先於章節 2.1 說明先前學者利用深度學習模型於電子健康紀錄的研究文獻概況；章節 2.2 是遷移學習的運作方法說明及其應用於電子健康紀錄的研究概況；本研究的研究方法設計，主要是利用深度學習的時間序列模型提取時序資料中各特徵在時間上的關係，並透過注意力機制對重要資訊給予更多的關注，用以提升模型的預測準確率與解釋力，因此於章節 2.3 介紹處理時序資料的時間序列模型及其在電子健康紀錄的應用；在最後的章節 2.4 則是對注意力機制的基本原理運作與實際應用說明。

### 2.1. 電子健康紀錄利用深度學習模型進行分析

本研究因使用電子健康紀錄為研究資料並利用深度學習模型進行特徵提取與分析預測，因此將先對電子健康紀錄的結構特性做介紹，後說明深度學習模型於電子健康紀錄的應用概況。

#### 2.1.1. 電子健康紀錄

EHRs 包含病患的人口統計資訊和就診紀錄，就診紀錄中涵蓋入院時間、出院時間、基本身體測量數據、診斷、手術、用藥和醫學影像...等[9]。隨著時間的累積與儲存技術的進步，EHRs 雖然提供了大量的病患數據以充分表示患者及醫生的歷史資訊，但 EHRs 具有高維度、雜訊、異質性、稀疏性、不完整性、隨機性及系統誤差[10-12]，隨著不同診所的紀錄方式不同，甚至有非結構化的診斷內容。如圖 1 所示，相較於其他醫學資料利用人工智慧進行 EHRs 資料分析的研究明顯較少，因此，可以得知 EHRs 預測模型的建立及資料的利用上挑戰性較高；在實際的應用上，常見的問題如診斷和用藥在不同醫院間有不同的規格，用藥是根據醫生的判斷

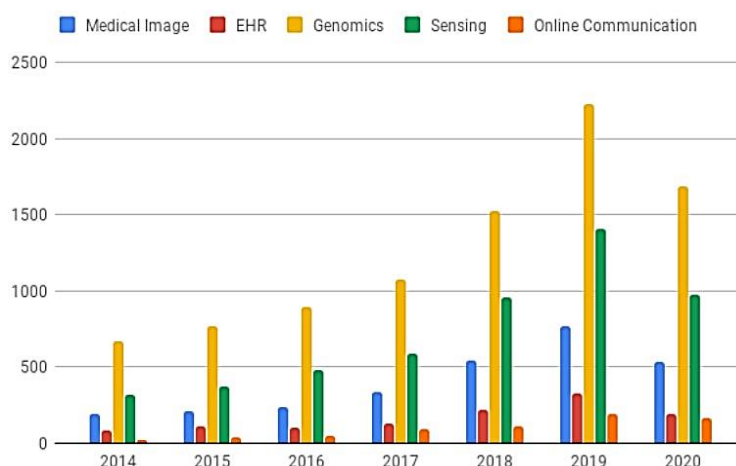


圖 1 PubMed 上利用人工智慧進行健康資訊學研究的已發表論文分布[13]

與該醫院所使用的藥物種類，如在治療類風濕性關節炎的疾病調節抗風濕藥物 (DMARDs)，其組成藥品的成分有 leflunomide 和 ciclosporin...等，每種成分在藥品製作上加入的劑量不同，能給予的治療程度也相對不同，需要在進行分析前將每位病患所服用的藥物進行當量轉換，才能確保代表病程狀態的用藥資訊有可比較性，進而對病患的病程發展進行分析。因此，本研究在 EHRs 資料前處理的部分會針對病患的用藥資訊，使用世界衛生組織(World Health Organization, WHO)所訂製的每日標準劑量(Defined Daily Dose, DDD)進行量轉換。

### 2.1.2. 利用深度學習模型於電子健康紀錄的相關文獻

隨著電腦計算能力與深度學習的蓬勃發展，深度學習被廣泛利用在各個領域，在醫學上也不例外[13]；醫療紀錄的應用，以預測死亡率、罹病機率、和非計畫內入院機率的研究最多，在 Avati 等人的研究中[14]，使用前幾年患者的 EHRs 數據建立深度神經網路模型(Deep Neural Network, DNN)，用以預測患者未來 3 至 12 個月的死亡率，在罹病機率預測上，Choi 等人[15]將就診紀錄以倒序的方式輸入時間序列模型並搭配注意力機制，預測患者未來心臟衰竭的機率，而用於預測非計畫性

再入院機率的 DeepCare 和 DeepR[16,17]，則是分別利用長短期記憶模型(Long Short Term Memory, LSTM) 搭配詞嵌入(Word Embedding)和卷積神經網路(Convolutional Neural Networks, CNN)進行 EHRs 的分析研究，另外，在 EHRs 數據的應用上，也有少數研究的目標為下次就診時的用藥種類與時間的預測，如 DoctorAI[18]。

從上述的文獻回顧中可以發現，深度學習被廣泛地利用在 EHRs 數據的預測模型建立上，並且都有十分顯著的效果；然而，從預測任務的方向來看，現有的研究中大多以單一疾病的罹病機率與較廣泛的概率預測任務為主，鮮少以慢性疾病的病程發展作為預測目標，因此，本研究將透過深度學習從病患的就診紀錄資訊捕捉類風濕性關節炎的病程發展特徵，並進行預測。

## 2.2. 遷移學習應用於電子健康紀錄

本研究主要利用遷移學習作為研究方法的框架，進行不同醫療任務預測模型的特徵共享，因此將先介紹遷移學習的運作原理，後說明遷移學習應用於電子健康紀錄的文獻研究狀況。

### 2.2.1. 遷移學習

以往傳統的機器學習中，為了使模型預測有效，對於現有預訓練的數據與未來將收集到的數據的假設為有相似特徵、在相同的空間分布並且是大量帶有標記的，但是在實際應用情況下，此條件並不容易符合，尤其對於新興領域，因此發展出遷移學習來解決此問題，遷移學習主要的概念是將已訓練好的模型參數遷移至新的預測目標模型進行訓練，如圖 2 為傳統機器學習與遷移學習訓練過程示意圖，傳統機器學習主要為一個任務建立一個預測模型，而要使預測模型能夠達到有效的準確率，需要有足夠的訓練資料供模型進行學習，但此條件在實際場景應用並不容易達成，因此，利用遷移學習將與目標任務資料有相關的來源任務預測模型



的參數遷移至目標任務預測模型，除了達到共享參數之外，還能使目標任務預測模型能以最低限度的有效標記之數據來達到預測，減輕目標任務的訓練階段對有標記數據的依賴，並加快模型建置與學習效率，使網路架構不需要從零開始訓練。因此，遷移學習可以被利用於不同但有相似分布的數據，對其特徵做不同任務預測模型間參數的遷移[20]，透過 Yosinski 等人的研究[21]可以了解遷移學習在深度學習中的可行性與其效果。

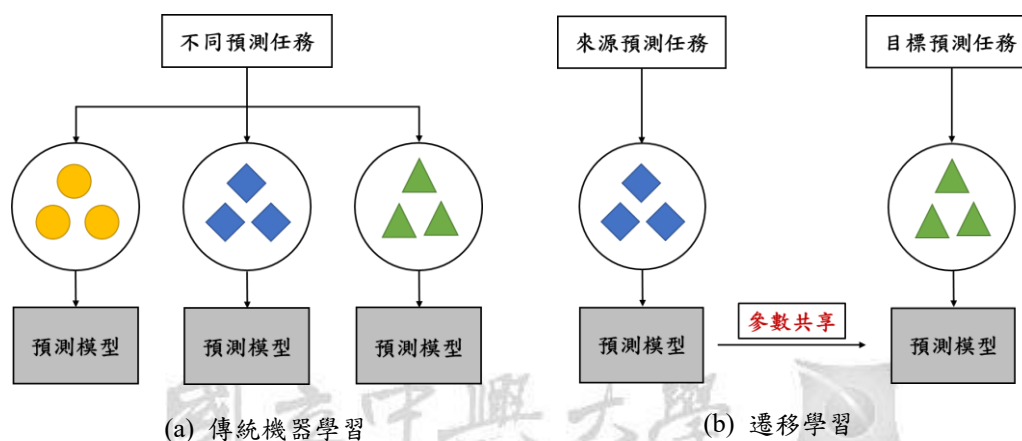


圖 2 傳統機器學習與遷移學習訓練過程示意圖[19]

而此方法的特性正好與醫療資料特性相符，醫療資料大致分為醫療影像及醫療數據，醫療影像包含超音波影像、電腦斷層掃描與核磁共振掃描等，透過完善的醫療體系累積出龐大的醫療影像資料量，但卻缺乏對影像的標記和診斷，而這些標記工作的進行也將耗費許多時間與人力，在標記的準確度上也會因長時作業而下降，因此在深度學習的應用發展上有著一定的挑戰性；醫療數據則為病患醫療紀錄的所有身體測量數據、診斷及用藥，但在這些醫療數據在應用上也存在限制，如不同醫院的醫生在就診紀錄上的習慣差異，或者特定的醫療任務預測會因疾病的病徵不易被察覺，導致病患未能在一定的時間內就醫治療，造成能進行完整訓練過程的數據缺乏，進而導致預測模型的效果不好。

因此，可以將遷移學習的概念應用在醫療資料上，使因數據缺乏造成預測效果表現不好的醫療任務能夠以最低限度的有效標記之數據進行預測模型的訓練。在

遷移學習中共享參數運用到目標任務預測模型的方式可以分為凍結(Frozen)與微調(Fine-tune)，如圖 3 為兩種方法的訓練過程示意圖，凍結是將來源任務預測模型訓練好的參數遷移到目標任務預測模型的前 n 層，再將前 n 層凍結起來，不進行參數的訓練，接著對目標任務預測模型的其餘層進行初始化，並針對目標任務進行訓練，達到目標的預測效果，而微調則是將目標任務預測模型訓練時的誤差，反向傳播至從來源任務預測模型複製來的 n 層進行權重更新動作，讓複製來的參數調整至符合目標任務的預測。兩種方法的選擇是取決於目標任務數據集的大小和從來源任務預測模型遷移而來的參數數量，如果目標數據集小且來源任務預測模型預訓練的參數量較多，通常會為避免模型過擬合(Overfitting)而選擇凍結；若目標任務數據集大或來源任務預測模型預訓練的參數量少，則可以將基本參數微調至符合目標任務的預測以提性能。

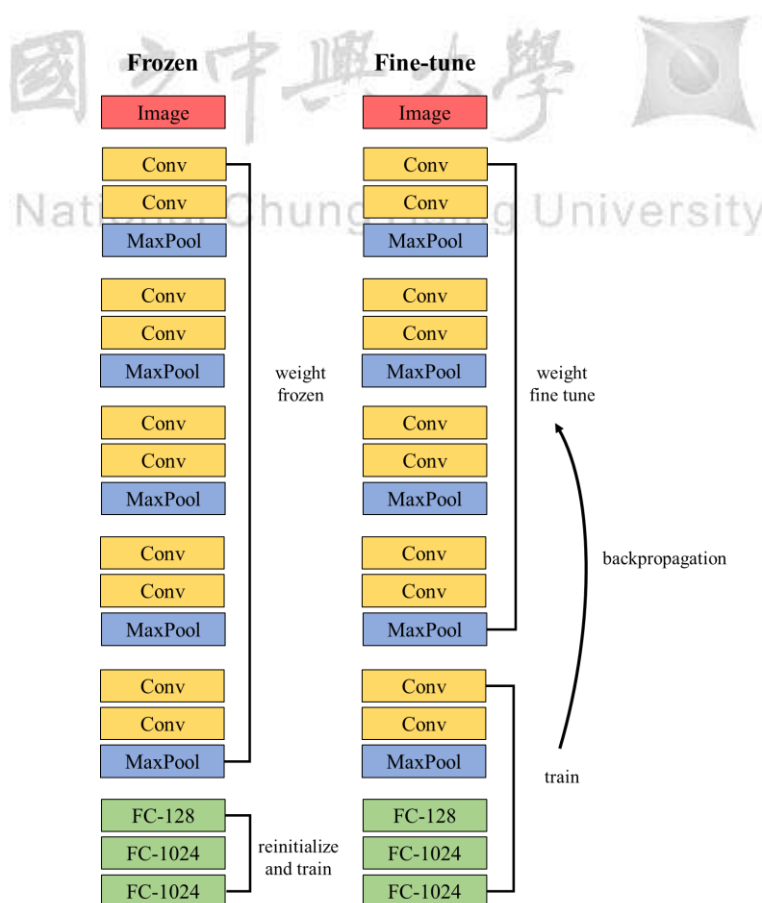


圖 3 遷移學習的凍結與微調方法訓練過程示意圖

### 2.2.2. 遷移學習應用於電子健康紀錄

在以往的研究中，遷移學習多被利用在影像分析上，在醫學領域也不例外，如 [22-25]，皆在醫學影像辨識應用的準確率上有明顯的進步；在 EHRs 資料的應用上，有利用跨醫院數據在相同預測任務模型的參數遷移[18]，Lee 等人[26]以人工神經網路(Artificial Neural Networks, ANNs)為模型基礎進行遷移學習，用於小 EHRs 數據集訓練去識別化的模型，Desautels 等人[27]用遷移學習搭配決策樹，在預訓練模型參數共享下，幫助資料數少的醫院之死亡率預測模型準確率改善，Priyanka 等人[28]利用基於遞迴神經網路(Recurrent Neural Network, RNN)的時間序列特徵提取模型 TimeNet[29]和 HealthNet[28]對生理訊號資料進行特徵提取，接著將預訓練模型的參數遷移至目標模型上進行二分類的死亡率預測。

從以上的研究文獻中可以發現，遷移學習在 EHRs 資料的應用，多以針對單一疾病預測或特定任務數據量不足的模型建立問題，從數據量較多的醫院進行相同任務模型的參數遷移，再微調以符合目標數據集的特徵分布，在目標任務的準確率都有明顯的改善；然而，在遷移學習的目標任務上，卻鮮少利用不同疾病間相互影響的關係進行跨疾病的分析，因此，本研究將利用遷移學習對有關連性之疾病進行跨疾病的病程發展預測分析。

### 2.3. 遞迴神經網路(Recurrent Neural Network, RNN)

本研究利用 RNN 進行時序資料的在時間關係中的特徵提取，RNN 是透過週期性地連結隱藏層節點來捕捉時序資料內的動態資訊[30]，因此常被用於時序或空間上有強烈關係的任務[31]；然而，在 RNN 的模型訓練情況中，會出現梯度消失的問題，意即模型會遺忘早期的資訊，造成後續預測錯誤，因此發展出有額外記憶機制的長短期記憶(Long Short Term Memory, LSTM)[32]和 Gated Recurrent Unit(GRU)[33]，能夠比基礎的 RNN 記憶更多先前的資訊，進而得到較好的預測

準確率，因此在研究方法設計中使用這兩種 RNN 進行特徵提取及比較。本章節將針對這兩種時間循環網路進行介紹，並說明應用的研究概況。

### 2.3.1. LSTM

如上所訴，LSTM 主要是為了解決長時間的時間序列訓練時所造成的梯度消失的問題，意即模型訓練的過程中，處理後面時間的節點對於前面時間所捕捉到的資訊感知力會下降，無法學習到早期的訊息。LSTM 利用兩種傳輸狀態紀錄資訊，分別為單元狀態(Cell State)與隱藏狀態(Hidden State)，單元狀態主要是由上一個時間點的有用單元狀態資訊與當前輸入的有用資訊組成，是利用網路學習到的矩陣權重與組合向量相乘得出的控制閥決定哪些為有用資訊，組合向量則由當前輸入及前一時間點的隱藏狀態拼接而成；隱藏狀態則是當前單元狀態透過控制閥留下來的有用資訊，且參與了當前輸出的計算。另外，有名為雙向長短期記憶(Bi-directional Long Short-Term Memory, Bi-LSTM)[34]的 LSTM 變形，主要是由前向 LSTM 與後向 LSTM 組成，再將兩個方向的輸出結果拼接成為最終的輸出，意即不只捕捉過去對未來的關係，也捕捉未來對過去資訊的關係，例如在自然語言的語意判斷任務上，”這杯飲料不好喝”，此句話的不好喝用來形容前面的飲料，就能夠利用 Bi-LSTM 有效地捕捉與前面文字的語意關係。

透過以上的說明，可以發現 LSTM 能有效地將不需要的資訊遺忘，留下有用的資訊進行訓練，因此，在各領域被廣泛地使用且得到很好的效果，如在自然語言處理[35、36]、多變量時序預測[37]、生物學[38]和臨床醫學數據應用[39-41]。

### 2.3.2. GRU

GRU 的輸出及輸入結構與 RNN 的相似，但其中的單元計算是與 LSTM 相似，因此，使用 GRU 能夠達到 LSTM 相同的效果，並大幅地降低訓練成本。GRU 是由隱藏狀態(hidden state)與兩種門控訊號對過去的訊息進行篩選，分別是

重置門控(Reset Gate)和更新門控(Update Gate)，都是由矩陣權重與組合向量相乘得出，組合向量是當前輸入與前一時間點的隱藏狀態拼接而成。首先，將前一時間點的隱藏狀態透過重置門控得出新的前一時間點隱藏狀態表示，再將當前輸入和新的前一時間點隱藏狀態拼接起來與矩陣權重相乘，得到包含當前輸入的新隱藏狀態，最後透過更新門控對前一時間點隱藏狀態資訊的遺忘及新隱藏狀態資訊的記憶，進而獲得當前隱藏狀態，並利用當前隱藏狀態計算當前的輸出。GRU 的遺忘與記憶是只由更新門控進行控制，代表遺忘了多少前一時間的隱藏狀態資訊，會由包含當前輸入的新隱藏狀態補足，以保持恆定的狀態。

這種計算方式在計算資源和時間有限制的情況下，成為了處理時間序列資料的有效解決方法運用在各個領域上，如自然語言處理應用[42]、股票指數預測[43]與電子健康紀錄的 Doctor AI[18]。

## 2.4. 注意力機制(Attention Mechanism)

注意力機制是模仿人類大腦中的訊號處理機制[44]，當人類的視覺快速掃描過全域性資料，會將目光聚焦在需要關注的重點資訊上，對其投以更多注意力，而忽略其他無用的資訊，透過此機制可以提升對大量資訊的處理效率與準確性，並且提供對資訊的可解釋性。而注意力機制可以大致分為全局注意(Global Attention)和局部注意(Local Attention)，全局注意是將注意力放在全部的資料，意即若輸入有 4 個時間，就 4 個時間點皆投以關注，而局部注意只將關注放在少部分的輸入。注意力機制可以看成由一組鍵值(Key-Value)資料中的 Key 與目標的搜尋(Query)計算相似性得出權重值，意即各輸入對目標輸出的影響程度，再將權重值與 Value 進行加權求和，計算可以分成以下三個步驟[45]:

$$e_{ij} = a(s_{i-1}, h_j) \quad (2-1)$$

公式(2-1)利用 RNN 在時間  $i-1$  的隱藏狀態  $s_{i-1}$  與輸入第  $j$  個位置的  $h_j$  計算輸入  $j$  與輸出  $i$  的相似程度，其中  $a$  為計算相似性的函數，常見的計算方法可以分為以下

幾種[45]:

$$a(s_{i-1}, h_j) = \begin{cases} s_{i-1}^\top h_j & \text{dot} \\ s_{i-1}^\top W_\alpha h_j & \text{general} \\ v_\alpha^\top \tanh(W_\alpha [s_{i-1}; h_j]) & \text{concat} \end{cases}$$

公式(2-2)中 $\alpha_{ij}$ 為使用 $Softmax$ 標準化後的權重值，

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_z} \exp(e_{ik})} \quad (2-2)$$

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j \quad (2-3)$$

最後根據 $\alpha_{ij}$ ，進行加權求和得到輸出 $c_i$ 。

除了 Bahdanau 等人提出的基礎注意力機制結構[45]與 Luong 等人在注意力機制基礎上進行權重計算的變化[46]，還有隨著任務不同進而發展出注意力機制的變形[47-49]，應用在各類預測任務中，如多變量時序預測[37、50]，EHRs 資料應用如 Choi 等人[15]將病患就診紀錄以倒序方式輸入時序模型並搭配兩層注意力機制，用以預測病患未來心臟衰竭的機率，在 Sha 等人[51]的研究中利用 GRU 搭配注意力機制預測病患未來的死亡率，而 Yang 等人[4]與 Kaji,等人[52]則是利用 LSTM 和雙向 LSTM 搭配注意力機制分析加護病房病患的數據建立預測未來疾病的模型。

由此可見，注意力機制可以靈活地變化且使用在不同類型的任務上，因此，本研究將利用遷移學習搭配注意力機制，在預訓練模型的參數遷移至目標任務模型後，加入注意力機制進行微調訓練，使模型能更加專注於重要的資訊且預測的結果更富有可解釋力。

## 第三章 研究設計與實施

在目標疾病病程發展預測的分析中，需要解決長時序、高維度特徵、系統性誤差與資料樣本少的問題。因此本研究針對患者的歷史就診紀錄，透過當量轉換成為適合進入分析模型之資料格式，使用遷移學習搭配時間序列模型有效地提取特徵並解決資料樣本數少的問題，最終得到目標疾病病程發展預測之結果；以上所訴，即為利用遷移學習進行跨疾病分析方法之框架。本章節將介紹系統架構，並說明各步驟的設計與實行。

### 3.1. 研究流程

圖 4 為本研究提出的一個基於電子健康紀錄資料的研究流程，本研究想透過分析過去的歷史就醫紀錄來了解病患身體健康狀態並預測與其有相關的疾病之未來病程發展。第一步是為了解決資料的高維度、隨機性及系統誤差性而進行的資料預處理，資料分為兩種，第一種為來源域疾病的就醫紀錄，第二種是目標域疾病的就醫紀錄，將兩個種資料進行當量轉換並加總成固定時間間隔且適合預測模型分析之各病患的病程表示矩陣，第二步是跨疾病的病程發展預測方法(Transfer Learning-Cross Disease Course, TL-CDC)，使用來源疾病的病程表示矩陣建構來源疾病病程發展預測模型，再利用來源疾病病程發展預測模型的權重於目標疾病病程發展預測模型且針對目標任務進行預測模型訓練，最後輸出目標疾病的病程發展預測。將於下一節介紹本研究所使用的資料集，3.3 節說明資料預處理的過程，並於 3.4 節詳細介紹跨疾病的病程發展預測方法。

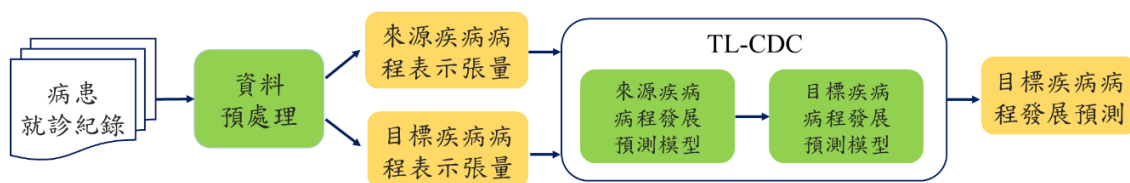


圖 4 研究流程圖

## 3.2. 資料集

台灣全民健康保險(健保)在行政院衛生署的監督下於 1995 年 3 月 1 日全面實施，是屬於全國統一費率級距的強制性社會保險，主要目標是使全民可以獲得公平的醫療資源，在此制度下，納保率高達 99% 以上，使得健保資料成為醫療研究中極具代表性的資料，因此，將以健保研究資料庫為資料來源進行研究。本研究是以遷移學習為主要研究方法，因此，此章節分為來源域資料與目標域資料進行說明。

### 3.2.1. 來源域資料

來源域資料使用 1999 至 2010 年牙周病患者的就醫紀錄，本研究使用如圖 5 所示的門診處方及治療明細檔(Ambulatory care expenditures by visits, CD)、門診處方醫令明細檔(Details of ambulatory care orders, OO)，資料記錄格式如圖 5 的表格所呈現，本研究再來源域資料所使用之 CD1999 檔案代表所有牙周病患者在 1999 年進行門診的 CD 檔，圖中列出幾個本研究使用的欄位，ID 代表各病患已轉碼的國民身分證統一編號，SEQ\_NO 是案件流水號與 OO 檔案的流水號一致，可以串起兩個檔案，用以得知就診藥品代碼(DRUG\_NO)及藥品用量(DRUG\_USE)，DRUG\_DAY 則為給藥日份，ACODE\_ICD9\_1 國際疾病分類號一是用以記錄傷病名稱，總共有 3 欄進行記錄。

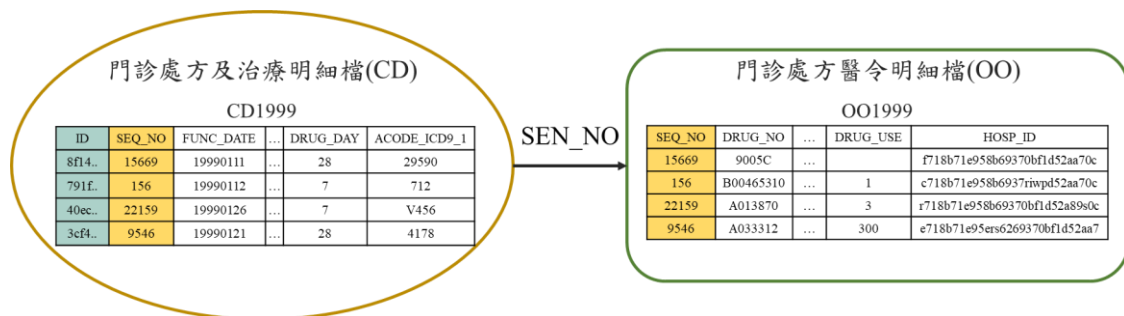


圖 5 就醫紀錄資料表示圖



本研究將透過此經過串檔的門診診斷資料了解病患的病程發展；而牙周病患者的篩選條件是 2000 年位於 CD 檔的國際疾病分類號(ACODE\_ICD9)1-3 任一欄診斷代碼為 5233-5235 兩次或兩次以上，共 864,478 名，其中有 728,353 人擁有 12 年完整的門診紀錄，為了準確評估患者之病程發展，本研究刪除牙周病用藥相關數值紀錄有異狀的 1,868 人，包含給藥日分紀錄為使用頻率的 129 人，如紀錄為 TIW 代表每星期 3 次，與藥品代碼是為牙周病用藥但給藥日份為空值的 1,739 人，而本研究也將用藥數值為離群值的 2,330 人刪除，其 3 個月內用藥天數超過 100 天，最終為本研究來源域資料共有 724,155 人。

### 3.2.2. 目標域資料

目標域資料則是從 724,155 名患者在 1999 至 2010 年間國際疾病分類號 1-3 任一欄診斷代碼為 7140，共 31,160 人，刪除類風濕性關節炎將用藥數值為離群值的 2 人與病患使用的藥物種類無法進行病程評估比較的 2,855 人，如 ATC 碼為 L01BA01、L01AA01...等，此類藥物沒有製訂 DDD 導致無法將用藥劑量進行後續的當量轉換動作，以至於無法評估該病患的病程發展，最終為本研究目標域的資料共 28,303 人。

### 3.3. 資料預處理

由於健保資料庫之數據特徵欄位維度高且病患歷史就醫間隔時間長度不一致，意即不容易了解各病患在各個時間點的病程發展，而此問題會對預測模型的訓練結果造成影響，導致預測效果不準確，因此，需要透過本章節的資料預處理方法，利用適合表示病患病程發展的特徵進行資料轉換與計算，再進入分析模型中。

本研究使用圖 6 所示的流程，透 SEQ\_NO 串連 CD 及 OO 檔，整理出 1999 至 2010 年共 12 份的 CD&OO 檔，再利用 ID 整理出屬於各病患 1999 至 2010 年的所有門診紀錄。隨著來源與目標疾病的病程表示特徵不同，所需要的資料預處理方式

也不同，因此，本節將分為來源域資料預處理方法及目標域資料預處理方法兩個部分詳細說明。

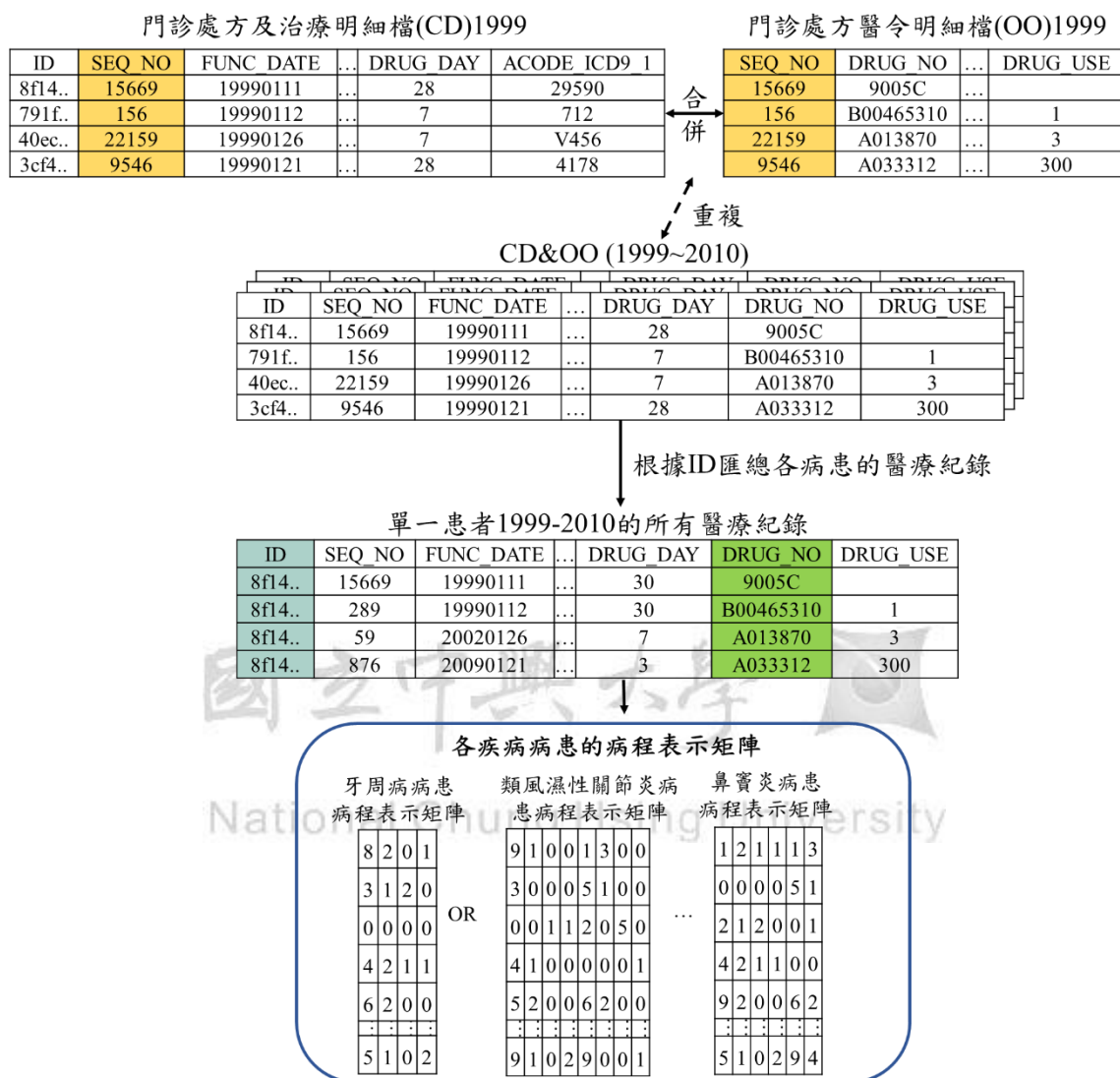


圖 6 資料前處理流程圖

### 3.3.1. 來源域資料預處理方法

本研究在來源域疾病的病程表示特徵的是使用專業醫師建議之治療牙周病的抗生素用藥天數(PD\_DAY)、用藥種類數(PD\_DRUG)、牙周病相關手術次數(PD\_DISPOSAL)與洗牙次數(PD\_SCALING)來表示牙周病患者的病程發展。而資料的預處理方法是使用先前整理出的 724,155 份牙周病患者 1999 年至 2010 年的

所有門診紀錄，透過表 1 的牙周病病程表示矩陣計算方式來比對 DRUG\_NO 是否為牙周病治療的代碼。

表 1 牙周病病程表示矩陣計算方法

**牙周病用藥天數計算(PD\_DAY):**

IF DRUG\_NO = PD\_DRUG\_CODE :

PD\_DAY += DRUG\_DAY

PD\_DRUG\_CODE = [A000080277, A000081100, ..., A00054110]

**牙周病用藥種類數計算(PD\_DRUG):**

IF DRUG\_NO = PD\_DRUG\_CODE :

IF DRUG\_NO NOT IN PD\_DRUG\_LIST :

PD\_DRUG\_LIST.APPEND(DRUG\_NO)

PD\_DRUG = LEN(PD\_DRUG\_LIST)

**牙周病相關手術次數計算(PD\_DISPOSAL):**

IF DRUG\_NO = PD\_DISPOSAL\_CODE :

PD\_DISPOSAL += 1

PD\_DISPOSAL\_CODE = [91001C, 91006C-910012C, 92015C, 92016C]

**洗牙次數(PD\_SCALING):**

IF DRUG\_NO = PD\_SCALING\_CODE :

PD\_SCALING += 1

PD\_SCALING\_CODE = [91003C, 91004C]

本研究使用一個月的時間長度來加總牙周病病程發展特徵的數值，以表 2 ID 為 8f14 患者的 4 筆醫療紀錄進行說明，先用 FUNC\_DATE 欄位判別是否為 1999 年 1 月的門診紀錄，並且比對到 DRUG\_NO 為牙周病治療的抗生素代碼 A000080277 時，記錄 DRUG\_DAY 欄位的 7，並且將用藥種類數加一，如表 3 TIME 為 1 的 PD\_DAY 和 PD\_DRUG 欄位，當比對到 DRUG\_NO 為牙周病手術代碼

91001C 時，將牙周病手術次數加一，其手術日期為 1 月 18 日，因此紀錄在 TIME 為 1 代表手術次數的 PD\_DISPOSAL 欄位上；而 2 月沒有比對到牙周病治療的相關代碼，因此 TIME 為 2 的數值皆為 0；當 3 月醫療紀錄的 DRUG\_NO 欄位比對到代表洗牙的代碼 91003C，將 TIME 為 3 代表洗牙次數的 PD\_SCALING 欄位數值加一。

表 2 ID 為 8f14 的病患之 1999-2010 年醫療紀錄

ID	SEQ_NO	FUNC_DATE	...	DRUG_DAY	DRUG_NO	DRUG_USE
8f14	15669	19990111	...	7	A000080277	5
8f14	7	19990118	...		91001C	
8f14	56	19990202	...	7	A013870	3
8f14	996	19990321	...		91003C	

表 3 ID 為 8f14 的病患之病程表示矩陣(一個月為一筆的時序資料)

TIME	PD_DAY	PD_DRUG	PD_DISPOSAL	PD_SCALING
1	7	1	1	0
2	0	0	0	0
3	0	0	0	1
⋮	⋮	⋮	⋮	⋮
144	14	2	1	0

### 3.3.2. 目標域資料預處理方法

本研究在目標域資料預處理方法使用目標域疾病的病程表示特徵除了牙周病病程表示的四個欄位之外，加入專業醫師建議之類風濕性關節炎口服與注射藥物的使用量與種類數來表示患者的病程發展，分別為類風濕性關節炎口服藥物使用量(RA\_USE\_ORAL)、口服藥物種類數(RA\_DRUG\_ORAL)、注射藥物使用量(RA\_USE\_INJ)和注射藥物種類數(RA\_DRUG\_INJ)。而資料預處理方法是使用先前整理出的 28,303 份類風濕性關節炎患者 1999 年至 2010 年的所有門診紀錄，透過比對 DRUG\_NO 是否為治療類風濕性關節炎的用藥代碼，再進行藥物使用量的當

量轉換。

在藥物的使用量紀錄上，本研究使用 DDD 進行藥物的統一度量，根據 WHO 的定義，DDD 為平均 70 公斤成人針對主要適應症假定的每天平均維持劑量，以類風濕性關節炎治療用藥 DMARDs 組藥物為例，若 A 醫院內的藥物為 leflunomide，其 ATC 碼為 L04AA13，WHO 訂定的 DDD 是 20mg，院內最小包裝為 1cap，含量為 15mg/cap，醫生開立的處方為 2cap/day，則藥物使用量為 1.5DDD，而 B 醫院內的藥物為 ciclosporin，其 ATC 碼為 L04AD01，WHO 訂定的 DDD 是 250mg，藥物最小包裝為 1cap，含量為 125mg/cap，醫生的處方為 3cap/day，藥物使用量為 1.5DDD，由此可見，在不同醫院的同類型不同藥物間可以透過 DDD 進行當量換算，將藥物使用量標準化，並了解與比較患者的病情嚴重程度。

以表 4 ID 為 791f 患者的 4 筆醫療紀錄進行說明，先用 FUNC\_DATE 欄位判斷是否為 1999 年 1 月的門診紀錄，並且比對到 DRUG\_NO 為類風濕性關節炎治療用藥代碼 B023615100 時，將其 DRUG\_USE 除以 B023615100 的 DDD 為 20mg，等於 1.5DDD 與其用藥代碼一起記錄下來，再比對到就診日期為 1 月 16 日的 B018142100，將其 DRUG\_USE 除以 B018142100 的 DDD 為 250mg，等於 0.2DDD 與其用藥代碼記錄下來，最後將 1 月的所有 DDD 加總起來並且計算用藥代碼數目，輸出如表 5 TIME 為 1 的 RA\_USE\_ORAL 和 RA\_DRUG\_ORAL 欄位；而 2 月比對到 AC45781100，將其 DRUG\_USE 除以 AC45781100 的 DDD 為 150mg，等於 0.667DDD，因此 TIME 為 2 的 RA\_USE\_ORAL 和 RA\_DRUG\_ORAL 欄位數值為 0.667 與 1；當 3 月醫療紀錄的 DRUG\_NO 欄位比對到代表洗牙的代碼 91003C，將 TIME 為 3 代表洗牙次數的 PD\_SCALING 欄位數值加一。

表 4 ID 為 791f 的病患之 1999-2010 年醫療紀錄

ID	SEQ_NO	FUNC_DATE	...	DRUG_DAY	DRUG_NO	DRUG_USE
791f	159	19990105	...	5	B023615100	30
791f	66	19990116	...	2	B018142100	50
791f	355	19990220	...	7	AC45781100	100
791f	458	19990330	...		91003C	

表 5 ID 為 791f 的病患之病程表示矩陣(一個月為一筆的時序資料)

TIME	1	2	3	...	144
PD_DAY	0	0	0	...	3
PD_DRUG	0	0	0	...	1
PD_DISPOSAL	0	0	0	...	0
PD_SCALING	0	0	1	...	0
RA_USE_ORAL	1.7	0.667	0	...	12.5867
RA_DRUG_ORAL	2	1	0	...	4
RA_USE_INJ	0	0	0	...	0
RA_DRUG_INJ	0	0	0	...	0

### 3.3.3. 病程發展表示矩陣

透過比對計算與當量轉換將原始資料進行標準化及加總以表示患者的疾病病程狀態，新的資料會因各病患有不同的就醫時間間隔長度而難以表現其當前的健康狀態，導致時間序列模型無法有效捕捉疾病發展的時序特徵，進而影響任務的預測準確度，因此，本研究將病患的就醫紀錄根據研究所定的特徵，將用藥天數、用藥種類數及相關手術等欄位進行加總，藉由時間內的用藥加總使病患的用藥紀錄能有統一且適合後續時序關係特徵捕捉的呈現方式，來源疾病病患病程表示矩陣為四欄，目標疾病病患病程表示矩陣為八欄，記錄其一個月內的數值，共 144 筆時序資料，如表 5，本研究將以序列是一個月的資料為基礎做不同序列的長度加總，了解序列長度對病程發展預測的影響，將加總的序列長度設為  $k$ ，當  $k=3$  時代表加總成三個月為一筆序列資料的病患特徵矩陣，亦代表病患一季的病程發展，如將表

5 TIME 為 1, 2, 3 的數值相加為表 6 的 TIME 為 1，在此僅以目標疾病的病程表示矩陣為例。

表 6 ID 為 791f 的病患之病程表示矩陣(k=3)

TIME	1	2	3	...	48
PD_DAY	0	0	0	...	5
PD_DRUG	0	0	0	...	1
PD_DISPOSAL	0	0	0	...	0
PD_SCALING	1	0	0	...	2
RA_USE_ORAL	2.367	5.3	0	...	49.1734
RA_DRUG_ORAL	3	1	0	...	4
RA_USE_INJ	0	0	0	...	0
RA_DRUG_INJ	0	0	0	...	1

### 3.4. TL-CDC

本研究使用遷移學習做為跨疾病的病程發展預測方法的框架，在來源疾病病程發展預測模型中，本研究使用 RNN 作為提取特徵的深度學習模型，包含 LSTM、Bi-LSTM、GRU，這三種專門處理時序資料的時間序列模型增加了額外的記憶機制來進行時序關係的特徵提取，主要會由 Input gate、Forget gate、Output gate 以及 Memory cell candidate 元件組成；而目標疾病病程發展預測模型則是將來源疾病病程發展預測模型訓練出的時序特徵參數遷移至目標疾病病程發展預測模型的時間序列模型中，再加入注意力機制的計算進行微調，最後輸出下一個時間點的病程預測。以下分為來源疾病病程發展預測模型與目標疾病病程發展預測模型進行架構流程講解。

#### 3.4.1. 來源疾病病程發展預測模型

本研究使用的來源疾病病程發展預測模型架構如圖 7。為了有效捕捉時間上的關係，選擇基於 RNN 的 LSTM、Bi-LSTM、GRU，並加入注意力機制的計算

進行來源疾病病程的特徵提取，而本研究也針對加入注意力機制對於參數遷移後的目標疾病預測任務效果進行比較。本小節將針對來源疾病病程發展預測模型的時間序列模型與注意力機制進行詳細講解。

#### 3.4.1.1 時間序列模型

本研究使用時間序列模型中的 Many-to-One 架構，首先，將過去四個時間點的牙周病病程表示輸入時間序列模型，若是以  $k$  為 3 的病程矩陣之四個時間點資料輸入模型，代表是輸入患者一年的病程資料，並輸出下一時間點的病病程表示，亦為後續三個月的病程發展預測，如圖 7 中上方多位病患資料組成的來源疾病病程表示張量作為輸入資料，而其中一個二維的矩陣代表一位患者的病程發展，綠色處一列代表一個時間點的病程表示，首先將要輸入的資料加 1 取  $\log_{10}$ ，是為避免因各病患的用藥量差異大所造成的模型收斂不完全問題，再將處理完的四個時間點資料點輸入時間序列模型進行訓練並預測紅色處代表第五個時間點的病程發展，接著再向下滑動至藍色方框處作為係一次進入模型輸入的四個時間點，預測下一時間點淺紅色疾病表示處；透過固定的序列長度輸入時間序列模型，再經由網路中 Input gate、Forget gate、Output gate 以及 Memory cell candidate 元件進行時間關係的特徵捕捉，而時間循環網路在訓練的過程中需要透過這些元件去計算權重，太多層的時間循環網路會造成訓練速度緩慢或擬合不足 (Underfitting) 的問題，且本研究所使用的來源疾病樣本數多，鑒於時間與硬體成本的上考量，本研究設定一層的時間循環網路及 128 個隱藏層神經元進行時間序列模型的建置，再將網路的最後一個輸出進入全連接層預測出下一個時間點的病程發展表示，以圖 7 的虛線表示時序模型輸出後的運作流程，在此僅以 LSTM 為示意圖。



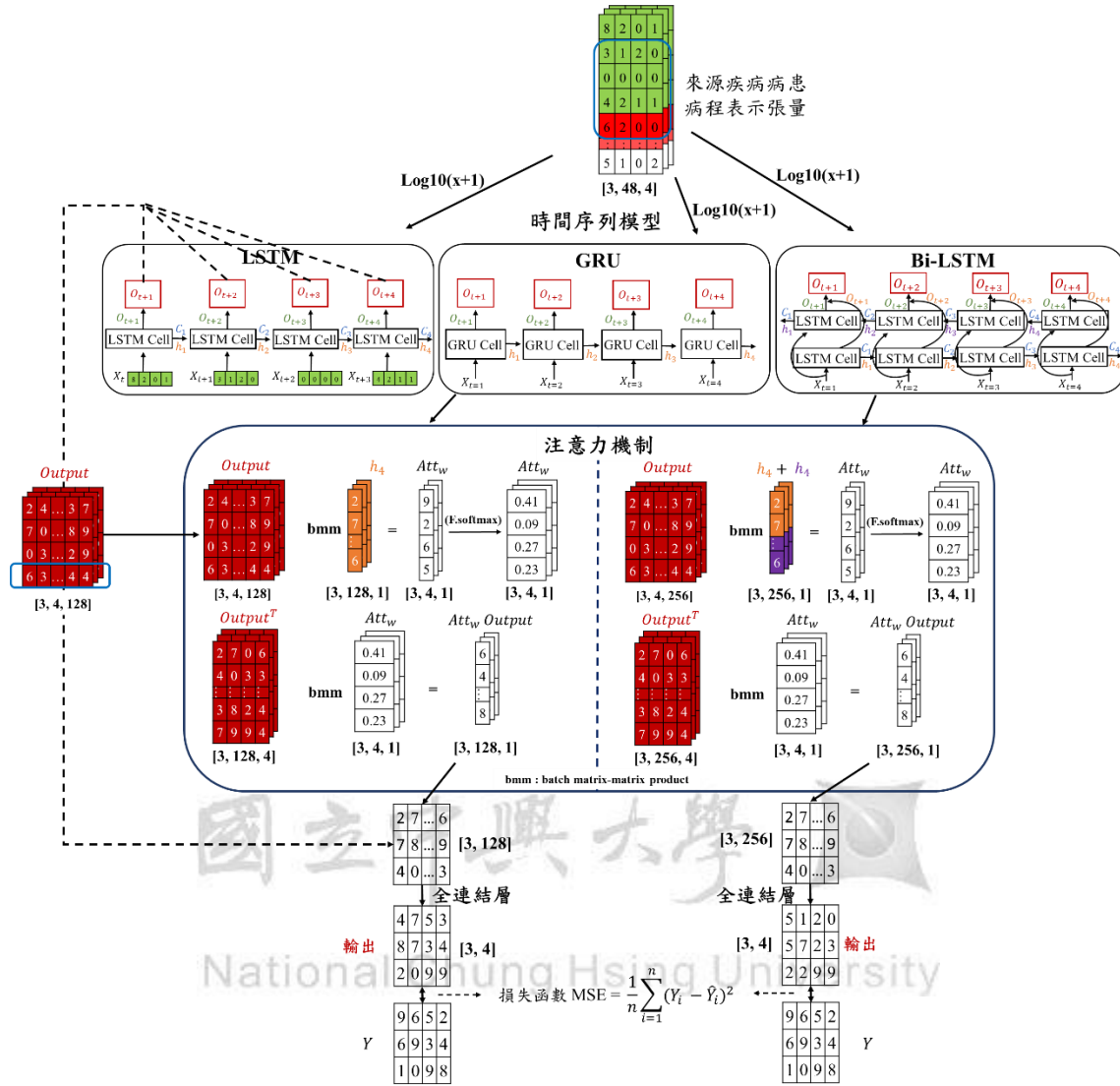


圖 7 來源疾病病程發展預測模型

### 3.4.1.2 注意力機制

本研究在注意力機制的設計是使用 Global Attention 來關注 4 個時間點對預測時間點病程發展的重要程度，如圖 7 中的注意力機制部分，主要是透過時序模型 4 個時間點的輸出(圖示虛線接實線處)與能夠間接表示預測時間點輸出的最後一個隱藏狀態計算出各時間點的資訊對預測時間點輸出的重要程度，圖  $Att_w$  代表經過計算後的注意權重，計算如公式(3-1)，

$$e_{ij} = a(\text{Output}, h_4) \quad (3-1)$$

公式(3-2)將各時間點的注意權重進行標準化，使其相加起來等於 1，

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_z} \exp(e_{ik})} \quad (3-2)$$

公式(3-3)再將時序模型的輸出與權重相乘得出加權過後的輸出向量 $c_i$ ，如圖示  $Att_w$  Output，

$$c_i = \sum_{j=1}^{T_x} Output^T \alpha_{ij} \quad (3-3)$$

最後進入全連接層預測下一個時間點的病程發展表示，圖 6 中注意力機制的左半部為完整的輸入與輸出計算過程。。

而注意力機制的右半部計算，是 Bi-LSTM 搭配注意力機制的計算示意圖，透過 Bi-LSTM 計算前向與後向時間循環網路得到的隱藏狀態拼接起來，如公式 (3-4)  $[\overrightarrow{h_{t-1}}; \overleftarrow{h_{t-1}}]$  所表示，得到間接代表預測時間點的隱藏狀態，再與 Bi-LSTM4 個時間點的輸出計算注意權重，

$$e_{ij} = a(Output, [\overrightarrow{h_{t-1}}; \overleftarrow{h_{t-1}}]) \quad (3-4)$$

公式(3-5)將各時間點的注意權重進行標準化，使其相加起來等於 1，

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_z} \exp(e_{ik})} \quad (3-5)$$

公式(3-6)再將時序模型的輸出與權重相乘得出加權過後的最終輸出向量 $c_i$ 。

$$c_i = \sum_{j=1}^{T_x} Output^T \alpha_{ij} \quad (3-6)$$

### 3.4.2. 目標疾病病程發展預測模型

本研究方法的目標疾病病程發展模型框架如圖 8。為了使目標預測任務能夠以最低限度的有效訓練資料量進行病程發展預測模型訓練，利用與其有相似分布的來源疾病病患的就醫紀錄所建置之病程發展預測模型。首先，透過線性轉換將目標疾病病程表示張量轉成與來源疾病病程表示張量相同特徵維度的資料，再將來源疾病預測模型預訓練好的參數遷移至目標疾病病程發展預測模型的時間序列

模型上，再針對目標任務進行訓練。

遷移來的參數分為兩種，第一種是時間序列模型捕捉時序關係後直接預測下一時間點的來源疾病病程發展，如圖 8 來源疾病病程發展預測模型的藍色框參數遷移所示，第二種則是時間序列模型加入注意力機制計算的參數，如上節所述，加入注意力機制計算對下一時間點的來源疾病病程發展預測有影響之過去時間點資訊增加較高的權重配置，進而預測下一時間點的來源疾病病程發展預測，模型訓練後含有注意力機制計算的時間序列模型參數遷移如圖的紅色框所示。

時間序列模型的輸出分為兩種針對目標任務的預測方式，第一種是輸出直接進入全連結層進行目標預測任務的學習訓練並預測，第二種是將輸出經過針對目標任務之注意力機制的加權計算後再進入全連結層對目標預測任務進行預測。而目標疾病病程發展預測模型的注意力機制計算，如圖 8 所示，與來源疾病病程發展預測模型的計算方式相同，將對下一時間點的目標疾病病程發展預測有較大影響之過去時間點資訊賦予較高權重，使對預測任務有幫助的資訊能有更大的影響力，進而增加目標疾病預測任務的效果

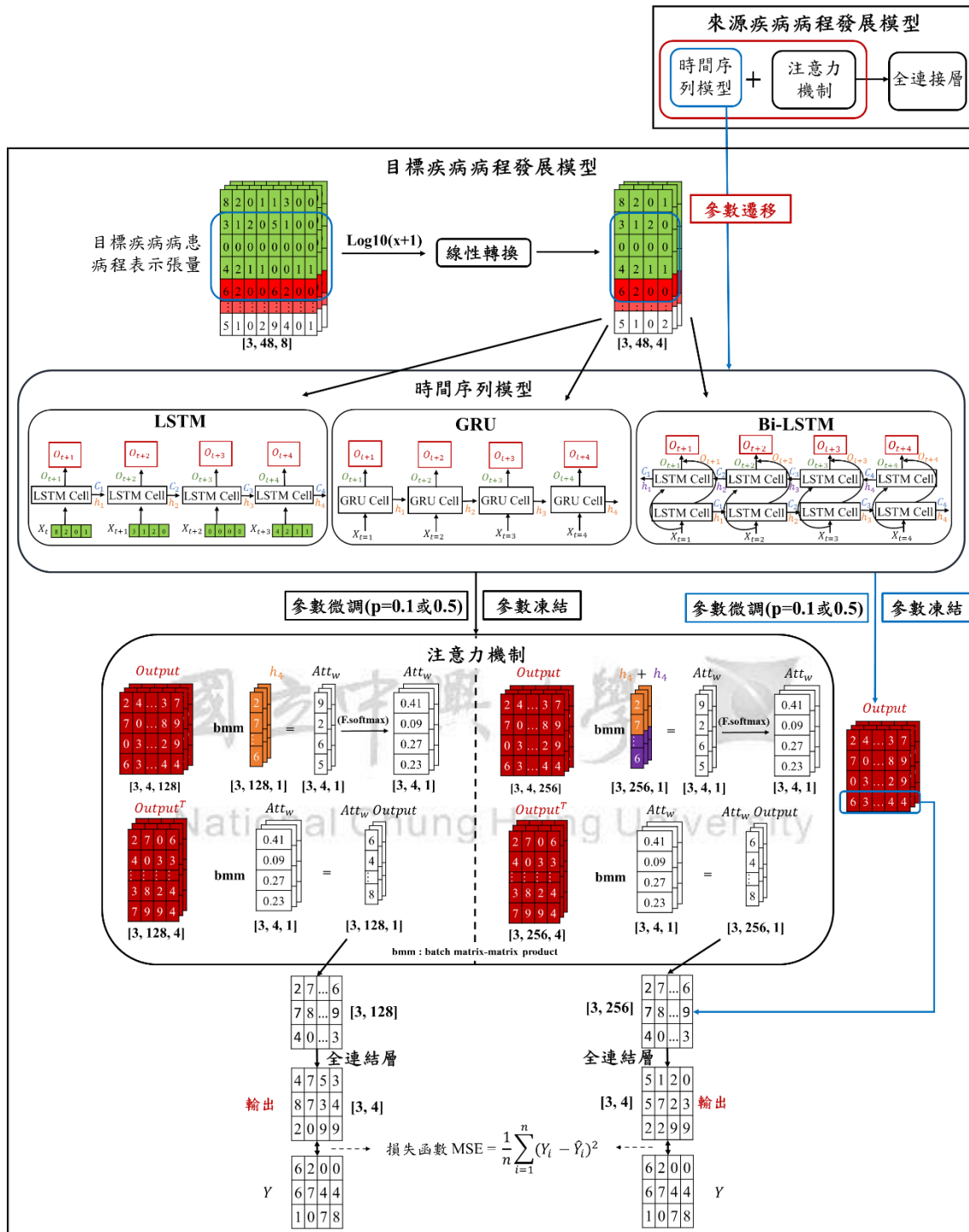


圖 8 目標疾病病程發展預測模型

## 第四章 實驗結果與分析

本章節在探討應用遷移學習於跨疾病病程發展預測的效能，將先描述訓練數據設置與實驗結果評估定義，後針對效能評估進行說明。

### 4.1. 實驗數據設置

本研究利用遷移學習作為研究方法的框架，因此將分別對來源疾病病程發展預測模型與目標疾病病程發展預測模型的訓練數據設置製作說明。

來源疾病病程發展預測模型是使用 724,155 位牙周病患者就醫紀錄所製成的病程表示張量進行訓練，三成患者的病程表示張量作為測試資料，其餘七成的訓練資料在每個 epoch 訓練隨機抽出兩成作為驗證資料，用以評估模型的效能，模型訓練的 batch size 設定為 1000，epoch 設定為 50，學習率為  $1e-4$ ，優化器使用 Adam。

目標疾病病程發展預測模型是使用 28,303 位類風濕性關節炎患者就醫紀錄所製成的病程表示張量進行訓練，三成患者的病程表示張量作為測試資料，其餘七成的訓練資料在每個 epoch 訓練隨機抽出兩成作為驗證資料，用以評估模型的效能；目標疾病患者資料量少，為避免太大的 batch size 會造成模型訓練不穩定，將 batch size 設定為 300，又因利用遷移學習能有效減少目標疾病預測任務的訓練，將 epoch 減少為 20，學習率為  $1e-4$ ，優化器使用 Adam。。

### 4.2. 實驗結果評估函數

深度學習的損失函數(Loss Function)又稱成本函數(Cost Function)，是用來評估模型的預測值與真實值之間的差異程度，因此，亦成為神經網路訓練過程要優化的目標函數，神經網路訓練的過程就是使損失函數最小化的過程，當損失函數越小代表預測值與真實值的差異越小，意即模型的預測效果越好。

損失函數可以大致分為分類與回歸兩個面向，而模型的主要任務是預測用藥的相關數值，所以是屬於回歸面向。本研究使用在回歸預測上最常被使用的均方誤差(Mean square error, MSE)作為訓練時的損失函數，計算如公式(4-1)，

$$MSE = \frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n (Y_{j,k} - \hat{Y}_{j,k})^2 \quad (4-1)$$

其計算方法是求預測值( $Y_i$ )與真實值( $\hat{Y}_i$ )之間差異的均方值，公式中的平方是為了避免相減後的數值有負號而影響累加後的損失值，m 代表患者人數，n 代表要預測的目標數目，均方的計算是因為模型訓練時樣本及預測目標數不為 1，因此必須計算所有樣本的平均值。透過公式(4-1)可以知道計算出來的 MSE 越趨近 0 為最好的預測效果。而本研究在結果的評估函數亦直接使用模型訓練時的損失函數值進行預測效果的評估。

### 4.3. 實驗結果

本節將針對研究所提出的 TL-CDC 方法對於目標疾病病程發展預測效果進行呈現與討論，將實驗數據序列長度 k 定為 2、3、4 進行目標疾病病程預測比較，以了解序列長度對預測的影響，並分為 TL-CDC 方法的實驗比較及與其它處理多變量時列資料模型的外部實驗比較。

#### 4.3.1. TL-CDC 實驗結果

此小節將探討本研究方法的預測效果，將先討論遷移學習的預測有效性，再說明遷移學習的參數凍結與微調方法對病程預測的影響，後比較加入注意力機制計算對病程預測的有效性。

#### 4.3.1.1. 探討遷移學習對目標疾病病程發展預測的影響

表 7 中來源疾病病程預測模型的時序模型分為有無加入注意力機制計算進行病程發展預測分析，FZ 代表凍結來源疾病病程發展預測模型中的時序模型訓練出來的參數再遷移至預測目標疾病病程發展預測模型，意即來源疾病病程發展預測模型的參數將完全不參與目標疾病病程發展預測任務訓練的反向傳播計算，而 TA 則是初始化權重的訓練，意即只有使用目標疾病的病程表示矩陣進行病程發展預測分析。

從表 7 可以看出在不同序列長度的資料上，將來源模型的參數凍結遷移至目標模型上進行目標疾病病程發展預測所表現出的效果，大部分都沒有直接使用目標疾病的病程表示矩陣進行預測分析來的好，在三種不同序列長度中只有使用不加入注意力機制計算的 Bi-LSTM 與 GRU 進行參數遷移，有較好的預測效果表現。我們認為來源疾病的注意力機制計算是針對來源疾病的序列資料加以關注，因此在直接凍結來源模型參數的情況下，不適合將有加入注意力機制計算的參數遷移至目標疾病的預測模型。

表 7 參數凍結方法與初始化權重對目標疾病病程預測效果比較表

來源模型 \	k=2		k=3		k=4	
	FZ	TA	FZ	TA	FZ	TA
<b>LSTM</b>	0.0602	0.0561	<b>0.0662</b>	<b>0.0879</b>	0.0947	0.094
<b>LSTM w/ Att</b>	0.0701	0.0694	0.089	0.0866	0.01	0.094
<b>Bi-LSTM</b>	<b>0.0671</b>	<b>0.0687</b>	0.0624	0.0602	0.0103	0.0945
<b>Bi-LSTM w/ Att</b>	0.0606	0.0556	0.0611	0.0596	0.0967	0.0942
<b>GRU</b>	0.0721	0.0686	<b>0.0602</b>	<b>0.0611</b>	<b>0.1061</b>	<b>0.1457</b>
<b>GRU w/ Att</b>	0.0588	0.0556	0.0642	0.0596	0.0981	0.0939

#### 4.3.1.2 探討遷移學習的凍結與微調參數對目標疾病病程發展預測的影響

對於遷移來的參數，本小節將探討凍結與微調在目標疾病病程發展預測模型預測效果的差異，在微調比較上，本研究討論 10%與 50%的機率參與反向傳播計算對預測效果的影響。

表 8 的 FT 代表微調方法且分為 10%與 50%的機率進行，從凍結與微調的比較表格發現，使用目標疾病的資料對遷移來的參數進行部分微調能有效提升預測效果，尤其參數遷移利用在 LSTM 與 Bi-LSTM 上，不論是否在來源疾病病程發展預測模型中加入注意力機制，利用微調方法進行參數遷移都能提升效果。基於 GRU、k=3,4 的參數遷移效果沒有明顯提升，我們認為是當序列的資訊含量變多，在針對目標疾病預測任務進行微調訓練時，GRU 這種記憶單元計算較簡單的時序模型，無法有效地保留從來源疾病預測模型遷移來對目標疾病預測任務有幫助的參數，導致微調方法在 GRU 的效果並無提升。

表 8 參數凍結與微調方法對目標疾病病程預測效果比較表

來源模型	k=2			k=3			k=4		
	FT		FZ	FT		FZ	FT		FZ
	0.1	0.5		0.1	0.5		0.1	0.5	
LSTM	0.0599	<b>0.0561</b>	0.0602	0.062	<b>0.0593</b>	0.0662	<b>0.0925</b>	0.095	0.0947
LSTM w/ Att	0.0567	0.0589	0.0701	0.0606	0.0602	0.089	0.1004	0.0942	0.1
Bi-LSTM	0.072	0.0709	0.0671	0.0598	0.0596	0.0624	0.0936	<b>0.0928</b>	0.103
Bi-LSTM w/ Att	<b>0.0574</b>	0.058	0.0606	0.0602	<b>0.0593</b>	0.0611	0.0942	0.0973	0.0967
GRU	0.0685	0.079	0.0721	0.0721	0.067	0.0602	0.1095	0.1099	0.1061
GRU w/ Att	0.0573	<b>0.0562</b>	0.0588	0.0611	0.061	0.0642	0.0938	0.0967	0.0981

#### 4.3.1.3 探討有無加入注意力機制對目標疾病病程發展預測的影響

本小節針對加入注意力機制的預測效果比較在下表，其中的 TA w/ Att 表示對目標任務預測加入注意力機制計算，以 LSTM w/Att (TA w/ Att)為例，代表遷移來



的參數有加入針對來源任務的注意力機制計算並且在目標疾病的預測模型訓練也有加入對目標任務的注意力機制計算。

表 9 注意力機制對目標疾病病程發展預測效果比較表

來源模型(目標模型)	k=2				k=3				k=4			
	FT		FZ	TA	FT		FZ	TA	FT		FZ	TA
	0.1	0.5			0.1	0.5			0.1	0.5		
LSTM(TA wo Att)	0.0599	0.0561	0.0602	0.0561	0.062	0.0593	0.0662	0.0879	<b>0.0925</b>	0.095	0.0947	0.094
LSTM w/ Att(TA wo Att)	0.0567	0.0589	0.0701	0.0694	0.0606	0.0602	0.089	0.0866	0.1004	0.0942	0.1	0.094
LSTM(TA w/ Att)	0.056	<b>0.0552</b>	0.0679	-	0.0594	<b>0.0573</b>	0.0587	-	0.1036	0.0926	0.0965	-
LSTM w/ Att(TA w/ Att)	<b>0.0552</b>	0.0557	0.0559	-	0.0599	0.0596	0.0616	-	0.0946	0.0969	0.0985	-
Bi-LSTM(TA wo Att)	0.072	0.0709	0.0671	0.0687	0.0598	0.0596	0.0624	0.0602	0.0936	<b>0.0928</b>	0.103	0.0945
Bi-LSTM w/ Att(TA wo Att)	0.0574	0.058	0.0606	0.0556	0.0602	0.0593	0.0611	0.0596	0.0942	0.0973	0.0967	0.0942
Bi-LSTM(TA w/ Att)	0.0564	0.0578	0.057	-	<b>0.0581</b>	0.0589	0.0609	-	0.0941	0.0938	0.0985	-
Bi-LSTM w/ Att(TA w/ Att)	<b>0.0562</b>	0.0578	0.0575	-	0.0692	0.0612	0.0888	-	0.0956	0.0948	0.0999	-
GRU(TA wo Att)	0.0685	0.079	0.0721	0.0686	0.0721	0.067	0.0602	0.0611	0.1095	0.1099	0.1061	0.1457
GRU w/ Att(TA wo Att)	0.0573	0.0562	0.0588	0.0556	0.0611	0.061	0.0642	0.0596	<b>0.0938</b>	0.0967	0.0981	0.0939
GRU(TA w/ Att)	0.06	0.0576	0.0633	-	0.0595	<b>0.0589</b>	0.0592	-	0.0989	0.0992	0.0983	-
GRU w/ Att(TA w/ Att)	0.0583	0.0585	<b>0.0557</b>	-	0.064	0.06	0.0608	-	0.0967	0.094	0.0988	-

National Chung Hsing University

表 9 呈現了注意力機制對目標任務的預測效果，可以發現參數遷移搭配針對目標任務的注意力機制對目標任務的預測效果有明顯的提升。以三種時間序列模型來看，k=2 的最低 MSE 是在目標任務預測加入注意力機制計算，且以 LSTM 的預測效果表現最好，在使用 Bi-LSTM 和 GRU 兩種時間序列模型的最好預測效果是在初始化權重模型上，而本研究提出之方法能在預測效果與其相差不大的情況下達到跨疾病的病程發展預測應用；在 k=3 的最低 MSE 皆是在來源模型的參數沒有加入針對來源任務的注意力機制計算再搭配針對目標任務的注意力機制計算進行微調訓練，且損失皆較使用初始化權重模型來的低，其中以 LSTM 的預測表現最好；而 k=4 的最低 MSE 是在目標模型的參數沒有加入針對目標任務的注意力機制計算，針對目標任務預測進行微調訓練，損失皆較使用初始化權重模型來的低，最好的預測效果是使用 LSTM 進行預測。從以上的結果可以得出，不同

序列長度資料適合使用不同的遷移方法，而序列長度為 3 的資料，在是否有進行參數遷移的預測效果有較明顯地差異，使用不同時序模型進行參數遷移時，遷移方法的表現也較有一致性。

### 4.3.2. TL-CDC 與其他時間序列模型比較結果

鑒於過去研究沒有適合多變量和長時序疾病資料的病程發展預測模型，本研究將用能夠處理多特徵和長時序資料之模型，對來源疾病病程發展矩陣分析預測，再將其訓練出的參數遷移至使用目標疾病病程發展矩陣進行分析之相同架構的模型來與本研究方法做預測效果及模型訓練時間的比較。

#### 4.3.2.1 效能評估

此小節針對本研究方法預測表現最好的 LSTM 之預測結果與傳統回歸、TPA-LSTM[37]和 DSANet[50]進行比較。由表 10 可以發現使用傳統線性回歸、TPA-LSTM 和 DSANet 的初始化權重之預測效果大部分都比使用參數遷移方法來的好，代表計算簡單的線性回歸和專門發展來捕捉時序關係的複雜模型方法皆不適合應用於遷移學習框架進行分析；在三種不同序列長度的資料上，本研究方法的 MSE 皆較其他方法低，代表其能有效地利用兩種相關但不同的疾病之病程資料在遷移學習框架下進行跨疾病病程發展預測分析。

表 10 TL-CDC 與其他時間序列模型對目標疾病病程發展預測效果比較表

	k=2				k=3				k=4			
	FT		FZ	TA	FT		FZ	TA	FT		FZ	TA
	0.1	0.5			0.1	0.5			0.1	0.5		
TL-CDC	0.0552	0.0557	0.0559	0.0694	0.0594	0.0573	0.0587	0.0866	0.0925	0.095	0.0947	0.094
Linear Regression	0.0805	0.0625	0.0805	0.0601	0.0648	0.0716	0.0815	0.0809	0.1103	0.0996	0.152	0.098
TPA-LSTM	0.0576	0.0559	0.0784	0.0566	0.0594	0.0591	0.0931	0.059	0.0971	0.0956	0.1539	0.954
DSANet	0.0672	0.0703	0.247	0.0624	0.0807	0.0824	0.0766	0.0754	0.1206	0.1077	0.8275	0.1322

#### 4.3.2.2 模型訓練時間評估

本研究在顯示卡 NVIDIA GeForce RTX 3090 Ti、24GB G6X，顯示卡加速軟體 CUDA11.2, CUDNN8.1.1 進行  $k=4$  的模型訓練，表 9 是模型訓練時間的比較，來源疾病病程預測表示利用 LSTM 作為來源疾病病程發展預測模型所花費的訓練時間，而目標疾病病程預測代表將來源模型參數遷移至目標模型後，目標疾病病程發展預測模型花費的訓練時間；由表 11 內的數據可以得知 TL-CDC 的預測效果表現最好且模型訓練時間相較於大部分模型減少許多。在此模型訓練時間比較上，以  $k=4$  進行比較，但當  $k$  為 2 或 3 時，序列資料數會倍數增加，而模型訓練時間成指數增加，在衡量硬體、時間成本與預測準確度之下，TL-CDC 為跨疾病病程發展預測的最佳選擇。

表 11 模型訓練時間比較表

訓練時間 (min)	k=4	
	來源疾病 病程預測	目標疾病 病程預測
TL-CDC	19.158	1
Linear Regression	10.826	0.593
TPA-LSTM	99.607	3.82
DSANet	6320.214	118.655

#### 4.3.3 注意力機制之解釋力呈現

為了證明注意力機制在預測任務中的效果，本研究分析了注意權重的分布。如圖 9 所示，橫軸代表輸入模型的四個時間點之病程資料，縱軸代表三名不同的病患，而圖為  $k=3$  時目標疾病病程發展預測模型訓練後的權重分布，亦表示三名患者輸入模型四個時間的病程資料對下一時間點目標疾病病程預測的影響程度，可以發現不同的病人在不同時間所呈現出的注意權重是不相同的，而三名患者表

現出三種影響程度。

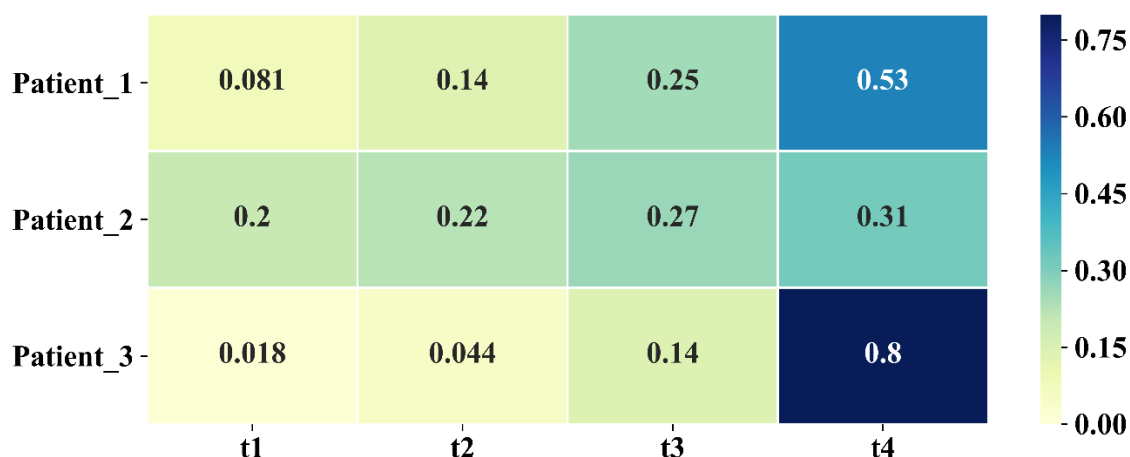


圖 9 注意力機制個案分析

首先，以病患一進行說明，病患一在 t3 與 t4 兩個時間點的注意力權重相較另外兩個時間點高出許多，代表後兩個時間點對預測時間點的病程發展影響較高，病患一的四個時間點之原始病程表示如表 12 所示，預測目標為下一時間點

的後四欄的用藥數值，可以由表格看出 t4 的類風濕性關節炎口服用藥量較多，且牙周病相關治療次數也增加；t3 的牙周病與類風濕性關節炎相關治療較少，因此注意權重減少。病患二的四個時間點之權重皆介於百分之二十至三十之間，是屬於四個時間的病程對預測時間點的病程發展預測有較一致的影響程度，其原始病程表示如表 13 所示，可以發現 t1 和 t2 的數值皆為 0，而 t3 與 t4 則是類風溼性關節炎口服用藥量有些微增加，因此後兩個時間點的權重較前兩個時間點的權重高一點。病患三以 t4 的權重高出另外三個時間點許多，其原始病程表示如表 14 所示，從表中可以看出 t4 在牙周病用藥天數與類風濕性關節炎治療用藥上除了口服用藥種類數增加之外還增加了注射用藥，因此最後時間點的病程會給予最多關注權重。另外，從三位患者的權重分布可以發現每一位患者越靠近預測時間點的注意權重會越高，此現象也合理解釋了未來的病程發展會與最靠近的時間點之病程發展有較強烈關係。

透過圖表的比較，可以證明模型能夠學習病程發展表示對預測時間點的預測任務之不同權重，且在前面章節的實驗結果也證明適當地加入注意力機制計算能夠提升對目標任務預測的效果。

表 12 Patient\_1 的病程發展表示矩陣(k=3)

TIME	1	2	3	4
PD_DAY	10	71	89	18
PD_DRUG	2	9	8	3
PD_DISPOSAL	2	3	0	1
PD_SCALING	0	1	0	1
RA_USE_ORAL	2.2833	2.9667	3.4833	22.7
RA_DRUG_ORAL	5	8	8	8
RA_USE_INJ	0	0	0	0
RA_DRUG_INJ	0	0	0	0

表 13 Patient\_2 的病程發展表示矩陣(k=3)

TIME	1	2	3	4
PD_DAY	0	0	0	0
PD_DRUG	0	0	0	0
PD_DISPOSAL	0	0	0	0
PD_SCALING	0	0	2	0
RA_USE_ORAL	0	0	6.917	6.667
RA_DRUG_ORAL	0	0	2	1
RA_USE_INJ	0	0	0	0
RA_DRUG_INJ	0	0	0	0

表 14 Patient\_3 的病程發展表示矩陣(k=3)

TIME	1	2	3	4
PD_DAY	0	3	3	8
PD_DRUG	0	1	1	2
PD_DISPOSAL	0	0	0	0
PD_SCALING	0	0	0	0
RA_USE_ORAL	8.917	52.6	90.1	60.433
RA_DRUG_ORAL	3	3	3	6
RA_USE_INJ	0	0	0	3.333
RA_DRUG_INJ	0	0	0	1

## 第五章 結論與未來研究方向

### 5.1. 結論

本研究主要目的是利用全民健保資料進行跨疾病病程發展預測模型建置，以慢性疾病患者的就醫紀錄資料實現，目的是為協助醫生進行疾病病程發展之推估。根據實驗結果能夠看出，本研究所提出之 TL-CDC 方法擁有模型訓練時間上的優勢，且不同序列長度的就醫資料皆能達到好的預測效果，其中以序列長度為 3 亦代表一季的資料最能完整表示病患的病程發展與有效地應用於本方法進行跨疾病的病程發展預測，而本研究方法加入的注意力機制能提升預測任務的效果，並且有效地增加模型訓練與結果的解釋力。本研究成功地利用就醫紀錄資料進行跨疾病病程發展預測的模型建立，使資料量少的疾病能利用有相關但不同的疾病預測模型的參數進行資料量少的疾病之預測模型建立，利用醫學上證實的疾病間之相關性，將其關係應用在跨疾病的預測上，除了達到目標疾病有好的預測效果之外，也增加了醫療資料的利用率。

### 5.2. 未來研究

本研究所提出利用遷移學習框架進行跨疾病病程發展預測的模型建立，現階段已能夠應用有效地應用於有關之慢性疾病的就醫資料，但還有許多需要優化的部分，針對未來研究我列出以下兩個方向：

1. 本研究只利用了患者的用藥資訊，而病患的病程發展會與其他個人資訊或與疾病更直接之治療的紀錄，如年齡、性別、共病數目與住院資訊...等變數，因此若能多增加病患資訊，能夠使病程發展預測分析更加準確。
2. 因就診紀錄涵蓋了許多與病程發展不相關的資訊，導致資料前處理部分

需要花費許多時間，因此未來希望能優化資料前處理的方式，提升處理速度。



## 參考文獻

- [1] World Health Organization (2020 ,December 9). WHO reveals leading causes of death and disability worldwide: 2000-2019 [Online].Available: <https://www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019>
- [2] KWAK, Gloria Hyunjung; HUI, Pan. DeepHealth: Review and challenges of artificial intelligence in health informatics. *arXiv preprint arXiv:1909.00384*, 2019.
- [3] RAVÌ, Daniele, et al. Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 21.1: 4-21, 2016
- [4] Yang Yang, Xiangwei Zheng, and Cun Ji. Disease prediction model based on bilstm and attention mechanism.In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, p.1141–1148, 2019.
- [5] He, R., Liu, Y., Wang, K., Zhao, N., Yuan, Y., Li, Q., & Zhang, H.. Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional LSTM. *IEEE Access*, 7, p.102119-102135, 2019.
- [6] Choi, E., Bahadori, M. T., Song, L., Stewart, W. F., & Sun, J. GRAM: graph-based attention model for healthcare representation learning. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 787-795), 2017.
- [7] Okada, M. et al. Periodontal Treatment Decreases Levels of Antibodies to Porphyromonas Gingivalis and Citrulline in Patients With Rheumatoid Arthritis and Periodontitis. *J. Periodontol.* 84, 74–84 ,2013.
- [8] Erciyas K, Sezer U, Ustün K, et al. Effects of periodontal therapy on disease activity and systemic inflammation in rheumatoid arthritis patients. *Oral Dis*; 19: 394– 400,



2013.

- [9] G. S. Birkhead, M. Klompas, and N. R. Shah, “Uses of electronic health records for public health surveillance to advance public health,” *Annu. Rev. Public Health*, vol. 36, pp. 345–59, 2015.
- [10] Jensen, P. B., Jensen, L. J. & Brunak, S. Mining electronic health records: towards better research applications and clinical care. *Nat. Rev. Genet.* 13, 395–405, 2012.
- [11] Weiskopf, N. G., Hripcsak, G., Swaminathan, S. & Weng, C. Defining and measuring completeness of electronic health records for secondary use. *J. Biomed. Inform.* 46, 830–836, 2013.
- [12] Weiskopf, N. G. & Weng, C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J. Am. Med. Inform. Assoc.* 20, 144–151, 2013.
- [13] KWAK, Gloria Hyunjung; HUI, Pan. DeepHealth: Review and challenges of artificial intelligence in health informatics. *arXiv preprint arXiv:1909.00384*, 2019.
- [14] Avati, A., Jung, K., Harman, S., Downing, L., Ng, A., & Shah, N. H. Improving palliative care with deep learning. *BMC medical informatics and decision making*, 18(4), 55-64, 2018.
- [15] Choi, E., Bahadori, M. T., Kulas, J. A., Schuetz, A., Stewart, W. F., & Sun, J. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *arXiv preprint arXiv:1608.05745*, 2016.
- [16] Pham, T., Tran, T., Phung, D., & Venkatesh, S. Deepcare: A deep dynamic memory model for predictive medicine. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 30-41). Springer, Cham, 2016.
- [17] Nguyen, P., Tran, T., Wickramasinghe, N., & Venkatesh, S. Deepr: a convolutional net for medical records. *IEEE journal of biomedical and health informatics*, 21(1),

22-30, 2016.

- [18] Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., & Sun, J. Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine learning for healthcare conference* (pp. 301-318). *PMLR*, 2016.
- [19] Pan, S. J., & Yang, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359, 2009.
- [20] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22(10):1345–1359, 2010.
- [21] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. How transferable are features in deep neural networks?. In *Advances in neural information processing systems* ,p. 3320-3328, 2014.
- [22] Shie, C. K., Chuang, C. H., Chou, C. N., Wu, M. H., & Chang, E. Y. Transfer representation learning for medical image analysis. In *2015 37th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 711-714), 2015.
- [23] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., & Summers, R. M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5), 1285-1298, 2016.
- [24] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Dong, J. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5), 1122-1131, 2018.
- [25] Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., ... & De Albuquerque, V. H. C. A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Applied Sciences*, 10(2), 559, 2020.

- [26] Lee, J. Y., Dernoncourt, F., & Szolovits, P. Transfer learning for named-entity recognition with neural networks. *arXiv preprint arXiv:1705.06273*, 2017.
- [27] Desautels, T., Calvert, J., Hoffman, J., Mao, Q., Jay, M., Fletcher, G., ... & Das, R. Using transfer learning for improved mortality prediction in a data-scarce hospital setting. *Biomedical informatics insights*, 9, 1178222617712994, 2017.
- [28] Gupta, P., Malhotra, P., Narwariya, J., Vig, L., & Shroff, G. Transfer learning for clinical time series analysis using deep neural networks. *Journal of Healthcare Informatics Research*, 4(2), 112-137, 2020.
- [29] Malhotra P, TV V, Vig L, Agarwal P, Shroff G .TimeNet: Pre-trained deep recurrent neural network for time series classification. In: *25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, p607–612, 2017.
- [30] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. LSTM: A Search Space Odyssey. *arXiv:1503.04069*, 2015.
- [31] CHO, Kyunghyun, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.
- [32] HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. *Neural computation*, 9.8: 1735-1780, 1997.
- [33] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014
- [34] Schuster, M., & Paliwal, K. K. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11), 2673-2681, 1997.
- [35] HERMANTO, Andi; ADJI, Teguh Bharata; SETIAWAN, Noor Akhmad. Recurrent neural network language model for english-indonesian machine translation:

- Experimental study. In: *2015 International conference on science in information technology (ICSITech)*. IEEE, p. 132-136, 2015.
- [36] AO, Lirong; GUAN, Yazhuo. An improved LSTM structure for natural language processing. In: *2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*. p. 565-569, 2018.
- [37] Shih, S. Y., Sun, F. K., & Lee, H. Y. Temporal pattern attention for multivariate time series forecasting. *Machine Learning*, 108(8), 1421-1441, 2019.
- [38] HOCHREITER, Sepp; HEUSEL, Martin; OBERMAYER, Klaus. Fast model-based protein homology detection without alignment. *Bioinformatics*, 23.14: 1728-1736, 2007.
- [39] LIPTON, Zachary C., et al. Learning to diagnose with LSTM recurrent neural networks. *arXiv preprint arXiv:1511.03677*, 2015.
- [40] LEI, Liqi, et al. An effective patient representation learning for time-series prediction tasks based on EHRs. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, p. 885-892, 2018.
- [41] DA SILVA, Denise Bandeira, et al. DeepSigns: A predictive model based on Deep Learning for the early detection of patient health deterioration. *Expert Systems with Applications*, 165: 113905, 2021.
- [42] Athiwaratkun, B., & Stokes, J. W. (2017, March). Malware classification with LSTM and GRU language models and a character-level CNN. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2482-2486), 2017.
- [43] Chen, W., Zhang, Y., Yeo, C. K., Lau, C. T., & Lee, B. S. Stock market prediction using neural network through news on online social networks. In *2017 international smart cities conference (ISC2)* (pp. 1-6). IEEE, 2017.

- [44] Mnih, V., Heess, N., Graves, A., & Kavukcuoglu, K. Recurrent models of visual attention. *arXiv preprint arXiv:1406.6247*, 2014.
- [45] Bahdanau, D., Cho, K., & Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- [46] Luong, M. T., Pham, H., & Manning, C. D. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.
- [47] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... & Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning* (pp. 2048-2057). PMLR, 2015.
- [48] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.
- [49] Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv preprint arXiv:1901.02860*, 2019
- [50] Huang, S., Wang, D., Wu, X., & Tang, A. Dsanet: Dual self-attention network for multivariate time series forecasting. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 2129-2132), 2019.
- [51] Sha, Y., & Wang, M. D. Interpretable predictions of clinical outcomes with an attention-based recurrent neural network. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 233-240), 2017
- [52] Deepak A Kaji, John R Zech, Jun S Kim, Samuel K Cho, Neha S Dangayach, Anthony B Costa, and Eric K Oermann. An attention based deep learning model of clinical events in the intensive care unit. *PloS one*, 14(2):e0211057, 2019