天主教輔仁大學資訊管理學系碩士論文

指導教授 ： 黃曜輝 博士

應用機器學習與深度學習模型於登革熱預測:氣候與空氣品質的整合分析

Application of Machine Learning and Deep Learning Models for Dengue Fever Prediction: An Integrated Analysis of Climate and Air Quality

研究生：慎世煒 撰

中 華 民 國 113 年 7 月

論文題目：應用機器學習與深度學習模型於登革熱預測：氣候與空氣品質的整合分析

校(院)系所組別：輔仁大學資訊管理學系碩士班

研究生：慎世煒　　**指導教授：**黃曜輝

論文頁數：67

關鍵詞：登革熱、空氣品質、氣候、機器學習、深度學習、風險預測、支持向量機、隨機森林、人工神經網路、長短期記憶網路、門控循環單元

論文摘要內容：

本研究目的在針對雲林、嘉義、台南、高雄和屏東等台灣南部五個地區的氣候和空氣品質數據，預測登革熱疫情的爆發情況。研究將登革熱的確診數量劃分為三個等級：Green（確診數:0~3）、Yellow（確診數:4~30）和Red（確診數:31以上）。研究採用支持向量機（SVM）、隨機森林（RandomForest）、人工神經網路（ANN）、長短期記憶網路（LSTM）和門控循環單元（GRU）五種機器學習和深度學習模型進行兩階段預測。第一階段判斷疫情是否發生（Green/Not Green），若判斷為Not Green，則進行第二階段預測，評估疫情的嚴重程度（Yellow/Red）。這種分階段的預測方法，有助於提高疫情預測的準確性，為政府和公眾提供有效的預警資訊，從而及時採取防控措施，減少登革熱疫情的影響。

Title: Application of Machine Learning and Deep Learning Models for Dengue Fever Prediction: An Integrated Analysis of Climate and Air Quality

Keywords: Dengue Fever; Air Quality; Climate; Machine Learning; Deep Learning; Risk Prediction; SVM; Random Forest; ANN; LSTM; GRU

This study aims to predict the outbreak of dengue fever using climate and air quality data from five regions in southern Taiwan: Yunlin, Chiayi, Tainan, Kaohsiung, and Pingtung. We categorize the number of confirmed dengue cases into three levels: Green (0-3 cases), Yellow (4-30 cases), and Red (more than 31 cases). The study employs five machine learning and deep learning models: Support Vector Machine (SVM), Random Forest, Artificial Neural Network (ANN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) for a two-stage prediction process. The first stage determines whether an outbreak will occur (Green/Not Green). If classified as Not Green, the second stage assesses the severity of the outbreak (Yellow/Red). This phased prediction approach enhances the accuracy of dengue outbreak predictions, providing effective early warning information for the government and the public, thereby enabling timely preventive measures to mitigate the impact of dengue fever.

謝詞

在這篇論文完成之際，我要向所有在我碩士學習過程中給予我幫助和支持的教授、秘書和同學們表達我最誠摯的感謝。

首先，我要特別感謝我的指導教授：黃曜輝教授。您的專業指導、耐心教誨和不懈支持，使我在學術研究的道路上獲益良多。您的嚴謹治學態度和深厚的學識對我有著深遠的影響，您總是不厭其煩地解答我在研究中遇到的困惑，幫助我提升學術能力，並在研究過程中不斷鼓勵我前行。沒有您的指導，這篇論文無法順利完成。

同時，我也要感謝碩士班秘書：羅淑貞秘書。您在行政和後勤上的支持對我的學習和研究提供了極大的幫助。無論是課程安排、助教任用，還是各種手續的辦理，您的熱心幫助使我能夠專心於學術研究，減少了許多行政程序上的後顧之憂。

此外，我還要感謝所有與我一同度過碩士生涯的同學們。你們的陪伴、鼓勵和幫助使這段學習旅程充滿了溫暖和動力。特別感謝與我同一指導教授的同學們：蘇彥碩、林子軒、林浩景以及洪子堯，我們一起討論問題、分享經驗、互相支持，你們的友情和合作讓我受益匪淺。

最後，感謝我的家人和朋友們在我求學期間的理解和支持。你們的鼓勵和愛護是我克服困難、追求夢想的堅強後盾。

再次向所有在這兩年碩士生涯中幫助過我的人致以衷心的感謝！

目錄

表目錄

圖目錄

第壹章　緒論

第一節　研究背景與動機

登革熱被認為是近年來傳播速度最快的由節肢動物帶然散播的病毒(Guzman & Harris, 2015)。登革熱病毒主要有四種血清類型：DEN-1、DEN-2、DEN-3和DEN-4 (Changal et al., 2016)，主要由兩種蚊子傳播，分別是埃及斑蚊（Aedes aegypti）和白線斑蚊（Aedes albopictus）(Kraemer et al., 2019)。有研究估計全球每年約有3.9億例登革熱 (Bhatt et al., 2013；Murray et al., 2013)。

最初，登革病毒只在熱帶和亞熱帶地區流行，研究估計每年有39億人面臨感染風險 (Brady et al., 2016)。但如今登革熱遍布全球128個國家，其中包括歐洲和北美洲的國家，這兩個地區主要均為非熱帶地區 (Murray et al., 2013)。這種疾病可能升級為登革出血熱，在少數人中可能導致死亡，每年的登革死亡人數約在20,000至25,000之間 (Gui et al., 2021)。登革熱流行造成的經濟負擔已對許多國家產生重大影響。有研究顯示，美洲國家在2000年至2007年期間因登革熱造成的經濟損失高達平均每年21億美元 (Shepard et al., 2011)，這一數字突顯了登革熱對流行地區造成了巨大經濟負擔。因此，如何有效地預防和控制登革熱成為全球衛生領域的迫切課題。

熱帶和亞熱帶地區的人口集中，使得登革熱在這些區域更容易爆發 (World Health Organization, 2023)。這些地區的氣候和環境為登革熱病毒傳播提供了理想的條件。高溫和潮濕的氣候促進了病媒蚊的繁殖，這些蚊子又是登革熱病毒的傳播媒介。因此，人口密集區和熱帶地區成為登革熱爆發的主要戰場。

自1995年以來，發生登革熱疫情的國家數量增加了三倍，估計每年在120多個國家中發生1億至4億次感染 (Brady ＆Hay, 2020)。受影響國家的經濟負擔也正不停增加，特別是美洲、東南亞和西太平洋等登革熱高風險地區 (World Health Organization, 2023)。登革熱在全球範圍內持續擴散，威脅著全球公共衛生安全。2023年夏天台灣南部地區面臨著疫情升溫的嚴重挑戰，截至12月10日累計共26047例本土病例，另累計死亡病例56例。顯示在新冠肺炎疫情結束後，隨著經

濟活動的復甦，人們對於防疫已漸漸鬆懈，現行關於登革熱防治的觀念和做法已無法有效抑制登革熱的傳播。而且隨著都市化的快速推進、交通便捷以及旅遊業復甦等因素的增加 (Gubler, 2011)，導致本土登革熱和外來登革熱相互交替感染，傳染的風險日益升高，疫情防治形勢變得更加嚴峻。衛福部疾管署已將登革熱列為法定傳染病中的第二類，代表一旦發現確診病例，就需要由醫療和保健機構立即向上級機構通報，並迅速提出相應的防疫策略。為了有效遏制疫情擴散，急需發展登革熱爆發預測系統，並實施有效的防疫措施，從而直接減少登革熱爆發的可能性。

台灣地理位置處於熱帶及亞熱帶地區交界，氣候潮濕悶熱，登革熱病毒主要透過埃及斑蚊及白線斑蚊的傳播，而這些蚊子主要分布於南部地區及野外。台灣的地理位置以及氣候條件為登革熱病媒蚊提供了理想的生存環境。加上交通便捷、國際交流頻繁、氣候暖化等因素，使得台灣的登革熱疫情在夏季時爆發流行的機率大幅增加。此外，已有研究發現，當每次登革熱疫情有增加趨勢時，都可以透過空氣品質監測的相關資料發現一些端倪，兩者存在著某些關聯 (Lu et al., 2023)。因此在透過現有的空氣品質資料如何預測登革熱爆發，提早應對並盡快進行清理積水容器、噴藥等防治措施，以避免登革熱在台灣根深蒂固，引發自發的本土病例持續發生，這也應成為目前台灣應重視的公共衛生議題。

第二節　研究目的

本研究旨在探討登革熱與氣候與空氣汙染因子之間的相關性，特別聚焦台灣西南地區，包括雲林、嘉義、台南、高雄、屏東等縣市，以期深入了解這一區域的氣候、空氣汙染與登革熱之間的關聯。

首先，台灣西南部地區是登革熱的高發區域。這一地區的氣候、環境生態、人口密度和人口的移動都使其成為病媒蚊滋生的理想場所。隨著全球暖化造成的平均氣溫上升和極端氣候增加，病媒蚊活動的範圍也逐漸擴大，預期將進一步增加受登革熱影響的地區。因此，研究這一地區的登革熱爆發與空氣汙染因子之間的關係，對於發展相對應的預測、預防和控制策略至關重要。

其次，空氣汙染因子在台灣西部地區也是一個日益嚴重的問題。特別是以重工業發展為核心產業的高雄市，工業污染排放、交通運輸和都市化發展都是

潛在的汙染源，這些因子可能對環境和人類健康產生廣泛的影響。空氣汙染不僅直接影響呼吸系統，還可能對自然環境和病媒蚊的生態產生間接影響。因此，有充足理由相信登革熱與空氣汙染因子之間可能存在著複雜的相互作用。

雲林、嘉義、臺南、高雄、屏東等縣市被選為本研究的主要研究地點，這是因為這些地區不僅有著相對較高的登革熱風險，同時也受到空氣汙染因子的多方面影響。透過對這些地區的空氣品質資料與登革熱確診資料的綜合研究，有望藉由大量的氣候、空氣汙染與登革熱數據，進一步深入分析登革熱傳播的預測因子之間的關係，以及尋找預測變數的最佳組合並且評估外部資料的預測能力。

總而言之，這項研究旨是在根據已有的研究成果進行延伸，為台灣西南部地區的登革熱預防和預測提供更多的氣候及空氣汙染相關的科學依據。另外也可以加入在其他研究中所提到的經濟發展、人口流動以及人口密度等因素，共同探討與登革熱的關聯，並提供未來的政府或後繼的研究者有關於登革熱預測的建議。

第三節　論文架構圖

本研究旨在探討登革熱與氣候與空氣汙染因子之間的相關性，特別聚焦台灣西南地區，包括雲林、嘉義、台南、高雄、屏東等縣市，以期深入了解這一區域的氣候、空氣汙染與登革熱之間的關聯。

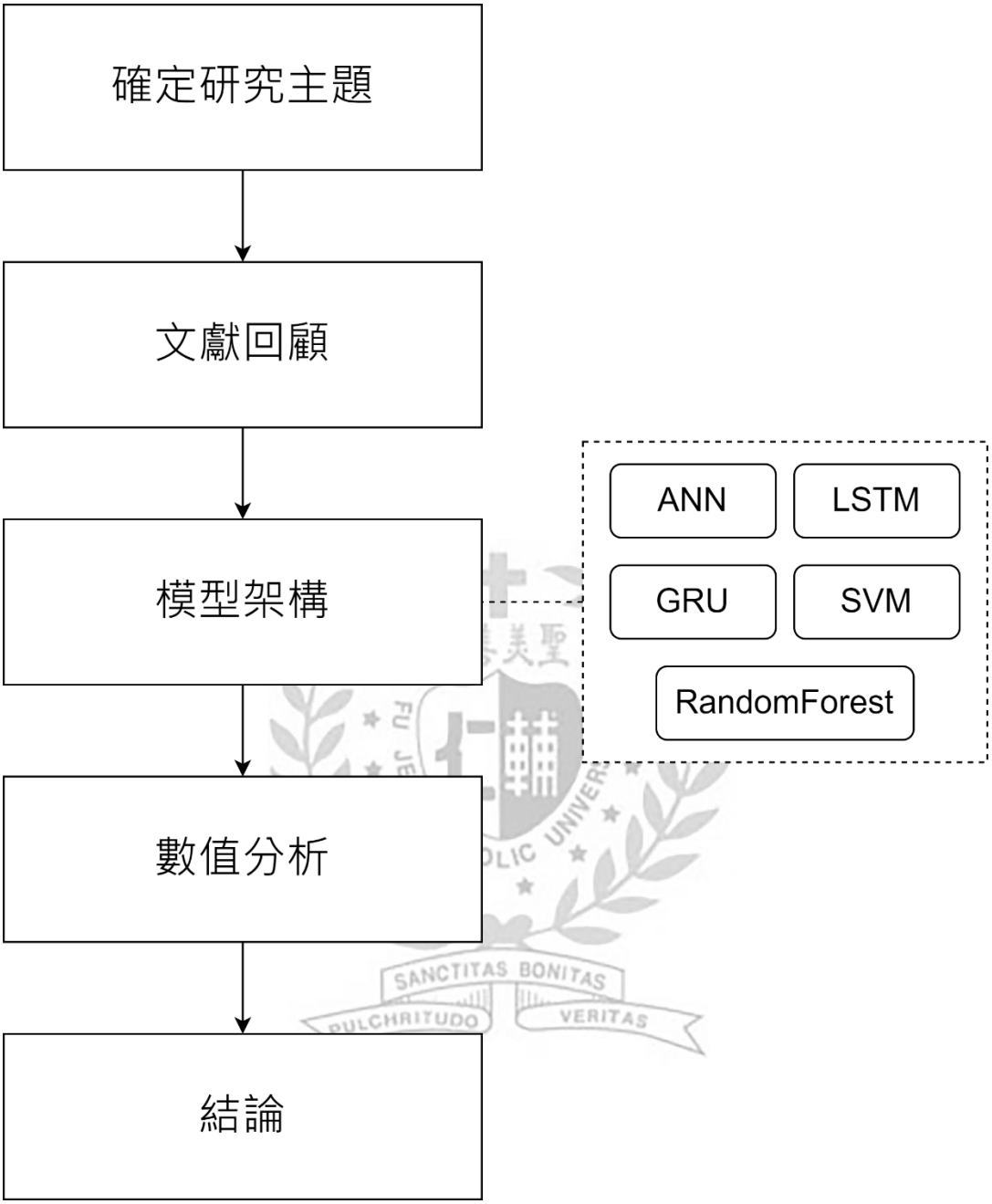

圖 1-1 論文架構圖

第貳章　文獻探討

第一節　以氣候因素為背景之登革熱相關研究

氣候變遷是人們需要面對的重要的環境變遷之一。隨著全球暖化和極端氣候等現象越來越頻繁的出現，本來受溫度限制的微生物和其他病毒媒介的影響範圍也將擴大 (McMichael, 2003)。氣候與其他傳播方式，例如水、食物、土壤和空氣傳播與疾病之間的聯繫已被確定（Colwell and Patz, 1998；Epstein, 2001），其中又以氣候與病媒蚊傳播之間的連結最強。也就是說，氣候會對登革熱的流行和強度產生影響。

登革熱病毒主要由兩種蚊子傳播，分別是埃及斑蚊（Aedes aegypti）和白紋斑蚊（Aedes albopictus）(Kraemer et al., 2019)。其中埃及斑蚊因在城市登革熱傳播研究中表現出較佳的傳播能力 (Kek et al., 2014)，且在埃及斑蚊身上檢出的登革熱病毒量較多，因此埃及斑蚊是登革熱的主要媒介(Chung & Pang, 2002)。先前在泰國以及肯亞的研究顯示(Ngugi et al., 2017；Wongkoon et al., 2007)，在城市環境中，埃及斑蚊在戶外積水容器中孳生的現象說明氣候因素是如何影響病媒蚊生長的環境。研究發現，溫度上升至34°C會增加埃及斑蚊所有發育階段的速率，從而導致病媒蚊數量增長 (Morin et al., 2013)。在較溫暖的溫度下增加登革熱傳播，可能是由於在媒介中更快的病毒複製，短暫的外在孵化期，以及埃及斑蚊的進食率增加(Morin et al., 2013；Lai, 2018；Ebi and Nealon, 2016)。

除了溫度，濕度也會增加登革熱病毒的傳播和埃及斑蚊卵的孵化率(Xu et al., 2014；Thu et al., 1998)。巴西的一項研究發現，每周最低溫度超過18°C對埃及斑

蚊密度有正面影響，而濕度超過75%則有負面影響(Ferreira et al., 2017)。人類行為與氣候因素相結合，例如在干燥氣候中的水儲存，可能促使蚊子的生產和登革傳播(Aziz et al., 2012)。Schmidt等人的研究結果證實，在缺乏自來水供應的農村地區，登革傳播風險較高 (Schmidt et al., 2011)。最後，風速較強 (Lu et al., 2009；Xiang et al., 2017) 和風速（風速x持續時間）也同樣有抑制登革熱傳播的作用 (Ehelepola et al., 2015)。

下表為本次研究所使用之氣候測項:

表2-1 氣候測項對照表

測項簡稱	單位	測項名稱
AMB_TEMP	℃	大氣溫度
RAINFALL	mm	雨量
RH	%	相對溼度
WIND_SPEED	m/sec	風速(以每小時最後 10 分鐘計算平均)
WS_HR	m/sec	風速小時值(以每個小時計算平均)

第二節　以空氣汙染為背景之登革熱相關研究

近年來，都市化快速發展等因素使得空氣污染成為一個嚴重的公共衛生問題。對於城市與鄉村的能源消耗研究發現，城市居民相較農村居民會消耗更多的能源 (Gong et al., 2012；Zhao et al., 2012；Zheng et al., 2014)。這種情況導致城市空氣污染物的人均排放量以及單位土地面積的排放量更高 (Han et al., 2014)。針對馬來西亞兩座城市的一項研究顯示，病媒蚊幼蟲密度與上週的空氣污染指數呈現中度負相關，但與本週的濕度和溫度呈正相關 (Ahmad et al., 2018)。在中美洲的瓜地馬拉城也有研究顯示，傳統的柴火烹飪與蚊媒傳染病，包括登革熱的發生率減少有關 (Madewell et al., 2020)。

焚燒植物所產生之煙霧與蚊蟲似乎有著某種關聯。有許多研究證明煙霧能夠驅趕影響蚊子 (Davis & Bowen, 1994； Biran et al., 2007)。為了驅趕蚊蟲、避免蚊子叮咬，燃燒植物以產生煙霧是一種常見的做法 (Tabuti, 2008)。Dulhunty等人(2000)在太平洋島國所羅門群島的馬萊塔島訪問了124位當地居民，其中有52%的受訪者報告使用火來保護自己免受蚊子叮咬。在非洲，Pålsson &Jaenson (1999)調查了幾內亞比索23個鄉村的人們用來減少蚊子叮咬活動的植物物種和植物衍生產品，作者發現當地居民會焚燒 Hyptis suaveolens Poit (脣形科) 以及 Daniellia oliveri Rolfe (雲實科)兩種植物來驅趕病媒蚊。

除了城市化、工業發展等原因，農業上的「刀耕火種」也被認為是空氣汙染的形成因素之一 (Latif et al., 2018)。除此之外，霧霾現象也可能與登革熱或病媒蚊有關。霧霾是因燃燒煤炭、交通工具廢氣、工業污染以及建築業粉塵等因素產生的化合物加上季風等其他氣候因素而產生 (Pu, 2017)。其中懸浮在空氣中的灰塵、化合物和其他乾燥顆粒物（PM）會使空氣品質變差。而霧霾中的部分化合物與蚊香、薰香等有驅蚊以及滅蚊功效的產品燃燒產生之煙霧的成分相似 (Hogarh et al., 2018)。

Massad et al. (2010) 對2006年新加坡登革熱病例數低於預期，以及2007年登革熱病例數高於預期的現象進行研究。他們的結論是，2006 年新加坡登革熱病例數低於預期，是由於當年霧霾影響該國導致蚊子死亡率增加所致。另外，也有其他研究發現嚴重乾旱既會增加霧霾的發生頻率和嚴重程度，也會降低蚊子的生存能力 (Luz et al., 2003；Forattini et al., 1993；Forattini et al., 1995)。這些文獻可用於證實霧霾與病媒蚊的死亡率有關。

以下為本次研究所使用之空氣汙染測項：

表2-2 空氣汙染測項對照表

測項簡稱	單位	測項名稱
SO_2	ppb	二氧化硫
CO	ppm	一氧化碳
O_3	ppb	臭氧
PM_{10}	$\mu g/m^3$	懸浮微粒
$PM_{2.5}$	$\mu g/m^3$	細懸浮微粒
NO_X	ppb	氮氧化物
測項簡稱	單位	測項名稱
NO	ppb	一氧化氮
NO_2	ppb	二氧化氮

第三節　使用機器學習模型進行疾病預測

在過去的二十年中，機器學習 (ML) 方法已應用於許多學科，例如地理、環境和流行病學，以便從高度異質的資料中得出有意義的發現(Zhao et al., 2020)。機器學習有助於包含大量相關變量，從而能夠對變量之間複雜的交互作用進行建模，並且可以在不預設函數形式（例如線性、指數和邏輯）的情況下擬合複雜模型，為疾病預測提供更靈活的方法(Breiman, 2001 ; Murphy, 2012)。

決策樹、支援向量機、淺層神經網路、K 最近鄰、梯度提升和樸素貝葉斯是登革熱預測研究中常用的機器學習方法 (Guo et al., 2017 ; Scavuzzo et al., 2018)。與上述機器學習方法相比，一種常見的機器學習演算法隨機森林（RandomForest）已被證明在預測方面相當準確，因為它能夠透過使用引導聚合來克服過度擬合的常見問題(Raczko & Zagajewski , 2017 ; Meyer et al., 2016 ; Rodriguez-Galiano et al., 2015 ; Statnikov et al., 2008) 。

SVM模型應用於分類問題具有與其他模型相比有較高準確性、可以輕鬆處理複雜的非線性資料以及過擬合問題較其他模型少等優點(Leopord et al., 2016)。

Nordin et al. (2020) 提出了一種使用支持向量機對登革熱流行病例進行分類的預測模型。作者使用馬來西亞吉蘭丹州衛生局所提供之登革熱患者資料，並提出了一個使用支援向量機（SVM）預測未來登革熱爆發的預測模型。而核函數為RBF的SVM模型的表現優於其他模型。

第四節　使用ANN Model進行疾病預測

ANN（Artificial Neural Network）的概念於1943年首次由 McCulloch, W. S., & Pitts, W.提出。作者用數學搭配閥值（Threshold）邏輯來描述生物大腦的運作過程，論文中提出了「ANN的概念」和「神經元數學模型」。基於前人提出的想法，Frank Rosenblatt在1957年時，發明了人類史上第一個能模擬人類感知的神經網絡，名為「感知器（Perceptron）」。

ANN使用預測變數（例如環境因素）的組合來模擬與目標變數（例如登革熱爆發風險）的關係。疾病發生模型可以基於線性和非線性方法，模擬短期和長期（氣候）環境變數與登革熱發生率之間的複雜關係（Bhatt et al., 2013 ; Racloz et al., 2012 ; Medeiros et al., 2012）。

但線性模型通常無法模擬這些因素之間複雜的相互作用，結果往往較差（Laureano-Rosario et al., 2014）。非線性方法通常比線性模型表現出更大的功效（Parham & Michael, 2010）。例如，泰國、新加坡和馬來西亞的研究也使用人工神經網路 (ANN) 模型來預測登革熱病例，準確率超過 80%（Rachata et al., 2008 ; Aburas et al., 2010 ; Hwang et al., 2016）。

第五節　使用LSTM Model進行疾病預測

Hochreiter & Schmidhuber（1997）首次發表LSTM模型，並應用其來預測時間序列中間隔和延遲非常長的重要事件問題，這些問題使用過往的循環網絡演算法難以解決。通過截斷梯度避免梯度爆炸（gradient explosion）或梯度消失（gradient vanishing）等問題，LSTM可以學習跨越更多離散的時間步驟，而不會受到梯度問題的影響。因此，LSTM被視為時間序列學習中最先進的深度學習模型之一，適用於具有長期依賴性的傳染病等時間序列學習（Xu et al., 2020）。

近年來LSTM模型在預測傳染病方面受到了廣泛的關注與應用。例如，LSTM已被用來預測流感、COVID-19和登革熱（Chimmula & Zhang, 2020；Zhang & Nawata, 2018）。在各式各樣的統計模型中，大多數現有模型的缺點是為線性時間序列數據而設計的。然而，由於季節性疾病（如登革熱）既包含趨勢也包含不規則模式，所以並不適合使用傳統時間序列模型進行預測（Mussumeci & Coelho, 2020）。因此，LSTM被引入來解決這些預測問題。與傳統的自迴歸模型相比，LSTM的主要優勢在於它們能夠處理發生率的非線性、不穩定和重尾分布（Mussumeci & Coelho, 2020）。

儘管目前少有研究在登革預測中應用LSTM，但LSTM表現出比其他模型更好的性能。Mussumeci & Coelho（2020）預測了巴西790個城市在2010年至2018年間的每週登革熱病例數。與兩個機器學習模型（隨機森林迴歸和LASSO迴歸）相比，LSTM模型表現最佳。儘管在需要訓練數百個模型時，LSTM的計算成本可能非常高，但結果顯示，在不重新訓練的情況下，這些模型仍然在兩年多的時間內保持準確。除此之外，Xu et al.（2020）也透過LSTM來預測中國大陸20個城市在2005年至2018年間的每月登革病例數量。簡而言之，與其他候選機器學習模型相比，LSTM模型將預測的平均均方根誤差（RMSE）降低了12.99％到26.82％。然而，

作者發現在登革發病率較低的城市中，LSTM模型顯示出較差的性能。

第六節　使用GRU Model進行疾病預測

GRU（Gated Recurrent Unit）由韓國學者於2014年提出（Chung et al., 2014）。GRU是另一種改進的RNN，與LSTM相似。GRU僅由更新門（Update gate）和重置門（Reset gate）組成，比起原始的LSTM不但結構較為簡單，計算上也更加容易。由於結構簡單，GRU在某些情況下的訓練時間可以比LSTM更快，且對於時間序列資料的訓練結果也能優於LSTM（Fu et al., 2016）。

GRU 模型在預測傳染病方面同樣受到了廣泛的關注與應用。例如，GRU同樣已被用來預測流感、COVID-19和登革熱（Ma et al., 2022；Yang et al., 2023）。GRU的內部結構更簡單，也更容易訓練。與LSTM相比，GRU 具有調節單元內部資訊流的門控單元，但沒有單獨的儲存單元，但其結構更簡單，計算量也更小（Chung et al., 2014）。

GRU模型已被應用在不少疾病預測，且GRU在某些時間序列研究中表現出比其他模型更好的性能。Sathler & Luciano (2017) 為波多黎各聖胡安市創建登革熱病例預測模型。與另兩個模型（貝葉斯迴歸和LSTM）相比，GRU模型表現最佳。在訓練伊基多斯和聖胡安兩座城市的數據時 GRU 模型都能很快達到收斂狀態。除此之外，Muthamizharasan & Ponnusamy（2022）也透過 CNN 和 GRU 兩種模型混合使用來預測肯亞馬爾堡病毒 (MarburgVirus Disease)一種透過果蝠傳播的疾病。該模型的正確率達99.1%。


FU JEN CATHOLIC UNIVERSITY
SANCTITAS BONITAS
PULCHRITUDO
VERITAS

第參章　研究方法

第一節　研究架構

本節將對本研究之預測流程進行詳細介紹。預測流程如圖3-1所示。

首先，本研究收集了台灣在2010年至2023年間的登革熱病例數據以及相關的氣候和空氣品質數據。登革熱數據來自衛生福利部疾病管制署，空氣品質與氣候資料來自環境部公開資料。對收集到的數據進行了整理，以確保其格式一致性。再來將其資料頻率從小時轉換為週，這樣能夠更好地捕捉數據的長期趨勢和週期性變化。

在轉換過程中，計算了每週的平均數、中位數、最大值和最小值。這些統計特徵不僅可以提供數據的基本描述，還可以幫助理解數據的分佈情況和變化趨勢。例如，平均數和中位數可以顯示數據的中心趨勢，而最大值和最小值則可以反映數據的極端情況。這些特徵對於後續的模型訓練具有重要意義。

由於登革熱病例數據存在極端值，在將確診數轉換為類別後會產生不平衡的特性，因此在進行模型訓練之前對數據進行了過採樣處理。實驗使用了ADASYN（Adaptive Synthetic Sampling Approach for Imbalanced Learning）方法來生成合成樣本，以此來平衡訓練數據集中的正負樣本比例。過採樣處理後，又對數據進行了標準化處理，以消除不同特徵之間的量級差異，從而提高模型的訓練效果。

在數據處理完成後，將數據集按照80:20的比例劃分為訓練集和測試集。接著，建立登革熱風險預測模型。在模型訓練過程中，改用Focal Loss取代傳統的交叉熵損失函數（Cross-Entropy Loss）。Focal Loss可以更好地處理不平衡數據問題，通

過降低容易分類樣本的權重，強調難以分類樣本的損失，從而提升模型對少數類別的辨識能力。

最後，使用了一系列指標對預測模型的效能進行評估。這些指標包括準確率（Accuracy）、召回率（Recall）、F1-Score、ROC曲線（Receiver Operating Characteristic curve）和AUC（Area Under Curve）。通過這些指標，可以全面地評估模型的預測能力和穩定性。準確率可以衡量模型的總體正確率，召回率則側重於模型對少數類別的識別能力，F1-Score綜合了準確率和召回率，ROC曲線和AUC則提供了模型在不同閾值下的性能表現。

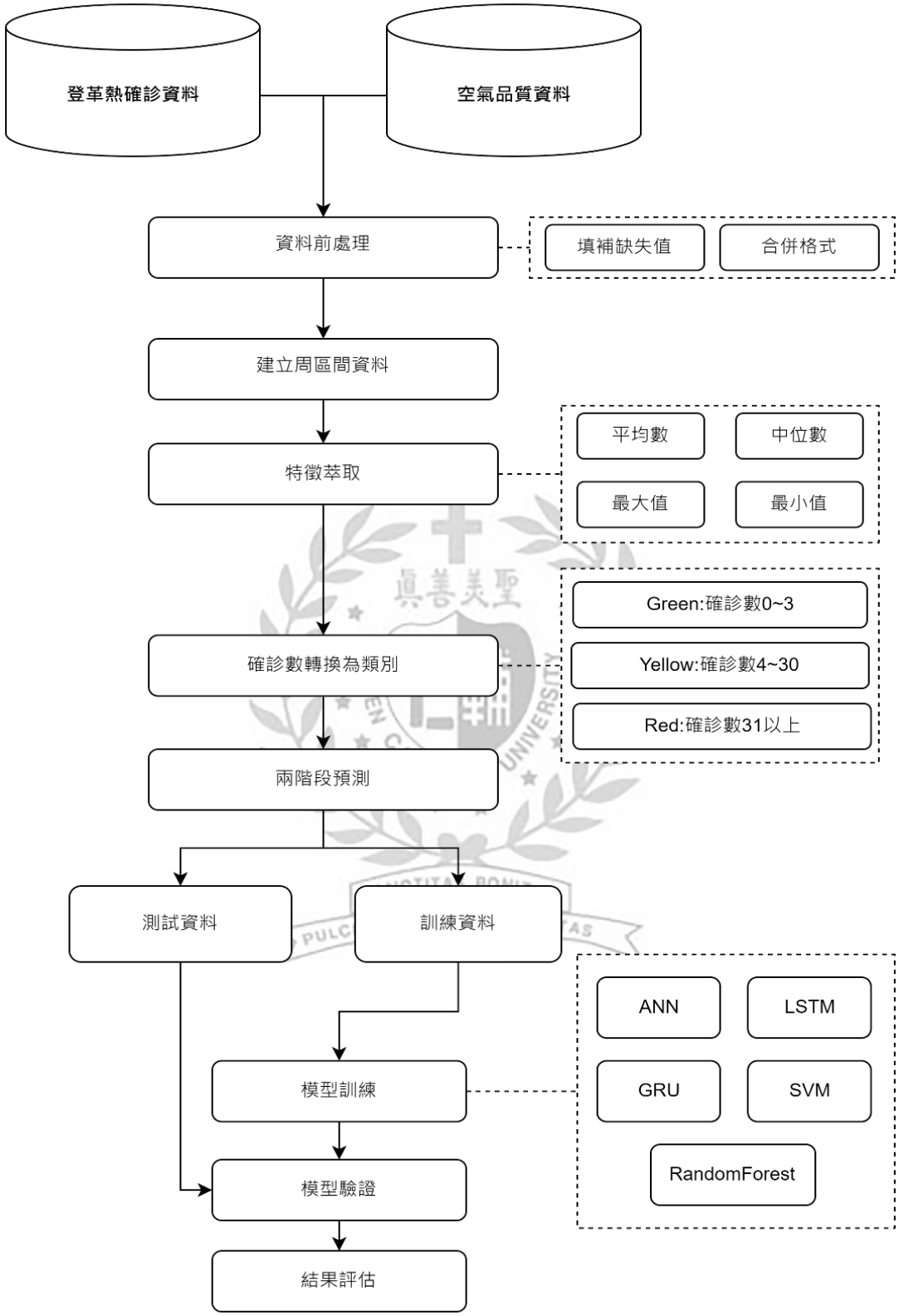
登革熱確診資料
空氣品質資料
資料前處理
填補缺失值
合併格式
建立周區間資料
特徵萃取
平均數
中位數
最大值
最小值
確診數轉換為類別
Green:確診數0~3
Yellow:確診數4~30
Red:確診數31以上
兩階段預測
測試資料
訓練資料
模型訓練
ANN
LSTM
GRU
SVM
RandomForest
模型驗證
結果評估

圖3-1 研究架構圖

第二節　地點說明

考慮到南部地區是台灣主要的登革熱病例發生地區，本研究將聚焦於台南、高雄和屏東。然而，考慮到登革熱有隨著全球暖化等氣候變遷因素北移之趨勢。因此，也將雲嘉南地區納入本次研究的範圍，用於評估自雲林縣至屏東縣之間的季風氣候區發生登革熱的風險。

圖3-2顯示了2023年至2024年間，台灣南部地區的登革熱疫情熱區以及病例數量。圖中顯示雲林、嘉義和台南、高雄是台灣登革熱病例的主要發生地區。因此，雲林以南的縣市病例被選為研究對象，以評估雲林、嘉義、台南、高雄和屏東等地之空氣品質對預測登革熱疫情風險的貢獻。

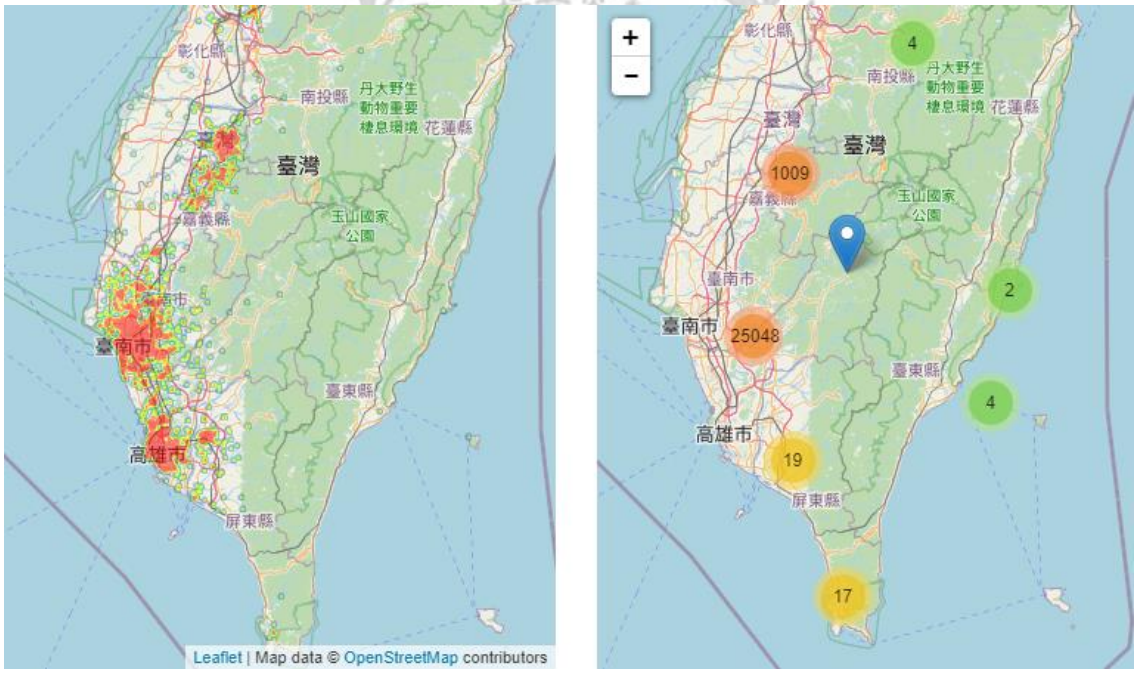

圖3-2 2023/1/1~2024/1/1登革熱地圖(衛生福利部 疾病管制署，2024)

第三節　機器學習模型

一、SVM

支持向量機（Support Vector Machine, SVM）是一種監督式機器學習算法，

廣泛應用於分類和回歸問題中。其核心思想是通過在特徵空間中找到一個最優的超平面，以最大化不同類別之間的間隔，使模型具有良好的泛化能力。因此，SVM模型適合處理高維數據和非線性分類問題。

SVM的目的是在特徵空間中找到一個超平面，使得各類樣本與超平面之間的間隔最大。這些支持向量在決定超平面的位置和方向上起著關鍵作用。對於線性可分的數據，超平面可以表示為：

$$w * x + b = 0$$

其中，w是超平面的法向量，x 是偏置項，b 是特徵向量。

對於線性不可分的情況，SVM 通過引入核函數（Kernel Function）將原始特徵空間映射到更高維的空間，使得在這個高維空間中可以找到線性可分的超平面。常見的核函數包括徑向基核函數（RBF）、多項式核函數（Polynomial Kernel）和 Sigmoid 核函數等。

二、RandomForest

隨機森林（Random Forest）是一種基於決策樹的集成學習方法。集成學習是一種通過結合多種學習算法來提升預測性能的方法，相較於單獨使用一種算法，集成學習通常能夠提供更優異的預測結果（Breiman, 2001）。

隨機森林通過 Bagging 演算法生成多個訓練數據集，然後在這些數據集上構建多棵決策樹。最終，隨機森林利用測試數據來評估和驗證模型的效果。通過構建多個決策樹並將其輸出進行平均或投票來提高預測的準確性和穩定性。這讓隨機森林在處理高維數據和防止過擬合方面具有優勢，是一種常用的分類算法。

第四節　深度學習模型與Focalloss

一、　ANN Model

ANN（Artificial Neural Networks）是一種計算處理系統，其靈感主要來自於生物神經系統（例如人腦）的運作方式。ANN主要由大量相互連接的計算節點（稱為神經元）組成，這些神經元以分佈式的方式共同工作，從輸入中學習以最佳化最終輸出。

ANN的基本結構如下圖所示。當數據傳送到輸入層，輸入層將會把數據分配到隱藏層。隱藏層根據前一層的輸出進行決策，並評估自身的隨機變化如何影響最終輸出，這個過程被稱為學習。當多個隱藏層層疊在一起時，這通常被稱為深度學習。

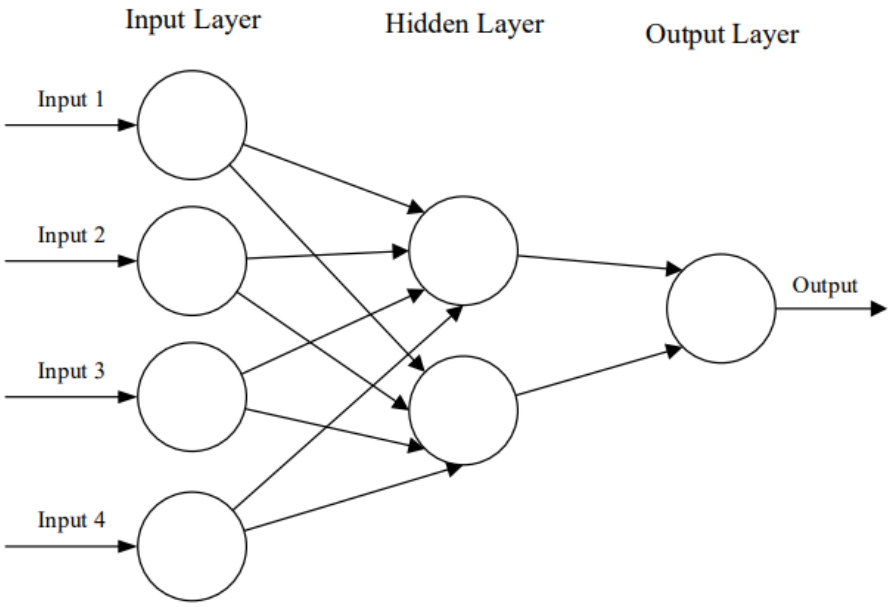

圖3-3 ANN Model 架構圖 (O'shea & Nash, 2014)

ANN在許多領域中得到了廣泛應用，包括圖像識別、語音識別和自然語言處理等。隨著計算能力和數據量的增加，ANN的性能得到了顯著提升，特別是在解決複雜問題方面。

此外，ANN的訓練過程涉及大量的數學運算和最佳化技術。例如，反向傳播

算法（backpropagation）便是一種常用的訓練方法，它通過計算誤差梯度來調整神經元的權重，從而最小化輸出和目標值之間的誤差。這種迭代調整的過程，使ANN模型能夠不斷提高其預測準確性。

二、 LSTM Model

長短期記憶（Long Short-Term Memory，LSTM）是深度學習模型中遞迴神經網路（Recurrent Neural Network，RNN）的一種變體。它通過遞迴神經網絡結構來記住長期的依賴關係和短期的變化趨勢。LSTM最早由Hochreiter & Schmidhuber於1997年提出，作為RNN的一種改進版本。

LSTM的核心結構包括四個單元（cell），分別是：輸入門（input gate）、記憶單元（memory cell）、遺忘門（forget gate）和輸出門（output gate）。這些單元協同工作，使得LSTM能夠有效地處理長序列的信息。四個單元的具體功能以及公式如下:

1. 輸入門控制是否允許特徵值進入記憶單元，以保留重要信息。

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{1}$$

$$\tilde{C}_t = tanh(w_C \cdot [h_{t-1}, x_t] + b_C) \tag{2}$$

2. 記憶單元儲存經過計算的值，使下一個步驟能夠使用。

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{3}$$

3. 遺忘門決定是否清除或遺忘記憶單元中的資料。

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f \tag{4}$$

4. 輸出門控制是否將計算出的值輸出。

$$o_t = \sigma(w_0[h_{t-1}, x_t] + b_0) \tag{5}$$

$$h_t = o_t * tanh(C_t) \tag{6}$$

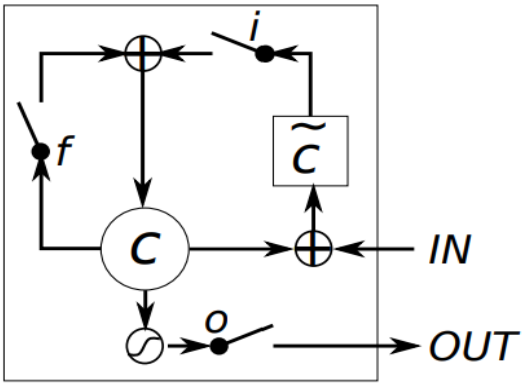

圖3-4 LSTM Model 架構圖 (Chung et al., 2014)

隨著資料輸入模型，透過四種單元交互作用，LSTM中的單元 會對該訊息進行判斷，符合規則的訊息會被儲存，不符合的訊息會被清除，以此設計原則，LSTM成功克服了標準RNN模型的長期依賴（long-term dependency）和梯度消失問題 (Le & Zuidema, 2016)。近年來，LSTM因被驗證比標準RNN具有更高的驗證精度和預測精度(Noh, S. H. , 2021)，已成為深度學習應用於時間序列資料分析中主要的訓練模型，並在金融、供需預測以及健康監測等領域得到廣泛應用(Kader et al., 2022)。

三、 GRU Model

閘門循環單元（Gated Recurrent Unit, GRU）是一種改進的循環神經網路（RNN），旨在解決標準RNN在處理長序列數據時存在的梯度消失問題。GRU通過引入更新門（Update Gate）和重置門（Reset Gate）來控制信息的流動，從而提高模型在長期依賴問題上的表現。GRU的核心在於其兩個主要的門機制：更新門和重置門。兩個單元的具體功能以及公式如下:

1. 更新門（Update Gate）：控制著多少先前的隱藏狀態將保留到當前的隱藏狀態。

$$z_t = \sigma(w_z \cdot [h_{t-1}, x_t]) \tag{7}$$

2. 重置門（Reset Gate）：決定了多少先前的隱藏狀態將被遺忘。

$$r_t = \sigma(w_r \cdot [h_{t-1}, x_t]) \tag{8}$$

GRU的核心在於其兩個主要的門機制：由上述公式得出當前細胞狀態$\tilde{h}_t$。具體過程如下：首先，將重置r_t乘以上一個時間點的輸出值h_{t-1}，決定要遺忘的比例。然後，將當前輸入值 x_t與$r_t \cdot h_{t-1}$相結合，經過激活函數 tanh 轉換得到$\tilde{h}_t$。接著，用$(1 - z_t)$確定上一個輸出值h_{t-1}的保留比率，並用$z_t \cdot \tilde{h}_t$決定當前細胞狀態 $\tilde{h}_t$ 的保留率，將兩者相加即可獲得當前輸出值 h_t ，如公式 (9) 和公式 (10) 所示。

$$\tilde{h}_t = tanh((w \cdot [r_t * h_{t-1}, x_t])) \tag{9}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{10}$$

其中，σ是 sigmoid 激活函數，tanh是雙曲正切激活函數，x_t 是當前輸入，$h_(t-1)$是前一時間步的隱藏狀態，z_t和 r_t分別是更新門和重置門，$\tilde{h}_t$ 是候選隱藏狀態，W_z 和 、W_r和W是權重矩陣。

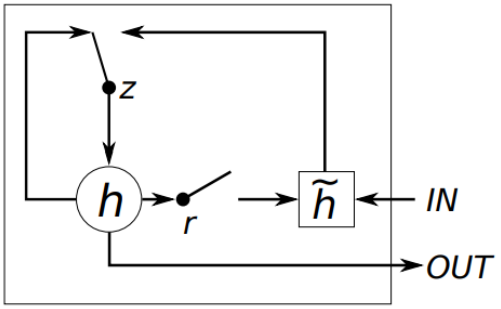

圖3-5 GRU Model 架構圖 (Chung et al., 2014)

四、 Focal Loss

Focal Loss 用於目標檢測任務，如物體檢測和實例分割，尤其是在資料類別比例極度不平衡的情況下。它也適用於分類問題，特別是在二分類和多分類任務中，當類別極度不平衡時，Focal Loss 能有效提升模型的表現。

其基本原理是對難以分類的樣本賦予更高的權重，減少易分類樣本對損失的影響，從而使模型更專注於難分類的少數類別。Focal Loss 的公式如下：

$$FL(p_t) = -a_t(1-p_t)^{\gamma}\log(p_t)$$

其中p_t是模型對正確標籤的預測概率，a_t是平衡因子，用來平衡正負樣本的影響，γ 是調節因子，用來調整易分類樣本和難分類樣本的影響。當 γ=0時，Focal Loss 退化為交叉熵損失；隨著 γ值增加，Focal Loss 更加關注那些難以分類的樣本。

具體來說，p_t表示對正類的預測概率。如果預測正確，p_t接近 1，此時 $(1-p_t)$ 接近 0，導致 $(1-p_t)^{\gamma}$也接近 0，從而降低了易分類樣本的損失值。相反，如果預測錯誤，p_t接近 0，此時 $(1-p_t)$ 接近 1，$(1-p_t)^{\gamma}$ 也接近 1，從而較高地保留了難分類樣本的損失值。

通過上述公式，Focal Loss 在處理類別不平衡問題時，能更好地關注難以分類的樣本，從而提升少數類別的分類效果。在實際應用中，a_t 和 γ的值需要根據具體問題和數據集進行調整，以達到最佳效果。

第五節 模型評估

一、 Recall

召回率是一個評估模型對於正樣本的識別能力的指標。它表示在所有實際為正的樣本中，被模型正確識別為正的比例。召回率的計算公式為：

$$Recall = TP/(TP+FN)$$

其中，TP（True Positives）是模型正確預測為正的樣本數，FN（False Negatives）是模型錯誤預測為負的正樣本數。

二、 F1-score

F1-score是一個綜合了精確率（Precision）和召回率（F1-score）的指標，用於評估模型的整體表現。它是精確率和召回率的調和平均數，當模型需要在精確率和召回率之間取得平衡時，F1-score特別有用。其計算公式為：

$$\text{F1score} = 2 * (Precision * Recall)/(Precision + Recall)$$

三、 AUC-ROC

AUC-ROC是一個評估模型分類性能的指標。ROC曲線是一個反映模型對不同閾值下分類效果的曲線，其中橫軸為假陽性率（False Positive Rate, FPR），縱軸為真陽性率（True Positive Rate, TPR）。AUC（Area Under Curve）是ROC曲線下的面積，取值範圍在0到1之間。AUC值越大，模型的分類效果越好。計算公式為：

$$TPR = TP/(TP + FN)\ ，FPR = FP/(FP + TN)$$

其中，TN（True Negatives）是模型正確預測為負的樣本數。

四、 混淆矩陣

混淆矩陣是一個矩陣表格，用來表示模型預測結果的真實標籤和預測標籤之間的對比。它包含四個指標：TP、FP、TN和FN。通過混淆矩陣，可以詳細觀察模型在各類別上的預測準確性。混淆矩陣的形式如下：

表3-1 混淆矩陣示意

	實際為正	實際為負
預測為正	TP	FP
預測為負	FN	TN


聖美善眞
FU JEN CATHOLIC UNIVERSITY
SANCTITAS BONITAS
PULCHRITUDO
VERITAS

第肆章　研究過程

第一節　預測流程

本節將詳細介紹本研究的預測流程，如圖4-1所示。本研究收集了2010年至2023年間台灣的登革熱病例數據及相關氣候和空氣品質數據。數據來自衛生福利部疾病管制署和環境部公開資料，並進行整理以確保格式一致性。將數據頻率從小時轉換為週，以捕捉長期趨勢和週期性變化。在資料轉換的過程中計算了每週的平均數、中位數、最大值和最小值，這些統計特徵有助於理解數據的分佈和變化趨勢，對模型訓練具有重要意義。

由於登革熱病例數據存在極端值且不平衡，在模型訓練前使用ADASYN方法進行過採樣處理，生成合成樣本以平衡數據集中的正負樣本比例。過採樣後對數據進行標準化處理，以消除不同特徵之間的量級差異。數據處理完成後，將數據集按照80:20的比例劃分為訓練集和測試集。建立登革熱風險預測模型時，使用Focal Loss取代傳統的交叉熵損失函數，以更好地處理不平衡數據，提升模型對少數類別的辨識能力。

最後，使用準確率（Accuracy）、召回率（Recall）、F1-Score、ROC曲線（Receiver Operating Characteristic curve）和AUC（Area Under Curve）等指標對預測模型的效能進行評估。這些指標全面衡量模型的預測能力和穩定性，其中準確率衡量模型的總體正確率，召回率側重於模型對少數類別的識別能力，F1-Score綜合準確率和召回率，ROC曲線和AUC提供模型在不同閾值下的性能表現。

表4-1 氣象、空氣品質與登革熱確診數資料介紹表

資料	來源	時間	資料區域	資料筆數
氣象、空氣品質	環境部	2010~2023	雲嘉南高屏共 26 個測站	3675
登革熱	衛福部疾管署	2010~2023	雲嘉南高屏	3675

第二節　資料前處理

首先，由於嘉義市的土地面積較小且僅有零星的登革熱確診病例，將嘉義市與嘉義縣的數據合併為一個區域，這樣的處理不僅能夠增加數據的量級，還能減少由於數據稀疏帶來的分析偏差，使合併後的數據更能反映該地區的整體疫情情況，提高了數據的可靠性。

再來，合併資料時發現2017年以前的數據格式與2018年之後的數據格式有所不同，這可能是由於設備更新或數據收集方法的變化所致。為了解決這一問題，需要先對數據進行了格式統一處理，重新整理和合併數據，以確保所有年份的數據可以合併在一起。這一步驟至關重要，因為一致的數據格式才能方便確保後續資料建模、分析以及評估等步驟。

為了更加全面地捕捉數據中的長期趨勢和週期性變化，研究將原始資料頻率從小時轉換為週。這一處理旨在平滑短期波動，方便更清晰地觀察數據中的長期模式。在進行頻率轉換的過程中，同時計算了每週的多個統計特徵，包括平均數、中位數、最大值和最小值。每週平均數和中位數提供了每週數據的中心趨勢，幫助理解整體數據的變化，可幫助識別極端情況，避免了極端值對結果的影響。每週最大值以及最小值則有助於幫助識別極端情況，了解數據的上下限。通過這些統計特徵的計算，能夠更好地理解數據的分佈和變化趨勢。這些特徵的引入不僅豐富了數據的描述，還對模型的訓練過程提供了更多的訊息。有助於模型更好地捕捉數據的內在規律和變化趨勢，提高模型的預測準確性和穩定性，從而提升對登革熱發病率的預測能力。

此外，為了更好地理解和分析登革熱確診數據，以及避免資料轉換為類別後面臨的極度不平衡問題。將資料轉換為等級劃分的形式，具體而言，將確診數劃分為三個等級：0到3例為Green，4到30例為Yellow，31例以上為Red。

表4-2 確診數、類別對照表

確診數	類別
0~3	Green
4~30	Yellow
31 以上	Red

這種劃分方式不僅有助於簡化數據，還能夠突顯不同區域的疫情嚴重程度，有助於後續模型的分類和預測。最後，將2010年到2023年期間的各區域空氣品質資料進行了特徵提取，具體來說，將每周的空氣品質數據計算出平均值、中位數、最大值和最小值。這些統計特徵可以提供數據的基本描述，並幫助捕捉數據的趨勢和變化。例如，平均值可以反映出一周內空氣品質的整體狀況，中位數則能夠減少極端值對數據的影響，而最大值和最小值可以提供關於空氣品質波動範圍的有用信息。這些資料前處理步驟不僅為研究提供了清晰而一致的數據基礎，幫助更好地理解數據的內在結構和特徵，為後續的登革熱風險預測模型的建立和訓練奠定了堅實的基礎。

第三節　二階段預測

本研究共使用了五個模型進行預測，分別是人工神經網絡（ANN）、長短期記憶網絡（LSTM）、門控循環單元（GRU）、支持向量機（SVM）和隨機森林（RandomForest）。目標變數被分為三個類別：Green、Yellow 和 Red。然而，由於資料中各類別的比例嚴重不平衡，為避免少數類別之預測結果不佳，預測流程需要分兩個階段進行。其中，類別Green有3192筆資料，佔總資料的86.86%；類別Yellow有283筆資料，佔總資料的7.7%；類別Red則有200筆資料，佔總資料的5.4%。如下表4-3可見資料比例多達約87%為Green，佔絕大多數資料比例。

表4-3 類別資料不平衡比例表

類別	資料比數	類別比例
Green	3192	86.86%
Yellow	283	7.7%
Red	200	5.4%

在第一階段，模型的分類任務是區分 green 和 not green 兩種類別。這一步驟旨在先排除掉那些明顯不屬於 green 類別的數據，從而減少後續預測的難度。第二階段則進一步對 not green 類別進行預測，將其分為 yellow 和 red 類別。這種分階段預測的方法類似於決策樹的概念，通過逐步細化分類，使模型能夠更精確地處理數據不平衡的問題。

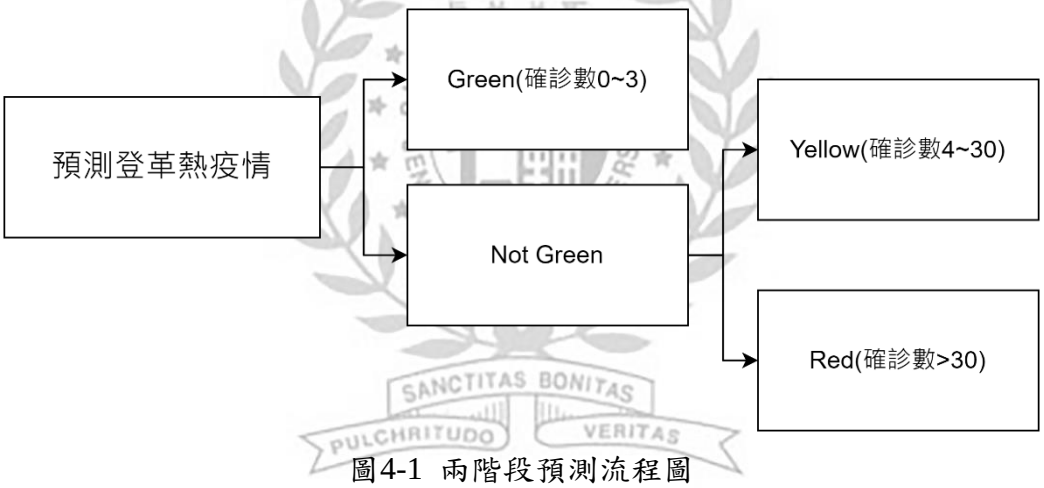

圖4-1 兩階段預測流程圖

這種雙階段預測策略充分利用了每個模型的優勢，在第一階段簡化分類問題，第二階段針對剩餘的更難分類的數據進行細化預測，從而提高整體預測的準確性和穩定性。每個模型在這兩個階段都進行訓練和測試，以確保最終的預測結果具有較高的可靠性和有效性。

第四節　預測模型

一、 ANN Model

本研究所使用之 ANN 模型是一個多層神經網路，層數一共有 10 層，每層之間使用 ReLU 激活函數，最終輸出層使用 Softmax 激活函數來進行分類。具體的層配置包括逐漸減少的神經元數量，從輸入層的 input_dim 開始，經過 512、256、128、64、32、16 等多層，最終輸出 2 個分類結果。在訓練過程中，epochs 設定在 30。使用 Focal Loss 損失函數（γ=3.0，α=[0.15,0.85]）來處理類別不平衡問題，並且使用 Adam 最佳化函數（學習率 0.00006）來最佳化模型的參數。

層數

這個模型總共有10層全連接層（Fully Connected Layers, FCL），具體如下：

表4-4 ANN模型設定說明

Layers	Number of neurons
FCL 1	512
FCL 2	256
FCL 3	128
FCL 4	64
FCL 5	32
FCL 6	16
FCL 7	16
FCL 8	16
FCL 9	16
FCL 10	2

激活函數

在前9層（fc1到fc9），使用了ReLU（Rectified Linear Unit）激活函數。在最後一層（fc10），使用了Softmax激活函數。

表4-5 ANN模型激活函數表

Layers	Activation Function
FCL 1~9	ReLU
FCL 10	Softmax

最佳化函數

使用Adam優化函數來優化模型的參數。學習率設置為0.00006。

表4-6 ANN模型最佳化函數表

Optimization Function	Learning Rate
Adam	0.00006

損失函數

使用了Focal Loss損失函數來處理類別不平衡問題。具體參數設置如下：

表4-7 ANN模型損失函數表

參數	Value
γ	3.0
α	[0.15, 0.85]

二、 LSTM Model

本研究所使用之LSTM模型由一層輸入層、一層隱藏層和一層輸出層組成。隱藏層有 50個神經元，輸出層有 1 個神經元。啟動函數設定為ReLU。在訓練過程中，epochs設定在26。使用Focal Loss損失函數（γ=3.0，α=[0.12,0.88]）來處理類別不平衡問題，並且使用Adam優化函數（學習率為0.0005）來優化模型的參數。

層數

本模型總共有1層隱藏層，具體如下：

表4-8 LSTM模型介紹表

Layers	Units
1	50

激活函數

在隱藏層中，使用了sigmoid和tanh兩種激活函數。在最後一層（fc10），使用了Sigmoid激活函數。

表4-9 LSTM模型激活函數表

Layers	Activation Function
FCL 1	Tanh, Sigmoid

最佳化函數

使用Adam優化函數來優化模型的參數。學習率設置為0.00006。

表4-10 LSTM模型最佳化函數表

Optimization Function	Learning Rate
Adam	0.0005

損失函數

使用了Focal Loss損失函數來處理類別不平衡問題。具體參數設置如下：

表4-11 LSTM模型損失函數表

參數	Value
γ	3.0
α	[0.12, 0.88]

三、 GRU Model

本研究所使用之GRU模型由一層輸入層、一層隱藏層和一層輸出層組成。隱藏層有 64個神經元，輸出層有 1 個神經元。啟動函數設定為ReLU。在訓練過程中，epochs設定在30。使用Focal Loss損失函數（γ=3.0，α=[0.12,0.88]）來處理類別不平衡問題，並且使用Adam優化函數（學習率為0.0005）來優化模型的參數。

層數

這個模型總共有1層隱藏層，具體如下：

表4-12 GRU模型設定

Layers	Units
1	64

激活函數

使用了ReLU（Rectified Linear Unit）激活函數。在最後一層（fc10），使用了

Sigmoid激活函數。

表4-13 GRU模型激活函數表

Layers	Activation Function
FCL 1	Tanh, Sigmoid

最佳化函數

使用Adam優化函數來優化模型的參數。學習率設置為0.00006。

表4-14 GRU模型最佳化函數表

Optimization Function	Learning Rate
Adam	0.0005

損失函數

使用了Focal Loss損失函數來處理類別不平衡問題。具體參數設置如下：

表4-15 GRU模型損失函數表

參數	Value
γ	3.0
α	[0.12, 0.88]

四、 SVM

本研究使用支持向量機（Support Vector Machine, SVM）模型進行分類任務。SVM模型擅長於處理高維數據，並在各類分類任務中表現出色。但由於模型的參數均有其淺在價值，因此使用網格搜尋法（GridSearch）搜尋kernel、C、γ等參數之最佳組合。最後，為了評估模型，使用了交叉驗證（Cross-Validation）K-fold=3，並設置隨機種子以確保結果的可重複性（random_state=1）。

表4-16 SVM模型GridSearch之超參數範圍

Hyperparameter	Value
Kernel	RBF, Poly, Sigmoid
C	(1, 11, 1)
γ	(1, 11, 1)

表4-17 SVM模型GridSearch之最佳超參數組合

Hyperparameter	Value
Kernel	Poly.
C	1
γ	1

五、 Random Forest

本研究使用隨機森林（RandomForest, RF）模型進行分類任務。RF模型通過構建多個決策樹並將其結合來提高模型的分類性能和穩定性。使用網格搜尋法（GridSearch）搜尋n_estimators、min_samples_split等參數之最佳組合。最後，為了評估模型，使用了交叉驗證（Cross-Validation）K-fold=3，並設置隨機種子以確保結果的可重複性（random_state=1）。

表4-18 Random Forest模型GridSearch之超參數範圍

Hyperparameter	Value
n_estimators	(5, 100, 1)
min_samples_split	(2, 30, 1)

表4-19 Random Forest模型Grid Search之最佳超參數組合

Hyperparameter	Value
n_estimators	95
min_samples_split	4

第五節 實驗結果

以下介紹本研究使用的軟硬體環境。作業系統為Windows 11 家用版，程式語言使用Python 3.11。主要使用的Python套件包括Pandas 2.1.4、NumPy 1.26.4、Matplotlib 3.8.0、Scikit-learn 1.4.2、PyTorch 2.3.0+cpu和Imbalanced-learn 0.11.0。此外， IDE方面，使用Jupyter Notebook。硬體資源包括Intel Core i5-1240P @

1.70GHz處理器、16 GB DDR5記憶體、Intel Iris X_e圖形處理器、512GB NVMe SSD硬碟。

以下為第一階段預測結果:

表4-20 ANN模型第一階段預測結果

類別	Accuracy	Recall	F1-score	Support	AUC
Green	82%	82%	89%	639	88%
Not Green		81%	54%	96	

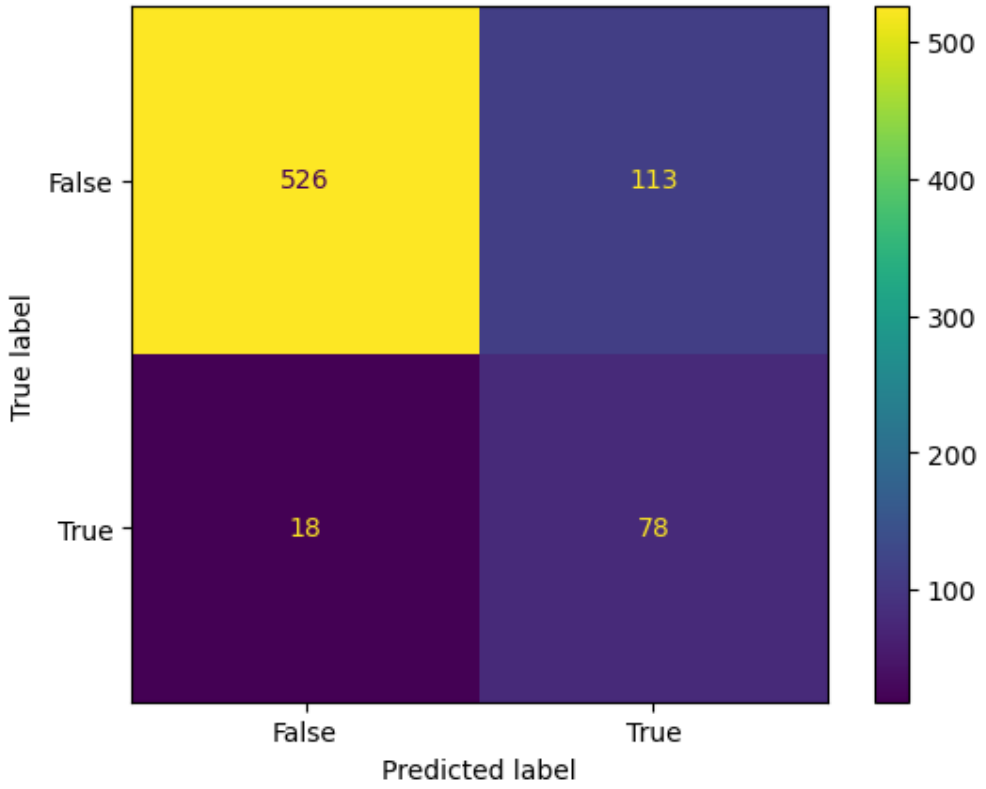

圖4-2 ANN模型第一階段預測混淆矩陣

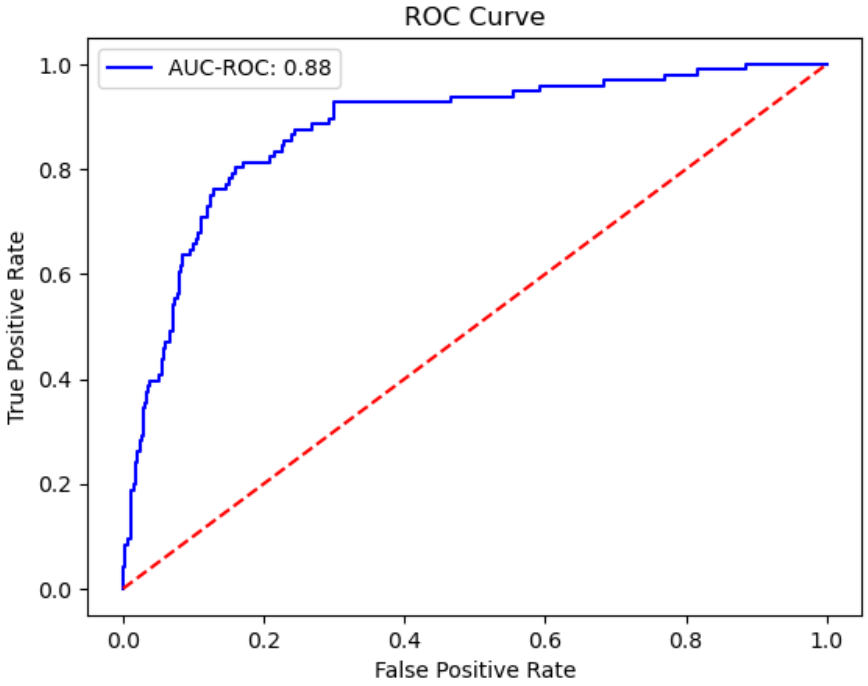

圖4-3 ANN模型第一階段預測ROC圖

表4-21 LSTM模型第一階段預測結果

類別	Accuracy	Recall	F1-score	Support	AUC
Green	81%	81%	88%	639	88%
Not Green		81%	53%	96	

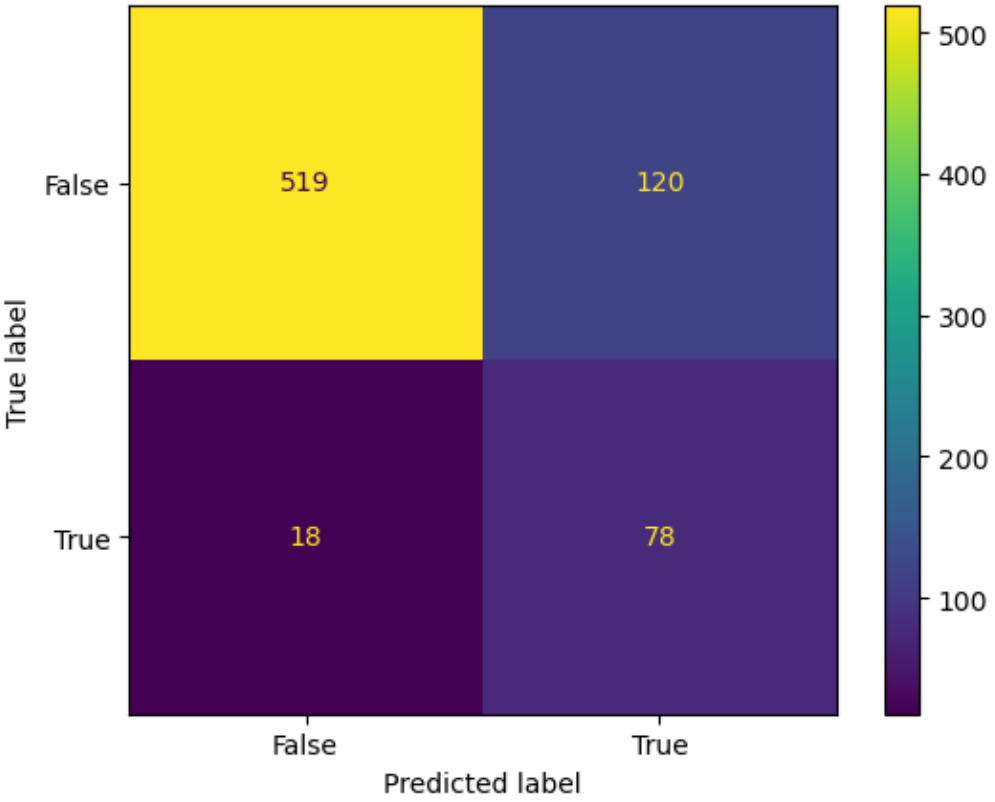

圖4-4 LSTM模型第一階段預測混淆矩陣

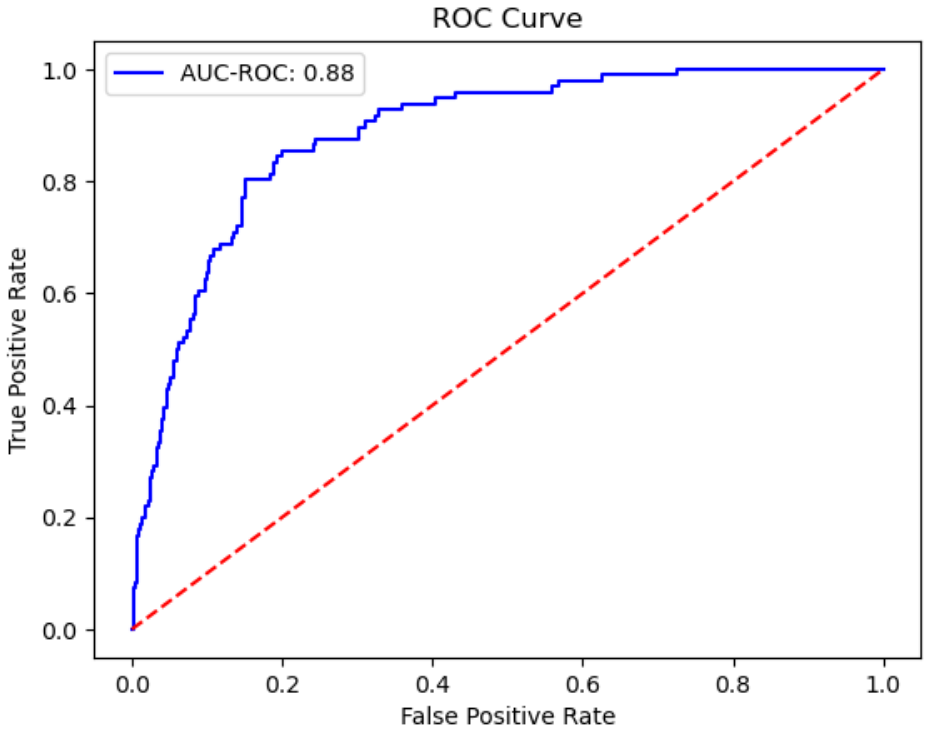

圖4-5 LSTM模型第一階段預測ROC圖

表4-22 GRU模型第一階段預測結果

類別	Accuracy	Recall	F1-score	Support	AUC
Green	80%	80%	87%	639	89%
Not Green		83%	52%	96	

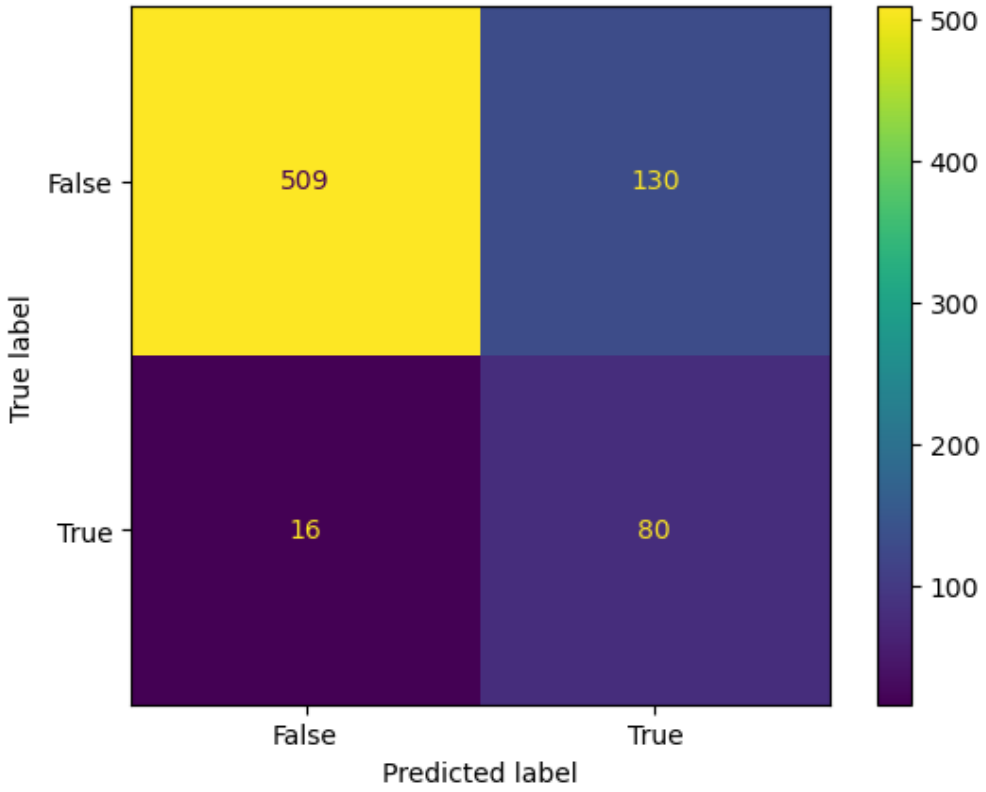

圖4-6 GRU模型第一階段預測混淆矩陣

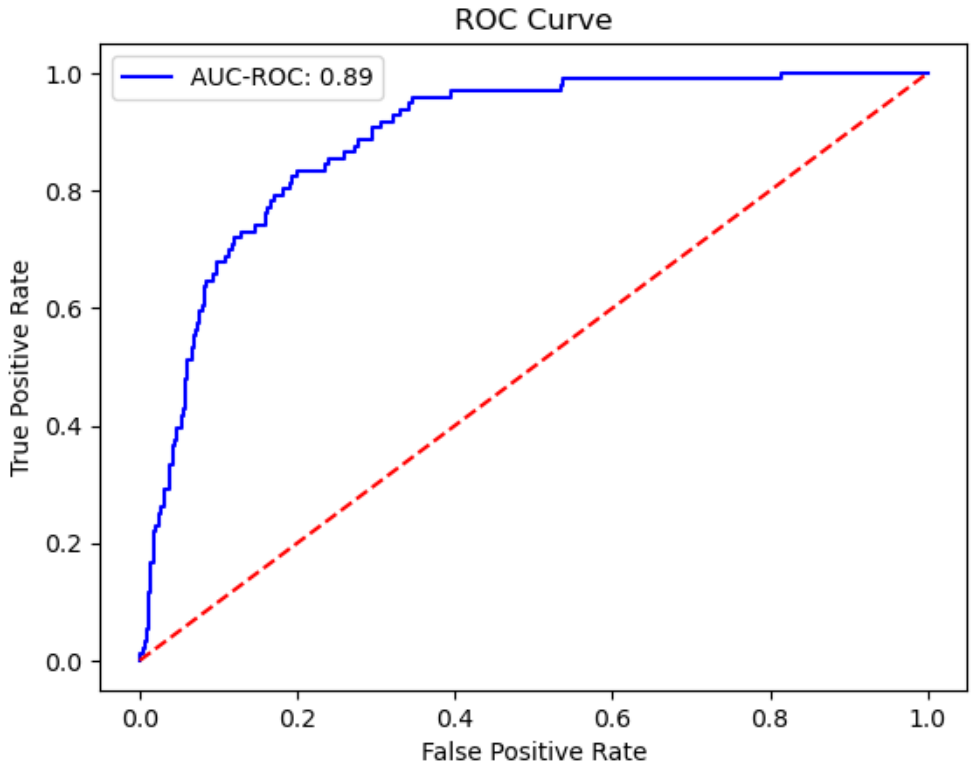

圖4-7 GRU模型第一階段預測ROC圖

表4-23 SVM模型第一階段預測結果

類別	Accuracy	Recall	F1-score	Support	AUC
Green	82%	82%	89%	639	88%
Not Green		81%	54%	96	

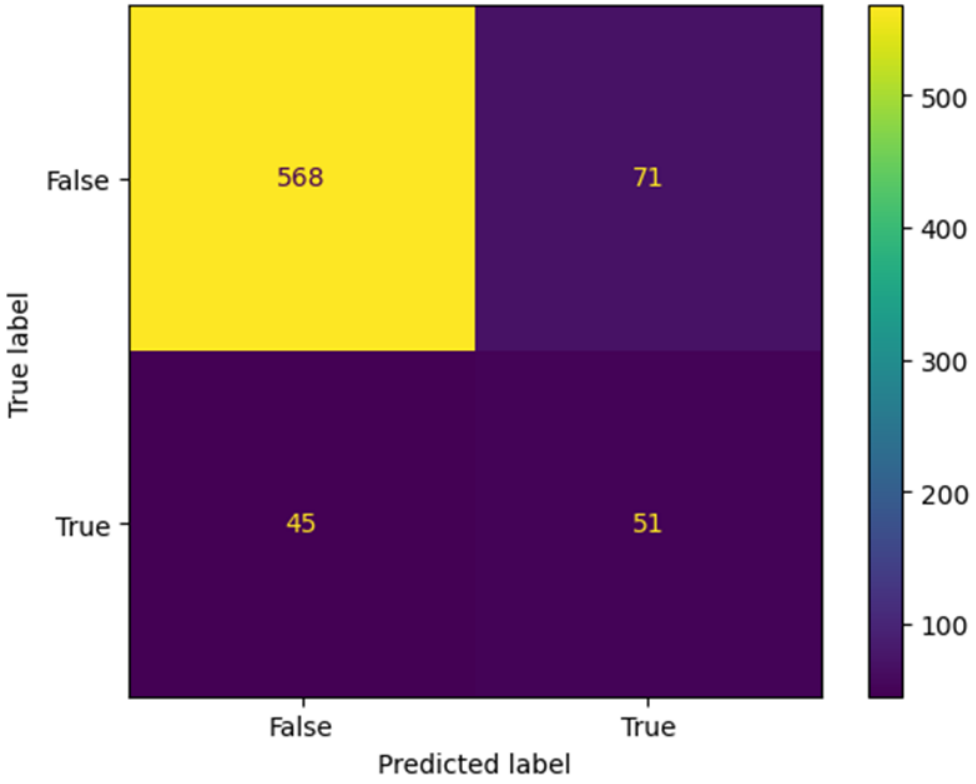

圖4-8 SVM模型第一階段預測混淆矩陣

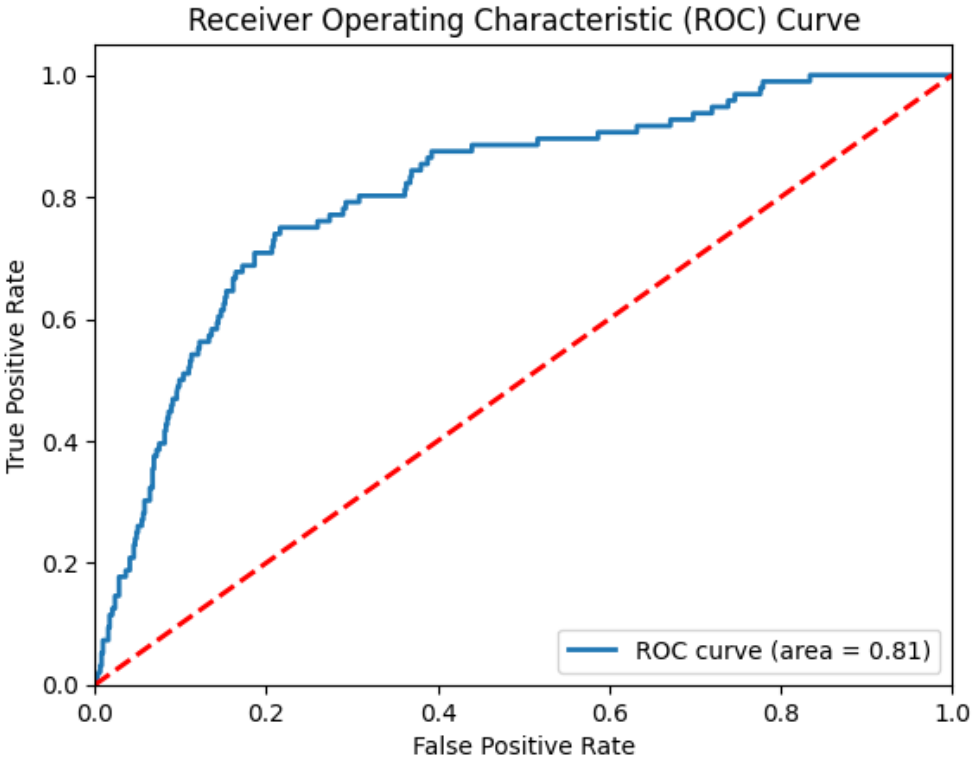

圖4-9 SVM模型第一階段預測ROC圖

表4-24 RandomForest模型第一階段預測結果

類別	Accuracy	Recall	F1-score	Support	AUC
Green	82%	82%	89%	639	88%
Not Green		81%	54%	96	

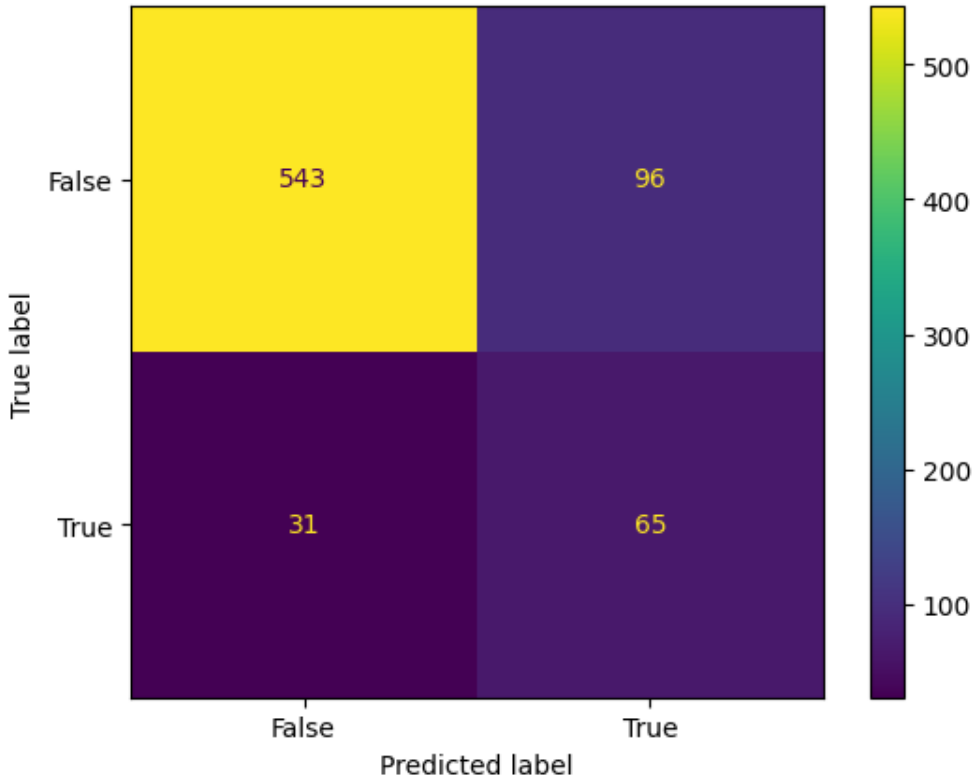

圖4-10 RandomForest模型第一階段預測混淆矩陣

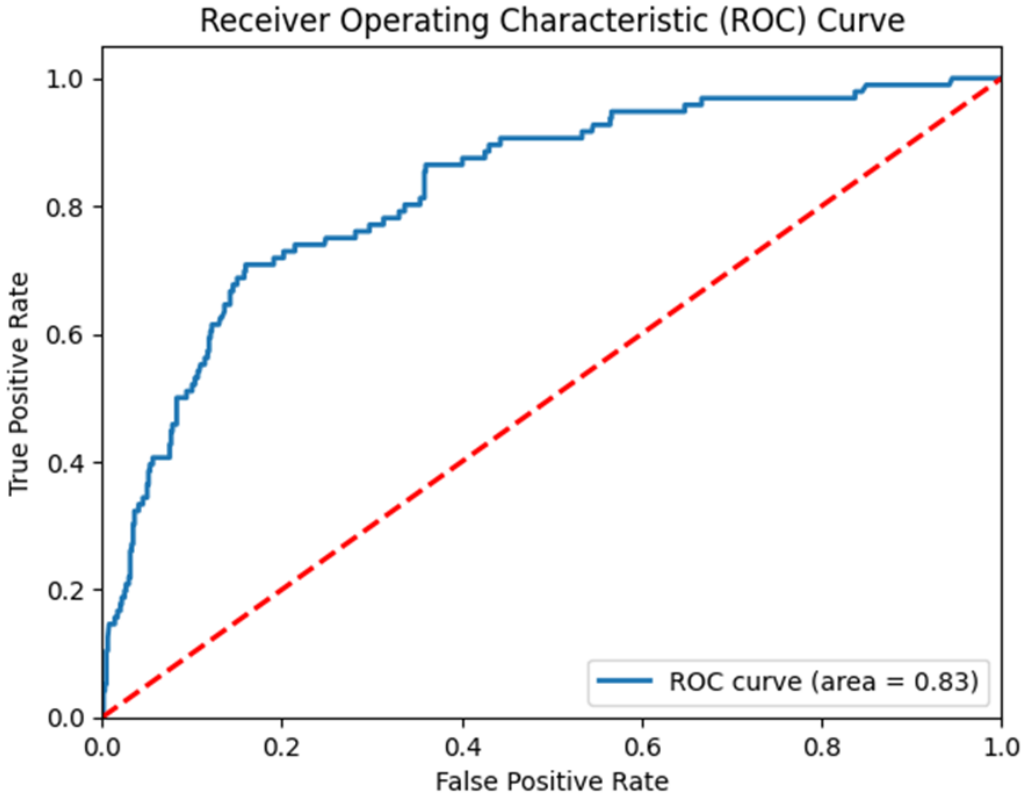

圖4-11 RandomForest模型第一階段預測ROC圖

第二階段預測結果

表4-25 ANN模型第二階段預測結果

類別	Accuracy	Recall	F1-score	Support	AUC
Green	72%	72%	74%	54	81%
Not Green		72%	70%	43	

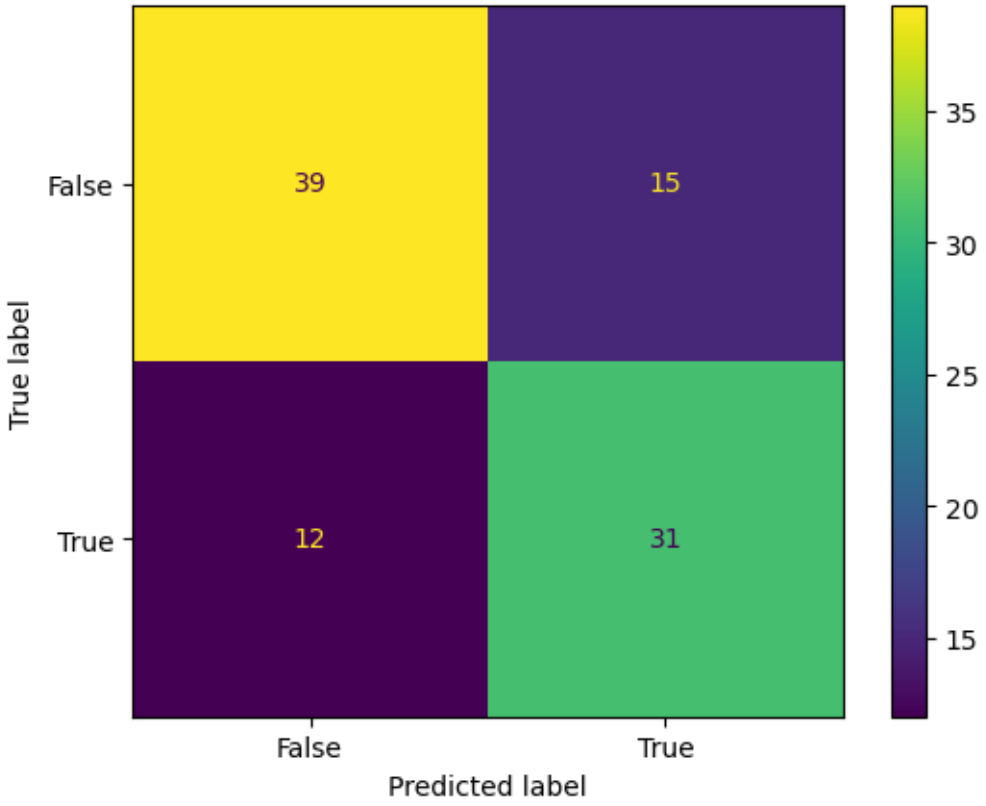

圖4-12 ANN模型第二階段預測混淆矩陣

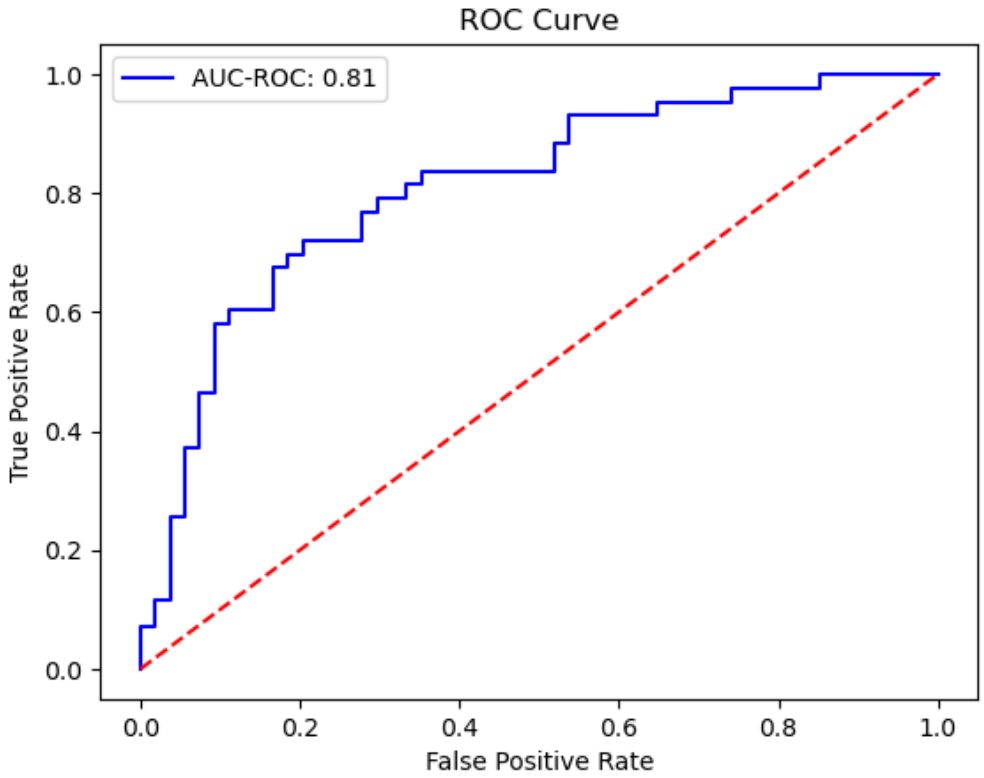

圖4-13 ANN模型第二階段預測ROC圖

表4-26 LSTM模型第二階段預測結果

類別	Accuracy	Recall	F1-score	Support	AUC
Green	75%	70%	76%	54	84%
Not Green		81%	74%	43	

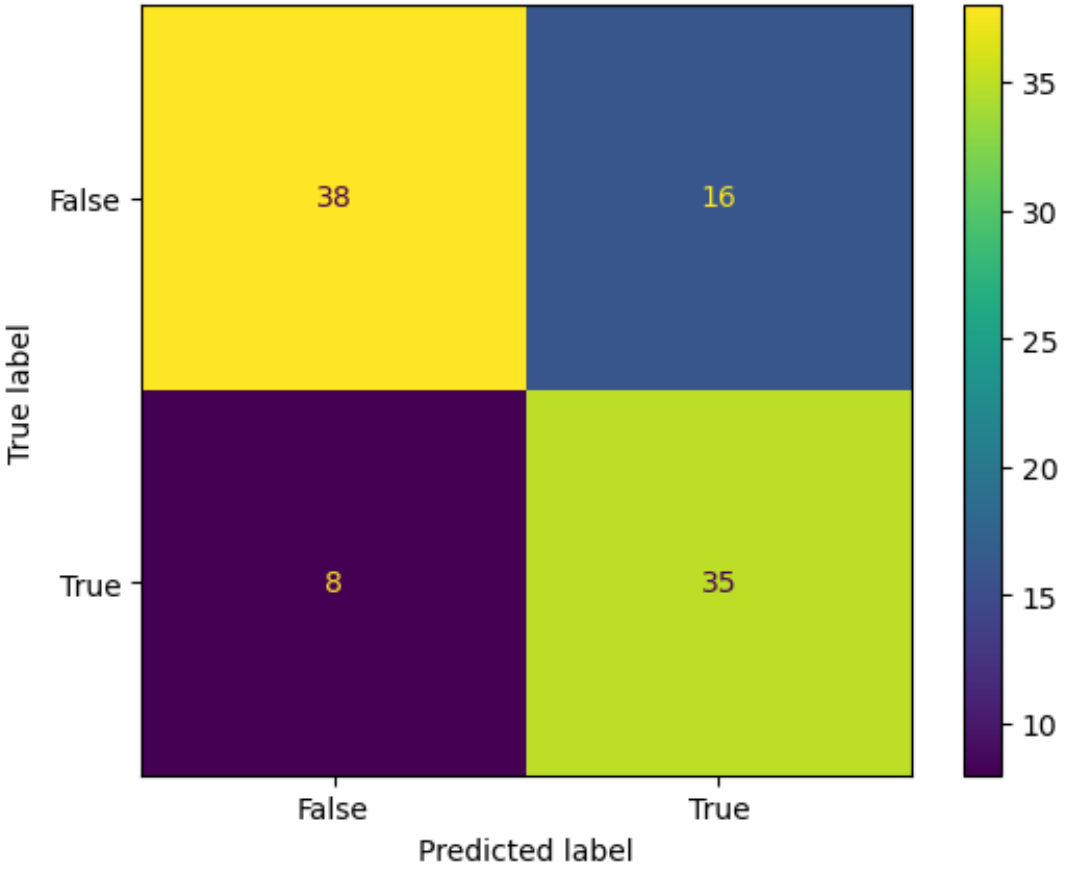

圖4-14 LSTM模型第二階段預測混淆矩陣

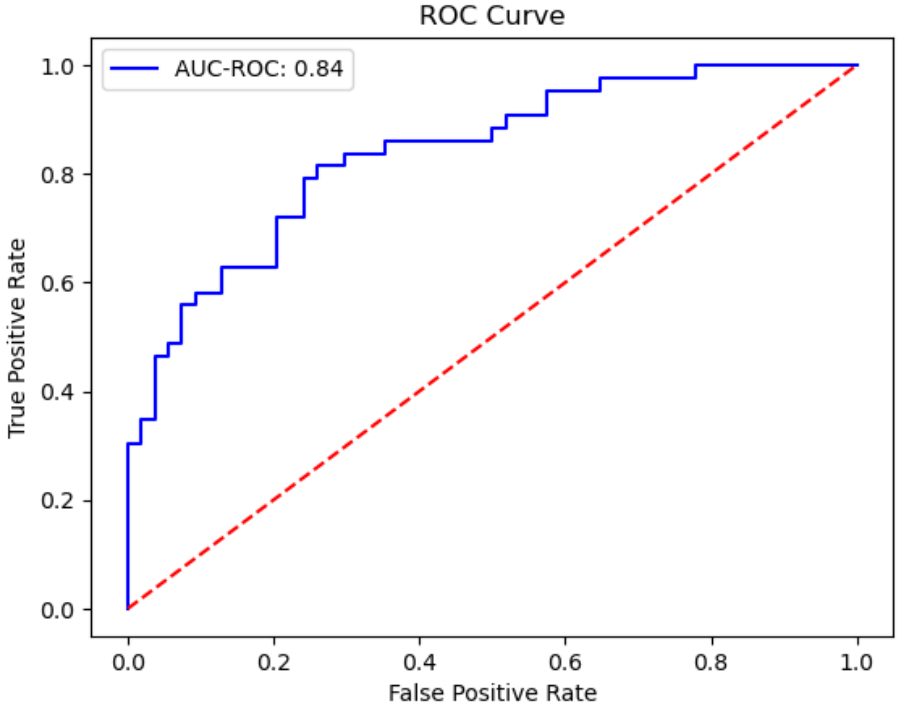

圖4-15 LSTM模型第二階段預測ROC圖

表4-27 GRU模型第二階段預測結果

類別	Accuracy	Recall	F1-score	Support	AUC
Green	76%	80%	79%	54	85%
Not Green		72%	73%	43	

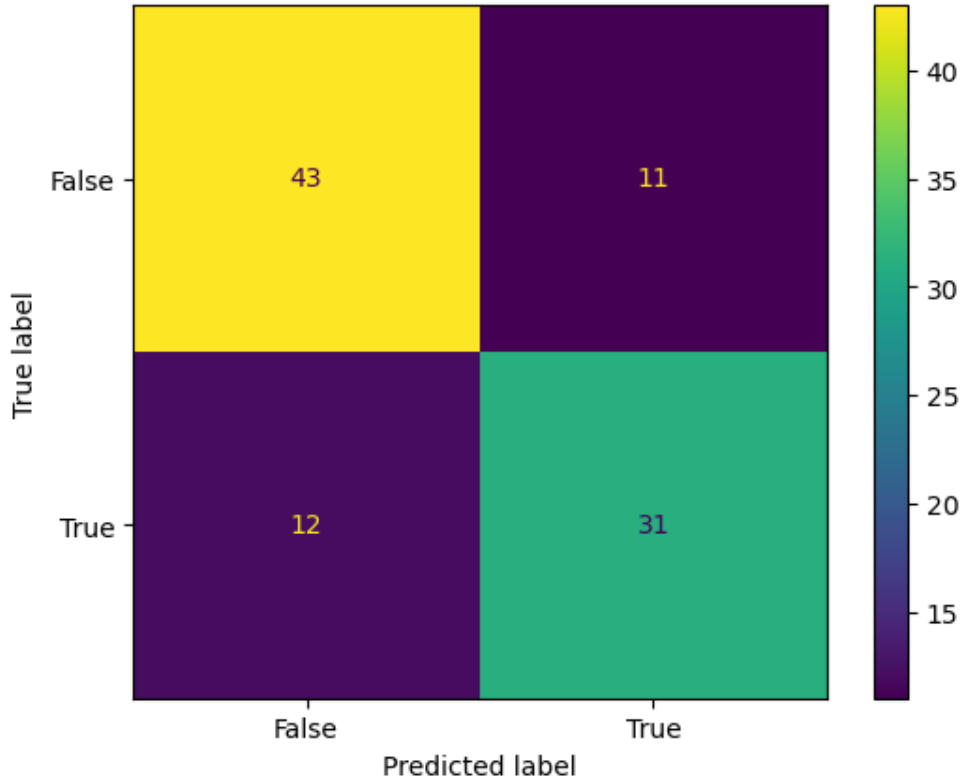

圖4-16 GRU模型第二階段預測混淆矩陣

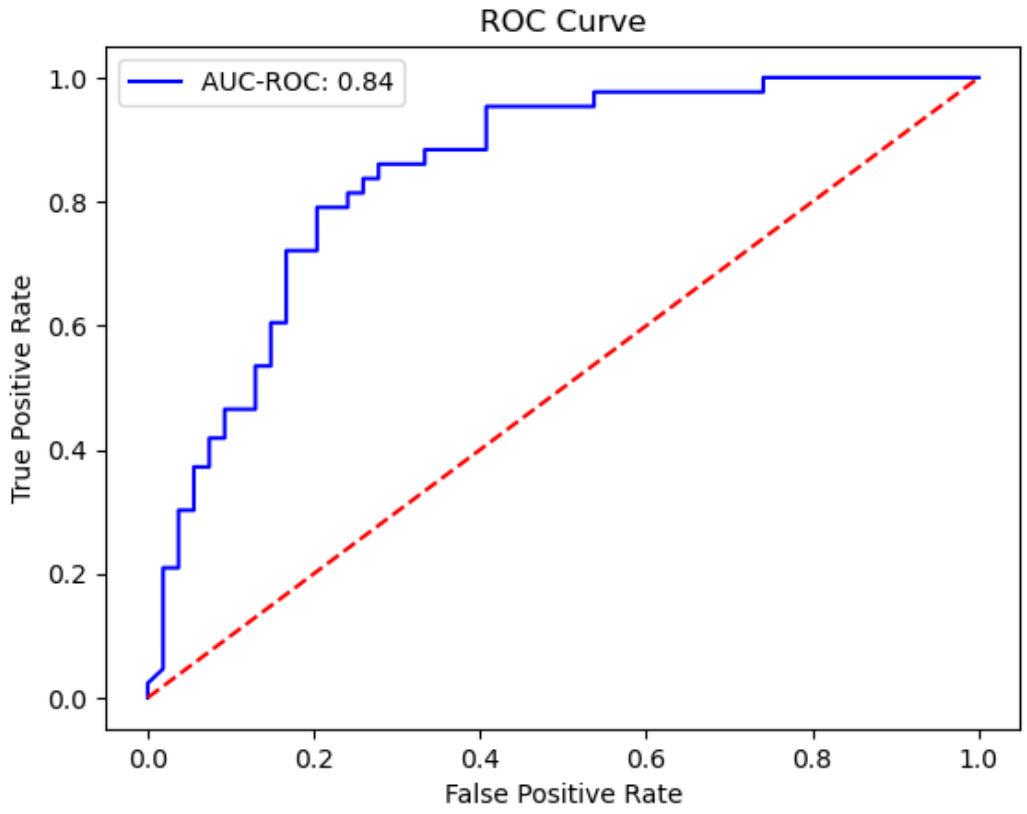

圖4-17 GRU模型第二階段預測ROC圖

表4-28 SVM模型第二階段預測結果

類別	Accuracy	Recall	F1-score	Support	AUC
Green	76%	70%	77%	54	81%
Not Green		84%	76%	43	

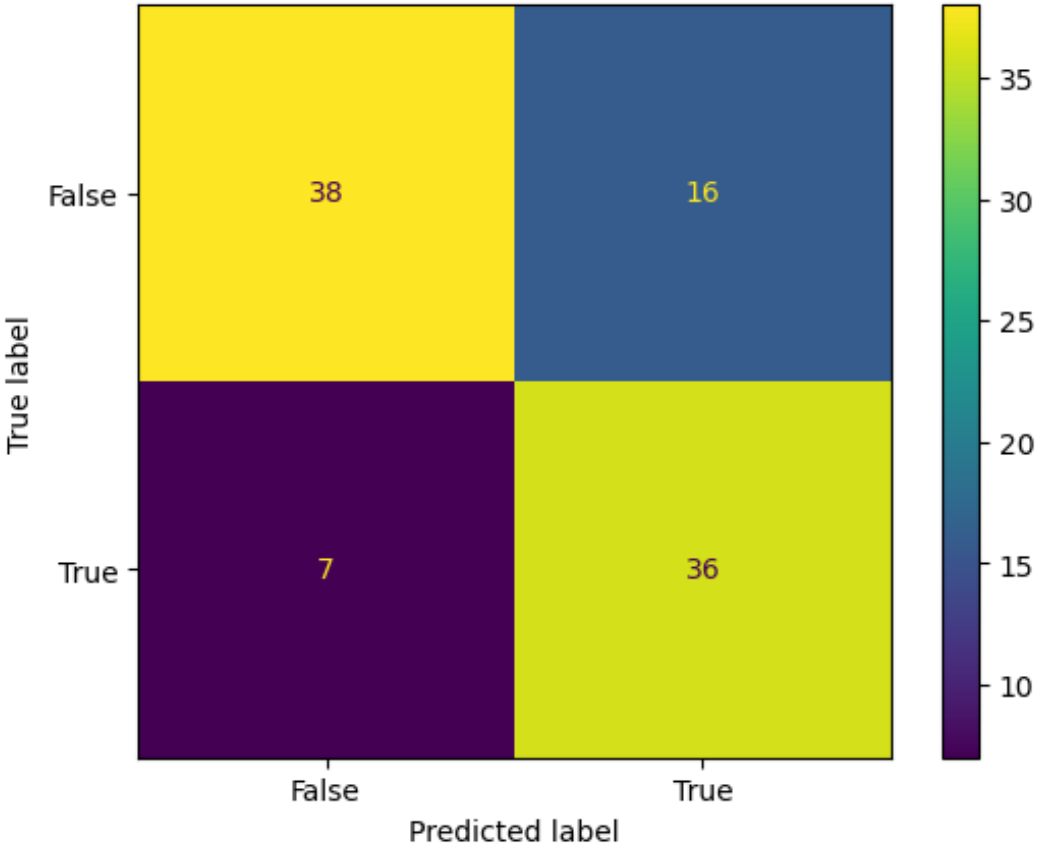

圖4-18 SVM模型第二階段預測混淆矩陣

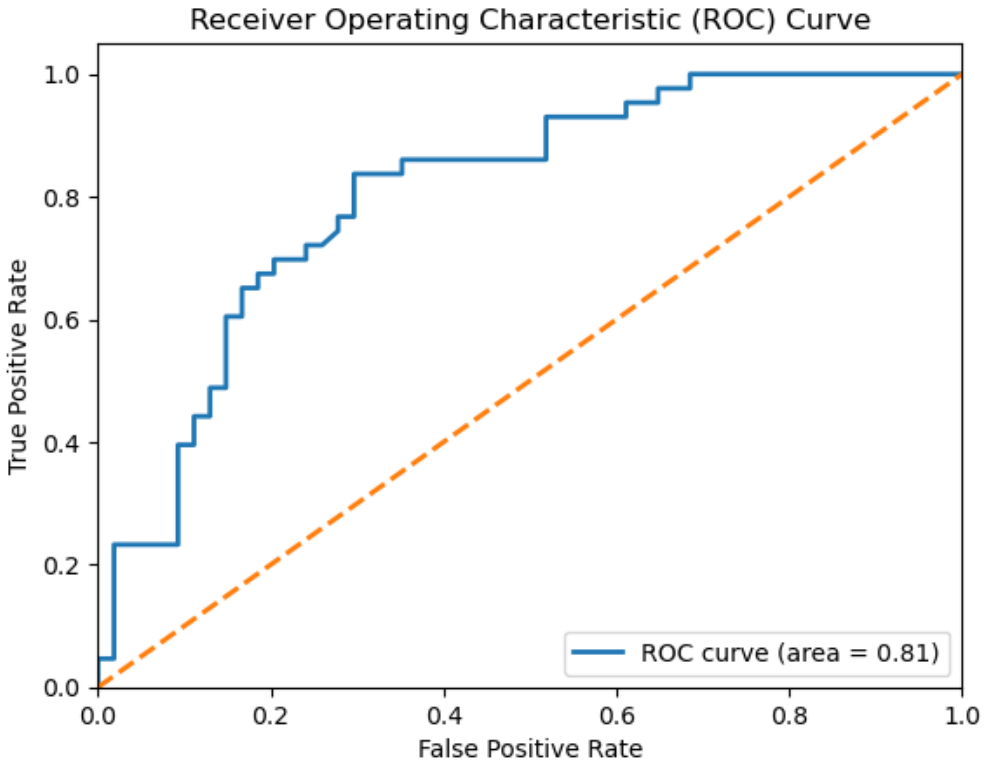

圖4-19 SVM模型第二階段預測ROC圖

表4-29 RandomForest模型第二階段預測結果

類別	Accuracy	Recall	F1-score	Support	AUC
Green	72%	70%	74%	54	81%
Not Green		74%	70%	43	

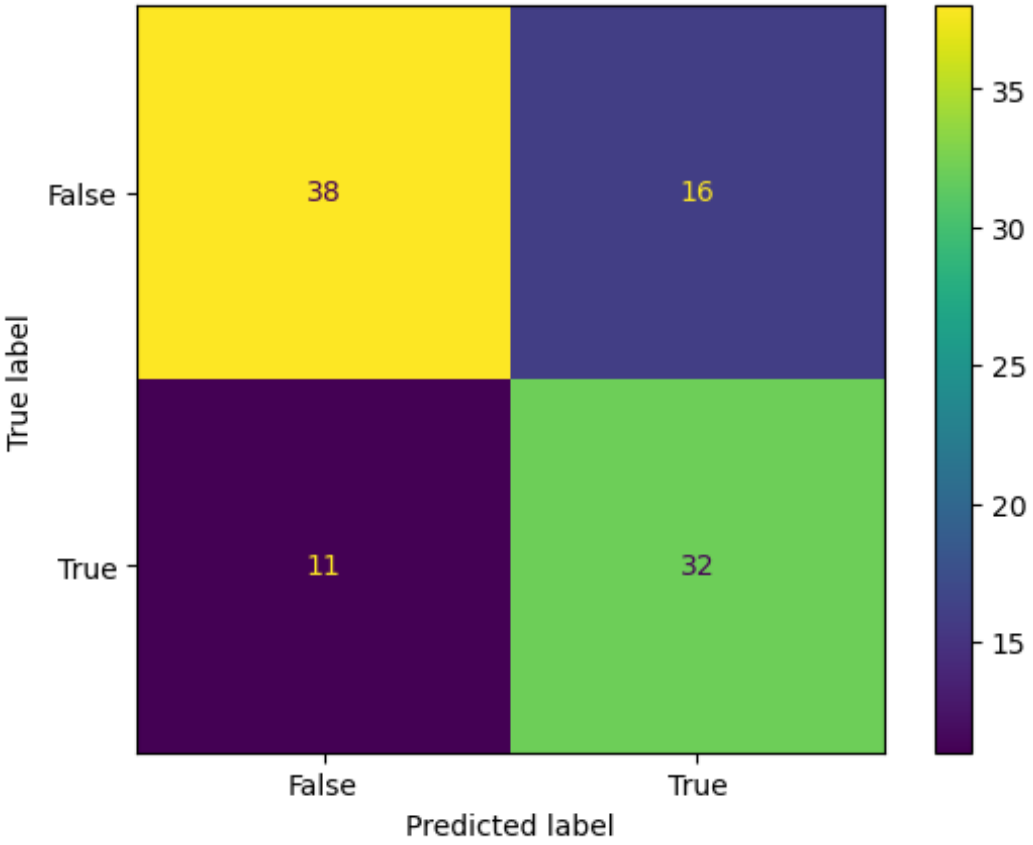

圖4-20 RandomForest模型第二階段預測混淆矩陣

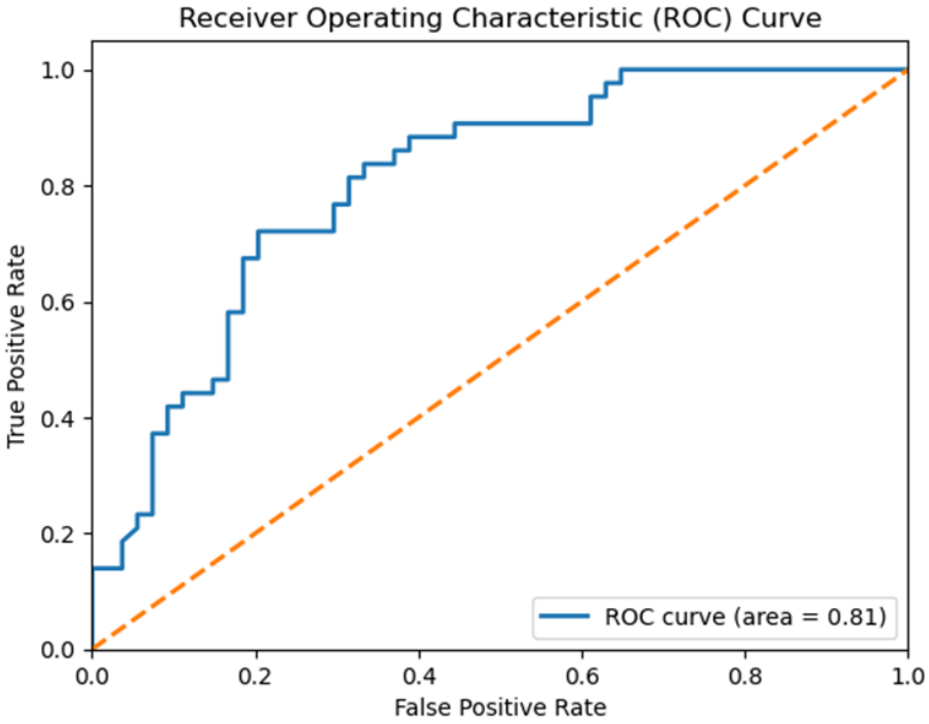

圖4-21 RandomForest模型第二階段預測ROC圖

結果統整

表4-30顯示了在第一階段分析中不同模型的預測結果。被評估的模型包括人工神經網絡（ANN）、長短期記憶網絡（LSTM）、門控循環單元（GRU）、支持向量機（SVM）和隨機森林（Random Forest）。所使用的評估指標包括準確率（ACC）、召回率（Rec）、精確率（PRE）、F1分數（F1s）和支持數（Sup），評估的類別為「綠色」和「非綠色」。其結果分別說明如下：

在人工神經網絡（ANN）中，「綠色」類別的準確率為82%，召回率也為82%，精確率達到97%，F1分數為89%，支持數為639。而「非綠色」類別的準確率為81%，召回率為41%，精確率為54%，F1分數為96%，支持數為96。

對於長短期記憶網絡（LSTM）模型，「綠色」類別的準確率為81%，召回率同樣為81%，精確率為97%，F1分數為88%，支持數為639。「非綠色」類別的

準確率也為81%，但召回率下降到39%，精確率為53%，F1分數為96%，支持數同樣為96。

門控循環單元（GRU）模型在「綠色」類別中的準確率為80%，召回率也為80%，精確率為97%，F1分數為87%，支持數為639。而在「非綠色」類別中，準確率為83%，召回率為38%，精確率為52%，F1分數為96%，支持數為96。

支持向量機（SVM）在「綠色」類別中的準確率為84%，召回率為89%，精確率為93%，F1分數達到91%，支持數為639。然而，「非綠色」類別的準確率僅為53%，召回率為42%，精確率為47%，F1分數為96%，支持數為96。

最後，隨機森林（Random Forest）模型在「綠色」類別中的表現最佳，準確率高達90%，召回率為95%，精確率為92%，F1分數為93%，支持數為639。「非綠色」類別的準確率為54%，召回率為62%，精確率為58%，F1分數為96%，支持數為96。

這些結果顯示了不同機器學習模型在預測「綠色」和「非綠色」類別時的效果，隨機森林模型在「綠色」類別中的準確率最高，而支持向量機模型在「非綠色」類別中的準確率相對較高。所提供的各項指標幫助理解每個模型在不同情況下的有效性。

表4-30 第一階段預測結果統整

模型	類別	ACC	Rec	PRE	F1s	Sup
ANN	Green	82%	82%	97%	89%	639
	Not Green		81%	41%	54%	96
LSTM	Green	81%	81%	97%	88%	639
	Not Green		81%	39%	53%	96
GRU	Green	80%	80%	97%	87%	639
	Not Green		83%	38%	52%	96
SVM	Green	84%	89%	93%	91%	639
	Not Green		53%	42%	47%	96
Random Forest	Green	90%	95%	93%	94%	639
	Not Green		54%	62%	58%	96

表4-31 第二階段預測結果統整

模型	類別	ACC	Rec	PRE	F1s	Sup
ANN	Yellow	72%	72%	76%	74%	54
	Red		72%	67%	70%	43
LSTM	Yellow	75%	70%	83%	76%	54
	Red		81%	69%	74%	43
GRU	Yellow	76%	80%	78%	79%	54
	Red		72%	74%	73%	43
SVM	Yellow	76%	70%	84%	77%	54
	Red		84%	69%	76%	43
Random Forest	Yellow	77%	74%	83%	78%	54
	Red		81%	71%	76%	43

結論

由於此類別資料存在不平衡問題，根據文獻建議不應以正確率來評判模型的好壞，而是應該改為使用Recall、F1-Score、AUC等指標做為參考依據（Kulkarni et al., 2020；Liu et al., 2023）。

在第一階段預測中，判斷區分Green（確診數0-3）和Not Green（確診數超過3）的分類中，ANN、LSTM和GRU模型在兩個類別上的召回率較為均衡，較適合作為預測登革熱有無爆發跡象的模型。

在第二階段預測登革熱的嚴重程度中，Yellow（確診數4-30）和Red（確診數超過30）兩類別的資料比例分別為56%和44%，不平衡並不明顯，因此以正確率評斷模型結果，以RandomForest為表現最佳的模型。


聖美善眞
FU JEN CATHOLIC UNIVERSITY
SANCTITAS BONITAS
PULCHRITUDO
VERITAS

第伍章　討論與未來展望

第一節　討論

大多數登革熱來源國之傳統監測系統大約可分為三類：第一，事件型監測系統，如ProMED-mail和HealthMap，專注於來自新聞、報告和官方信息的非結構化報告，以提高對新興公共衛生問題的認識（Milinovich et al., 2014；Yan et al., 2017）。第二，搜索查詢監測系統，如Google Trends和百度，依賴於互聯網用戶對網絡搜索引擎特定信息的請求。通過追蹤與疾病相關的搜索詞頻率變化來估算發病率（Milinovich et al., 2014；Yan et al., 2017）。第三，社交媒體監測系統，如X、Facebook和微博，支持用戶生成內容的生產和共享，以類似於新聞報告的方式獲取與疾病活動相關的訊息，這是社交媒體平台發展之後的結果（Yan et al., 2017）。

大多數來源國傳統監測系統的登革熱報告存在較大延遲，因此無法及時明確來源國的登革熱發病率與非流行國家輸入風險之間的關係，從而無法及時預警。因此，本研究旨在調查氣候與空氣品質數據是否能夠幫助及時預測登革熱疫情爆發的風險程度。使用了人工神經網絡（ANN）、長短期記憶網絡（LSTM）、門控循環單元（GRU）、支持向量機（SVM）和隨機森林（Random Forest）等模型進行訓練，並以訓練結果進行綠色、黃色和紅色三階段的登革熱預警。

本次研究使用了2010年至2023年的登革熱確診數據，這些數據顯示出確診數量的顯著波動。2010年至2014年間確診數相對平穩，但在2015和2016年，

登革熱疫情達到高峰。此後的2017年至2022年，由於政府在登革熱大流行後加強了水溝清淤和噴藥等防治措施，民眾也自發清理積水容器，疫情有所緩解。再加上2020年後新冠肺炎疫情導致的戶外活動減少，人們的環境清潔和衛生意識提升，這段時間的登革熱疫情幾乎被忽視。

然而，隨著新冠疫情結束，人們的生活回歸正常，或開始報復性旅遊，人口的大量流動與登革熱疫情之間的密切關係再次顯現。這推測可能導致了2023年台灣南部的再次大流行。

這些問題導致在進行登革熱確診數據分類預測時遇到了類別不平衡的挑戰。由於數據比例懸殊，多分類問題中模型的各項評估指標始終難以突破。因此，本研究改為兩階段的二分類問題。在第一階段，區分Green（確診數0-3）和Not Green（確診數超過3）的分類中，五種模型的Recall指標均達到八成以上。在第二階段，進行Yellow（確診數4-30）和Red（確診數超過30）的分類預測時，五種模型的Recall指標也均達到七成以上。儘管結果不盡理想，但這些模型預測結果仍可提供做為登革熱疫情爆發預警的參考工具。

第二節　未來展望

登革熱的主要病媒蚊包括「白線斑蚊」和「埃及斑蚊」。白線斑蚊在全台灣均有分布，而埃及斑蚊主要集中在嘉義布袋以南地區。國家蚊媒傳染病防治研究中心接受環境部委託，在全台監測這兩種蚊子的分布情況。根據調查推估，隨著全球氣候變暖，登革熱病媒蚊的分布範圍可能會向北擴展，埃及斑蚊可能跨越北回歸線，擴散至台中和東部的花蓮等地區。

由國家蚊媒傳染病防治研究中心技術助理研究員黃旌集所帶領的團隊依

據不同的升溫情境進行推估：未來氣溫若升高1.5℃，埃及斑蚊的北界可能會延伸至台中市太平區和花蓮縣瑞穗鄉；若升溫2.0℃，北界可能進一步擴展至台中市北屯區和花蓮縣花蓮市；若升溫2.5℃，北界可能擴展至台中市潭子區和花蓮縣花蓮市；在最極端的情況下，若氣溫升高至4.4℃，埃及斑蚊的北界可能會擴展至苗栗縣竹南鎮和花蓮縣秀林鄉，分布區域將占台灣總面積的43.3%。這項監測調查及推估有助於應對氣候變遷，提升對登革熱疫情風險的適應能力。除了持續監測外，最重要的是做好環境管理，清除病媒蚊的孳生源，防止埃及斑蚊向北擴散。

未來的研究應該著重於多方面的發展。一方面，應該進一步完善氣候和空氣品質數據的收集和分析方法，提升預測模型的精度。另一方面，應加強與公共衛生部門的合作，及時更新和共享數據，以便制定更有效的防治策略。此外，探索新的數據來源和技術，如遙感技術和物聯網設備，將有助於實時監控蚊媒的動態分布。

未來，通過不斷完善數據收集、分析和預測模型，加強跨部門合作，並探索創新技術，相信未來將能更有效地預測和應對登革熱疫情，保護公共健康。

參考文獻

1. 國家蚊媒傳染病防治研究中心 (2024)。暖化效應：登革熱「北漂」？！。取自 https://nmbdcrc.nhri.edu.tw/2024/05/15/%e6%9a%96%e5%8c%96%e6%95%88%e6%87%89%ef%bc%9a%e7%99%bb%e9%9d%a9%e7%86%b1%e3%80%8c%e5%8c%97%e6%bc%82%e3%80%8d%ef%bc%9f%ef%bc%81/
2. Aburas, H. M., Cetiner, B. G., & Sari, M. (2010). Dengue confirmed-cases prediction: A neural network model. Expert Systems with Applications, 37(6), 4256-4260.
3. Ahmad, R., Suzilah, I., Wan Najdah, W. M. A., Topek, O., Mustafakamal, I., & Lee, H. L. (2018). Factors determining dengue outbreak in Malaysia. PloS one, 13(2), e0193326.
4. Aziz, A.T., Dieng, H., Ahmad, A.H., A Mahyoub, J., Turkistani, A.M., Mesed, H., Koshike, S., Satho, T., Salmah, C., Ahmad, H., Zuharah, W.F., Ramli, A.S., Miake, F. (2012). Household survey of container-breeding mosquitoes and climatic factors influencing the prevalence of Aedes aegypti(Diptera: Culicidae) in Makkah City, Saudi Arabia. Asian Pac. J. Trop. Biomed. 2012, 2, 849–857.
5. Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., ... & Hay, S. I. (2013). The global distribution and burden of dengue. Nature, 496(7446), 504-507.
6. Biran, A., Smith, L., Lines, J., Ensink, J., & Cameron, M. (2007). Smoke and malaria: are interventions to reduce exposure to indoor air pollution likely to increase exposure to mosquitoes?. Transactions of the Royal Society of Tropical Medicine and Hygiene, 101(11), 1065-1071.

7. Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science, 16(3), 199-231.
8. Changal, K. H., Raina, A. H., Raina, A., Raina, M., Bashir, R., Latief, M., ... & Changal, Q. H. (2016). Differentiating secondary from primary dengue using IgG to IgM ratio in early dengue: an observational hospital based clinico-serological study from North India. BMC infectious diseases, 16(1), 1-7.
9. Chimmula, V. K. R. & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, solitons & fractals, 135, 109864.
10. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.
11. Chung, Y. K. & Pang, F. Y. (2002). Dengue virus infection rate in field populations of female Aedes aegypti and Aedes albopictus in Singapore. Tropical Medicine & International Health, 7(4), 322-330.
12. Colwell, R. R. & Patz, J. A. (1998). Climate, infectious disease and health: an interdisciplinary perspective.
13. da Cruz Ferreira, D. A., Degener, C. M., de Almeida Marques-Toledo, C., Bendati, M. M., Fetzer, L. O., Teixeira, C. P., & Eiras, Á. E. (2017). Meteorological variables and mosquito monitoring are good predictors for infestation trends of Aedes aegypti, the vector of dengue, chikungunya and Zika. Parasites & vectors, 10(1), 1-11.
14. Davis, E. E., & Bowen, M. F. (1994). Sensory physiological basis for attraction in mosquitoes. Journal of the American Mosquito Control Association, 10(2 Pt 2), 316-325.
15. Dulhunty, J. M., Yohannes, K., Kourleoutov, C., Manuopangai, V. T., Polyn, M. K., Parks, W. J., & Bryan, J. H. (2000). Malaria control in central Malaita, Solomon Islands 2. Local perceptions of the disease and practices for its treatment and

prevention. Acta Tropica, 75(2), 185-196.

16. Ebi, K. L., & Nealon, J. (2016). Dengue in a changing climate. *Environmental research*, *151*, 115-123.
17. Ehelepola, N. D. B., Ariyaratne, K., Buddhadasa, W. M. N. P., Ratnayake, S., & Wickramasinghe, M. (2015). A study of the correlation between dengue and weather in Kandy City, Sri Lanka (2003-2012) and lessons learned. Infectious diseases of poverty, 4, 1-15.
18. Epstein, P. R. (2001). Climate change and emerging infectious diseases. Microbes and infection, 3(9), 747-754.
19. Forattini, O. P., Kakitani, I., Massad, E., & Marucci, D. (1993). Studies on mosquitoes (Diptera: Culicidae) and anthropic environment: 4-Survey of resting adults and synanthropic behaviour in South-Eastern, Brazil. *Revista de saude publica*, *27*, 398-411.
20. Forattini, O. P., Kakitani, I., Massad, E., & Marucci, D. (1995). Studies on mosquitoes (Diptera: Culicidae) and anthropic environment: 9-Synanthropy and epidemiological vector role of Aedes scapularis in South-Eastern Brazil. *Revista de Saúde pública*, *29*, 199-207.
21. Fu, R., Zhang, Z., & Li, L. (2016, November). Using LSTM and GRU neural network methods for traffic flow prediction. In 2016 31st Youth academic annual conference of Chinese association of automation (YAC) (pp. 324-328). IEEE.
22. Gong, P., Liang, S., Carlton, E. J., Jiang, Q., Wu, J., Wang, L., & Remais, J. V. (2012). Urbanisation and health in China. The lancet, 379(9818), 843-852.
23. Gubler, D. J. (2011). Dengue, urbanization and globalization: the unholy trinity of the 21st century. Tropical medicine and health, 39(4SUPPLEMENT), S3-S11.
24. Gui, H., Gwee, S., Koh, J., & Pang, J. (2021). Weather factors associated with reduced risk of dengue transmission in an urbanized tropical city. International

Journal of Environmental Research and Public Health, 19(1), 339.

25. Guo, P., Liu, T., Zhang, Q., Wang, L., Xiao, J., Zhang, Q., ... & Ma, W. (2017). Developing a dengue forecast model using machine learning: A case study in China. PLoS neglected tropical diseases, 11(10), e0005973.
26. Guzman, M. G., & Harris, E. (2015). Dengue. The Lancet, 385(9966), 453-465.
27. Han, L., Zhou, W., Li, W., & Li, L. (2014). Impact of urbanization level on urban air quality: A case of fine particles (PM2.5) in Chinese cities. Environmental Pollution, 194, 163-170.
28. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
29. Hogarh, J. N., Agyekum, T. P., Bempah, C. K., Owusu-Ansah, E. D., Avicor, S. W., Awandare, G. A., ... & Obiri-Danso, K. (2018). Environmental health risks and benefits of the use of mosquito coils as malaria prevention and control strategy. Malaria Journal, 17, 1-12.
30. Hwang, S., Clarite, D. S., Elijorde, F. I., Gerardo, B. D., & Byun, Y. (2016). A web-based analysis for dengue tracking and prediction using artificial neural network. Adv Sci Technol Lett, 122, 160-4.
31. Kader, N. I. A., Yusof, U. K., Khalid, M. N. A., & Husain, N. R. N. (2022, September). A review of long short-term memory approach for time series analysis and forecasting. In International Conference on Emerging Technologies and Intelligent Systems (pp. 12-21). Cham: Springer International Publishing.
32. Kek, R., Hapuarachchi, H. C., Chung, C. Y., Humaidi, M. B., Razak, M. A. B., Chiang, S., ... & Ng, L. C. (2014). Feeding host range of Aedes albopictus (Diptera: Culicidae) demonstrates its opportunistic host-seeking behavior in rural Singapore. Journal of Medical Entomology, 51(4), 880-884.
33. Kraemer, M. U., Reiner Jr, R. C., Brady, O. J., Messina, J. P., Gilbert, M., Pigott, D.

M., ... & Golding, N. (2019). Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. Nature microbiology, 4(5), 854-863.

34. Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. In Data democracy (pp. 83-106). Academic Press.
35. Lai, Y. H. (2018). The climatic factors affecting dengue fever outbreaks in southern Taiwan: an application of symbolic data analysis. Biomedical engineering online, 17(2), 1-14.
36. Latif, M. T., Othman, M., Idris, N., Juneng, L., Abdullah, A. M., Hamzah, W. P., ... & Jaafar, A. B. (2018). Impact of regional haze towards air quality in Malaysia: A review. Atmospheric Environment, 177, 28-44.
37. Laureano-Rosario, A. E., Garcia-Rejon, J. E., Gomez-Carro, S., Farfan-Ale, J. A., & Muller-Karger, F. E. (2017). Modelling dengue fever risk in the State of Yucatan, Mexico using regional-scale satellite-derived sea surface temperature. Acta tropica, 172, 50-57.
38. Le, P., & Zuidema, W. (2016). Quantifying the vanishing gradient and long distance dependency problem in recursive neural networks and recursive LSTMs. arXiv preprint arXiv:1603.00423.
39. Leopord, H., Cheruiyot, W. K., & Kimani, S. (2016). A survey and analysis on classification and regression data mining techniques for diseases outbreak prediction in datasets. Int. J. Eng. Sci, 5(9), 1-11.
40. Liu, S., Roemer, F., Ge, Y., Bedrick, E. J., Li, Z. M., Guermazi, A., ... & Sun, X. (2023). Comparison of evaluation metrics of deep learning for imbalanced imaging data in osteoarthritis studies. Osteoarthritis and Cartilage, 31(9), 1242-1248.
41. Lu, H. C., Lin, F. Y., Huang, Y. H., Kao, Y. T., & Loh, E. W. (2023). Role of air pollutants in dengue fever incidence: evidence from two southern cities in Taiwan.

Pathogens and Global Health, 117(6), 596-604.

42. Lu, L., Lin, H., Tian, L., Yang, W., Sun, J., & Liu, Q. (2009). Time series analysis of dengue fever and weather in Guangzhou, China. BMC Public Health, 9, 1-5.
43. Luz, P. M., Codeço, C. T., Massad, E., & Struchiner, C. J. (2003). Uncertainties regarding dengue modeling in Rio de Janeiro, Brazil. Memórias do Instituto Oswaldo Cruz, 98(7), 871-878.
44. Ma, J., Wang, X., Wang, Y., Wang, J., Chu, X., & Zhao, J. (2022). Enhancing online epidemic supervising system by compartmental and gru fusion model. Mobile Information Systems, 2022(1), 3303854.
45. Madewell, Z. J., López, M. R., Espinosa-Bode, A., Brouwer, K. C., Sánchez, C. G., & McCracken, J. P. (2020). Inverse association between dengue, chikungunya, and Zika virus infection and indicators of household air pollution in Santa Rosa, Guatemala: A case-control study, 2011-2018. PLoS One, 15(6), e0234399.
46. Massad, E., Coutinho, F. A. B., Ma, S., & Burattini, M. N. (2010). A hypothesis for the 2007 dengue outbreak in Singapore. Epidemiology & Infection, 138(7), 951-957.
47. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5, 115-133.
48. McMichael, A. J. (Ed.). (2003). Climate change and human health: risks and responses. World Health Organization.
49. Medeiros, L. C. D. C., Castilho, C. A. R., Braga, C., de Souza, W. V., Regis, L., & Monteiro, A. M. V. (2011). Modeling the dynamic transmission of dengue fever: investigating disease persistence. PLOS neglected tropical diseases, 5(1), e942.
50. Meyer, H., Kühnlein, M., Appelhans, T., & Nauss, T. (2016). Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. Atmospheric research, 169, 424-433.
51. Milinovich, G. J., Williams, G. M., Clements, A. C., & Hu, W. (2014). Internet-

based surveillance systems for monitoring emerging infectious diseases. The Lancet infectious diseases, 14(2), 160-168.

52. Morin, C. W., Comrie, A. C., & Ernst, K. (2013). Climate and dengue transmission: evidence and implications. Environmental health perspectives, 121(11-12), 1264-1272.
53. Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.
54. Murray, N. E. A., Quam, M. B., & Wilder-Smith, A. (2013). Epidemiology of dengue: past, present and future prospects. Clinical epidemiology, 299-309.
55. Mussumeci, E., & Coelho, F. C. (2020). Large-scale multivariate forecasting models for Dengue-LSTM versus random forest regression. Spatial and Spatio-temporal Epidemiology, 35, 100372.
56. Muthamizharasan, M., & Ponnusamy, R. (2022, December). A Hybrid CNN and GRU-based Spatial-temporal Marburg Virus Disease Hotspot Association Mining for Health Management in Kenya. In 2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI) (Vol. 1, pp. 1-6). IEEE.
57. Ngugi, H. N., Mutuku, F. M., Ndenga, B. A., Musunzaji, P. S., Mbakaya, J. O., Aswani, P., ... & LaBeaud, A. D. (2017). Characterization and productivity profiles of Aedes aegypti (L.) breeding habitats across rural and urban landscapes in western and coastal Kenya. Parasites & vectors, 10, 1-12.
58. Noh, S. H. (2021). Analysis of gradient vanishing of RNNs and performance comparison. Information, 12(11), 442.
59. Nordin, N. I., Sobri, N. M., Ismail, N. A., Zulkifli, S. N., Abd Razak, N. F., & Mahmud, M. (2020, March). The classification performance using support vector machine for endemic dengue cases. In Journal of Physics: Conference Series (Vol. 1496, No. 1, p. 012006). IOP Publishing.
60. O'shea, K., & Nash, R. (2015). An introduction to convolutional neural networks.

arXiv preprint arXiv:1511.08458.

61. Pålsson, K., & Jaenson, T. G. (1999). Plant products used as mosquito repellents in Guinea Bissau, West Africa. Acta Tropica, 72(1), 39-52.
62. Parham, P. E., & Michael, E. (2010). Modeling the effects of weather and climate change on malaria transmission. Environmental health perspectives, 118(5), 620-626.
63. Pu, C. (2017). Causes of Haze Pollution Under the Regional Compound Environment and Legal Governance Countermeasures. Nature Environment & Pollution Technology, 16(3).
64. Rachata, N., Charoenkwan, P., Yooyativong, T., Chamnongthal, K., Lursinsap, C., & Higuchi, K. (2008, October). Automatic prediction system of dengue haemorrhagic-fever outbreak risk by using entropy and artificial neural network. In 2008 International Symposium on Communications and Information Technologies (pp. 210-214). IEEE.
65. Racloz, V., Ramsey, R., Tong, S., & Hu, W. (2012). Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. PLoS neglected tropical diseases, 6(5), e1648.
66. Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. European Journal of Remote Sensing, 50(1), 144-154.
67. Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71, 804-818.
68. Sathler, C., & Luciano, J. (2017). Predictive modeling of dengue fever epidemics: A

neural network approach.

69. Scavuzzo, J. M., Trucco, F., Espinosa, M., Tauro, C. B., Abril, M., Scavuzzo, C. M., & Frery, A. C. (2018). Modeling Dengue vector population using remotely sensed data and machine learning. Acta tropica, 185, 167-175.
70. Schmidt, W. P., Suzuki, M., Dinh Thiem, V., White, R. G., Tsuzuki, A., Yoshida, L. M., ... & Ariyoshi, K. (2011). Population density, water supply, and the risk of dengue fever in Vietnam: cohort study and spatial analysis. PLoS medicine, 8(8), e1001082.
71. Shepard, D. S., Coudeville, L., Halasa, Y. A., Zambrano, B., & Dayan, G. H. (2011). Economic impact of dengue illness in the Americas. The American journal of tropical medicine and hygiene, 84(2), 200.
72. Tabuti, J. R. (2008). Herbal medicines used in the treatment of malaria in Budiope county, Uganda. Journal of ethnopharmacology, 116(1), 33-42.
73. Thu, H. M., Aye, K. M., & Thein, S. (1998). The effect of temperature and humidity on dengue virus propagation in Aedes aegypti mosquitos. Southeast Asian J Trop Med Public Health, 29(2), 280-284.
74. Tsai, J. F., Chu, T. L., Cuevas Brun, E. H., & Lin, M. H. (2022, January). Solving patient allocation problem during an epidemic dengue fever outbreak by mathematical modelling. In Healthcare (Vol. 10, No. 1, p. 163). MDPI.
75. Wongkoon, S., Jaroensutasinee, M., Jaroensutasinee, K., & Preechaporn, W. (2007). Development sites of Aedes aegypti and Ae. albopictus in Nakhon Si Thammarat, Thailand.
76. World Health Organization. Dengue and severe dengue. (2023). Available online: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue (accessed on 30 Jan 2024).
77. Xiang, J., Hansen, A., Liu, Q., Liu, X., Tong, M. X., Sun, Y., ... & Bi, P. (2017).

Association between dengue fever incidence and meteorological factors in Guangzhou, China, 2005–2014. Environmental research, 153, 17-26.

78. Xu, J., Xu, K., Li, Z., Meng, F., Tu, T., Xu, L., & Liu, Q. (2020). Forecast of dengue cases in 20 Chinese cities based on the deep learning method. International journal of environmental research and public health, 17(2), 453.
79. Yan, S. J., Chughtai, A. A., & Macintyre, C. R. (2017). Utility and potential of rapid epidemic intelligence from internet-based sources. International Journal of Infectious Diseases, 63, 77-87.
80. Yang, L., Li, G., Yang, J., Zhang, T., Du, J., Liu, T., ... & Yang, W. (2023). Deep-learning model for influenza prediction from multisource heterogeneous data in a megacity: Model development and evaluation. Journal of Medical Internet Research, 25, e44238.
81. Zhang, J., & Nawata, K. (2018). Multi-step prediction for influenza outbreak by an adjusted long short-term memory. Epidemiology & Infection, 146(7), 809-816.
82. Zhao, J., Chen, S., Wang, H., Ren, Y., Du, K., Xu, W., ... & Jiang, B. (2012). Quantifying the impacts of socio-economic factors on air quality in Chinese cities from 2000 to 2009. Environmental pollution, 167, 148-154.
83. Zhao, N., Charland, K., Carabali, M., Nsoesie, E. O., Maheu-Giroux, M., Rees, E., ... & Zinszer, K. (2020). Machine learning and dengue forecasting: Comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia. PLoS neglected tropical diseases, 14(9), e0008056.
84. Zheng, X., Wei, C., Qin, P., Guo, J., Yu, Y., Song, F., & Chen, Z. (2014). Characteristics of residential energy consumption in China: Findings from a household survey. Energy Policy, 75, 126-135.