龔千芬、龔嘉德、蘇志民、郝沛毅、林奕儒 (2021),「早期偵測敗血症不良結果:利用深度學習與模糊支持向量機」,*資訊管理學報*,第二十八卷,第四期,頁 445-476。

早期偵測敗血症不良結果:利用深度學習與模糊支持向量機

龔千芬
國立高雄科技大學 智慧商務系

龔嘉德
高雄長庚紀念醫院 急診醫學部

蘇志民
高雄長庚紀念醫院 急診醫學部

郝沛毅*
國立高雄科技大學 智慧商務系

林奕儒
國立高雄科技大學 智慧商務系

摘要

敗血症是全世界的主要死亡原因，敗血性休克的死亡率高達 50%。根據世界衛生組織估計，每年有超過 600 萬人死於敗血症，早期診斷和治療可以預防大多數的敗血性休克發病與死亡，但是，目前缺乏可靠的早期敗血症智能預測系統。時至今日，在大數據分析的快速發展和重症監護室的豐富醫療數據不斷累積的推波助瀾下，讓人們對於開發智能模型來早期預測敗血性休克等急性醫療狀況產生了極大的興趣。

本研究開發一套嶄新的智能敗血症早期預測技術，它包含長短期記憶體，卷積神經網路、完全連接網路與模糊支持向量機，以實現敗血症的早期預測。我們使用 2010 至 2018 年在長庚醫院接受醫療的 17 歲以上病患的電子健康記錄，提取的數據包含靜態特徵集，例如人口統計數據和過去病史；以及動態特徵集，例如帶有時間戳的生命徵象與生物標記。本研究提出一個混合的深度神經網路模型來自動學習關鍵特徵，第一個組件是卷積神經網路，它可以獲得動態信息的局部特徵。第二個組件是完全連接的神經網路，它可以擷取靜態信息內的隱含特徵，

* 本文通訊作者。電子郵件信箱：haupy@nkust.edu.tw
2021/5/3 投稿；2021/6/28 修訂；2021/8/11 接受

此外，長短期記憶體用於補抓動態信息的時間依賴性特徵，最後，深度模型學習得到的特徵將餵入到嶄新的模糊孿生支持向量機中，以預測敗血症不良結果。本研究提出的智能系統可以提前 28 天預測敗血症發作，這將為減輕敗血症發作的風險，提供早期解決方案。

關鍵詞：敗血症早期預測、深度學習、模糊支持向量機、臨床決策支持系統、醫學資訊學

Kung, C.F., Kung, C.T, Su, C.M., Hao, P.Y & Lin, Y.J. (2021). Early detection of sepsis utilizing deep learning and fuzzy support vector machine. *Journal of Information Management*, 28(4), 445-476.

Early detection of sepsis utilizing deep learning and fuzzy support vector machine

Chien-Feng Kung
Department of Intelligent Commerce, National Kaohsiung University of Science and Technology

Chia-Te Kung
Department of Emergency Medicine, Kaohsiung Chang Gung Memorial Hospital

Chih-Min Su
Department of Emergency Medicine, Kaohsiung Chang Gung Memorial Hospital

Pei-Yi Hao *
Department of Intelligent Commerce, National Kaohsiung University of Science and Technology

Yi-Ju Lin
Department of Intelligent Commerce, National Kaohsiung University of Science and Technology

Abstract

Sepsis is a leading cause of death over the world and septic shock, the most severe complication of sepsis, reaches a mortality rate as high as 50%. The World Health Organization estimates that more than six million people die of sepsis annually, and early diagnosis and treatment can prevent most morbidity and mortality. However, reliable and intelligent systems for predicting sepsis are scarce. The rapid development in big data analytics and the data-rich environment of intensive care units has generated great interest in developing models to predict acute medical conditions such as septic shock.

This study proposes a novel technique that combines long short-term memory (LSTM), convolutional neural network (CNN), fully connected neural network, and

* Corresponding author. Email: haupy@nkust.edu.tw

2021/5/3 received; 2021/6/28 revised; 2021/8/11 accepted

fuzzy twin support vector machine to achieve early prediction of sepsis. We used data from the electronic health records (EHRs) of all patients above 17 years old admitted to the medical ICU at Chang-Geng Memorial Hospital between January 2010 and December 2018. The extracted data contained sets of static features, such as demographic and clinical information, and temporal features such as time-stamped vital signs. In this study, we propose a general deep neural network framework that incorporates two additional components with the aim of improving LSTM to automatically extract important features. The first component, a CNN, is added before LSTM to obtain local characteristics of EHRs. The second component, a fully connected neural network, introduces static information (e.g., age) to LSTM. Finally, a LSTM is applied to handle dynamic information (e.g., the lab result). The features learned by the deep learning model are fed to a novel fuzzy twin support vector to predict sepsis onset in patients admitted to an intensive care unit. Using EHRs data, sepsis onset can be predicted up to 28 d in advance. Our findings will offer an early solution for mitigating the risk of sepsis onset.

Keywords: Sepsis early prediction, deep learning, fuzzy support vector machines, clinical decision support systems, medical informatics

壹、前言

敗血症(sepsis)是一種具有很高死亡率和發病率的臨床併發症,根據世界衛生組織估計,在過去十年中,全球每年的敗血症發病率為每 10 萬人中 437 人。此外,每年大約有 3000 萬人(包括 420 萬新生兒與兒童)罹患敗血症,導致 600 萬人死亡。在美國,每年至少有 170 萬人被診斷為敗血症,導致近 27 萬例死亡,超過前列腺癌,乳腺癌和愛滋病的總和(Hatfield et al. 2018)。敗血病最嚴重的併發症是敗血症性休克,其死亡率高達 50%,而且年發病率持續上升。從經濟的角度來看,它是美國醫院治療費用最高的疾病,美國每年與敗血症相關的醫療費用約為 237 億美元,佔全國醫療系統總費用的 6.2%,超過任何其他的疾病(Torio & Moore 2016)。值得注意的是,造成嚴重死亡與經濟負擔的重要原因,是由於不能及時診斷敗血症。先前研究表明,及時識別與給予適當的治療,例如「早期目標導向治療(early goal-directed therapy)」,是對抗敗血症高成本和高死亡率的關鍵(Paoli et al. 2018)。早期診斷和治療可以預防多達 80%的敗血症死亡。相反地,敗血症休克患者給予抗生素治療的時間每延遲 1 個小時,死亡率就會增加 8% (Kumar et al. 2006)。在醫院中,廣泛使用各種基於規則的疾病嚴重度評分系統,以嘗試鑑定敗血症患者。這些評分,例如急性生理和慢性健康評估(Acute Physiology and Chronic Health Evaluation, APACHE II),簡化的急性生理評分(Simplified Acute Physiology Score, SAPS II)和順序器官衰竭評估評分(Sequential Organ Failure Assessment Scores, SOFA),都是臨床手動列出的,但在敗血症診斷中缺乏準確性(Despins 2017)。早期預警評分的一個嚴重局限性,是它們的範圍很廣,並非專門針對敗血症開發的,這意味著許多其他疾病也可能產生很高的評分。這些特異性低的警報,可能會導致高度警報疲勞與醫療體系的崩潰。此外,臨床實踐中使用的預警評分也過於簡單化,它將每個變量分配獨立評分,而忽略不同變量之間的複雜關係及其隨時間的演變(Smith et al. 2013)。

每年,全球有成千上萬的成年人被送入重症監護病房(intensive care units, ICU)。ICU 為每個患者收集並儲存了大量信息,包括生命徵象,實驗室檢查結果和人口統計學詳細信息。機器學習與數據分析技術的快速發展和重症監護室的數據豐富環境,共同為重症監護領域的醫學突破,提供了前所未有的機會。機器學習透過考慮來自患者電子病歷(electronic health records, EHR)的大量輸入變量之間的相關性,來預測感興趣的結果(例如,敗血症),提供了克服基於啟發式警告評分的局限性的可能性(Nemati et al. 2018)。許多經典的機器學習模型已被應用於與敗血症/敗血性休克有關的任務,例如邏輯斯回歸、支持向量機(Gultepe et al. 2014)、隨機森林分類器(Taylor et al. 2016)。用於敗血症檢測的 ML 模型遠遠超出了現有臨床預警系統評分的預測能力(Islam et al. 2019)。最近,Shimabukuro et al. (2017)在一項隨機試驗中,使用 ML 模型進行敗血症檢測,他們證明了幾種積極的效果,病患的住院死亡率降低了 12.4 個百分點($p = 0.018$),平均住院時間從 13.0 天減少至 10.3 天($p = 0.042$)。然而,經典的機器學習模型

的預測效能有很大程度是取決於特徵工程(feature engineering)。由於生理過程的複雜性和醫療事件之間的非線性關係，選擇強大的特徵是一項具有挑戰性的任務。創建、分析、選擇和評估適當特徵的工程過程是既費力又費時。

深度學習技術能夠自動進行特徵提取和選擇過程，從而提升了預測效能，因此，越來越多深度學習模型用於檢測各種疾病(Bhandary et al. 2020)。「深度學習」可以說是類神經網路的品牌重塑，Hinton & Salakhutdinov (2006)在《Science》期刊上的論文闡述：具有多個隱藏層的人工神經網路擁有優異的特徵學習能力，有利於視覺化或分類；換句話說，深度學習乃是透過組合低層特徵，來形成更加抽象的高層特徵，以得到資料的分層特徵表示。因此，「深度模型」是手段，「特徵學習」才是目的。深度學習的關鍵優勢，是具有自動抽取特徵(feature extraction)的能力，它可以取代傳統專家在特徵工程所花費的時間，並且對於圖像、語音這種特徵不明顯的資料類型，能夠在大規模訓練數據上取得優異的效果。由於深度學習的優異特徵學習能力，很適合處理生命徵狀的醫療時間序列類型的資料，並且從中自動學習關鍵特徵，取代傳統專家根據經驗手動建立的特徵，所以近年來最頂尖的敗血症早期診斷預測模型，都是以深度學習為基礎(Lauritsen et al. 2020; Rafiei et al. 2021)。

深度學習的崛起，也取代了支持向量機(support vector machine, SVM)在機器學習領域霸主的地位，但是，還是有一些研究顯示，支持向量機的預測效能是優於深度學習的(Xu et al. 2018)。這樣的結果並不讓人意外，因為支持向量機雖然沒有學習特徵的能力，但是如果在給定良好特徵的情況下，基於統計學習理論創建的支持向量機仍然是最佳的分類器。Huang & LeCun (2006)首度發表結合卷積神經網路(convolutional neural network, CNN)與SVM的方法，文章中提到，CNN很適合學習圖像中位置或方向不變性的特徵(invariant feature)，但是它並不是最適合於分類任務的方法，相反的，SVM在給定良好特徵的條件下，它是最佳分類器，但是SVM無法學習圖像中位置或方向不變性的特徵(invariant feature)。所以，他們使用CNN學習不變性特徵，再將特徵交給SVM建立分類器，獲得正確率明顯地提升。這也是說，人工智慧的大師，CNN的創始人 Yann LeCun也為SVM背書，所以SVM的研究並沒有因為深度學習而落幕，許多學者跟隨著大師的腳步，使用深度學習來擷取特徵，再將該特徵交給SVM來學習分類器，並且應用到各種醫療任務中(Ye et al. 2020)。

雖然，利用深度學習擷取特徵，再將特徵餵給支持向量機學習分類器，可以融合這兩種方法的優點，並且廣泛應用在各種醫療應用的預測任務上(Bhandary et al. 2020; Ye et al. 2020)，然而，在大量的文獻分析與回顧中，發現目前並沒有結合 CNN 與 LSTM 與模糊支持向量機的模型應用在敗血症不良後果預測的研究，這激發本研究的主要動機: 在本研究中，我們將開發嶄新的結合深度學習與模糊支持向量機的敗血症不良結果預測模型，我們將結合 CNN 與 LSTM 當作特徵擷取器，以擷取與時間無關的模式，以及這些模式的時間依賴性與因果關係，並且以嶄新的模糊孿生支持向量機當作最頂層的分類器(Hao et al. 2020)。以統計

學習理論為基礎的支持向量機，擁有優異的推理能力，而模糊理論則適合處理現實中不確定與不精確的模糊訊息。而完美融合兩者優點的模糊支持向量機，特別適合處理醫療資料中大量的雜訊與不精確的資料。醫療數據集經常會遭遇到由於大量缺失值(missing value)導致的樣本不精確與不確定的問題，通常，我們會使用向前/向後填補或是中位數填補來處理缺失值，對於缺失值比率較高的樣本，我們可以指派較低的模糊歸屬權重，使得它們可以允許被分類錯誤，而不會產生過度學習(overfitting)的問題，以免分類模型因為要過度擬合這些擁有大量缺失值的不確定的訓練樣本，而導致沒有缺失值的重要訓練樣本的影響力被稀釋。此外，決策必須考慮到不確定性、風險與信心。通常，決策涉及的風險越高，所需要的信心程度也就越大。決策的目的是降低風險，如果沒有考慮信心程度就制定決策，反而會提高風險。模糊支持向量機不僅可以提供敗血症不良結果的預測，還能提供新進樣本屬於不良結果的歸屬程度(亦即是，對於預測結果的信心程度)，因此，特別適合於敗血症不良結果的早期預警的決策任務中。

貳、文獻探討

Mao et al. (2018)提出基於基於機器學習的敗血症預測算法（InSight），他們僅使用六個生命體徵與梯度增強樹(gradient tree boosting)來預測敗血症，在發病前四個小時，InSight 預測敗血症性休克的 AUC (Area under the ROC Curve)為 0.96，嚴重膿毒症的 AUC 為 0.85。結果顯示，在識別和預測敗血症與敗血症性休克方面，InSight 的表現優於現有的敗血症評分系統。InSight 也是僅使用生命徵象輸入的敗血症篩查系統，其 AUC 第一個超過 0.90。Taneja et al. (2017)使用自適應增強(Adaptive Boosting, adaboost)機器學習技術，將單個血液樣本中的多個生物標誌物測量結果與電子病歷數據（electronic health records, EHR）相結合，以便早期辨識從敗血症早期至高峰期的患者。實驗結果顯示，結合生物標誌物和 EHR 數據可得到 ROC 曲線下面積（AUC）為 0.81，而僅使用 EHR 數據僅可達到 0.75 的 AUC，結果表明，生物標誌物有助於早期識別這些患者。

深度學習是近來熱門的機器學習技術，其中，長短期記憶體（Long-Short Term Memory, LSTM）在各種順序標籤的預測應用任務（例如，氣候變化、醫療記錄、交通監控）中顯示出廣闊的前景，與遞迴神經網絡等其他順序模型相比，LSTM 能夠記憶長時間的依賴性。Saqib, Sha & Wang (2018)提出使用機器學習（machine learning, ML）技術與敗血症患者休克發作前的 24 小時與 36 小時的生命體徵與實驗室結果，來建構敗血症預測模型。他們使用的 ML 模型包括傳統方法（即，隨機森林（random forest, RF）和邏輯回歸（logistic regression LR））以及深度學習技術（即，長短期記憶體（LSTM）），並且使用 MIMIC3（Medical Information Mart for Intensive Care III）的數據集來測試 ML 技術，實驗結果顯示，RF 的效果為最好，他的曲線下面積（AUC-ROC）得分為 0.696，而 LSTM 網路的預測性能不超過 RF。卷積神經網路(convolutional neural network, CNN)在圖像識別領域獲得重大的成功， Kok et al. (2020)提出使用深度時空卷積網路(temporal

convolutional network)的自動診斷工具來預測敗血症。它所使用的數據集為《2019 年心臟病學 PhysioNet 計算》挑戰賽發布的開源數據集。每位患者的記錄包含 40 個特徵：每小時記錄 8 個生命體徵，26 個實驗室測量值和 6 個人口統計學變量。並且使用高斯程序回歸（Gaussian Process Regression, GPR）用於預測每個特徵的可能值的分佈，以緩解缺失值(missing value)的問題，結果顯示，其預測準確率(accuracy)與 AUC 分別為 95.5%和 0.91。

由於 LSTM 與 CNN 各有其優點，將它們組合在一起來建構敗血症預測模型，是近來流行的研究趨勢，Lin et al. (2018)提出一個通用的深度神經網路框架，他第一個組件是長短期記憶體(LSTM)與卷積神經網路（CNN），用於處理動態信息（例如實驗室結果）。第二個組件是完全連接的神經網路，它引入靜態信息（例如年齡）。結果表明，合併 LSTM、CNN 和靜態信息，其預測效能優於原始的 LSTM。Lauritsen et al. (2020)提出一種用於早期檢測敗血膿毒症的深度學習系統，該系統可以從原始醫療事件序列(electronic health record event sequences)數據本身中，學習關鍵因素的特徵和相互作用，而無需依賴勞動強度大的特徵提取工作。他們的敗血症檢測系統是由卷積神經網路(CNN)和長短期記憶網路(LSTM)組成，並且使用來自多家丹麥醫院的數據來建立模型，實驗結果顯示，預測表現範圍從 AUC 0.856（敗血症發作前 3h）到 AUC 0.756（敗血症發作前 24h）。Rafiei et al. (2021)提出一種稱為智能敗血症預測器（Smart Sepsis Predictor, SSP）的新技術，它可用於預測重症監護病房患者的敗血症發作。SSP 是一種深度神經網路結構，包含長短期記憶（long short-term memory, LSTM），卷積層（convolutional layer）和完全連接層，SSP 使用 ICU 數據，包含人口統計數據、生命體徵與實驗室測試結果，SSP 可以提前 12 小時預測敗血症發作，為減輕敗血症發作的風險，提供了早期的解決方案。為了評估 SSP，他們使用 2019 年 PhysioNet / CinC Challenge 數據集，其中包括 40,366 名入住 ICU 的患者的記錄。在敗血症發作前 4 小時、8 小時與 12 小時的 AUC 分別為 0.87、0.86 和 0.84。

上述研究顯示，由於深度網路結構能夠自動學習關鍵特徵，而無需依賴費時費力的特徵工程，因此，最近的研究都是使用深度網路於敗血症不良結果的早期預測，然而，由於醫療資料集的異質結構，同時包含動態訊息與靜態訊息，所以，僅使用單一種深度模型，是無法有效地分析醫療資料內異質且交互作用錯綜複雜的資料。其中，全連結神經網路適合處理靜態資料，CNN 與 LSTM 則是擅長處理動態資料，不過，CNN 適合發掘時間序列中，與時間無關的模式，以圖像辨識為例，CNN 可以找出圖像中具有「鳥腳」的關鍵特徵，進而幫助辨識出圖像中的物件為「鳥」，而且不論「鳥腳」這個關鍵特徵出現在圖像中哪個位置。而 LSTM 則是適合分析時間序列中，出現模式的時間依賴關係。唯有混合這些模型，才能全面地分析醫療異質資料，找出預測敗血症不良結果的蛛絲馬跡。因此，本研究將結合 CNN、LSTM 與全連結網路，以自動學習關鍵特徵，並且將它餵入到頂層的模糊支持向量機，來早期預測敗血症的不良反應。

參、研究方法

敗血症(Sepsis)是一種當病原體入侵血液循環之後，引起全身性發炎反應並可能造成後續的器官多功能障礙或敗血性休克的疾病，重度敗血症的死亡率甚至高達 50%；目前已有臨床證據表示，敗血症的病患在病況惡化之前的一段時間，其生命徵象便會反應出來，使醫師們可以及早對敗血症進行處置，進而降低病患的死亡率，但由於生命徵象的變化非常細微，以至於醫師們有時會難以察覺而導致錯過早期治療的時間；此外，由於醫師在進行診斷時，並不會只採用單面向的資料來做診斷，而是會綜合評估病患的各種資訊(如：過去病史、血液檢測報告等)後才進行對應的處置，但是，結合後的龐大的病患資訊，讓醫師不容易在短時間之內就可以掌握這些龐大資料中的脈絡，進而造成一些診斷上的盲點。在急診室中，醫生會根據大量複雜的生理和臨床數據做出挽救生命的決定，同時還要在高度不確定性和嚴格的時間限制下進行操作。由於病患在進入急診室之後，通常 3~6 小時之內醫師便會決定該病患是要進入觀察室或者是離院，因此，本研究主要以病患進入急診後 3 小時之整合性資料，預測敗血症病患在住院期間是否會發生不良結果;本研究將長庚醫院 8 院區 2010~2018 年的資料處理成一個整合性資料（生命徵象、血液檢測報告、過去病史與其他靜態資料）後，結合卷積神經網路、長短期記憶與全連結神經網路的複合式神經網路模型來擷取關鍵特徵，並且透過嶄新的模糊孿生支持向量機來預測敗血症病患是否會在 28 天內出現不良結果。

一、資料蒐集

本研究原始資料來源為長庚醫院 8 個院區從西元 2010 年至 2018 年病患資料，其中會先進行初步的篩選，以篩選出有敗血症的病患。篩選方式為病患同時符合以下條件：

- 有進行血液培養● 有施打抗生素
- 17 歲以上 ● 入院 28 天

本研究的病患族群為急診醫師懷疑有敗血症的病人，由於敗血症是一個動態變化的疾病，在急診醫師有限的資料下，一般會以臨床判斷作為族群匡列的依據，而使用的方法則以病患有接受血液培養測試、接受抗生素的治療及辦理住院來認定，另外，排除了小兒及青少年的病患，因此只使用超過 17 歲的病患資料(長庚醫院是以 17 歲來區分成人科及小兒科)，除此之外，並無其他的排除條件。本研究使用長庚醫院研究資料庫的「去連結病患資料」，並且通過人體試驗委員會(IRB)同意於「人工智慧技術在急診感染症的鑑別診斷及預後風險評估之決策輔助系統運用」的研究使用，IRB 通過案號為 201801713B0，篩選後，有不良結果及無不良結果之樣本數量如表 1 所示。其中可發現，無不良結果的樣本數量，佔所有樣本的比例高達 80%，有著嚴重的資料不平衡的問題。

表 1: 樣本數量(蒐集資料時間為 2010～2018)

	數量	佔總體比例
有不良結果	70528	20%
死亡	6048	2%
呼吸衰竭	15350	4%
休克	20151	6%
轉入 ICU	28979	8%
無不良結果	295797	80%
總計	366325	100%

二、特徵選擇和數據處理

敗血症是多方面的，並且表現出許多初步的體徵，例如引起極度發燒，血壓下降和心率飆升的症狀(Fairchild & O'Shea 2010)。目前，醫生們根據這些症狀以及專門的生物標記物進行診斷和預測。然而，敗血膿毒症的複雜性和不同身體系統中大量器官功能障礙，可能會導致每種情況下的分子生物學標誌物不同。例如，急性期蛋白與多種類型的炎症有關，包括敗血症，但由於它亦與類似的炎症併發症有關，所以僅使用急性期蛋白，對於敗血症的預測並不準確，唯有結合多個生物標誌物與生命徵象，才能確保正確的預測(Pierrakos & Vincent 2010)。因此，為了建構敗血症不良結果預測工具，本研究提取的資料包含靜態特徵集，例如人口統計、臨床信息與過去病史，以及動態特徵集，例如帶有時間戳的生命徵象與生物標記。靜態特徵（例如，性別）每次訪問僅收集一次，並且在訪問期間保持不變； 而動態特徵（例如，生命徵象或實驗室血液測試）是在患者就診期間多次收集的。一般而言，靜態信息用向量表示，而動態信息用時間序列表示。在本研究中，我們蒐集的靜態特徵包含二個描述性特徵（性別與年齡）與過去病史，動態特徵則包含四個生命徵象：收縮壓（SBP），舒張壓（DBP），心律（HR），呼吸速率（RR）和體溫（BT）以及六種生物標記：肌酸酐、乳酸、C-反應蛋白、白血球計數、多型核白血球與淋巴球，底下分別介紹這些特徵的處理方式。

(一)生命徵象(Vital Signs)

生命徵象(Vital Signs)亦稱為生命跡象，係指一組 4~6 個對維持人體生命最重要的表徵，這些徵狀的測量結果除了提供潛在疾病的線索外，也可用於評估病患的恢復程度。生命徵象包含體溫(Blood Temperature, BT)、 血壓(Blood Pressure, BP)、心律(Heart Rate, HR)與呼吸速率(Respiration Rate, RR)這四種，如圖 1 所示。

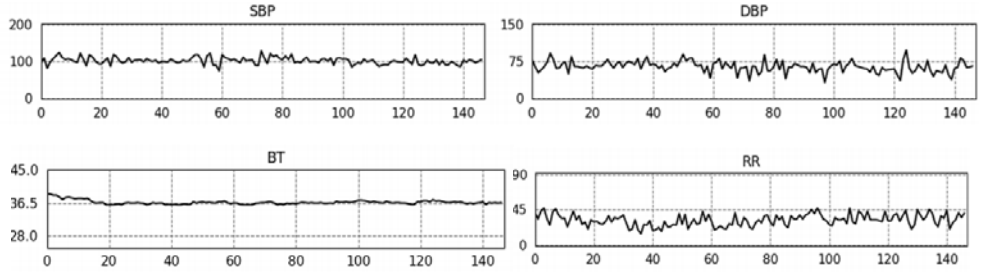

圖 1: 生命徵象

本研究選擇生命徵象作為客觀的預測變量，因為無論臨床情況如何，生命徵象都是常規且頻繁地從所有患者中收集的，並且這些值很少受到檢查者的影響。在本研究中，每位病患的生命徵象數據，包含住院期間內 BT、HR、NBP D、NBP S 及 RR 的所有測量數據，若是病患有缺少任何一種生命徵象的測量，該病患便不會納入樣本之中。表二顯示生命徵象的預處理步驟，首先，我們對遺失值進行填補(遺失值填補策略在後續章節會詳細介紹)，再使用最大最小標準化對生命徵象進行最終處理，以消除資料內部變異範圍不同所造成的影響。

表 2: 生命徵象預處理過程

病患編號	測量日期	測量時間	BT	HR	NBPD	NBPS	RR
A0001	20200101	0800	37	80	200	100	30
A0001	20200101	1600	38	60	180	110	40
A0001	20200102	0000	（遺漏）	90	210	80	20

⇩ （遺漏值填補）

病患編號	測量日期	測量時間	BT	HR	NBPD	NBPS	RR
A0001	20200101	0800	37	80	200	100	30
A0001	20200101	1600	38	60	180	110	40
A0001	20200102	0000	38	90	210	80	20

⇩ （最大最小標準化）

病患編號	測量日期	測量時間	BT	HR	NBPD	NBPS	RR
A0001	20200101	0800	0	0.67	0.67	0.67	0.5
A0001	20200101	1600	1	0	0	1	1
A0001	20200102	0000	1	1	1	0	0

雖然生命徵象對於敗血症的預測是一個很重要的因素，但是，目前研究結果也表明，要確保模型在預測敗血症有高靈敏度的情況下，幾乎都會造成許多的偽陽性，因此，本研究除了會使用生命徵象之外，也會使用其他如生物標誌與過去病史等數據，藉以確保模型在保有高靈敏度的情況下，盡可能的減少偽陽性的數量，使模型在臨床上可以更為實用。

(二)生物標記(Biomarker)

生物標記又稱為生物指標或生物標誌物，在醫學上通常是指血液中的某種蛋白質，藉由測量該蛋白質的數據，便可以反應出病患是否發生某種疾病，或者是顯示某疾病的嚴重程度。本研究採用下列六種臨床上醫師們較常使用的生物標記，來做為本研究整合性資料的來源:

- **肌酸酐**(Creatinine): 又稱肌酐，為肌酸和磷酸肌酸代謝的最終產物。Froon et al. (1994) 的研究中，利用傳統統計證明了因為敗血症而死亡的病患，其肌酸酐濃度明顯比倖存者更高，而 Doi et al. (2009) 的研究中，他利用統計的方式證明在敗血症病患之中肌酸酐的變化，與病患不良預後結果有一定程度的關係。
- **乳酸 (Lactate)**: 它是身體組織灌流不足而導致無氧呼吸的產物，可用於檢查疾病的嚴重程度，在 2016 年的敗血症治療處理指南中指出，乳酸的升高會導致病患的不良結果，Nguyen et al. (2004)使用邏輯斯回歸證實，清除乳酸與死亡率有顯著的反比關係。

- **C-反應蛋白(C-Reactive Protein, CRP):** 它是人體肝臟細胞所產生的特殊蛋白，可作為發炎反應的一種指標。當病患發生了急性發炎反應時，CRP 會經由肝臟大量分泌而迅速上升，以協助其他分子和受傷組織、細胞核內抗原與病原體進行結合。但是，因為 CRP 檢查特異度較低，因此，通常用做篩檢和監測組織的損傷。
- **白血球計數(White Blood Count, WBC):** 當身體出現發炎反應或遭受感染時，白血球的數量便會增加。在臨床意義上，WBC 若是過高的話，表示身體有出現因細菌感染所造成的發炎反應，或是患上白血病，反之則是有可能因為病毒感染、肝硬化、造血功能障礙或因藥物副作用而造成 WBC 過低。
- **多形核白血球 (Segment):** 係指白血球中的嗜中性球，也是人體中含量最高的一種白血球。當身體發生感染之後，嗜中性球便會透過游移的方式迅速穿過血管內皮細胞進入感染部位，對入侵的病原體進行吞噬。由於許多生理、病理因素，皆可能會造成嗜中性球的數量變化，故在臨床上除了要測量 WBC 之外，同時也會測量其他各血球的數量。
- **淋巴球 (Lymphocyte):** 它是白血球中體積最小的一種，是一種具有免疫識別功能的細胞系。淋巴球為白血球中佔比第二多的，故當有感染發生時，淋巴球的數量也會產生較大的變化，因此本研究便將 WBC、Segment 與 Lymphocyte 這三種血液檢測數據一同納入整合性的資料之中。

由於有些病患並不會對每一項生物標記都進行測量，而且，通常也不會持續的進行檢測，因此，只要病患於該次入院，有進行過上述六種生物標誌的任一種的檢測，便會將該病患納入樣本集合，反之則不會納入。

(三)動態特徵的量測時間間距與遺失值處理策略

在實際臨床狀況中，生命徵象與生物標記量測的時間間距並不一樣，生命徵象是常規且頻繁地從所有患者中收集的，ICU 患者的大多數生命徵象至少每小時監測一次，並且通過監測設備自動測量，並將這些值記錄在電子病歷中。相反地，生物標記則通常不是常規且週期地量測，而且也不是每個小時量測一次。為了能將這兩種不同類型的動態資料共同餵入我們的預測模型進行分析與訓練，在資料預處理階段，本研究將它們的時間間距設定為一致，每筆動態資料的時間總長度是 3 小時，每筆動態資料的時間間距設定為 1 小時，也由於實際的生命徵象與生物標記的量測的時間間距並不一樣，所以會有缺失值的問題。為了解決缺失值的問題，參考相關研究文獻的策略(Rafiei et al. 2021; Saqib et al. 2018)，本研究在每一個時間間距(1 小時)中，如果該動態特徵有量測多次，則我們取用平均值(average values)來做為該時段的代表性數值，如果在某個時間間距出現缺失值，則我們先使用前項填補(forward-fill)的策略，將前面最近的值填補至隨後時段的缺失值(根據前後時段是因果程序的假設)，最後，我們使用後項填補(backward-fill)的策略，來填補其他的缺失值。如果某個動態特徵在整個時段(3 個小時)都沒有量測值(例如，對於生物標記特徵，很少有病患會同時量測那 6 種生物標記，所

以我們只要病患有至少量測 1 種生物標記，就把他納入到樣本集合中)，則是對於沒有量測值的那個動態特徵，採取中位數填補(median-fill)策略，這是因為缺少該項動態特徵量測值的原因，很可能是由於醫生認為該特徵對患者來說是正常的，並且無須量測與記錄。

由於在實際臨床狀況下，生命徵象與生物標記量測的時間間距，並不是本研究能夠操弄的變項，所以本研究無法分析實際量測時的時間間距長短對於預測效能之靈敏度的影響，不過在預先實驗中，本研究有嘗試在資料預處理階段時，將處理後的時間間距設定為 0.5 小時、1 小時、2 小時與 3 小時，結果顯示將時間間距設定為 1 小時的預測效果最好，這是因為將時間間距設定為 1 小時，所提供的訊息量比時間間距為 2 或 3 小時更豐富，而時間間距設定為 0.5 小時，雖然提供訊息量更多，但都是重複的訊息，所以預測效能反而會下降。

(四)過去病史

病患的過去疾病史也提供了預測敗血症不良反應的蛛絲馬跡，但是，由於病患的過去病史分散於醫院原始資料的各檔案之中，因此本研究會將急診、住院與門診的病患過去病史資料個別進行處理，並將屬於同類型疾病的 ICD9 與 ICD10 整合為同一種代號。本研究的過去病史主要以肝病、偏癱、充血性心力衰竭、周邊血管疾病(中風)、失智症、慢性阻塞肺病、 結蹄組織疾病、糖尿病、腎臟疾病、惡性腫瘤、白血病及愛滋病共 12 種慢性疾病為主，其餘疾病則暫不列入本研究之病患過去病史紀錄。整合後的過去病史表如表 3 所示，其中，1 表示該病患患有該疾病，0 表示病患無該疾病。

表 3: 過去病史表

病患編號	紀錄日期	紀錄時間	肝病	腎臟病	糖尿病		愛滋病
A0001	20200101	0800	1	0	0		0
B0001	20200101	0800	1	1	0	…	0
B0001	20200201	0800	1	1	1		0
C0001	20200101	0800	1	1	1		0

三、結合 LSTM、CNN 與全連結層之神經網路模型來擷取關鍵特徵

長短期記憶體(Long short-term memory, LSTM)是一種循環神經網路，他是專門設計用於避免深度循環網路中梯度消失和梯度爆炸的問題。 LSTM 使網路能夠將隱藏狀態的先前信息，保留在內部儲存器。因此，它特別適用於處理事件之間存在長期時間依賴性的任務。LSTM 網路的架構如圖 2 所示。它是由儲存單元狀態(memory cell state)，表示為 C_t，與以下三個閘門構成：遺忘閘門 $f_t \in [0, 1]$，輸入閘門 $i_t \in [0, 1]$，以及輸出閘門 $o_t \in [0, 1]$。這三個閘門會相互影響，以控制信息的流動。在訓練過程中，網路會學習記住什麼，以及何時允許讀/寫，以最大程度地減少分類錯誤。更具體地說，「忘記」閘門決定來自先前儲存單元狀態的哪些信息已過期，並且應該予以刪除；輸入閘門從候選儲存單元狀態 C_t^* 中選擇信息，以更新單元狀態；輸出閘門會過濾儲存單元中的信息，因此模型僅考慮與預測任務相關的信息。

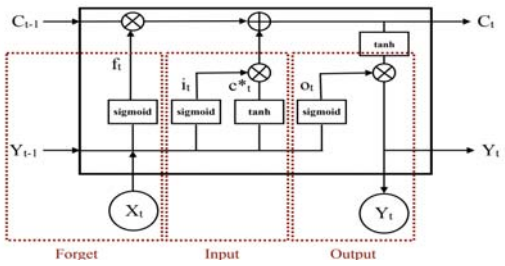

圖 2: LSTM 網路，圖片來源(Lin et al. 2018)

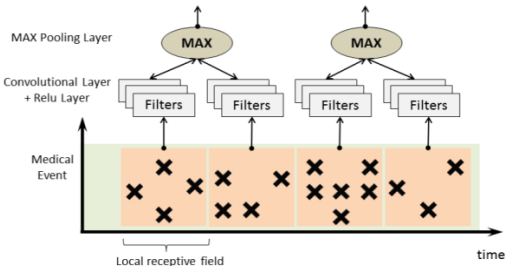

圖 3: 卷積神經網路從動態特徵中擷取狀態表示

每個閘門的數值的計算如下，其中 $W_{[i,f,C,o]}$是權重矩陣，$b_{[i,f,C,o]}$是偏差向量：

$$i_t = sigmoid(W_i \cdot [y_{t-1}, X_t] + b_i), \quad f_t = sigmoid(W_f \cdot [y_{t-1}, X_t] + b_f)$$
$$C_t^* = \tanh(W_C \cdot [y_{t-1}, X_t] + b_C), \quad o_t = sigmoid(W_o \cdot [y_{t-1}, X_t] + b_o) \quad (1)$$

LSTM 模組中的儲存單元值 C_t和輸出標籤 y_t則是使用以下公式計算

$$C_t = C_{t-1} \cdot f_t + C_t^* \cdot i_t, \qquad y_t = o_t * \tanh(C_t) \quad (2)$$

對於我們的任務，只有在最後一個時間步伐 T 的輸出標籤，是我們做出的預測。LSTM 可以有效地捕獲時間序列數據中的潛在時間結構，因此特別適合於對電子病歷（electronic health records, EHR）中的動態信息進行建模，在 EHR 中，長時間間隔內醫療事件之間，存在很強的統計依賴性。而這種依賴性，是識別生理惡化的早期跡象的關鍵特徵。

卷積神經網路(Convolutional Neural Network, CNN)已經在許多與圖像/視頻相關的任務中取得了顯著成果，因為它可以有效地捕獲短時間區域內的「時不變(time-invariant)」特徵 (這些特徵是很難被 LSTM 捕抓到的)。這些特徵是時不變(time-invariant)的，因為它們可以在任何時間點出現，並且可以從部分記錄中提取出來，而不是依賴於全部記錄。基於合理的假設，如果具有某個模式存在於某個時間位置，那麼，它也可能存在於其他時間位置。因此，在 LSTM 之前引入 CNN，可以將原始 EHR 的局部特徵提取為緊湊狀態表示形式，以便 LSTM 可以更輕鬆地識別時間依賴性。本研究採用如圖 3 所示的 CNN 模組，來提取局部與時間不變的依賴性。

CNN 的輸入是表示為時間序列的動態事件信息；CNN 的輸出則是緊湊的嵌入(embedding)狀態表示向量，並將在下一步中，將此狀態向量饋入到 LSTM 模組中。卷積層採用了一組卷積濾波器，它將神經元僅連接到時間區域中的一個局部範圍，濾波器連接的時間局部範圍稱為局部接收場。圖三顯示一個具有大小為 3 的局部接收場的卷積層的範例。濾波器執行一組計算以確定該區域是否顯示出特定模式(pattern)。然後，濾波器沿序列的時間方向移動，產生指示出不同模式發生位置的輸出序列。令 $\mathbf{x}_{i:i+j}$ 表示事件時間串列中 $x_i, x_{i+1}, \ldots, x_{i+j}$ 的子串列，卷積運算會使用一個濾波器 $\mathbf{w} \in \mathrm{R}^{h \times k}$，將它應用到一個內含 h 個事件的窗口中(h 即為局部接收場的大小，k 是事件 x_i 的維度)，以產生一個新的特徵。例如，特徵 c_i 是將濾波器套用到時間窗口 $\mathbf{x}_{i:i+h-1}$ 而生成的，計算公式為

$$c_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+h-1} + b) \quad (3)$$

其中，b 是一個偏差項，f 是一個非線性激發函數，例如 tanh 或 relu。將濾波器 $\mathbf{w} \in \mathrm{R}^{h \times k}$ 應用於事件串列 $\{\mathbf{x}_{1:h}, \mathbf{x}_{2:h+1}, \ldots, \mathbf{x}_{n-h+1:n}\}$ 後，將產生一個特徵圖 $\mathbf{c} = [c_1, c_2, \ldots, c_{n-h+1}]$。最後，池化層透過取平均值或使用最小/最大運算來執行時間採樣，從而降低了模型對於偏移(shifts)和扭曲(distortion)的敏感性。請注意，CNN 可以通過倒傳遞學習(backpropagation)來自動擷取特徵。而 CNN 輸出的狀態表示向量可以明確地表達局部特徵，因此，使得 LSTM 可以更輕鬆地識別其時間依賴性與模式。

雖然 CNN 與 LSTM 適合處理動態的時間序列資料，但是卻不適用於諸如描述性特徵（年齡和性別）與過去病史的靜態資料，相反地，全連結神經網路 (fully connected neural network)更適合處理靜態資料，因此，在本研究中，我們使用圖 4 的架構，我們使用全連結神經網路處理靜態信息，並且與 LSTM 最後一個時間步伐的輸出向量串聯在一起，連接而成向量即視為透過混合深度網路自動學習到的關鍵特徵向量，並將該特徵向量餵入到最頂層模糊孿生支持向量機，以用於做出敗血性不良結果的最終預測。

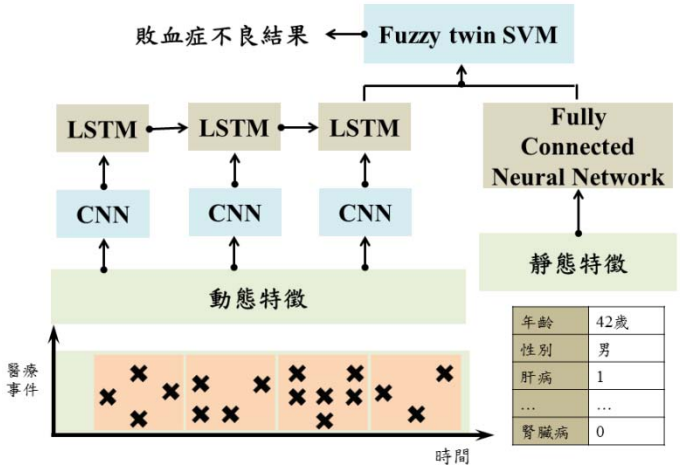

圖 4: 結合 CNN、LSTM 與 FCNN 來從動態與靜態特徵中擷取關鍵特徵

四、運用模糊孿生支持向量機作為最頂層的分類器

在本研究中，我們將以 Hao et al. (2020)提出模糊孿生支持向量機作為最頂層的分類器，以判別敗血症患者的不良結果。模糊孿生支持向量機(fuzzy twin support vector machine)將模糊集合理論融合至孿生支持向量機(twin SVM)當中，並且完美結合支持向量機與模糊理論的優點，其中，基於統計學習理論的支持向量機擁有極佳的推理能力，而模糊理論則是擅長處理現實世界中複雜與不精確的資訊。為了建構模糊孿生支持向量機，我們需要使用下述定理:

定理 1: 令 $X=(m,c)$ 代表一個對稱三角形模糊數，其中心點為 m 與模糊範圍半徑為 c，根據定理 1，對任何三角形模糊數 $A=(m_A,c_A)$ 與 $B=(m_B,c_B)$，我們有

$$A \underset{f}{\geq} B \quad \text{iff} \quad m_A + c_A \geq m_B + c_B \quad \text{與} \quad m_A - c_A \geq m_B - c_B. \quad (4)$$

其中"$\underset{f}{\geq}$" 代表*模糊大於*(fuzzy larger than)。

定理 2: 給定一個模糊權重向量 $\widetilde{\mathbf{W}}=(\mathbf{w},\mathbf{c})$ 與模糊偏移量 $\widetilde{\mathbf{B}}=(b,d)$，其中，模糊權重向量 $\widetilde{\mathbf{W}}$ 內的每一項 $\widetilde{\mathbf{W}}_i=(w_i,c_i)$ 是模糊數(fuzzy numbers)。它可以表示為向量形式 $\mathbf{w}=[w_1,\ldots,w_n]^T$ 與 $\mathbf{c}=[c_1,\ldots,c_n]^T$，代表 "近似於 **w**"，其中 **w** 為中心點且 **c** 為模糊半徑。同樣地， $\widetilde{\mathbf{B}}=(b,d)$ 是模糊偏移量，代表"近似於 b"，其中 b 是中心點且 d 是模糊半徑。給定一筆資料向量 $\mathbf{x}=[x_1,\ldots,x_n]^T$，模糊超平面

$$\widetilde{Y}=\widetilde{\mathbf{W}}_1x_1+\cdots\widetilde{\mathbf{W}}_nx_n+\widetilde{\mathbf{B}}=\widetilde{\mathbf{W}}^T\mathbf{x}+\widetilde{\mathbf{B}}, \quad (5)$$

是由下列歸屬函數(membership function)來定義:

$$\mu_{\widetilde{Y}}(y)=\begin{cases}1-\dfrac{\left|y-\left(\mathbf{w}^T\mathbf{x}+b\right)\right|}{\mathbf{c}^T\mathbf{x}+d} & \mathbf{x}\neq 0\\ 1 & \mathbf{x}=0, y=0\\ 0 & \mathbf{x}=0, y\neq 0\end{cases} \quad (6)$$

其中 $\mu_Y(y)=0$ 當 $\mathbf{c}^T\mathbf{x}+d\leq\left|y-\left(\mathbf{w}^T\mathbf{x}+b\right)\right|$ 時， $|\cdot|$ 代表取絕對值。

(一)求解最佳化問題

考慮一組資料樣本集合 T= $\{(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),\ldots,(\mathbf{x}_N,y_N)\}$，其中 $\mathbf{x}_i\in R^n$ 是第 i 筆資料向量， $y_i\in\{-1,1\}$則是第 i 筆資料所對應的類別，而 $i=1,\ldots,N$。假設屬於正類別的樣本數目有 Np 筆，這 Np 筆正類別資料向量可以用矩陣 $\mathbf{A}\in R^{Np\times n}$ 表示，同樣地，假設屬於負類別的樣本數目有 Nn 筆，這 Nn 筆負類別資料向量可以用矩陣 $\mathbf{B}\in R^{Nn\times n}$ 表示。我們的模糊孿生支持向量機(fuzzy twin SVM)尋找兩個模糊超平面 $\widetilde{\mathbf{W}}_1^T\mathbf{x}+\widetilde{\mathbf{B}}_1=\Theta$ 與 $\widetilde{\mathbf{W}}_2^T\mathbf{x}+\widetilde{\mathbf{B}}_2=\Theta$，一個類別對應一個模糊超平面，使得每一個超平面能夠與對應的類別的樣本點能夠距離越近越好，同時要與另一個類別的樣本點能夠距離越遠越好，其中， $\widetilde{\mathbf{W}}_1=(\mathbf{w}_1,\mathbf{c}_1)$ 與 $\widetilde{\mathbf{W}}_2=(\mathbf{w}_2,\mathbf{c}_2)$ 分別是這兩個超平面的模糊權重向量，而 $\widetilde{\mathbf{B}}_1=(b_1,d_1)$ 與 $\widetilde{\mathbf{B}}_2=(b_2,d_2)$ 分別是這兩個超平面的模糊偏移量。Θ 表示"模糊零(fuzzy zero)"，它也是一個三角形模糊數，以零為中心點，而模糊半徑為 O_w。要延伸到非線性的模糊分類器，我們可以使用核心函數技巧，先將樣本點經由非線性轉換映射到一個高維度特徵空間(feature space)，然後在特徵空間估計最佳的模糊超平面，它對應到原來空間就是最佳的模糊曲面(fuzzy surface)，考慮下列由核心函數定義的模糊曲面

$$K(\mathbf{x}^T,\mathbf{C}^T)\widetilde{\mathbf{W}}_1+\widetilde{\mathbf{B}}_1=\Theta \quad \text{與} K(\mathbf{x}^T,\mathbf{C}^T)\widetilde{\mathbf{W}}_2+\widetilde{\mathbf{B}}_2=\Theta, \quad (7)$$

其中 $\mathbf{C}^T=[\mathbf{A}\ \mathbf{B}]^T$ 與 K 是適合的核心函數。要估計一對最佳的模糊曲面，我們可以求解下列最佳化問題:

$$\begin{aligned}\min\quad & \frac{1}{2}\left\|K(\mathbf{A},\mathbf{C}^T)\mathbf{w}_1+\mathbf{e}_1b_1\right\|^2+\frac{M_1}{2}\left(\left\|\mathbf{w}_1\right\|^2+\left\|\mathbf{c}_1\right\|^2+b_1^2+d_1^2\right)+C_1\mathbf{s}_2^T(\xi_{11}+\xi_{12})\\ s.t.\quad & -[(K(\mathbf{B},\mathbf{C}^T)\mathbf{w}_1+\mathbf{e}_2b_1)+(K(\mathbf{B},\mathbf{C}^T)\mathbf{c}_1+\mathbf{e}_2d_1)]+\xi_{11}\ge\mathbf{e}_2+\mathbf{e}_2\mathbf{I}_w,\\ & -[(K(\mathbf{B},\mathbf{C}^T)\mathbf{w}_1+\mathbf{e}_2b_1)-(K(\mathbf{B},\mathbf{C}^T)\mathbf{c}_1+\mathbf{e}_2d_1)]+\xi_{11}\ge\mathbf{e}_2-\mathbf{e}_2\mathbf{I}_w,\ \xi_{11},\xi_{12}\ge0\end{aligned}\tag{8}$$

與

$$\begin{aligned}\min\quad & \frac{1}{2}\left\|K(\mathbf{B},\mathbf{C}^T)\mathbf{w}_2+\mathbf{e}_2b_2\right\|^2+\frac{M_2}{2}\left(\left\|\mathbf{w}_2\right\|^2+\left\|\mathbf{c}_2\right\|^2+b_2^2+d_2^2\right)+C_2\mathbf{s}_1^T(\xi_{21}+\xi_{22})\\ s.t.\quad & [(K(\mathbf{A},\mathbf{C}^T)\mathbf{w}_2+\mathbf{e}_1b_2)+(K(\mathbf{A},\mathbf{C}^T)\mathbf{c}_2+\mathbf{e}_1d_2)]+\xi_{21}\ge\mathbf{e}_1+\mathbf{e}_1\mathbf{I}_w,\\ & [(K(\mathbf{A},\mathbf{C}^T)\mathbf{w}_2+\mathbf{e}_1b_2)-(K(\mathbf{A},\mathbf{C}^T)\mathbf{c}_2+\mathbf{e}_1d_2)]+\xi_{22}\ge\mathbf{e}_1-\mathbf{e}_1\mathbf{I}_w,\ \xi_{21},\xi_{22}\ge0\end{aligned}\tag{9}$$

QPP(8)與(9)的目標函數的第一項是模糊超平面到該對應類別(例如，類別 1)的樣本點的距離的總和。因此，最小化該項等同於確保模糊超平面與對應的類別越接近越好，最小化 QPP(8)與(9)的目標函數的第二項則是實現結構風險最小化(structure risk minimization)的精神。限制條件則是要求超平面與另一個類別(例如，類別 2)的樣本點的模糊距離大於模糊壹 1(根據定理 1 與 2)，其中模糊壹(fuzzy one)，它亦是一個三角形模糊數，以數字 1 為中心點，模糊半徑為 I_W，違反限制條件的誤差將被差額變數(slack variables) $\xi_{11}, \xi_{12}, \xi_{21}$ 與 ξ_{22} 量測，而參數 C_1 與 C_2 是用戶設定的懲罰參數，並在 QPP(8)與(9)的目標函數的第三項懲罰這些誤差(被差額變數所量測)。每一個訓練樣本都指派一個模糊歸屬度，代表該筆訓練樣本屬於對應類別的強度，向量 $\mathbf{s}_1$ 與 $\mathbf{s}_2$ 代表正類別與負類別的樣本點的模糊歸屬度所組成的向量，並在 QPP(8)與(9)的目標函數的第三項調控對於誤差的懲罰的強度，對於較為不確定的樣本點(例如，雜訊)，則指派較低的模糊歸屬程度，來減少它被分類錯誤時的懲罰強度。使用 Lagrangian 理論，我們得到下列對偶最佳化問題:

$$\begin{aligned}\max\begin{cases}\dfrac{-1}{2}(\boldsymbol{\alpha}_{11}+\boldsymbol{\alpha}_{12})^T\mathbf{R}(\mathbf{S}^T\mathbf{S}+M_1\mathbf{I})^{-1}\mathbf{R}^T(\boldsymbol{\alpha}_{11}+\boldsymbol{\alpha}_{12})\\ \dfrac{-1}{2M_1}(\boldsymbol{\alpha}_{11}-\boldsymbol{\alpha}_{12})^T\mathbf{R}\mathbf{R}^T(\boldsymbol{\alpha}_{11}-\boldsymbol{\alpha}_{12})+(1+\mathbf{I}_w)\boldsymbol{\alpha}_{11}^T\mathbf{e}_2+(1-\mathbf{I}_w)\boldsymbol{\alpha}_{12}^T\mathbf{e}_2\end{cases}\\ s.t.\quad \mathbf{0}\le\boldsymbol{\alpha}_{11},\boldsymbol{\alpha}_{12}\le C_1\mathbf{s}_2\end{aligned}\tag{10}$$

與

$$\begin{aligned}\max\begin{cases}\dfrac{-1}{2}(\boldsymbol{\alpha}_{21}+\boldsymbol{\alpha}_{22})^T\mathbf{S}(\mathbf{R}^T\mathbf{R}+M_1\mathbf{I})^{-1}\mathbf{S}^T(\boldsymbol{\alpha}_{21}+\boldsymbol{\alpha}_{22})\\ \dfrac{-1}{2M_1}(\boldsymbol{\alpha}_{21}-\boldsymbol{\alpha}_{22})^T\mathbf{S}\mathbf{S}^T(\boldsymbol{\alpha}_{21}-\boldsymbol{\alpha}_{22})+(1+\mathbf{I}_w)\boldsymbol{\alpha}_{21}^T\mathbf{e}_1+(1-\mathbf{I}_w)\boldsymbol{\alpha}_{22}^T\mathbf{e}_1\end{cases}\\ s.t.\quad \mathbf{0}\le\boldsymbol{\alpha}_{21},\boldsymbol{\alpha}_{22}\le C_1\mathbf{s}_2\end{aligned}\tag{11}$$

其中 $\mathbf{S}=\left[K(\mathbf{A},\mathbf{C}^T)\quad \mathbf{e}_1\right], \mathbf{R}=\left[K(\mathbf{B},\mathbf{C}^T)\quad \mathbf{e}_2\right]$，且

$$\mathbf{u}_1=[\mathbf{w}_1^T\ b_1]^T=-(\mathbf{S}^T\mathbf{S}+M_1\mathbf{I})^{-1}\mathbf{R}^T(\boldsymbol{\alpha}_{11}+\boldsymbol{\alpha}_{12})\qquad \mathbf{v}_1=[\mathbf{c}_1^T\ d_1]^T=\frac{-1}{M_1}\mathbf{R}^T(\boldsymbol{\alpha}_{11}-\boldsymbol{\alpha}_{12})\tag{12}$$

$$\mathbf{u}_2=[\mathbf{w}_2^T\ b_2]^T=(\mathbf{R}^T\mathbf{R}+M_2\mathbf{I})^{-1}\mathbf{S}^T(\boldsymbol{\alpha}_{21}+\boldsymbol{\alpha}_{22})\quad \mathbf{v}_2=[\mathbf{c}_2^T\ d_2]^T=\frac{1}{M_2}\mathbf{S}^T(\boldsymbol{\alpha}_{21}-\boldsymbol{\alpha}_{22}) \quad (13)$$

求解 QPPs (10) 與 (11)，我們得到模糊權重向量 $\widetilde{\mathbf{W}}_1=(\mathbf{w}_1,\mathbf{c}_1)$ 與 $\widetilde{\mathbf{W}}_2=(\mathbf{w}_2,\mathbf{c}_2)$，模糊偏移量 $\widetilde{\mathbf{B}}_1=(b_1,d_1)$ 與 $\widetilde{\mathbf{B}}_2=(b_2,d_2)$，一旦得到模糊曲面，我們根據新進樣本點 $\mathbf{x}\in R^n$ 與哪一個模糊曲面最接近，來決定樣本點 $\mathbf{x}\in R^n$ 是屬於哪一個類別，對於任意樣本點 $\mathbf{x}_i$ ，它與模糊曲面 $K(\mathbf{x}^T,\mathbf{C}^T)\widetilde{\mathbf{W}}_1+\widetilde{\mathbf{B}}_1=\Theta$ 的距離也是一個對稱三角形模糊數，其中心點為 $|K(\mathbf{x}_i^T,\mathbf{C}^T)\mathbf{w}_1+b_1|$ 與模糊半徑為 $K(\mathbf{x}_i^T,\mathbf{C}^T)\mathbf{c}_1+d_1$。樣本點 $\mathbf{x}_i$ 與模糊曲面 $K(\mathbf{x}^T,\mathbf{C}^T)\widetilde{\mathbf{W}}_2+\widetilde{\mathbf{B}}_2=\Theta$ 的距離也是一個對稱三角形模糊數，其中心點為 $|K(\mathbf{x}_i^T,\mathbf{C}^T)\mathbf{w}_2+b_2|$ 與模糊半徑為 $K(\mathbf{x}_i^T,\mathbf{C}^T)\mathbf{c}_2+d_2$。要決定新進樣本點 $\mathbf{x}_i$ 是屬於哪一個類別，我們必須要判別它跟哪一個模糊曲面最接近，亦即是說，我們必須比較兩個對稱三角形模糊數的大小。對於任意兩個對稱三角形模糊數 $A=(m_A,c_A)$ 與 $B=(m_B,c_B)$，模糊數 A 大於模糊數 B 的程度可以使用下列歸屬函數來計算:

$$R_{\geq B}(A)=R(A,B)=\begin{cases}1 & \text{if } \alpha>0 \text{ and } \beta>0\\ 0 & \text{if } \alpha<0 \text{ and } \beta<0,\\ 0.5\left(1+\dfrac{\alpha+\beta}{\max(|\alpha|,|\beta|)}\right) & \text{o.w.}\end{cases} \quad (14)$$

其中 $\alpha=(m_A+c_A)-(m_B+c_B)$ 與 $\beta=(m_A-c_A)-(m_B-c_B)$。令對稱三角形模糊數 $\widetilde{D}_1=(|K(\mathbf{x}_i^T,\mathbf{C}^T)\mathbf{w}_1+b_1|,K(\mathbf{x}_i^T,\mathbf{C}^T)\mathbf{c}_1+d_1)$ 代表模糊曲面 $K(\mathbf{x}^T,\mathbf{C}^T)\widetilde{\mathbf{W}}_1+\widetilde{\mathbf{B}}_1=\Theta$ 與樣本點 $\mathbf{x}_i$ 的模糊距離。令對稱三角形模糊數 $\widetilde{D}_2=(|K(\mathbf{x}_i^T,\mathbf{C}^T)\mathbf{w}_2+b_2|,K(\mathbf{x}_i^T,\mathbf{C}^T)\mathbf{c}_2+d_2)$ 代表模糊曲面 $K(\mathbf{x}^T,\mathbf{C}^T)\widetilde{\mathbf{W}}_2+\widetilde{\mathbf{B}}_2=\Theta$ 與樣本點 $\mathbf{x}_i$ 的模糊距離，我們的模糊支持向量機的決策函數為:

$$f(\mathbf{x})=R_{\geq\widetilde{D}_1}\left(\widetilde{D}_2\right)=R\left(\widetilde{D}_2,\widetilde{D}_1\right). \quad (15)$$

此模糊決策函數傳回值是界於0與1之間，代表樣本點 $\mathbf{x}$ 屬於正類別的模糊歸屬程度。在正類別與負類別之間存在一個模糊的邊界，此模糊邊界更能夠處理現實中不精確與模糊的特性。值得注意的是，模糊決策函數$f(\mathbf{x})$也提供了對於預測結果的信心程度，$f(\mathbf{x})$的傳回值越大，代表 $\mathbf{x}$ 屬於正類別的可能性越強，對於預測結果的信心程度也就越大。對於決策制定任務(例如，醫療早期診斷)，能夠提供預測結果的信心程度是很重要的優點。

(二)模糊歸屬函數計算

與傳統支持向量機不同，模糊支持向量機對每一個訓練樣本導入一個模糊歸屬程度s_i，並且允許不同訓練樣本在建構分類器時，具有不同的貢獻度，如QPP(8)與(9)中目標函數的第3項所示，模糊歸屬程度s_i決定對於分類誤差ξ_i懲罰的加權程度。對於較為不確定的樣本點(例如，雜訊)，則可以指派較低的模糊歸屬程度，來減少它被分類錯誤時的懲罰強度，如此可以避免過度學習(overfitting)那些雜訊

點，並且增加推理能力。因此，如何適當地定義模糊歸屬函數，是提升模糊支持向量機推理能力的關鍵因素。在本研究中，我們將定義適合於醫療任務的模糊歸屬函數，並解決在醫療預測任務上常見的兩個棘手的問題:樣本不平衡與缺失值眾多。

- **樣本不平衡(imbalance):** 醫療早期診斷任務有著嚴重的數據不平衡的問題，正類別(有不良結果的類別)與負類別(無不良結果的類別)的樣本數目差距非常巨大，這會導致分類模型靠近數量大的類別去傾斜，而得到一個有偏見的模型。傳統的解決方式是超取樣(oversampling)或次取樣(undersampling)，超取樣是將正類別(數量較少的類別)的樣本，重復(採樣)出現在訓練樣本中數次，使得正負類別的樣本數目趨向平衡；次取樣則是從負樣本(數量較多的類別)中隨機刪除樣本，使得正負類別的樣本數目趨向平衡；超取樣的缺點是，由於它增加了訓練樣本的數目(正類別的樣本重復出現數次)，因此會增加它的訓練時間；次取樣的缺點是，它可能刪除了關鍵的負樣本，而導致訓練效果不彰。模糊支持向量機可以輕鬆地解決樣本不平衡的問題，我們可以將數量較少的正類別樣本，指派較高的歸屬程度，較不允許分類錯誤，而數量較多的負類別樣本，指派較低的歸屬程度，以避免得到一個有偏見的預測模型。值得注意的是，這種做法的效果與之前介紹的超取樣一樣，但是，模糊支持向量機的訓練可以比使用超取樣更有效率，因為超取樣將正類別訓練樣本數目增加而導致訓練時間大幅增加，而模糊支持向量機的訓練樣本數目維持不變，因此不會增加訓練負擔。
- **缺失值(missing value)眾多:** 醫療數據集通常會遭遇到由於大量缺失值(missing value)導致的樣本不精確與不確定的問題，通常，我們會使用「向前/向後填補」或是「中位數填補」來處理缺失值。對於缺失值比率較高的樣本，我們可以指派較低的模糊歸屬權重，使得它們可以允許被分類錯誤，以免分類模型因為要過度擬合這些擁有大量缺失值的不確定的訓練樣本，而導致沒有缺失值的重要訓練樣本的影響力被稀釋。

在本研究中，我們提出一種基於親和力(affinity)、類別機率(class probability)、類別不平衡比率與遺失值比率的新穎的模糊歸屬程度計算方式。

1. 基於支持向量數據描述的親和力確定

傳統以距離為基礎的模糊歸屬函數決定方式，大多是假設樣本為高斯分布(橢圓形)，並且在輸入空間進行聚類(clustering)，如果樣本點與群集中心點(cluster center)的距離越接近，則模糊歸屬程度值越高。然而，樣本的分布可能是任意的，簡單地假設它是橢圓形群集，將無法準確表達其所屬類別的歸屬程度。因此，本研究將使用支持向量數據描述機(support vector domain description, SVDD)來決定樣本屬於自身類別的親和力(affinity)。

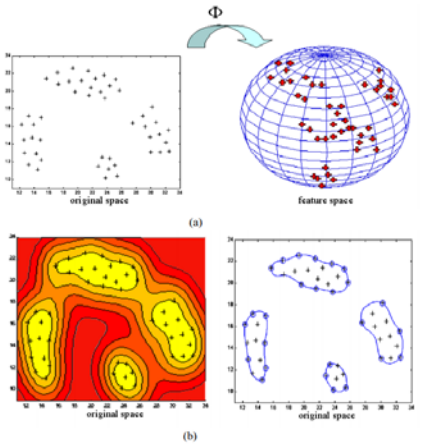

圖 5: SVDD 的概念圖

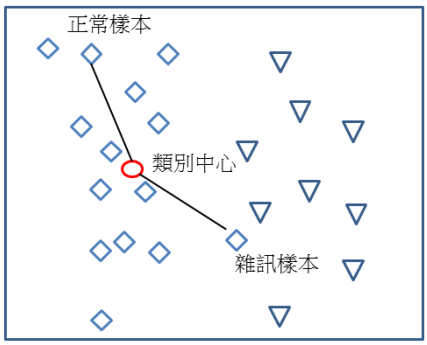

圖 6: 正常樣本與雜訊樣本到類別中心點的距離皆一樣，但是正常樣本與雜訊樣本對建構分類器做出不同的貢獻

首先，SVDD 將樣本點 $\{x_i\}, i=1,...,N$ 經由一個非線性映射 Φ 映射到一個高維度的特徵空間，並且尋找一個具有最小半徑的超球(hyper-sphere)來包圍住所有樣本點。而且，一個高維度特徵空間的超球，在原來輸入空間中可以是任意的形狀，SVDD 的概念圖如圖 5 所示，我們可以用一個最佳化的問題來描述它:

$$\min_{\mathrm{R,a},\xi_i} \quad R^2 + C\sum_i \xi_i \qquad s.t. \quad \left\|\Phi(x_i) - a\right\|^2 \le R^2 + \xi_i, \xi_i \ge 0 \qquad \forall i \tag{16}$$

其中 **a** 與 R 分別代表超球的球心與半徑，使用 Lagrangian Theorem ，我們可以得到它的對偶最佳化問題

$$\max_{\beta_i} \; W = \sum_i \Phi k(x_i, x_i)\beta_i - \sum_{i,j} \beta_i \beta_j k(x_i, x_j) \qquad st \; \sum_{i=1}^{N} \beta_i = 1, \;\; 0 \le \beta_i \le C \tag{17}$$

其中，$K(x_i, x_j) \equiv \Phi(x_i) \cdot \Phi(x_j)$ 是核心函數。在解出最佳解 β_i 後，我們可以由下面的公式，計算特徵空間中每一點到球中心點的距離

$$d_i^{svdd\,2} = \left\|\Phi(x_i) - a\right\|^2 = K(x_i, x_i) - 2\sum_j \beta_i K(x_i, x_j) + \sum_{j,k} \beta_i \beta_j K(x_j, x_k) \tag{18}$$

參考 Tao et al. (2020)學者的作法，對於每一個訓練樣本，我們可以透過以下算式計算它對於所屬類別的親和力

$$\mu^{affinity}(x_i) = \begin{cases} 0.8 \times \dfrac{d_i^{svdd} - \min_i(d_i^{svdd})}{\max_i(d_i^{svdd}) - \min_i(d_i^{svdd})} + 0.2 & d_i^{svdd} < \rho \times R \\ 0.2 \times \exp(\gamma(1 - {d_i^{svdd}}/{\rho \times R})) & d_i^{svdd} \ge \rho \times R \end{cases} \tag{19}$$

其中，$\rho \in (0, 1]$是一個權衡參數，用於控制可能的離群值和邊界樣本的大小，$\gamma > 0$ 是一個衰減權重參數，可以控制 d_i^{svdd} 的衰減程度。

2. 基於核心 *k* 最近鄰的類別機率確定

如圖 6 所示，與類別中心點具有相等距離的訓練樣本，可能對分類器的建構做出不同的貢獻，因此，僅考慮距離的模糊歸屬程度，並不足以準確描述其所屬類別的歸屬。為了解決此問題，我們使用核心 k 最近鄰（kernel k-nearest neighbor）方法，來決定訓練樣本屬於所屬類別的機率，對於訓練樣本 x_i，我們首先找出它

在特徵空間中的 k 個最接近的鄰居，樣本點 x_i 與 x_j 在特徵空間的距離可以根據下式計算出來

$$d_{ij}^{svdd2} = \left\|\Phi(x_i) - \Phi(x_j)\right\|^2 = K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j) \quad (20)$$

然後，我們計算這 k 個選定鄰居中，與 x_i 屬於相同類別的樣本數目。假設數目為 num_i^{own}。最終，樣本 x_i 屬於所屬類別的機率為

$$p_i = num_i^{own} / k \quad (21)$$

p_i 值越大，表示 x_i 鄰近的樣本點都與它屬於同一個類別，亦表示 x_i 屬於其自身類別的可能性越高，反之，p_i 值越小，表示 x_i 鄰近的樣本點都與它屬於不同的類別，亦表示 x_i 屬於其自身類別的可能性越低。

3.考慮類別不平衡比率與樣本遺失值比率的模糊歸屬函數

假設 n_{pos} 和 n_{neg} 分別是正樣本和負樣本的數量，類別不平衡比率（imbalance ratio, IR）定義為 $IR = n_{pos} / n_{neg}$。我們可以透過 IR，為大類別的樣本設置較低的歸屬程度，以避免分類模型只擬合大類別的數據。此外，假設樣本的特徵數目為 fn，而樣本 x_i 中具有缺失值的特徵數目為 $miss_i$，則樣本 x_i 的缺失值比率(missing ratio, MR)定義為 $MR_i = miss_i / fn$，我們可以透過 MR_i，為缺失值多的樣本設置較低的歸屬程度，以免分類模型因為要過度擬合這些擁有大量缺失值的不確定的訓練樣本，而導致沒有缺失值的重要訓練樣本的影響力被忽略。最終，本研究所使用的歸屬函數為

$$s(x_i) = \begin{cases} 1 & x_i \in \text{minority class} \\ \mu^{affinity}(x_i) * p_i * (1 - MR_i) * IR & x_i \in \text{majority class} \end{cases} \quad (22)$$

對於小類別的樣本點，由於他們的數量已經很稀少了，所以歸屬函數值設定為 1，至於大類別的樣本點，則是要考慮公式(19)的樣本親和力與公式(21)的類別機率，並且透過 IR 為大類別的樣本設置較低的歸屬程度，以及透過 MR_i 為缺失值多的樣本設置較低的歸屬程度。

肆、實驗結果

本研究主要以病患進入急診後 3 小時之整合性資料，預測敗血症病患在住院期間(28 天內)是否會發生不良結果；本研究的原始資料來源，是長庚醫院 8 個院區從西元 2010 年至 2018 年的病患資料(如表 1 所示)，本研究蒐集的病患資料的變項的敘述性統計分析如表 4 至 7 所示:

表 4:人口統計特徵(靜態資料)的敘述性統計分析

性別	男性				女性		
	56.7837%				43.2163%		
年齡	17-30 歲	31-40 歲	41-50 歲	51-60 歲	61-70 歲	71-80 歲	81 歲以上
	4.6793%	6.4526%	10.1188%	15.7076%	18.5306%	22.3294%	22.1816%

表 5:生物標記特徵(動態資料)中，各項生物標記有進行量測次數的統計分析

肌酸酐	乳酸	C-反應蛋白	白血球計數	多型核白血球	淋巴球
142049	63541	100158	156578	154786	154786

表 6:過去病史特徵(靜態資料)中，各項疾病出現次數的統計分析

糖尿病	惡性腫瘤	肝病	腎臟病	腦血管疾病	慢性阻塞肺病	充血性心力衰竭
43287	33708	28948	31522	32210	34885	28199
結蹄組織疾病	失智症	周邊血管疾病	心肌	白血病	偏癱	愛滋病
2509	10197	5380	4433	1210	3767	108

表 7:生命徵象特徵(動態資料)中，各項生命徵象的敘述性統計分析

	BT	HR	NBP D	NBP S	RR
平均數	25.12324	95.10605	68.17148 4	117.0685	19.18511
中間值	36.3	94	70	119	18
標準差	21.19934	205.9474	69.44189	42.7954	48.75312
變異數	449.4119	42414.34	4822.177	1831.446	2376.867
峰度	7358.574	161542.7	21824.27	1537.0 17	142522.9
偏態	47.40081	371.0113	129.8'708	12.62995	358.8382
範圍	3613	94121	18089	5284.5	20165
最小值	0	0	0	0	0
最大值	3613	94121	18089	5284.5	20165

本研究評估方法將會以 AUC、F1-Score 及 Sensitivity 為主要的衡量依據；由於醫學上的病例經常會出現類別不平衡的情況，假設有 100 例樣本，其中有 10 例為陽性，此時若模型全部預測為陰性，則此時模型的 Accuracy 便高達 90%，但是，此時模型的 Sensitivity 為 0，表示該模型並沒有辦法正確的分辨出陽性樣本，故在此種情況下，衡量模型效能時便不會僅參考 Accuracy；而 AUC(Area under the ROC Curve) 與 F1-Score 在衡量模型的效能上，因為他們已經考慮陽性預測值(Positive Predictive Value, PPV)與陰性預測值(Negative Predictive Value, NPV)，故較不會受到類別不平衡的影響，可以正確的評估模型的預測效能；而本研究由於較注重找出可能會發生不良結果的病患，因此將 Sensitivity 一併作為衡量的標準。Sensitivity 為衡量模型可以分辨出多少陽性樣本的一種指標。表 8 為混淆矩陣及各種相關指標之計算方式說明。

表 8:混淆矩陣以相關指標計算

	實際 Positive	實際 Negative
預測 Positive	True Positive (TP，真陽性)	False Positive (FP，偽陽性)
預測 Negative	False Negative (FN，假陰性)	True Negative (TN，真陰性)
Accuracy(準確度)	(TP + TN) / (TP + FP + TN + FN)	
Sensitivity(靈敏度)	TP / (TP + FN)	
Specificity(特異度)	TN / (FP + TN)	
PPV(陽性預測值)	TP / (TP + FP)	
NPV(陰性預測值)	TN / (TN + FN)	
F1-Score	(2 * PPV * Sensitivity) / (PPV + Sensitivity)	

在第一個實驗中，我們使用 Holdout 驗證法比較本研究提出的混合長短期記憶體(LSTM)、卷積神經網路(CNN)、全連結神經網路與模糊支持向量機的複合式預測模型，與下列最先進的敗血症不良結果早期預測演算法，對於本研究蒐集的長庚醫院病患的預測效能，其中，Gultepe et al. (2014) 提出基於線性支持向量機(Support Vector Machine, SVM)的敗血症早期不良結果預測演算法，Taylor et al. (2016) 提出使用基於隨機森林(Random Forest, RF)的敗血症智能預測算法，Saqib et al. (2018) 提出使用長短期記憶體（Long-Short Term Memory, LSTM）技術來建構敗血症預測模型，Kok et al. (2020)提出使用卷積神經網路(Convolutional Neural

Network, CNN)的自動診斷工具來預測敗血症，Rafiei et al. (2021)提出一種混合式智能敗血症預測器，它包含長短期記憶(Long short-term memory, LSTM)，卷積層(convolutional layer)和完全連接(fully connected layers, FCL)，在第一個實驗中，本研究採用 Holdout 驗證法，將病患資料分割為訓練數據（80%），驗證數據（10%）和測試數據（10%），訓練數據用於建立預測模型。驗證數據用於在訓練過程中，挑選最佳的模型參數(例如，類神經網路中的階層數目、每一層的神經元數目、激發函數與逞罰函數的選擇等)，測試數據則是在測試階段，對於使用訓練數據得到的最終預測模型，進行無偏頗的評估。實驗結果如表 9 所示，其中「CNN+LSTM+fuzzy twin SVM」表示本研究提出的 CNN 與 LSTM 的混合模型的最後一層改成模糊孿生 SVM 的預測方法，「CNN+LSTM+ SVM」表示 CNN 與 LSTM 的混合模型的最後一層改成傳統 SVM 的預測方法，從表 9 中可發現，本研究提出的方法在各項指標上表現皆是最優異的，值得注意的是，在 Specificity 與 PPV 指標上面，RF(Random Forest)的表現比 CNN 與 LSTM 的混合模型(CNN+LSTM)優異，該結果表明 RF 在預測無不良後果的病患方面，其預測效果較 CNN 與 LSTM 的混合模型準確。而 CNN 與 LSTM 的混合模型的 Sensitivity 與 NPV 皆比 RF(Random Forest)優異，代表 CNN 與 LSTM 的混合模型比 RF 更能夠找出會發生不良結果的病患。另外，從 Sensitivity 的表現上面，可以發現 CNN 與 LSTM 的混合模型是優於 RF，這表示在具有動態性質的資料上，僅是使用 CNN 與 LSTM 的混合模型，在預測有不良結果的病患時，便有不錯的成果，而若是將 CNN 與 LSTM 的混合模型的最後一層改成 fuzzy twin SVM 之後，便可以在不造成更多偽陽性的情況下有著更高的靈敏度，因此，本研究提出的 fuzzy twin SVM，CNN 與 LSTM 混合模型，不論是 Sensitivity 或 Specificity，皆是比其他方法更加優異。圖 7 則是顯示不同方法的 ROC 曲線(接收者操作特征曲線, receiver operating characteristic curve, ROC curve)，由圖 7 所示，本研究提出的方法的 AUC(Area under the ROC Curve)是比其他方法優異。

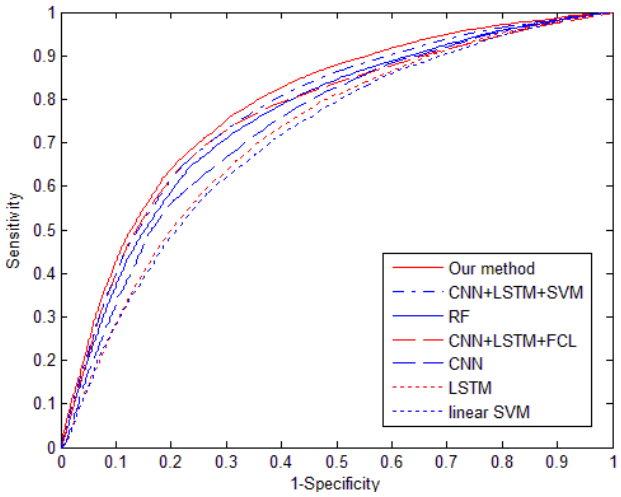

圖 7: 不同方法的 ROC 曲線圖

表 9:各種預測方法的校能比較(使用 Holdout 驗證方法:訓練集 80%/驗證集 10% /測試集 10%)

Method	Accuracy	Sensitivity	Specificity	PPV	NPV	F1_Score	AUC
CNN+LSTM+fuzzy twin SVM(本研究)	**0.729300**	**0.761450**	**0.691488**	**0.743780**	**0.711365**	**0.752511**	**0.788733**
CNN+LSTM+SVM	0.720034	0.739285	0.697393	0.741828	0.694592	0.740554	0.766248
RF(Taylor et al. 2016)	0.706003	0.721651	0.687599	0.730960	0.677451	0.726276	0.758848
CNN+LSTM(Rafiei et al. 2021)	0.713482	0.753000	0.667003	0.726746	0.696600	0.739640	0.76248
LSTM(Saqib et al. 2018)	0.683169	0.716997	0.643382	0.702797	0.659044	0.709826	0.719182
CNN(Kok et al. 2020)	0.693957	0.727039	0.655048	0.712554	0.671093	0.719724	0.740097
SVM(Gultepe et al. 2014)	0.66285	0.655033	0.672044	0.701416	0.623547	0.677432	0.71015

由於 Holdout 驗證方法(訓練集合 80% /驗證集合 10% / 測試集合 10%)可以畫出 ROC curve(如圖 7 所示),可以進一步用圖形化檢視分類模型的 1–Specificity 與 Sensitivity 之間效能的變化關係，所以常被許多生醫論文用來驗證分類效能，但是 *N* 折交叉驗證(*N*-fold cross-validation)則可以更公平地驗證預測效能，因為在 *N* 折交叉驗證中，每一筆樣本都會出現在訓練集合中，也會出現在測試集合中。所以第二個實驗中，我們使用 10 折交叉驗證法比較本研究提出的方法與其他最先進的敗血症預測演算法，實驗結果如表 10 所示，其中「CNN+LSTM+fuzzy twin SVM」表示本研究提出的 CNN 與 LSTM 的混合模型的最後一層改成模糊孿生 SVM 的預測方法,「CNN+LSTM+ SVM」表示 CNN 與 LSTM 的混合模型的最後一層改成傳統 SVM 的預測方法，如表 10 所示，在使用 10 折交叉驗證時，本研究提出的方法仍然是優於其他最先進的敗血症預測演算法。

表 10:各種預測方法的預測校能比較(使用 10 折交叉驗證法)

Method	Accuracy	Sensitivity	Specificity	PPV	NPV	F1_Score	AUC
CNN+LSTM+fuzzy twin SVM(本研究)	**0.763428**	**0.755738**	**0.77097**	**0.763924**	**0.763100**	**0.759749**	**0.793354**
CNN+LSTM+SVM	0.752392	0.743718	0.760898	0.753132	0.751832	0.748335	0.784308
RF(Taylor et al. 2016)	0.742297	0.74563	0.739028	0.737001	0.747613	0.74129	0.771279
CNN+LSTM(Rafiei et al. 2021)	0.743885	0.747233	0.740601	0.738586	0.749204	0.742884	0.772684
LSTM(Saqib et al. 2018)	0.702339	0.730386	0.674828	0.687861	0.718485	0.708451	0.732607
CNN(Kok et al. 2020)	0.722443	0.737610	0.707566	0.712140	0.733290	0.724651	0.754633
SVM(Gultepe et al. 2014)	0.685189	0.714665	0.65628	0.670932	0.701271	0.692045	0.715472

為了更嚴謹的證明本研究的方法是優於傳統的方法，本研究採用基於 *N* 折交叉驗證(*N*-fold cross-validation)的配對 *t* 檢驗(paired *t*-tests)(Demšar 2006)，首先，我們將樣本集拆分成 *N* 個大小相等的部分(*N*=10)，每一次，我們挑選其中一部分作為測試資料集合，其餘部分則是訓練資料集合，重覆此步驟 *N* 次，我們得到分類器在 *N* 個不同測試集合的分類效能，接著，我們使用配對 *t* 檢驗來驗證分類器 *A* 和 *B* 的預測效能之間是否存在統計上顯著的差異(基於它們在 *N* 個不同測試集上的預測效能的差異)，其中，虛無假設（Null hypothesis, H_0）是分類器 *A* 和 *B* 的預測效能沒有區別，在統計上，我們可以將分類器 *A* 和 *B* 的預測效能視為從某個分佈中抽取的隨機變量。H_0 表示分類器 *A* 和 *B* 的預測效能來自相同的分佈。如果 H_0 為真，則它們的預測效能之間的預期差異（在所

有可能的測試資料集合)為 0。一旦我們計算出 t 統計量,我們就可以計算 p 值並將其與我們選擇的顯著性水平進行比較,例如,$\alpha = 0.05$。如果 p 值小於 α,則我們拒絕虛無假設,並接受兩個模型之間存在顯著差異。表 11 顯示本研究提出的方法與其他先進敗血症預測方法的 t 檢驗的統計分析結果。例如,在比較本方法與 RF 在準確率(Accuracy)的 t 檢驗的 p 值為 1.8015e-05 (<0.05),這意味著我們的方法的準確率在 0.05 顯著水平上是優於 RF。如表 11 所示,本研究的方法在各項預測效能指標都是在 0.05 顯著水平上是優於其他的方法。

表 11:本研究提出的方法與其他先進敗血症預測模型執行 t 檢測的 p 值結果

Method	Accuracy	Sensitivity	Specificity	PPV	NPV	F1_Score	AUC
CNN+LSTM+SVM	0.00515	0.01649	0.00374	0.00243	9.6381e-04	7.1564e-04	8.1634e-04
RF(Taylor,et al. 2016)	1.801e-05	0.02755	1.1173e-05	6.639e-06	4.4523e-04	7.4361e-05	1.9560e-05
CNN+LSTM(Rafiei et al. 2021)	2.758e-05	0.04650	1.4486e-05	9.286e-06	7.5693e-04	1.2164e-04	3.0064e-05
LSTM(Saqib et al. 2018)	4.458e-08	6.6960e-04	2.9494e-08	2.053e-08	1.0165e-06	2.3977e-07	4.6719e-08
CNN(Kok et al. 2020)	2.748e-07	0.00205	1.8138e-07	1.054e-07	8.0829e-06	1.2495e-06	2.8829e-07
SVM(Gultepe et al. 2014)	3.733e-08	4.3919e-05	6.8429e-09	1.169e-08	4.6708e-07	1.5430e-07	3.9036e-08

在第三個實驗中,我們將驗證靜態特徵(包含人口統計特徵與過去病史)與動態特徵(包含生命徵象與生物標記)對於敗血症不良結果早期預測的效能貢獻度,其中, Dynamic 代表僅使用生命徵象與生物標誌的動態特徵資料集合;Vitals 代表僅使用生命徵象的特徵資料集合;Biomarkers 代表僅使用生物標誌的特徵資料集合; Static 代表僅使用人口統計特徵與過去病史的靜態特徵資料集合; Demographic 代表僅使用人口統計特徵的資料集合;History 代表僅使用過去病史的特徵資料集合,各種類型資料的預測效能顯示於表 12 中(這裡使用 10 折交叉驗證來衡量預測效能)。

表 12: 各種特徵集合的預測效能比較(使用 10 折交叉驗證法)

Data Source	Accuracy	Sensitivity	Specificity	PPV	NPV	F1_Score	AUC
Dynamic	**0.730384**	**0.73761**	**0.723297**	**0.723341**	**0.737568**	**0.730406**	**0.759283**
Vitals	0.717678	0.736006	0.699701	0.706217	0.729898	0.720804	0.741778
Biomarkers	0.695505	0.724985	0.666591	0.681235	0.711597	0.702358	0.725788
Static	0.602908	0.632379	0.574005	0.592854	0.61429	0.611902	0.633192
Demographic	0.562091	0.577211	0.547263	0.555662	0.568987	0.566142	0.592237
History	0.567651	0.588756	0.546953	0.560372	0.575649	0.57412	0.607855

由表 12 可知, Dynamic、Vitals 與 Biomarkers 比起 Static、Demographic 與 History 在各項指標上,皆具有更好的表現;這說明動態資料比起靜態資料包含更多關鍵因素,對於模型的預測效能具有重大的貢獻,也表明在僅使用動態資料的情況下,複合式的模型便已經有一定的預測效能。而靜態資料由於其所包含的訊息不多,且僅使用靜態資料時,也很難發現有不良結果與無不良結果之間的差異,因此,靜態資料只適合用於輔助動態資料,讓模型在預測時有更多的依據,從而使模型的預測可以更為的準確及穩定。實驗結果表明,同時考慮動態與靜態特徵時,能夠獲得最佳的預測效能。而在動態特徵當中,僅使用 Vitals 的預測效能比 Biomarkers 優異,可能是由於 Biomarkers 遭受到缺失值問題的影響,因為

病患通常會常規的量測各種生命體徵，所以缺失值比率較少。但是，病患通常不會對每一項生物標記都進行測量，而且，通常也不會持續的進行檢測，所以缺失值較多。在靜態特徵當中，過去疾病史特徵的預測效能比人口統計特徵優異。

為了更嚴謹的分析各種特徵在預測準確率的貢獻，我們採用基於 N 折交叉驗證(N-fold cross-validation)的配對 t 檢驗(paired t-tests)(Demšar 2006)，來驗證使用特徵集合 A 和 B 的預測效能之間是否存在統計上顯著的差異，其中，虛無假設（Null hypothesis, H_0）是特徵集合 A 和 B 的預測效能沒有區別， H_0 表示特徵集合 A 和 B 的預測效能來自相同的分佈。如果 H_0 為真，則它們的預測效能之間的預期差異（在所有可能的測試資料集合）為 0。一旦我們計算出 t 統計量，我們就可以計算 p 值並將其與我們選擇的顯著性水平進行比較，例如，$\alpha = 0.05$。如果 p 值小於 α，則我們拒絕虛無假設，並接受兩個特徵集合的預測效能之間存在顯著差異。表 13 顯示本研究使用的全部特徵集合(包含生命徵象、生物標記、過去病史與人口統計特徵的全部特徵集合)與各項特徵子集合的 t 檢驗的統計分析結果。例如，在比較全部特徵與動態特徵(Dynamic)在準確率(Accuracy)的 t 檢驗的 p 值為 1.1857e-06 (<0.05)，這意味著使用全部特徵的準確率在 0.05 顯著水平上是優於僅使用動態特徵(Dynamic)的。從表 13 中可以看出，本研究提出來的使用的全部特徵，在各項分類指標上，皆是在 0.05 顯著水平上面是優於僅使用部分特徵子集合。

表 13: 本研究提出的特徵集合(包含生命徵象、生物標記、過去病史與人口統計特徵的全部特徵集合)與各項特徵集合執行 t 檢測的 p 值結果

Data Source	Accuracy	Sensitivity	Specificity	PPV	NPV	F1_Score	AUC
Dynamic	1.1857e-06	0.00205	1.1436e-06	5.4824e-07	2.2943e-05	4.2852e-06	1.2592e-06
Vitals	1.2560e-07	0.00125	8.1409e-08	4.8985e-08	3.7951e-06	5.9387e-07	1.3144e-07
Biomarkers	4.7797e-08	5.2953e-05	7.9800e-09	1.4465e-08	5.8996e-07	1.9682e-07	4.9995e-08
Static	4.1690e-10	1.1021e-07	3.3051e-11	1.6400e-10	1.9144e-09	2.7132e-09	4.3842e-10
Demographic	8.4378e-12	1.4321e-08	5.0727e-12	3.4375e-12	5.6212e-11	1.8285e-10	9.1228e-12
History	1.8933e-11	5.3535e-08	9.8657e-13	4.2205e-12	1.8866e-10	5.4257e-10	2.0804e-11

在第四個實驗中，本研究針對不同的不良結果個別進行二元分類模型建構（例如，建構敗血症死亡預測模型、敗血症休克預測模型等..），來分析探討不同的不良結果的預測難度，實驗結果如表 14 所示，其中 Normal vs Dead 代表正常與死亡類別二元分類的結果；Normal vs Shock 代表正常與休克類別二元分類的結果；Normal vs RF 代表正常與呼吸衰竭類別二元分類的結果；Normal vs ICU 代表正常與轉入 ICU 類別二元分類的結果，如表 14 所示，預測效能最好的是 Normal vs RF，代表正常結果與呼吸衰竭的區隔是比較容易的，其次為 Normal vs Dead 與 Normal vs Shock，而 Normal vs ICU 的預測效能最差，這也代表要區隔正常與轉入 ICU 的病患是最不容易的，原因是病患轉入 ICU 可能是由於許多其他的病癥，所以特異性較差。

表 14: 針對各項不良結果單獨建立二元分類模型的預測效能比較

二元分類器	Accuracy	Sensitivity	Specificity	PPV	NPV	F1_Score	AUC
Normal vs Dead	0.760801	0.763427	0.758231	0.759482	0.762461	0.76132	0.78131
Normal vs Shock	0.743109	0.784805	0.701234	0.724474	0.765365	0.753291	0.76422
Normal vs RF	0.766903	0.77401	0.759867	0.763357	0.770832	0.768542	0.78661
Normal vs ICU	0.714521	0.73512	0.69412	0.706336	0.723691	0.720257	0.73645

在最後一個實驗中，本研究分析各種演算法參數調整對於模型效能的影響，本研究首先利用 CNN、LSTM 與全連結網路來學習關鍵特徵，並餵給模糊孿生 SVM 作為判別的依據，在深度神經網路學習特徵的演算法參數部分，學習率設為 1e-3，而 CNN 的卷積層使用 200 個濾波器(Filter)，濾波器尺寸(核心大小, Kernel Size)為 2，步長(Strides) 為 1，激勵函數(Activation)為 ReLU；池化層則是使用最大池化(Max Pooling)的方式；而 LSTM 網路則是使用 80 個隱藏單元(Units)，激勵函數為 Tanh；全連結網路(Full Connected network) 的隱藏層節點數目是 80 個。而在模糊孿生 SVM 進行分類的演算法參數部分，我們使用 RBF (radial basis function)核心函數，懲罰參數 C 值設定為 10。接下來的實驗，我們分析卷積神經網路(CNN)的中濾波器(filter)的尺寸與數量對於預測效能的影響，以及 SVM 中懲罰參數 C 對於預測效能的影響，實驗結果如表 15 至 17 所示。其中，CNN 的濾波器的主要目的是學習短時間區域內的「時不變(time-invariant)」的特徵，這些特徵是時不變的，因為它們可以在任何時間點出現，一個濾波器可以對應到一個「時不變」特徵，濾波器的數目越多，CNN 能夠捕獲的「時不變」特徵的種類也越多，也代表 CNN 的能力越強大；相反地，濾波器的數目越少，CNN 能夠捕獲的「時不變」特徵的種類也越少，也代表 CNN 的學習能力越小。CNN 的學習能力越小，則無法學習到充分的關鍵特徵，將導致分類器的訓練不足(under-fitting)，但是，CNN 的學習能力越強，也會使得 CNN 錯誤地學習到一些雜訊的關聯，將導致分類器過度訓練(over-fitting)，表 15 顯示不同濾波器數目時的預測效能，如表 15 所示，濾波器的數目過多或過少皆不好，最佳的濾波器數目為 200 個。

表 15: 使用不同濾波器數目的預測效能比較

濾波器數目	Accuracy	Sensitivity	Specificity	PPV	NPV	F1_Score	AUC
100	0.761518	0.768726	0.75445	0.754319	0.768993	0.761404	0.791588
150	0.762155	0.76215	0.762161	0.758609	0.765828	0.760314	0.792156
200	**0.763428**	**0.755738**	**0.77097**	**0.763924**	**0.763100**	**0.759749**	**0.793354**
250	0.763189	0.751406	0.774747	0.765876	0.760793	0.758513	0.793077
300	0.76295	0.749962	0.775689	0.766249	0.760034	0.757941	0.792826

濾波器的尺寸則是影響 CNN 學習到的「時不變」特徵的尺寸，濾波器尺寸越大，CNN 就去捕抓長時間範圍的「時不變」特徵，濾波器的尺寸越小，CNN 就去捕抓短時間範圍的「時不變」特徵，由於本研究為了克服急診醫生在少量資訊的情況下必須做出決定的困難，所以只使用病患在急診室 3 小時的動態特徵，來做出病患是否會在 28 天內有不良結果的預測，而且量測時間是每小時一次，所以醫療事件序列的長度只有 3，因此，本研究無法將濾波器的尺寸設定的太

大，表 16 表示濾波器尺寸分別設定為 1、2 與 3 時的預測效能，從表 16 可以看出預測效能會隨著濾波器的尺寸而改變，而且最佳的預測效能出現在濾波器尺寸設定為 2 時。

表 16: 使用不同濾波器尺寸的預測效能比較

濾波器尺寸	Accuracy	Sensitivity	Specificity	PPV	NPV	F1_Score	AUC
1	0.739193	0.772409	0.706613	0.720863	0.760093	0.745691	0.769511
2	**0.763428**	**0.755738**	**0.77097**	**0.763924**	**0.763100**	**0.759749**	**0.793354**
3	0.728469	0.776736	0.681127	0.704959	0.756788	0.739082	0.758932

SVM 的懲罰參數 *C* 控制對於訓練錯誤的懲罰程度，*C* 值設定的越小，則對於訓練錯誤的懲罰不夠，將導致分類器訓練不足(under-fitting)，相反地，*C* 值設定的越大，則對於訓練錯誤的懲罰越強烈，但是這也會使得分類器僅靠著記憶死背訓練資料集合而沒有學習得到泛化的知識，將導致分類器過度訓練(over-fitting)，表 17 顯示不同 *C* 值時的預測效能，如表 17 所示，*C* 值過大或過小，預測效果皆不好，最佳的 *C* 值為 10。

表 17: 使用 SVM 懲罰參數 *C* 值的預測效能比較

SVM 的 *C* 值	Accuracy	Sensitivity	Specificity	PPV	NPV	F1_Score	AUC
0.1	0.757313	0.788460	0.726762	0.738978	0.777980	0.762878	0.787611
1	0.760563	0.770332	0.750980	0.752131	0.769327	0.761093	0.790656
10	**0.763428**	**0.755738**	**0.770970**	**0.763924**	**0.763100**	**0.759749**	**0.793354**
100	0.762713	0.749480	0.775694	0.766144	0.759647	0.757653	0.792587
1000	0.759299	0.783968	0.735102	0.743785	0.776326	0.763320	0.789535

伍、結論與建議

由於病患在進入急診室之後，通常 3~6 小時之內醫師便會決定該病患是要進入觀察室或者是離院，因此本次研究主要以病患進入急診後 3 小時之整合性資料(生命徵象、生物標記與過去病史)，來預測敗血症病患在住院期間內(28 天)是否會發生不良結果；本研究提出的方法能夠達到 F1 score 為 0.75，AUC 為 0.78，證明使用病患進入急診後 3 小時之整合性資料預測敗血症病患在 28 天是否會發生不良結果的可行性。本研究為盡可能模擬早期診斷時的實際情況，使模型可在僅使用少量數據的情況下，預測病患於未來 28 天內是否可能會出現不良結果，以利醫師判斷該病患後續是否要進入觀察室抑或是離院。故本研究實驗僅以 3 小時的生命徵象與生物標記量測資料進行測試；除了在時間區段上做限制之外，本研究使用生命徵象、過去病史、部分血液檢測數據等在早期就能快速取得的資料為主，亦即本研究所取用的特徵量較少(本研究合計選用 25 種特徵)。由於這些限制，使得本研究相較於過去類似的研究，其結果仍舊有進步的空間。但同時也因為這些限制，使得本研究相比於其他研究，更加貼近臨床的實際情況，因為一位剛進入急診室的病患，並沒有辦法在短時間之內便取得大量數據。

另外從研究結果中也能發現，CNN 與 LSTM 的混合模型，其在 Sensitivity 的表現皆優於其他種機器學習模型，這表示 CNN 與 LSTM 的混合模型在探索有發生不良結果病患的數據特徵上，有著優於其他機器學習模型的效能，若若是將

CNN 與 LSTM 的混合模型的最後一層改成 fuzzy twin SVM 之後，模型的效能便會更上一層， 故以目前的結果來看，複合式的學習模型確實更優於僅使用單一模型的績效；另外，為了觀察資料對於模型效能的影響，本研究也以不同類型資料進行測試，可以發現在 3 小時的時間長度下，僅使用動態資料的衡量指標便有一定的水準，再加入靜態資料便可以有更好的成效；而僅使用靜態資料進行模型訓練其結果則不盡理想，這表示動態資料為模型的主要的判斷依據，靜態資料則是作為輔助判斷的特徵。

除此之外，本研究使用模糊支持向量機取代原本深度網路模型的最後一層，使得 F1 score 從 0.73 進步到 0.75，AUC 從 0.76 進步到 0.78，證明結合深度神經網路與模糊支持向量機，能夠融合兩者的優點。深度神經網路擁有優異的特徵學習能力，可以從動態資料中找出關鍵的特徵模式；而在給定良好特徵的情況下，基於統計學習理論建立的支持向量機則是最佳的分類器，此外，模糊理論則可以有效地解決醫療診斷應用中，由於缺失值導致的樣本不精確的問題，並且模糊理論更能夠避免由於雜訊樣本導致的過度學習(over-fitting)的問題，而本研究提出的方法混合了上述模型的優點，因此對於敗血症的早期預測的效能上，本研究提出的方法能夠優於其他最先進的敗血症早期預測演算法。

由於訓練樣本的模糊歸屬度估計方法會影響模糊支持向量機的分類效能，在未來，我們預期開發更適合敗血症不良結果早期預測應用的訓練樣本模糊歸屬度估計方法。除此之外，在未來我們也將運用更先進的深度學習技術，來從病患的動態與靜態特徵中，自動學習出更關鍵的特徵，以提高敗血症早期預測的效能。

參考文獻

Bhandary, A., Prabhu, G.A., Rajinikanth, V., Thanaraj, K.P., Satapathy, S.C., Robbins, D.E., Shasky, C., Zhang, Y.-D., Tavares, J.M.R.S. & Raja, N.S.M. (2020). Deep-Learning Framework to Detect Lung Abnormality – A study with Chest X-Ray and Lung CT Scan Images, *Pattern Recognition Letters*, 129, 271-278.

Demšar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets, *Journal of Machine Learning Research*, 7(1), 1−30.

Despins, L.A. (2017). Automated detection of sepsis using electronic medical record data: a systematic review, *Journal for Healthcare Quality*, 39(6), 322–333.

Doi, K., Yuen, P.S., Eisner, C., Hu, X., Leelahavanichkul, A., Schnermann, J. & Star, R.A. (2009). Reduced production of creatinine limits its use as marker of kidney injury in sepsis, *Journal of the American Society of Nephrology : JASN*, 20(6), 1217-1221.

Fairchild K.D. & O'Shea T.M. (2010). Heart rate characteristics: physiomarkers for detection of late-onset neonatal sepsis, *Clinics in Perinatology*, 37(3), 581-98.

Froon, A.H., Bemelmans, M.H., Greve, J.W., van der Linden, C.J. & Buurman, W.A. (1994). Increased plasma concentrations of soluble tumor necrosis factor

receptors in sepsis syndrome: correlation with plasma creatinine values, *Critical Care Medicine*, 22(5), 803-809.

Gultepe, E., Green, J.P., Nguyen, H., Adams, J., Albertson, T. & Tagkopoulos, I. (2014). From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system, *Journal of the American Medical Informatics Association*, 21(2), 315–325.

Hao, P.-Y., Kung, C.-F., Chang, C.-Y. & Ou, J.-B. (2020). Predicting Stock Price Trends Based on Financial News Articles and Using a Novel Twin Support Vector Machine with Fuzzy Hyperplane, *Applied Soft Computing*, accepted.

Hatfield, K.M., Dantes, R.B., Baggs, J., Sapiano, M.R.P., Fiore, A.E., Jernigan, J.A. & Epstein, L. (2018). Assessing variability in hospital-level mortality among US medicare beneficiaries with hospitalizations for severe sepsis and septic shock, *Critical Care Medicine*, 46(11), 1753-1760.

Hinton, G.E. & Salakhutdinov, R.R. (2006). Reducing the dimensionality of data with neural networks, *Science*, 313, 504–507.

Huang, F.J. & LeCun, Y. (2006). Large-scale Learning with SVM and Convolutional for Generic Object Categorization, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, USA, 284-291.

Islam, M.M., Nasrin, T., Walther, B.A., Wu, C.C., Yang, H.C. & Li, Y.C. (2019). Prediction of sepsis patients using machine learning approach: a meta-analysis, *Computer Methods and Programs in Biomedicine*, 170, 1-9.

Kok, C., Jahmunah, V., Oh, S.L., Zhou, X., Gururajan, R., Tao, X., Cheong, K.H., Gururajan, R., Molinari, F. & Acharya, U.R. (2020). Automated prediction of sepsis using temporal convolutional network, *Computers in Biology and Medicine*, 127(103957), 1-10.

Kumar, A., Roberts, D., Wood, K.E., Light, B., Parrillo, J.E., Sharma, S., Suppes, R., Feinstein, D., Zanotti, S., Taiberg, L., Gurka, D., Kumar, A. & Cheang, M. (2006). Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock, *Critical Care Medicine*, 34(6), 1589-1596.

Lauritsen, S.M., Kalør, M.E., Kongsgaard, E.L., Lauritsen, K.M., Jørgensen, M.J., Lange, J. & Thiesson, B. (2020). Early detection of sepsis utilizing deep learning on electronic health record event sequences, *Artificial Intelligence in Medicine*, 104(101820), 1-11.

Lin, C., Zhang, Y., Ivy, J., Capan, M., Arnold, R., Huddleston, J.M. & Chi M. (2018). Early Diagnosis and Prediction of Sepsis Shock by Combining Static and Dynamic Information Using Convolutional-LSTM, *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, New York, 219-228.

Mao, Q., Jay, M., Hoffman, J.L., Calvert, J., Barton, C., Shimabukuro, D., Shieh, L., Chettipally, U., Fletcher, G., Kerem, Y., Zhou, Y. & Das, R. (2018). Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU, *BMJ Open*, 8(1), e017833.

Nemati, S., Holder, A., Razmi, F., Stanley, M.D., Clifford, G.D. & Buchman, T.G.. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU, *Critical Care Medicine,* 46(4), 547–553.

Nguyen, H.B., River,s E.P., Knoblich, B.P., Jacobsen, G., Muzzin, A., Ressler, J.A. & Tomlanovich, M.C. (2004). Early lactate clearance is associated with improved outcome in severe sepsis and septic shock, *Critical Care Medicine*, 32(8), 1637-1642.

Paoli, C.J., Reynolds, M.A., Sinha, M., Gitlin, M. & Crouser, E. (2018). Epidemiology and costs of sepsis in the United States—an analysis based on timing of diagnosis and severity level, *Critical Care Medicine*, 46(12), 1889-1897.

Pierrakos, C. & Vincent, J.-L. (2010). Sepsis biomarkers: a review, *Critical Care,* 14(R15), 1-18.

Rafiei, A., Rezaee, A., Hajati, F., Gheisari, S. & Golzan, M. (2021). SSP: Early prediction of sepsis using fully connected LSTM-CNN model, *Computers in Biology and Medicine*, 128(104110), 1-10.

Saqib, M., Sha, Y. & Wang, M.D. (2018). Early Prediction of Sepsis in EMR Records Using Traditional ML Techniques and Deep Learning LSTM Networks, *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018 Jul, 4038-4041.

Shimabukuro, D.W., Barton, C.W., Feldman, M.D., Mataraso, S.J. & Das, R. (2017). Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial, *BMJ Open Respiratory Research*, 4(1), e000234.

Smith, G.B., Prytherch, D.R., Meredith, P. Schmidt, P.E. & Featherstone, P. I. (2013). The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death, *Resuscitation*, 84(4), 465–470.

Taneja, I., Reddy, B., Damhorst, G., Dave Zhao, S., Hassan, U., Price, Z., Jensen, T., Ghonge, T., Patel, M., Wachspress, S., Winter, J., Rappleye, M., Smith, G., Healey, R., Ajmal, M., Khan, M., Patel, J., Rawal, H., Sarwar, R., ... Zhu, R. (2017). Combining Biomarkers with EMR Data to Identify Patients in Different Phases of Sepsis, *Scientific Reports*, 7(10800), 1-12.

Tao, X., Li, Q., Ren, C., Guo, W., He, Q., Liu, R. & Zou, J. (2020). Affinity and class

probability-based fuzzy support vector machine for imbalanced data sets, *Neural Networks*, 122, 289-307.

Taylor, R.A., Pare, J., Venkatesh, A., Mowafi, H., Melnick, E., Fleischman W. & Hall M.K. (2016). Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach, *Academic Emergency Medicine*., 23(3), 269–278.

Torio, C.M. & Moore, B.J. (2016). National Inpatient Hospital Costs: the Most Expensive Conditions by Payer, 2013: Statistical Brief# 204, *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*, 2006–2016.

Xu, B., Shirani, A., Lo, D. & Alipour, M.A. (2018). Prediction of relatedness in stack overflow: deep learning vs. SVM: a reproducibility study, *In Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '18)*, ACM, New York, USA, 1-10.

Ye, Y., Tian, M., Liu, Q. & Tai, H.-M. (2020). Pulmonary Nodule Detection Using V-Net and High-Level Descriptor Based SVM Classifier, *IEEE Access*, 8, 176033-176041.