

龔千芬、郝沛毅 (2024), 「結合自調節生成對抗深層混和網路模型與深層模糊支持向量機來早期偵測敗血症不良結果」, 資訊管理學報, 第三十一卷, 第二期, 頁 239-276。

結合自調節生成對抗深層混和網路模型與深層模糊支持向量機來早期偵測敗血症不良結果

龔千芬

國立高雄科技大學智慧商務系

郝沛毅*

國立高雄科技大學智慧商務系

摘要

敗血症是一種可能導致死亡的嚴重疾病，根據世界衛生組織的統計，每年大約有 600 萬人因為敗血症休克而死亡，死亡率高達 50%。敗血症發作的早期預警和及時的干預治療，可以避免絕大多數的敗血症休克死亡的發生。人工智慧演算法的蓬勃發展與重症監護病房的大量病患臨床數據的蒐集，讓應用人工智慧技術開發早期敗血症預警系統，成為一個避免敗血症死亡的有前途的解決方案。支持向量機可以建立最佳的分類器，深度學習可以自動學習關鍵的特徵，本論文則是結合兩者的長處。本論文開發一種新穎的敗血症早期預警系統，它是由全連接神經網路、雙向長短期記憶體、卷積神經網路、注意力機制、合作網路、生成對抗網路與深層模糊支持向量機所組成。本論文使用 PhysioNet/Computing in Cardiology Challenge 2019 敗血症資料集開發一個嶄新的自調節生成對抗深度混合神經網路架構來自動擷取關鍵的特徵，其中，卷積神經網路可以擷取與位置無關的局部特徵，雙向長短期記憶體可以使用前向與後向運算來補抓動態訊息的時間依賴性特徵，全連接神經網路適合分析靜態訊息內的關鍵特徵，注意力機制可以更關注於能提升預測效能的重要特徵。此外，由合作網路與生成對抗網路所組成的自調節雙通道模組可以生成隨機雜訊，並且增強特徵的強健性和泛化性。最後，自調節生成對抗模型擷取的關鍵特徵，將交給新穎的模糊支持向量機來打造最佳的敗血症預警系統，另外，我們使用深層多核心學習演算法來將模糊支持向量機擴充到深層網路結構，以進一步提升支持向量機的推理能力。本論文開發的預警系統可以在 6 小時之前預測敗血症的發作，並為避免敗血症休克發作導致死亡的風險，提供理想的早期預警的解決方案。

關鍵詞：敗血症早期預測、深度學習、模糊支持向量機、醫學訊息學、生成對抗網路

* 本文通訊作者。電子郵件信箱：haupy@nkust.edu.tw
2023/6/2 投稿；2023/9/4 修訂；2023/9/30 接受

Kung, C.F. & Hao, P.Y. (2024). Early Detection of Sepsis Utilizing Self-Regulated Generative Adversarial Network and Deep Fuzzy Support Vector Machine. *Journal of Information Management*, 31(2), 239-276.

Early Detection of Sepsis Utilizing Self-Regulated Generative Adversarial Network and Deep Fuzzy Support Vector Machine

Chien-Feng Kung

Department of Intelligent Commerce, National Kaohsiung University of Science and
Technology

Pei-Yi Hao*

Department of Intelligent Commerce, National Kaohsiung University of Science and
Technology

Abstract

Sepsis is a serious disease that can lead to death. According to the statistics of the World Health Organization, about 6 million people die each year due to septic shock, and the mortality rate is as high as 50%. Early warning and early intervention of sepsis onset can avoid the vast majority of septic shock deaths. The vigorous development of artificial intelligence algorithms and the collection of clinical data from a large number of patients in intensive care units have made the application of artificial intelligence technology to develop an early warning system for sepsis a promising solution to avoid sepsis deaths.

Support vector machine can establish the best classifier. Deep learning algorithms can automatically learn key features. This research integrates the strengths of both models. This research proposes a novel sepsis early warning system, which is comprised of convolutional neural network, bidirectional long-term short-term memory, attention mechanism, fully connected neural network, generative adversarial network and deep fuzzy support vector machine. This research employs the PhysioNet/Computing in Cardiology Challenge 2019 sepsis dataset and develops a new self-regulating generative adversarial deep hybrid neural network architecture to automatically extract key features. Convolutional neural networks can capture position-invariant local features. Bidirectional long-term short-term memory can use forward and backward operations to capture time-dependent features of dynamic information.

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

2023/6/2 received; 2023/9/4 revised; 2023/9/30 accepted

Fully connected neural networks are suitable for analyzing static information. The attention mechanism can pay more attention to the important features that can improve the prediction performance. In addition, a self-regulated dual-channel module composed of a cooperative network and a generative adversarial network can generate random noise and enhance the robustness and generalization of features. Finally, the key features extracted by the self-adjusting generative adversarial model will be handed over to the novel fuzzy support vector machine to create the best sepsis early warning system. The early warning system developed in this paper can predict the onset of sepsis 6 hours earlier, and provide an ideal early warning solution to avoid the risk of death caused by septic shock.

Keywords: Sepsis early warning model, Deep learning, Fuzzy support vector machines, Medical informatics, Generative adversarial network

壹、前言

敗血症(Sepsis, 也稱膿毒症)是一種可能導致死亡的嚴重疾病, 根據國際共識會議(international consensus conference)的定義, 敗血症是由於患者對於感染的反應失調所引發的器官功能障礙, 嚴重的話會危及性命(Singer et al. 2016)。每年, 它影響著美國超過 100 萬患者和全球超過 3000 萬患者(De Backer & Dorman 2017), 相關死亡率約 20-30% 左右, 最嚴重的敗血症休克的死亡率更是高達 50%, 在過去的幾十年裡, 敗血症的發病率呈現穩步上升的趨勢(Elfeky et al. 2017)。此外, 敗血症給美國醫療保健系統帶來了沉重的經濟負擔。根據報導, 美國敗血症患者每年的醫療費用超過 240 億美元(Lagu et al. 2012)。另一個重要的事實是, 敗血症倖存者有更多機會遭受生活品質受損, 而且發生重大心血管病變的風險也更高(Ou et al. 2016)。

一般來說, 敗血症是由局部感染源的微生物侵入血液, 導致器官功能障礙而產生的。感染通常可以通過臨床症狀和微生物檢測來識別。因此, 目前敗血症診斷的黃金標準是血液培養。儘管如此, 血液培養的敏感性非常低, 超過一半的敗血症症狀患者的培養結果為陰性(Vincent et al. 2006)。此外, 血液培養需要執行許多步驟來分析, 因此成本高、耗時長, 並且需要專業操作(Kumar et al. 2019)。由於死亡率的顯著增加與治療的延遲有關, 血液培養的冗長時間導致廣譜抗菌藥物的早期使用。這導致一個不可避免後果, 就是不必要的使用廣譜抗菌藥物, 或是長期使用廣譜抗菌藥物, 這可能會造成一系列的副作用, 例如耐藥病原體的誕生和困難梭狀芽孢桿菌感染症(*Clostridium difficile* infection, CDI) (Dark et al. 2015)。另一種替代方案是採用高靈敏度生物傳感器來進行敗血症診斷, 這是可以克服血液培養缺點的方法, 並且為世界各地的臨床醫生提供幫助(Kumar et al. 2019)。近年來, 使用不同的生物標誌物開發了許多用於敗血症診斷的生物傳感器。生物標誌物被定義為作為病理或生理過程指標的生物分子或特徵 (Kumar et al. 2019)。敗血症生物標誌物的水平如何升高, 可用於回答三個臨床重要問題: (1)患者是否被感染; (2)病情是否嚴重; (3)患者是否對治療有反應(Vincent et al. 2006)。雖然敗血症已被證實與生物標誌物相關, 但敗血症的臨床表現複雜, 僅憑生物標誌物是很難判斷患者是否患有敗血症(Chen et al. 2017)。

在重症監護中, 早期識別敗血症發作對於控制疾病風險是至關重要的, 因為敗血症的治療具有高度時間敏感性。根據國際醫學指南, 建議在最初 3 小時內開始早期輸液治療, 以穩定敗血症引起的組織貫注不足(tissue hypoperfusion), 並建議儘早開始使用靜脈內抗菌藥物(intravenous antibacterial agents), 尤其是敗血症後 1 小時內。實際上, 治療每延遲一小時, 死亡率就會增加約 4-8%。為了把握控制疾病進展的“黃金時期”, 臨床上通常使用基於規則的敗血症評分系統, 包括全身炎症反應綜合徵(systemic inflammatory response syndrome, SIRS)標準、連續器官衰竭評估(sequential organ failure assessment, SOFA)評分、與改良早期預警評分(modified early warning score, MEWS), 以提醒可能發生敗血症。通

過提供及時的干預，這些早期預警評分可以幫助實施早期預警計劃，或具有高敏感性的特定院前治療。然而，這些標準的特異性較差。例如，病毒性流感的生理指標往往會引起誤報。

人工智能技術成為包括敗血症早期診斷在內的醫療預警的有效方法，通過整合電子病歷 (electronic health records, EHR)、醫學影像、病理生理等數據，發展出智能系統來分析和預測人體的健康狀況，並獲得準確的處方信息，幫助臨床醫生做出快速有效的判斷。近年來，已經有各種不同的人工智慧演算法被應用來建立適當的預測模型 (分類器)，以準確預測敗血症的發作，例如梯度提升樹(Barton et al. 2019)，支持向量機(Taneja et al. 2017)、隨機森林(Taylor et al. 2016)；人工神經網路(Saqib et al. 2018; Rafiei et al. 2021) 和邏輯回歸 (Faisal et al. 2018)。然而，這些方法有一些局限性。例如，隨機森林放棄了決策樹的主要優勢——可解釋性。人工神經網路通常被認為是「黑匣子」，很難解讀具體的分類決策是如何做出的。對於類別不是線性可分割的數據集，邏輯回歸可能具有較低的預測準確性。此外，傳統的機器學習模型嚴重依賴透過人工的特徵工程(feature engineering)來找出能夠幫助決策的關鍵特徵。但是，由於人體內部複雜的生理與化學反應，以及各種疾病與藥物之間複雜的互動關係，使得透過人工的特徵工程方式來建構強大的特徵組合，變成一件耗時、費力又困難的工作。

深度網路學習演算法具有自動學習重要特徵的優異能力，因此，越來越多深度網路學習演算法被應用於各種不同疾病的早期預警(Wang & Yao 2022; Apalak & Kiasaleh 2022)。深度學習教父 Hinton教授在《Science》期刊上闡述：深度網路的核心能力為透過逐層的組合低階層原始特徵，來獲得更加抽象的高級特徵，這樣可以提升預測效能 (Hinton & Salakhutdinov 2006)；也就是說，「深度網路」是過程，「特徵學習」才是目標。深度網路擁有自動學習重要特徵的優異能力，並且效能遠超過傳統專家費時耗力的特徵工程。由於深度網路在處理特徵隱晦不明確的現實應用問題獲得巨大的成功，近年來，越來越多專家使用深度網路處理生命體徵等醫療時間序列資料，來打造敗血症早期預警模型(Wang & Yao 2022; Apalak & Kiasaleh 2022)。

深度網路模型的蓬勃發展，超越支持向量機(support vector machine, SVM)在人工智慧研究領域的領先優勢，不過，依然有論文指出，支持向量機的分類正確率是超過深度網路模型 (Xu & Cohen 2018)。這其實並不會讓人感到驚訝，雖然支持向量機並不能夠自動學習重要特徵，然而若能夠事先給予支持向量機理想的特徵，那麼，根據統計學習理論而研發的支持向量機可以建立最佳的分類模型。Huang & LeCun (2006)最早提出融合卷積神經網路(convolutional neural network, CNN)與支持向量機的研究，該研究指出CNN在找出圖片與影像中位置無關的特徵 (position invariant feature)能獲得優秀的成績，但是CNN並沒有辦法建構最佳的分類模型，在另一方面，儘管SVM無法自動擷取重要的特徵，如果給予SVM理想的特徵，SVM可以建構最佳分類模型。所以，他們結合CNN與SVM的長處，透過CNN擷取位置無關的特徵，再將關鍵特徵餵給SVM，來打造最佳的分類模

型，讓預測效能可以進一步提升。換句話說，深度學習的先驅，CNN網路的研發者 Yann LeCun也替SVM發聲擔保，因此，SVM的相關研究沒有由於深度網路的興起而落寞，相反地，越來越多研究者追隨大師的步伐，運用深度網路來學習重要特徵，再把重要特徵餵給SVM來建立最佳的分類模型，並且用來解決各種醫療應用中的早期預警研究問題 (Ye et al. 2020)。所以，支持向量機並沒有被深度網路所淘汰，依舊有不少對於支持向量機的相關研究論文，能夠刊登在國際頂級期刊 (Hao 2021; Hao et al. 2022)，這表示支持向量機的研究還有很多發展空間。所以，本論文要使用深度網路來學習重要特徵，再把重要特徵交給支持向量機，來建構最佳的分類模型，以結合這兩種技術的長處。

由於引發敗血症的原因是錯綜複雜，並且表現出許多不同的體徵，因此需要考慮各方面的異質資料來源以評斷敗血症的併發，顯而易見地，單一個深度模型不可能同時捕抓以下三種特徵：局部特徵、時間依賴特徵、與長時間的全局特徵。其中，卷積神經網路 (convolutional neural networks, CNN) 擅長擷取與位置無關的局部模式特徵，雙向長短期記憶體 (bi-directional long short-term memory, BiLSTM) 可以由「從過去至未來」與「從未來至過去」這兩個方向來分析時間依賴性與因果關係特徵，注意力機制 (attention mechanism, AM) 可以對長時間的全局資料中的重要因素給予較多的關注，而較不關注於不重要的因素，以進一步提升預測準確率。因此，本論文結合CNN、BiLSTM與AM來擷取不同特質的特徵，來打造更準確的敗血症早期預警系統。

近年來，生成對抗網路在機器學習領域嶄露頭角(Xu et al. 2022)，生成對抗網路 (Generative Adversarial Network, GAN) 是透過讓兩個模型相互進行博弈的策略來進行學習。對抗學習可以有效地揭示模型的缺陷，並且提高模型的強健性。因為，最了解你的缺點的，就是你的敵人，也是由於敵人的存在，所以你需要不斷的精進自己，讓自己更加強大，對抗訓練就是根據這樣的理念，在學習過程中注入對手的設計，以增加模型的強健性。因此，本論文提出結合對抗網路與合作網路的雙通道模型的嶄新架構，稱為**自我調節的生成對抗網路(Self-regulated Generative Adversarial Network)**，合作網路可以幫助我們從「正面」的方向學習可以幫助預測敗血症的特徵，對抗網路則是幫助我們從「反面」的方向避免學習到錯誤的特徵，使得我們的模型可以有效避免雜訊的影響，並且捕抓到更穩健的特徵。

我們將結合這些深度模型的優點，並且將它們擷取到的特徵，餵入到申請者在 *Neural Networks* 期刊提出的新穎的模糊支持向量機來建構最佳的預測模型 (Hao et al. 2022)，使用模糊支持向量機當作頂層分類器，擁有以下優點：

- 早期預警等醫療任務通常具有極其嚴重的類別不平衡(class imbalance)問題，相較於正常樣本，異常樣本(敗血症發作)是非常罕見的，換句話說，正類別內的資料(有敗血症發作的病患)數目與負類別內的資料(無敗血症發作的病患)數目相差非常懸殊，這使得分類邊界會往樣本數量龐大的類別去偏斜，因此獲得一個具有偏見的分類模型。然而，模糊支持向量機可以對所有訓練

資料指派其模糊歸屬權重，模糊歸屬程度越大的樣本，代表它越重要，在訓練的過程中，越不允許分類錯誤。本論文將數量稀少的正類別資料(敗血症病患資料)，賦予較大的模糊歸屬程度，較不允許它們被錯誤分類，相反地，對於數量眾多的負類別樣本(正常病患)，賦予較小的模糊歸屬程度，適度地允許它們被錯誤分類，來避免建構出一個具有偏見的分類模型。

- 早期預警等醫療任務通常具有極其嚴重的遺失值(missing value)問題，由於量測頻率的不同或是量測項目的差異，導致醫療資料集通常具有極大量的遺失值(missing value)，代表該量測項目在這個時間窗口是沒有測量值的，通常是透過前向/後向填補遺失值，或是由量測項目的中位數來填補遺失值，但是，這將導致訓練樣本內充滿不精確與不確定的性質，模糊支持向量機很適合處理充滿模糊與不精確的現實問題，我們可以將遺失值比率較高的訓練資料，賦予較小的模糊歸屬程度，讓他們在訓練過程可以允許被適度的錯誤分類，以避免分類器要過度擬合(over-fitting)那些不精確的訓練資料，而遺失值較少的訓練資料(較為精確的樣本)可以賦予較高的模糊歸屬程度，使得它們在建構分類器時有更高的影響力。
- 理想的決策需要同時顧慮不確定性、信心與風險。決策的目標是減少風險，若是我們沒有顧慮到信賴程度，就輕率做出決策，可能會在不經意的情況下，造成風險提高。模糊支持向量機能夠分類敗血症的發作與否，還可以計算病患屬於敗血症「發作」類別的模糊歸屬程度，這可以解讀為對於預測結果的信賴指數，所以，模糊支持向量機很適合用來開發醫療決策支援系統中的敗血症預警模型。

此外，儘管根據統計學理論研發的支持向量機淺層網路模型能夠打造最佳的分類模型，仍然有許多學者想要進一步提升支持向量機的性能，尤其是透過將支持向量機從淺層網路模型擴展為深度網路模型。近來，有許多關於深層支持向量機網路模型的研究被提出來(Li et al. 2019; Wang et al. 2019)，證明擁有深層網路結構的支持向量機有廣闊的研究前景。因此，本論文將透過核心函數(kernel function)來逐層擷取特徵，以打造擁有深層網路架構的模糊支持向量機，來進一步提升敗血症早期預警的預測效能。

貳、文獻探討

Shashikumar et al. (2021)提出 DeepAISE (Deep Artificial Intelligence Sepsis Expert)，它是一種用於敗血症早期預測的循環神經模型。DeepAISE 自動學習與臨床風險因素之間存在的高階相互作用關係，並且與時間模式相關的預測特徵，這些特徵可以最大限度地提高觀察到的敗血症事件發生的可能性。使用來自三個不同醫療保健系統的住院患者數據集進行研究，DeepAISE 以最低的誤報率 (FAR 在 0.20 和 0.25 之間)產生最準確的預測(AUC 在 0.87 和 0.90 之間)，同時產生可解釋的臨床時間序列和危險因素表示。Liu et al. (2021) 提出名為 HeMA (hierarchically enriched machine learning approach)的即時管理誤報的機器

學習方法，並進行早期敗血症預測的案例研究。他們開發一個兩階段的框架，其中第一階段機器學習模型與第二階段的統計測試（特別是 Kolmogorov-Smirnov 測試）配對，以預測患者是否會發展為敗血症。與單獨的機器學習模型相比，該框架在不影響 F1 分數的情況下提高了特異性和陽性預測值。Nesaragi et al. (2021)開發一個機器學習框架，從電子健康記錄 (electronic health records, EHR) 異質性數據中捕獲線性關係和非線性相關性，以用於早期預測敗血症。他們使用逐點互信息 (pointwise mutual information, PMI) 矩陣描述每 6 小時窗口持續時間與 EHR 數據中選定的 24 個協變量的成對關聯統計。之後，從重塑的核心張量中獲得的 120 個潛在特徵被饋送到輕梯度提昇機器學習模型 (LightGBM) 進行二元分類，進一步減輕類別不平衡的問題。機器學習框架是通過貝葉斯優化設計的，根據公開可用的 PhysioNet/Computing in Cardiology Challenge 2019 訓練數據，接受者操作特徵曲線下面積 (AUROC) 為 0.8621。Yao et al. (2021)提出一種新的機率建模方法用於建立敗血症診斷的分類器，這種方法的特點是獨特的強解釋性，體現在三個方面：(1) 基於似然分析的證據獲取，(2) 基於機率規則的推理，以及 (3) 使用機器學習算法的參數優化。實驗結果表明，就分類能力而言，新方法建立的分類器在敗血症診斷方面的表現普遍優於大多數替代分類器。由於機率模型還具有固有的可解釋性，因此它可以用作為敗血症診斷中的決策支援的工具。El-Rashidy et al. (2022)建立一個更準確的預測模型來識別敗血症。首先，它們同時使用 NSGA-II (一種多目標遺傳算法優化方法) 和人工神經網路從患者數據中提取最佳特徵子集。在下一階段中，它們基於選擇的最優特徵集構建深度學習模型。他們提出的模型有兩層。第一層模型是用於預測敗血症的深度學習分類網路。這是神經網路模型的堆疊集合，可預測哪些患者會發展為敗血症。對於被預測患有敗血症的患者，將檢索他們進入 ICU 後前六個小時的數據，然後將此數據用於進一步的模型優化。在第二層的模型中，他們使用多任務迴歸深度學習模型來確定第一層預測患有敗血症的患者的敗血症發作時間和當時的血壓。使用來自重症監護 MIMIC III 真實世界數據集的醫學信息進行的測試，他們提出的分類模型的準確性、特異性、敏感性和 AUC 分別達到 0.913、0.921、0.832、0.906。此外，多任務迴歸模型預測敗血症發病時間和當時血壓的均方根誤差分別為 10.26 和 9.22。Wang & Yao (2022)提出一種採用多分支時間卷積網路 (Multi-Branching Temporal Convolutional Network, MB-TCN) 的嶄新預測框架，來對結構複雜的醫學數據進行建模，以實現敗血症的穩健預測。MB-TCN 框架不僅可以有效地處理遺失值和不平衡數據問題，還可以有效地捕獲時間模式和異構變量交互作用。他們使用來自 PhysioNet/Computing in Cardiology Challenge 2019 的真實醫學數據評估 MB-TCN 在預測敗血症方面的性能。實驗結果表明，MB-TCN 優於當前實踐中常用的預測方法。Liu et al. (2022a)提出基於 XGBoost 機器學習框架，以人口統計、生命體徵、實驗室檢測數據和醫療干預數據作為輸入特徵，提出一種動態預測敗血症和評估風險的嶄新模型。他們使用兩種特徵構建方法，對於觀察到的生命體徵和實驗室檢測的時間序列數據，他們執行時間相關

統計篩選後，構建時間相關的特徵。對於臨床干預數據，應用統計計數方法構建計數依賴特徵。此外，他們為 XGBoost 框架提出新的目標函數，並給出了目標函數的一階和二階梯度用於模型訓練。與目前最先進的預測方法相比，他們提出的模型具有最好的性能，AUROC 在 MIMIC-III 數據集上提高了 5.4%，在 PhysioNet Challenge 2019 數據集上提高了 2.1%。該模型的數據處理和訓練方法可以方便地應用於不同的電子健康檔案系統，具有廣闊的應用前景。Kim et al. (2022)提出一種神經結構搜索(neural architecture search, NAS) 模型，透過應用遺傳算法 (genetic algorithm, GA) 以低計算成本和高搜索性能預測敗血症的發作。他們提出的模型在神經網路內部共享所有可能的連接節點的權重。在外部，則是通過 GA 的基因型之間的權重共享效應降低了搜索成本。他們使用醫學時間序列數據集重症監護醫學信息資料集(Medical Information Mart for Intensive Care III, MIMIC-III) 進行預測分析，其主要目標是在敗血症發生前 3 小時預測發病。此外，在不同的預測時間(0-12 小時)下進行實驗以進行比較，他們所提出的模型在接受者操作特徵曲線下的面積 (AUROC) 得分為 0.94 (95% CI: 0.92–0.96)，比使用連續器官衰竭評估 (SOFA) 獲得的得分高 0.31–0.26。此外，與基於長短期記憶神經網路的簡單模型相比，他們提出的模型的 AUROC 值提高 12%。Liu et al. (2022b)開發一個稱為 MLePOMDP 的嶄新框架，他們使用隱藏馬爾可夫模型在高層次上描述敗血症的發展，並且從時間數據中進行高階“觀察”。此外，他們開發部分可觀察馬爾可夫決策過程 (partially observable Markov decision process, POMDP) 模型來做出分類決策。MLePOMDP 使用從床邊監視器收集的高頻生理數據進行校準和測試。與過去基於 POMDP 的框架不同，MLePOMDP 產生高度可解釋的結果，並解釋了一個新穎且具有臨床意義的動作空間。Sharma et al. (2022)提出一種以長短期記憶 (long short term memory, LSTM) 為核心模型的智能網路，用於對 ICU 收治的患者進行敗血症的早期預測。該網路對時間序列數據進行操作，並根據患者的歷史醫療數據預測患者發生敗血症的機率。它包括基於物聯網的設備，這些設備已被證明在獲取醫療保健領域的連續和即時數據方面具有影響力。實驗結果顯示 AUROC 為 0.864，準確率為 95.1%。Apalak & Kiasaleh (2022)設計一種基於深度神經網路的新方法，該方法改進重症監護病房 (ICU) 患者敗血症的早期預測。假設患者數據集因信息缺失而嚴重損壞，這對敗血症發作的檢測產生負面影響。他們提出一個生成學習框架來估計數據中缺失的信息。透過條件生成對抗網路 (Generative Adversarial Networks, GAN)，它在以類標籤為條件時利用長短期記憶 (LSTM) 網路作為生成器和鑑別器。深度 LSTM 網路也用於預測目的。預測網路使用條件 GAN 的輸出進行訓練，並在未見過的測試集上進行評估，以研究所提出模型的性能。他們提出的框架不僅可以識別長期時間依賴性，還可以發掘缺失的模式。對於敗血症的 4 小時、8 小時和 12 小時預測，他們所提出的方法的 AUROC 分別為 94.49、93.74 和 94.01。

上述文獻回顧結果顯示，因為深層網路模型具有自動擷取重要特徵的優異能力，特別適用於特徵不明顯的醫療早期預警任務，因此，近年來頂尖的敗血症早

期預警模型都是基於深層網路模型開發的，但是，因為醫療資料集同時包含動態與靜態資料，顯而易見的，單一個深層網路模型是不可能同時分析靜態與動態資料。其中，分析靜態資料的首選是全連結神經網路，分析動態資料是 CNN 與 BiLSTM 網路模型的強項，注意力機制(AM)可以讓模型關注更重要的因素，進而提升模型預測效能，生成對抗網路(GAN)近年來也在機器學習領域嶄露頭角，透過兩個神經網路相互博弈的方式來進行學習，可以有效地提高特徵學習的強健性。所以，本論文結合 CNN、BiLSTM、AM、GAN 與 FCN，來自動分析重要特徵。然後再將特徵交給最頂層的深層模糊支持向量機，來建構最佳的敗血症早期預警模型。

參、研究方法

敗血症(Sepsis，也稱膿毒症)是指當病原體入侵身體後，患者對於感染的反應失調所引發的器官功能障礙，嚴重的話會危及性命。根據統計，敗血症的死亡率高達 30%-50% (Martin et al. 2003)；早期預警是降低敗血症死亡率的最佳解決方案，每延後一小時進行抗生素治療，敗血症患者的死亡率就會升高 4-8%。此外，根據臨床證據顯示，病患在敗血症發作之前的一段期間內，生命徵象會產生一些細微的變化，讓醫師可以早期判斷敗血症的發作並且進行相關的醫療處置，並且降低敗血症病患的死亡機率。然而，因為生命體徵的數量眾多，而且變化極其細微，使得醫師時常會無法看出端倪，因而錯失早期預警並且進行干預醫療的黃金時間。另外，醫師在做出診斷決策時，並不是僅參考單方面的醫療資料就可以做出診斷決策，而是透過整合患者的各種醫療資訊(例如，生命體徵、實驗室的血液檢測報告等)，再做出適當的診斷決策。然而，當整合如此眾多的病患醫療資訊後，容易讓醫師無法在短時間內發掘龐大數據中的細微變化，因而導致診斷上的盲點。而要分析像病患資訊這種包含許多不同結構的資料，傳統的統計方法也是力有未逮。面對具有大量且複雜的變量(例如 ICU 數據)的任務，人工智慧是傳統的統計方法的一種實用的替代方法。利用 AI 來代替傳統統計進行異質資料的整合與分析，可以協助醫師消除診斷的盲點，並讓醫師在診斷上有更多可靠的依據，進而做出更加精確的診斷，也讓敗血症病患們可以有更多的治癒機會。因此，本論文提出結合卷積神經網路(Convolutional Neural Network, CNN)、雙向長短期記憶(Bi-directional Long Short-Term Memory, BiLSTM)與注意力機制(Attention Mechanism, AM)之混合模型來自動學習重要的特徵，並且運用自調適生成對抗模型的雙通道模型來獲得更穩健的特徵，其中，通道一的合作網路讓我們以「正面」的方向學習可以幫助預測的特徵，通道二的對抗網路讓我們以「反面」的方向避免學習到錯誤的特徵。最終，我們將捕獲的特徵餵給新穎的深層模糊支持向量機，來打造最佳的敗血症早期預警系統。

一、資料集與特徵

為了公平的與其他敗血症預警系統進行比較，本研究將使用敗血症公開資料集 PhysioNet/Computing in Cardiology Challenge 2019 (Reyna et al. 2019)來進行實驗比較。PhysioNet/Computing in Cardiology Challenge2019 挑戰賽資料集是從三個地理位置不同的美國醫院系統獲得病患數據，這些醫院系統包含：貝斯以色列女執事醫療中心 (Beth Israel Deaconess Medical Center, 醫院代號 A)、埃默里大學醫院 (Emory University Hospital, 醫院代號 B)與第三個匿名的醫院系統 (醫院代號 C)。這些病患數據是在過去十年中收集，並且獲得相應機構審查委員會的批准。PhysioNet/Computing in Cardiology Challenge2019 挑戰賽資料集使用 Sepsis-3 臨床標準對數據進行去識別化和標記。來自醫院系統 A 和 B 的 40,336 名患者的數據和標籤是公開發布並且可以自由下載，來自醫院系統 A、B 和 C 的 24,819 名患者的數據和標籤被隔離為隱藏測試集。

表 1：PhysioNet/Computing in Cardiology Challenge2019 挑戰賽資料集的變量及其定義

生命體徵 (第 1-8 列)					
編號	變數名稱	描述	編號	變數名稱	描述
1	HR	心跳 (每分鐘心跳次數)	5	MAP	平均動脈壓 (mm Hg)
2	O2Sat	脈搏血氧儀 (%)	6	DBP	舒張壓 (mm Hg)
3	Temp	體溫 (攝氏度)	7	Resp	呼吸頻率 (每分鐘呼吸次數)
4	SBP	收縮壓 (毫米汞柱)	8	EtCO2	呼氣末二氧化碳 (mm Hg)
實驗室值 (第 9-34 列)					
編號	變數名稱	描述	編號	變數名稱	描述
9	BaseExcess	過量的碳酸氫鹽 (mmol/L)	22	Glucose	血清葡萄糖 (mg/dL)
10	HCO3	碳酸氫鹽 (mmol/L)	23	Lactate	乳酸 (mg/dL)
11	FiO2	吸入氧分數 (%)	24	Magnesium	鎂 (mmol/dL)
12	PH	氫離子濃度指數	25	Phosphate	磷酸鹽 (毫克/分升)
13	PaCO2	動脈血二氧化碳分壓 (mm Hg)	26	Potassium	鉀 (毫摩爾/升)
14	SaO2	動脈血氧飽和度 (%)	27	Bilirubin total	總膽紅素 (mg/dL)
15	AST	天冬氨酸轉氨酶 (IU/L)	28	TroponinI	肌鈣蛋白 I (ng/mL)
16	BUN	血尿素氮 (mg/dL)	29	Hct	血細胞比容 (%)
17	Alkalinephos	鹼性磷酸酶 (IU/L)	30	Hgb	血紅蛋白 (g/dL)
18	Calcium	鈣 (毫克/分升)	31	PTT	部分凝血活酶時間 (秒)
19	Chloride	氯化物 (毫摩爾/升)	32	WBC	白細胞計數 (計數 *10 ³ /μL)
20	Creatinine	肌酐 (毫克/分升)	33	Fibrinogen	纖維蛋白原 (毫克/分升)
21	Bilirubin direct	直接膽紅素 (mg/dL)	34	Platelets	血小板 (計數 *10 ³ /μL)
人口統計 (第 35-40 列)					
編號	變數名稱	描述	編號	變數名稱	描述
35	Age	年齡	38	Unit2	ICU 的行政標識符
36	Gender	性別 女 (0) 或男 (1)	39	HospAdmTime	醫院入院和 ICU 入院之間的小時數
37	Unit1	ICU 的行政標識符	40	ICULOS	ICU 住院時間 (自 ICU 入院後的小時數)
類別標籤 (第 41 列)					
編號	變數名稱	描述			
41	SepsisLabel	對於敗血症患者，如果 $t \geq t_{\text{sepsis}} - 6$ ，SepsisLabel 為 1，如果 $t < t_{\text{sepsis}} - 6$ ，則為 0。對於非敗血症患者，SepsisLabel 為 0。 提示： t_{sepsis} 為敗血症發病時間，對於敗血症病患，發病前 6 小時內，類別標籤為 1，發病前 6 小時之外的時間，類別標籤為 0。			

病患數據包括每小時生命體徵摘要、實驗室值和靜態患者描述的組合，總共有 40 個臨床變量，表 1 描述 PhysioNet/Computing in Cardiology Challenge 2019 資料集的變量。在 PhysioNet/Computing in Cardiology Challenge 2019 挑戰賽資料集中，每一個病患的資料都是用一個表格儲存，如表 2 所示，其中包含病患隨時間變化的測量值。時間視窗是以小時為單位，表格中的每一個水平行，都代表在同一個時間視窗的一組量測值(例如，同一小時的心率與氧氣水平)，表格中的每一個垂直列，則是對應到一個量測變數(例如，HR 是心跳頻率)。而每一個垂直列的數值，則是提供該量測變數隨時間變化的一系列測量結果(例如，不同小時的心率量測結果)。值得注意的是，針對非敗血症病患，其類別標籤 SepsisLabel 值永遠為 0，而對於敗血症病患，只有其發病前 6 小時內的類別標籤 SepsisLabel 值為 1，其餘時間窗口的類別標籤 SepsisLabel 值為 0，所以在表 2 中，最後一個垂直列(類別標籤 SepsisLabel)一開始的值為 0，因為該時間窗口還未到發病前 6 小時內，等到第 3 水平行(第 3 個時間窗口)後，類別標籤 SepsisLabel 的值才變為 1，代表已經到了發病前 6 小時內的時間窗口。儲存格中的 NaN (不是數字)表示在該時間窗口內，沒有記錄變量的測量值(亦即是，遺失值)。

表 2：PhysioNet/Computing in Cardiology Challenge 2019 挑戰賽資料集的病患資料表格

HR	O2Sat	Temp	...	HospAdmTime	ICULOS	SepsisLabel
NaN	NaN	NaN	...	-50	1	0
86	98	NaN	...	-50	2	0
75	NaN	NaN	...	-50	3	1
99	100	35.0	...	-50	4	1

總結來說，PhysioNet/Computing in Cardiology Challenge 2019 挑戰賽資料集包含靜態資料(例如，人口統計數據)與動態資料(例如，在不同時間量測的生命體徵與實驗室數值)。靜態資料中的數值(另如，年齡與性別)，在每次就診期間只收集一次，而且於就診期間維持不變，動態資料中的數值(例如，生命體徵或實驗室數值)則是於病患就診期間進行多次量測與收集。通常，靜態資料用向量儲存，動態資料用時間序列儲存。

二、使用自我調節的生成對抗網路來學習關鍵特徵

本論文提出一個新穎的自我調節(self-regulated)的生成對抗網路，來更加強健的學習到重要的特徵，並且避免雜訊的干擾。自調節學習模型實際上是一個雙通道模型，如圖 1 所示，它是由合作網路(cooperative network, (Islam et al. 2003))與成對抗網路(generative adversarial network, GAN (Goodfellow et al. 2014))所組成。並採用記憶抑制器(Memory suppressor, S)來調節通道之間的通信，該通道負責減輕包含在身體訊號的隨機雜訊的干擾。

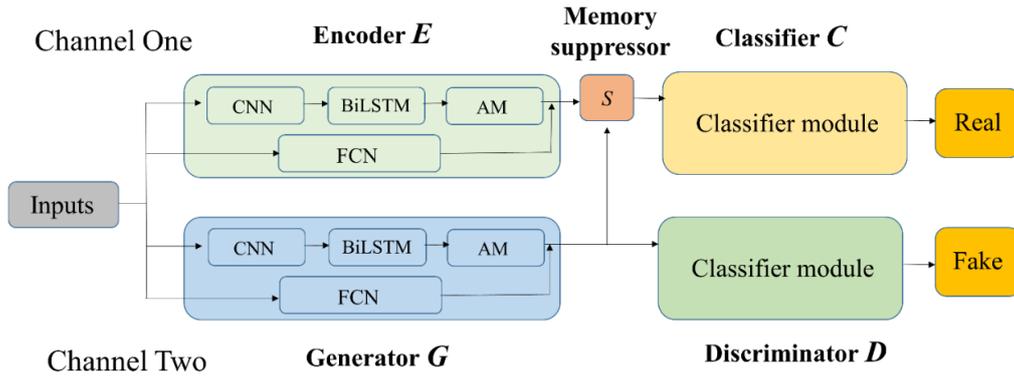


圖 1：本研究提出的自我調控的生成對抗模型的架構。

(一) 合作網路

通道一中的合作網路包含一個編碼器(Encoder, E)與分類器(Classifier, C)，其中，編碼器 E 是由卷積神經網路(Convolutional Neural Network, CNN)、雙向長短期記憶體(bi-directional long short-term memory, BiLSTM)、全連結網路 (fully connected network, FCN)與注意力機制(Attention mechanism, AM)等組件所組成(我們將在章節(四)詳細介紹各個模組的功能)。編碼器 E 具有可學習的模型參數 θ_E ，其目的是要捕捉病患信息(包含生命徵象、實驗室檢測數據與人口統計資料)的表示特徵 o_E ，換句話說：

$$o_E = E(p_{Vital}, p_{Laboratory}, p_{Demographics}; \theta_E) \quad (1)$$

其中， $E(\cdot)$ 表示編碼器的運算函式， p_{Vital} 表示病患的生命徵象數據， $p_{Laboratory}$ 表示病患的實驗室檢測數據， $p_{Demographics}$ 表示病患的人口統計數據， θ_E 是編碼器 E 的可學習的模型參數。通道一中的分類器 C 是一個單層感知器，分類器 C 具有可學習的模型參數 θ_C ，它的功能是在捕獲病患的表示特徵 o_E 後，試圖準確預測病患是否會在 6 小時內敗血症發作，也就是說：

$$\hat{y}_d = C(o_E; \theta_C) \quad (2)$$

並為第 d 個病患指派一個最可能的標籤(敗血症發作與否)，其中， $C(\cdot)$ 表示分類器的運算函式。因此，通道一中的編碼器 E 和分類器 C 在訓練過程中相互合作，為同一目標更新模型參數 θ_E 和 θ_C ，即是，最小化敗血症預測任務中的損失函數 $L(\hat{y}_d, y_d)$ ，所以，模型參數 θ_E 和 θ_C 的學習目標如下：

$$\theta_E, \theta_C = \operatorname{argmin} L(\hat{y}_d, y_d) \quad (3)$$

其中， y_d 表示二進制類別標籤的真實結果(one-hot 分佈)， $L(\cdot)$ 表示預測結果與真實結果的偏差。

(二) 生成對抗網路

在通道二中，生成器(Generator, G)和判別器(Discriminator, D)與編碼器 E 和分類器 C 具有相同的網路結構，它們也是基於可微函數 $G(\cdot)$ 與 $D(\cdot)$ 來完成下列運算：

$$o_G = G(p_{Vital}, p_{Biomarker}, p_{History}; \theta_G) \quad (4)$$

$$\hat{y}_d = D(o_G; \theta_D) \quad (5)$$

然而，根據生成對抗網路(GAN)的特點，它們陷入高度對抗性的競爭循環中。生成器 G 可以被視為「隨機性製造者」。它學習產生具有強烈隨機性的虛假特徵，這些特徵用於污染病患的表示特徵。因此，生成器 G 將真實的病患數據（包括生命徵象、實驗室檢測數據與人口統計資料）更改為假 z 。通過虛假的表示特徵 o_G ，生成器 G 反覆煽動判別器 D 產生錯誤的預測結果。同時，判別器 D 被視為「無助的醫師」，無論遇到真實的病患數據還是虛假的樣本 z ，都必須避免被欺騙，努力地正確預測敗血症發作與否。為了智取對手，生成器 G 還需要更新參數 θ_G 以最大化損失函數，但相反地，判別器 D 則是更新參數 θ_D 以最小化損失函數，在本研究中，我們使用交叉熵(Cross entropy)來作為衡量分類損失的目標函數，也就是說，我們的生成器 G 與判別器 D 的目標函數定義為

$$L(\hat{y}_d, y_d) = -\sum_{k=1}^M \sum_{j=1}^2 y_{d,k}^j \log(\hat{y}_{d,k}^j) = -\sum_{k=1}^M \sum_{j=1}^2 y_{d,k}^j \log(D(G(p_k, \theta_G); \theta_D)) ,$$

其中， y_d 表示二進制類別標籤的真實結果(one-hot 分佈)， p_k 代表病患資訊， M 是訓練資料的數目， $G()$ 是代表生成器的可微分函數， θ_G 是生成器 G 中可學習的模型參數， $D()$ 是代表判別器的可微分函數， θ_D 是判別器 D 中可學習的模型參數，生成對抗模型 GAN 中生成器 G 和判別器 D 之間的兩人最小-最大博弈，即是對於目標函數求最小-最大化(min-max)的混和問題，亦即是說， $\max_{\theta_G} \min_{\theta_D} L(\hat{y}_d, y_d)$ 。

所以，我們可以透過下列算式來學習生成器與判別器中的模型參數

$$\theta_G = \operatorname{argmax} L(\hat{y}_d, y_d) \quad (6)$$

$$\theta_D = \operatorname{argmin} L(\hat{y}_d, y_d) \quad (7)$$

已經有大量成功的研究證實，生成對抗模型 GAN 中 G 和 D 之間的兩人最小-最大博弈，有利於提高他們方法的性能(Xu et al. 2022)。

(三) 基於記憶抑制器的通信調節

透過記憶抑制器(Memory suppressor, S)的機制，我們嘗試優化編碼器 E ，如圖 1 所示。優化的目標是促進編碼器 E 與隨機生成器 G 盡可能地不同。更精確地說，抑制器 S 採用生成器 G 的輸出 o_G 作為參考，並假設其充滿隨機訊息。基於此設置，抑制器 S 檢查通道一中編碼器 E 的輸出 o_E ，以驗證 o_E 中的訊息是否與通道二中生成器 G 的 o_G 可以區分開來。如果 o_E 和 o_G 非常不同，則抑制器 S 允許編碼器 E 保存參數 θ_E ，否則，抑制器 S 可以強行擦除編碼器 E 中可能導致隨機性的部分記憶，這步驟也稱之為自調節(self-regulated)。本質上，抑制器 S 是一個沒有任何參數的虛擬模組，在數據流的前饋階段，我們只使用抑制器 S 來計算 o_E 和 o_G 的自調節損失。在反向傳播的參數更新階段，抑制器 S 可以看作是優化器，用來更新編碼器 E 的內存記憶。要保存或更新記憶體，則是利用 o_E 和 o_G 的向量逼近來檢查自調節損失函數 L_{diff} 。如果 o_E 和 o_G 之間的相似度越高，則自調節損失越大。在訓練過程中，編碼器 E 需要通過最小化自調節損失來更新參數 θ_E ，也就是說：

$$\theta_E = \operatorname{argmin} L_{diff}(o_E, o_G) \quad (8)$$

我們將在章節(六)詳細介紹自調節損失的計算過程。

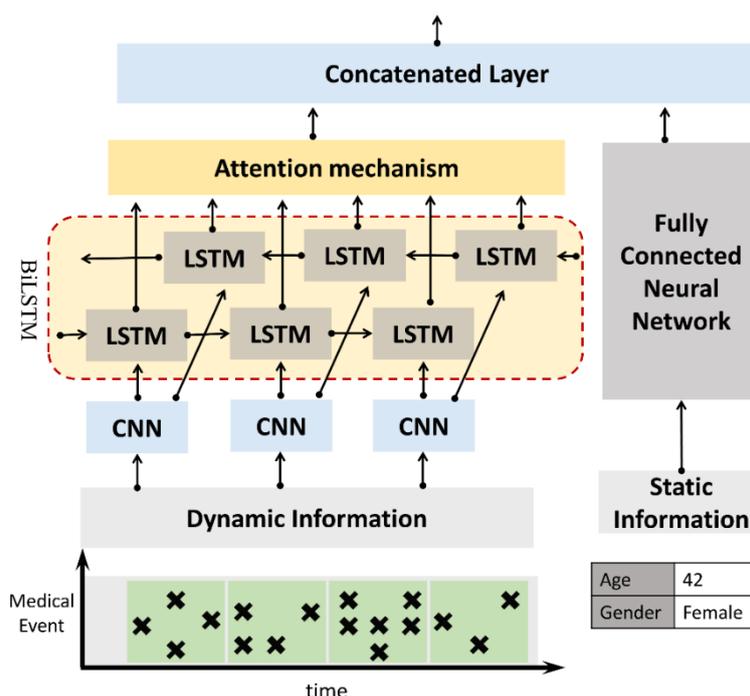


圖 2：結合 CNN、BiLSTM、AM 與 FCN 來擷取重要的特徵

(四) 基於記憶抑制器的通信調節

由於通道一（即合作網路）和通道二（即生成對抗網路 GAN）的結構完全相同，我們只介紹合作網路的組成元件，如圖 1 所示。合作網路由二個模組所組成，包括編碼器 E 與分類器 C ，我們將在以下幾個小節中介紹這些模組。

1. 編碼器模組

編碼器模組 E 是由卷積神經網路(Convolutional neural network, CNN)、雙向長短期記憶體(bi-directional long short-term memory, BiLSTM)、全連結網路 (Fully connected network, FCN)與注意力機制(Attention mechanism, AM)等組件所組成，如圖 2 所示。其中，CNN 用來擷取短期的局部特徵，BiLSTM 捕抓時間依賴特徵，AM 用來抓取全局中的重要特徵，FCN 則是從靜態資料中提取特徵。首先，我們把動態訊息的時間序列餵給卷積神經網路(Convolutional Neural Network, CNN)，CNN 在圖片/影像分類任務中取得顯著的勝利，這是因為 CNN 可以有效地捕獲與位置無關的局部特徵，這些特徵是與位置無關的，代表它們可能在圖像中任何位置出現，而且能夠從局部的圖像中擷取出來，而不需要使用全部圖像。CNN 是由三種不同類型的網路層組成：卷積層(Convolutional Layer)，線性整流單元(Rectified Linear Unit, ReLU)層和池化層(Pooling Layer)。卷積層採用一組卷積濾波器(filter)來尋找特定模式，將 CNN 應用於圖像分類任務時，濾波器可以在寬度和高度維度上移動。但是，當將 CNN 應用於 HER 的動態時間序列數據分析時，濾波器僅在時間維度上移動。濾波器橫跨整個時間特徵維度，但僅連接到數據的一小部分時間區域，濾波器對於該時間區域執行卷積運算來找出特定模式(pattern)。換句話說，濾波器讓 CNN 可以有效地學習局部和與時間不變特徵(time-

invariant features)。在輸入 HER 時間序列資料後，濾波器沿著時間序列移動，執行卷積運算來紀錄特定模式出現位置。假設 $\mathbf{x}_{i:i+j}$ 代表醫療時間串列中的一小部分時間區域 $x_i, x_{i+1}, \dots, x_{i+j}$ ，濾波器 $\mathbf{w} \in \mathbb{R}^{h \times k}$ 能夠對一個由 h 個事件組成的時間窗口執行卷積運算，以發掘特定的模式。 k 代表事件 x_i 的嵌入維度，卷積運算的公式為

$$c_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+h-1} + b) \quad (9)$$

其中， c_i 代表將濾波器運用到時間序列的局部區域 $\mathbf{x}_{i:i+h-1}$ 所產生的特徵， b 代表偏移量， f 表示非線性激發函數。將濾波器 $\mathbf{w} \in \mathbb{R}^{h \times k}$ 延著事件串列 $\{\mathbf{x}_{1:h}, \mathbf{x}_{2:h+1}, \dots, \mathbf{x}_{n-h+1:n}\}$ 執行卷積運算後，可以產生特徵圖像 $\mathbf{c} = [c_1, c_2, \dots, c_{n-h+1}]$ 。接著，池化

層執行「最小/最大運算」或「平均值」運算來進行下採樣。CNN 的輸出是原始 HER 序列中的與時間無關的局部特徵的緊湊潛在狀態表示向量，並在下一步中，將 CNN 的輸出餵給 BiLSTM 模型，以辨識其時間依賴性關係。

長短期記憶體(Long short-term memory, LSTM)是最著名的循環神經網路模型(Hochreiter & Schmidhuber 1997)，LSTM 可以把歷史信息的保留在內部儲存器，所以能夠有效地分析時間序列中的長程依賴性。LSTM 的組成元件包含輸入閘門 $i_t \in [0, 1]$ 、輸出閘門 $o_t \in [0, 1]$ 、遺忘閘門 $f_t \in [0, 1]$ 與儲存單元狀態(memory cell state, C_t)。這些閘門能夠彼此影響，來操控訊息的傳遞。在訓練階段中，「輸入」閘門控制候選儲存單元狀態 C_t^* 內的訊息傳遞，並且控制 C_t^* 對於新的儲存單元狀態 C_t 的影響程度，「輸出」閘門控制儲存單元的訊息 C_t 如何傳遞至輸出單元，「忘記」閘門控制儲存單元狀態的歷史信息 C_{t-1} 對於新的儲存單元狀態 C_t 的影響程度，如果歷史信息 C_{t-1} 已經變的不重要，則可以將它遺忘。這些閘門的運算公式如下所示：

$$i_t = \text{sigmoid}(W_i \cdot [y_{t-1}, X_t] + b_i), \quad (10)$$

$$f_t = \text{sigmoid}(W_f \cdot [y_{t-1}, X_t] + b_f) \quad (11)$$

$$C_t^* = \text{tanh}(W_c \cdot [y_{t-1}, X_t] + b_c), \quad (12)$$

$$o_t = \text{sigmoid}(W_o \cdot [y_{t-1}, X_t] + b_o) \quad (13)$$

其中， $W_{[i,f,c,o]}$ 為各個閘門的權重向量， $b_{[i,f,c,o]}$ 為各個閘門的偏差量。儲存單元 C_t 的內容，是使用下列公式計算

$$C_t = C_{t-1} \cdot f_t + C_t^* \cdot i_t \quad (14)$$

最終，LSTM 模型在時間 t 的輸出向量的計算公式如下

$$y_t = o_t * \text{tanh}(C_t) \quad (15)$$

LSTM 模型對於抓取時間序列數據中的遠距離的依賴性有優異的性能，所以經常被學者運用來分析電子病歷 (electronic health records, EHR) 中的時間序列，因為在電子病歷的時間序列中，不同量測變量之間細微的時間相關性，可能就是早期識別疾病惡化徵兆的重要線索。

然而，傳統的 LSTM 模型僅能透過單一方向，從歷史訊息來推估未來的結果，而無法使用反向的時間資訊。但是，如果可以同時從「過去到未來」與「未來到過去」來分析數據中的時間相關性，能夠得到更豐富的時間資訊。所以，

Graves & Schmidhuber (2005)研發雙向長短期記憶體(bi-directional long short-term memory, BiLSTM)，如圖 2 所示，BiLSTM 模型包含兩個 LSTM 元件，其中一個 LSTM 元件負責從「過去到未來」來分析時間序列資料，另外一個 LSTM 元件則負責「從未來到過去」來分析時間序列資料。在圖 2 中，BiLSTM 模型內的水平方向箭頭代表從不同的時間方向來分析數據流動，而 BiLSTM 模型內垂直方向箭頭代表資料從輸入層到輸出層的運算步驟。BiLSTM 已經在各項時序預測任務中顯示出卓越的性能，因此，本研究將使用 BiLSTM 來擷取時間序列中的因果關係。

假設第 k 個病患的動態時間序列數據餵給 CNN 網路後，它的輸出為 $x_k = [x_{k,1}, x_{k,2}, \dots, x_{k,T}]$ ，我們將 x_k 餵給 Bi-LSTM 網路，並且生成前後向時間表示 $h_{k,t}$ ，其計算步驟如下：

$$\vec{h}_{k,t} = \overrightarrow{LSTM}(\vec{h}_{k,t-1}, x_{k,t}) \quad (16)$$

$$\overleftarrow{h}_{k,t} = \overleftarrow{LSTM}(\overleftarrow{h}_{k,t+1}, x_{k,t}) \quad (17)$$

在此處，雙向時間級表示 $h_{k,t}$ 是結合 $\vec{h}_{k,t}$ 與 $\overleftarrow{h}_{k,t}$ ，亦即是

$$h_{k,t} = [\vec{h}_{k,t}, \overleftarrow{h}_{k,t}] \quad (18)$$

如此，我們得到雙向的時間表示序列為 $h_k = [h_{k,1}, h_{k,2}, \dots, h_{k,T}]$ ，接著，本研究將採用注意力機制(Attention mechanism, AM)來關注在全局時間序列中更重要的時間特徵。Treisman & Gelade (1980) 首先提出注意力機制的概念，AM 通過計算注意力的機率分佈，從大量的訊息中選出關鍵訊息，並且透過突出關鍵輸入數據，來對傳統模型進行優化。AM 的基本概念源自於人類的視覺過程，當人類的眼睛在觀察外在物件時，會先將視覺焦點放在其中某一個區域，不同的人關注的焦點並不相同。舉例來說，我們在看一個人的外表時，有的人會先將焦點放在他/她的臉蛋、身材或服裝等，在統整各個局部區域的訊息後，最終我們得到對於這個人的整體印象。也就是說，人類的視覺過程可以迅速尋找關鍵區域，而且將焦點放在該區域，以取得關鍵訊息。AM 也是基於這樣的概念，將注意力聚焦在能夠幫助判別的關鍵資訊，並且忽略不重要的資訊。本論文使用下列的注意力機制 AM 加權處理方式來獲得病患 k 最終的動態訊息表示特徵：

$$u_{k,t} = \text{sigmoid}(W_t \cdot h_{k,t} + b_t) \quad (19)$$

$$\beta_{k,t} = \frac{\exp(u_{k,t})}{\sum_j \exp(u_{k,j})} \quad (20)$$

$$x_k = \sum_t \beta_{k,t} \cdot h_{k,t} \quad (21)$$

其中 W_t 和 b_t 分別是可以學習的變換矩陣和偏置向量， $\beta_{k,t}$ 是 $h_{k,t}$ 的注意力權重分數， x_k 是經過注意力機制 AM 加權處理過的特徵向量。

儘管 CNN 與 BiLSTM 在處理時間序列等動態數據獲得巨大的成功，但是，他們卻不適合處理靜態資料（例如，年齡和性別等靜態特徵），在另一方面，全連結網路 (fully connected network, FCN)則是更適合去分析靜態數據，所以，本論文使用圖 2 的架構，年齡和性別等靜態數據是透過全連結神經網路來分析，最

後，將 FCN 的輸出向量 n_k 與 CNN-BiLSTM-AM 模組的輸出向量 x_k 串聯在一起，也就是說， $p_k = [x_k \ n_k]$ 。向量 p_k 即為編碼器模組所擷取的重要特徵，接著，編碼器產生的特徵向量 p_k 將餵給分類模組，以早期預警敗血症的發作。

2. 分類模組

在分類模組(classifier module)中，我們使用全連接層來執行最終的敗血症預測，計算步驟為：

$$\hat{y}_k^c = \sigma(W_k \cdot p_k + b_k) \quad (22)$$

其中 W_k 和 b_k 分別表示變換矩陣和偏置向量， \hat{y}_k^c 是目標病患的預測二元概率向量， $\sigma()$ 為 softmax 激發函數。同樣地，通道二中的判別器 D ，與通道一中的分類器 C 具有相同結構的全連接層，但是彼此不共享的權重和偏差。通道二中的判別器 D 的計算步驟為：

$$\hat{y}_k^d = \sigma(W_k^d \cdot p_k^d + b_k^d) \quad (23)$$

其中 W_k^d 和 b_k^d 分別是變換矩陣和偏置向量， \hat{y}_k^d 是通道二中的預測二元概率向量， $\sigma()$ 為 softmax 激發函數。

(五) 損失函數

本研究所提出的自我調節生成對抗網路的整體損失，是由「分類損失」和「自調節損失」兩部分組成。分類損失定義為預測的病患敗血症發作概率向量和真實的二進制類別標籤向量之間的交叉熵。假設有一批有 M 個樣本的訓練數據，我們使用下面公式來計算分類器 C 和判別器 D 的損失：

$$L(\hat{y}^c, y) = -\sum_{k=1}^M \sum_{j=1}^2 y_k^j \log(\hat{y}_k^{c,j}) \quad (24)$$

$$L(\hat{y}^d, y) = -\sum_{k=1}^M \sum_{j=1}^2 y_k^j \log(\hat{y}_k^{d,j}) \quad (25)$$

其中 y_k 是二進制 one-hot 向量，代表真實的敗血症發作與否。注意， \hat{y}^c 和 \hat{y}^d 分別是來自通道一中的分類器 C 和通道二中的判別器 D 的預測概率分佈。

假設編碼器 E 和生成器 G 的輸出分別為 $o_E \in R^{do}$ 和 $o_G \in R^{do}$ (do 為 o_G 的維度)。如果有 M 個訓練樣本， $o_E \in O_E$ 和 $o_G \in O_G$ ，基於相似度測量的自調節損失函數 $L_{diff}(O_E, O_G)$ 的計算如下：

$$L_{diff}(O_E, O_G) = \sum_{j=1}^M \left[\left\| O_E^j O_G^{j,T} \right\|_F \right]^2 \quad (26)$$

其中 $\left[\left\| \cdot \right\|_F \right]^2$ 表示 Frobenius 範數平方，用於計算矩陣或向量之間的相似性 (Bousmalis et al. 2016)。根據 $O_E^j O_G^{j,T}$ 的展開方程式 (Xu et al. 2022)， L_{diff} 可以改寫為：

$$L_{diff}(O_E, O_G) = \sum_{j=1}^M \left(\sum_{s=1}^{do} \sum_{l=1}^{do} |O_{E,s}^j O_{G,l}^j|^2 \right)^{\frac{1}{2}} \quad (27)$$

(六) 自調節生成對抗訓練

本研究透過最小化分類損失 $L(\hat{y}^c, y)$ 和自調節損失 $L_{diff}(O_E, O_G)$ 來訓練

我們的合作網路，編碼器 E 的目標函數為 $L(\hat{y}^c, y) + \lambda \cdot L_{diff}(O_E, O_G)$ ，其中 λ 是一個超參數，可用於調節分類損失 $L(\hat{y}^c, y)$ 和自調節損失 $L_{diff}(O_E, O_G)$ 。編碼器 E 需要更新參數 θ_E 以最小化目標函數，而分類器 C 的目標函數為 $L(\hat{y}^c, y)$ ，分類器 C 需要更新參數 θ_C 以最小化目標函數。所以，我們透過下列算式來學習編碼器與分類器中的模型參數：

$$\theta_E = \operatorname{argmin}(L(\hat{y}^c, y) + \lambda \cdot L_{diff}(O_E, O_G)) \quad (28)$$

$$\theta_C = \operatorname{argmin} L(\hat{y}^c, y) \quad (29)$$

在生成對抗網路中，生成器 G 與判別器 D 目標函數皆為 $L(\hat{y}^d, y)$ ，根據最小-最大博弈，生成器 G 需要更新參數 θ_G 以最大化目標函數，相反地，判別器 D 則是需要更新參數 θ_D 以最小化目標函數。所以，我們透過下列算式來學習生成器與判別器中的模型參數：

$$\theta_G = \operatorname{argmax} L(\hat{y}^d, y) \quad (30)$$

$$\theta_D = \operatorname{argmin} L(\hat{y}^d, y) \quad (31)$$

由於生成器與判別器的學習能力並不一樣，在實踐中，經常出現生成器迅速打敗判別器的情況，為了讓生成器與判別器能夠更均衡地一致收斂，Goodfellow et al. (2014) 建議判別器參數每調整 k 次後，調整一次生成器的參數。在本研究的實踐當中，我們將 k 值設定為 5 時有最佳的收斂效果。

在本研究提出的自調節生成對抗模型中，我們將合作網路與生成對抗網路 (GAN) 相結合，並通過聯合優化的策略，進一步提高它們的學習能力。對於圖 1 中所示的四個成員，即編碼器 E 、分類器 C 、生成器 G 和判別器 D ，我們所提出的自調節神經網路的主要受益者，是通道一中的編碼器 E 。它可以同時從通道一的合作網路和通道二的對抗網路中獲得收益，通道一的合作網路讓編碼器 E 從正面的方向學習可以幫助分類的特徵，通道二的對抗網路讓編碼器 E 從反面的方向避免學習到錯誤的特徵，並且可以消除可能的隨機性的雜訊（如通道二中的生成器 G 所做的那些污染）。最後，我們可以將編碼器 E 產生的特徵，交給頂層的嶄新模糊超平面支持向量機，透過模糊理論來處理資料的模糊性，使用支持向量機嚴謹的理論基礎來打造最佳的分類器。

三、運用以可能性理論為基礎的支持向量機作為最頂層的分類器

本論文以作者在 *Neural Networks* 期刊最新發表的模糊支持向量機做為核心分類器 (Hao et al. 2022)，模糊支持向量機將可能性理論 (Possibility theory) 導入支持向量機，並透過模糊數學規劃 (fuzzy mathematical programming) 來尋找出一個最佳的「模糊超平面 (fuzzy hyperplane)」，來使用最大的邊距 (margin) 將訓練樣本分為兩類，並且良好的處理資料內的不確定性。用於構建模糊超平面的參數，包括法向量 (normal vector) 和偏置項 (bias)，皆為對稱三角模糊數 (symmetric triangular fuzzy numbers)。為了找到一個最佳的模糊超平面來將訓練樣本分成兩類，我們使用以下定理：

定義 1. (Dubois & Prade 1998) 令 $P(X)$ 表示集合 X 的幂集(power set)，可能性度量(possibility measure)是集合函數 $Pos: P(X) \rightarrow [0, 1]$ ，並且具有以下屬性

$$(1) Pos(\phi)=0 \quad (2) Pos(X)=1 \quad (3) Pos(\cup_{t \in T} A_t) = \sup_{t \in T} Pos(A_t) \quad (32)$$

定義 2. (Dubois & Prade 1998) 為了比較兩個模糊數 \tilde{M} 和 \tilde{N} 的大小，Dubois 定義出以下指標來量化 \tilde{M} 大於或等於 \tilde{N} 的可能性：

$$Pos(\tilde{M} \geq \tilde{N}) = \sup\{\mu_{\tilde{M}}(x) \wedge \mu_{\tilde{N}}(x) | x \geq y \text{ and } x, y \in R\} \quad (33)$$

其中 $\mu_{\tilde{M}}$ 與 $\mu_{\tilde{N}}$ 分別是模糊數 \tilde{M} 與 \tilde{N} 的歸屬函數。根據此觀點，模糊數 \tilde{M} 大於或等於實數 N 的可能性定義為：

$$Pos(\tilde{M} \geq N) = \sup\{\mu_{\tilde{N}}(x) | x \geq N \text{ and } x \in R\} \quad (34)$$

定理 1. (Liu 2000) 令 $\tilde{M}=(a_1, a_2, a_3)$ 代表三角模糊數，給定一純量 $\lambda \in (0,1]$ ，我們有

$$Pos(\tilde{M} \geq 0) \geq \lambda \Leftrightarrow \frac{a_3}{a_3 - a_2} \geq \lambda \Leftrightarrow (1 - \lambda)a_3 + \lambda a_2 \geq 0 \quad (35)$$

定理 2. (Tankaka et al. 1982) 令 $\tilde{W} = (\mathbf{w}, \mathbf{c})$ 為模糊法向量與 $\tilde{B} = (b, d)$ 為模糊偏移量，其中，模糊法向量中每一個元素 $\tilde{W}_i = (w_i, c_i)$ 皆是對稱三角形模糊數，以 w_i 為中心與 c_i 為半徑，同樣地，模糊偏移量 $\tilde{B} = (b, d)$ 也是對稱三角形模糊數，以 b 為中心與 d 為半徑，則模糊超平面 $Y = \langle \tilde{W} \cdot \mathbf{x} \rangle + \tilde{B} = \tilde{W}_1 \cdot x_1 + \dots + \tilde{W}_n \cdot x_n + \tilde{B}$ 的歸屬函數定義為：

$$\mu_Y(y) = \begin{cases} 1 - \frac{|y - ((\mathbf{w} \cdot \mathbf{x}) + b)|}{\langle \mathbf{c} \cdot |\mathbf{x}| \rangle + d} & \mathbf{x} \neq 0 \\ 1 & \mathbf{x} = 0, y = 0 \\ 0 & \mathbf{x} = 0, y \neq 0 \end{cases} \quad (36)$$

其中 $\mu_Y(y) = 0$ 當 $\langle \mathbf{c} \cdot |\mathbf{x}| \rangle + d \leq |y - ((\mathbf{w} \cdot \mathbf{x}) + b)|$ 。

給定一個 2 元分類任務，其訓練樣本為 $\{(\mathbf{x}_i, y_i, s_i)\}, i=1, \dots, N$ ，其中 $\mathbf{x}_i \in R^n$ 代表第 i 個資料向量， $y_i \in \{-1, 1\}$ 表示對應的類別標籤， $s_i \in (0,1]$ 代表資料向量 \mathbf{x}_i 屬於類別 y_i 的模糊歸屬程度。我們的模糊支持向量機的目標是找出一個最佳的模糊超平面 $\langle \tilde{W} \cdot \mathbf{x} \rangle + \tilde{B} = 0$ 來分割這兩個類別的資料，其中 $\tilde{W} = (\mathbf{w}, \mathbf{c})$ 為模糊法向量與 $\tilde{B} = (b, d)$ 為模糊偏移量。當下列條件成立時，我們稱這兩類別的資料是基於可能性的(possibilistic)線性可分割的：

$$pos\{y_i(\langle \tilde{W} \cdot \mathbf{x}_i \rangle + \tilde{B}) \geq 1\} \geq s_i, i=1, \dots, N \quad (37)$$

換句話說，第 i 筆資料向量滿足線性可分割的可能性是大於或等於 s_i 。我們的模糊支持向量機是透過求解下列模糊數學規劃問題來找出最佳的模糊超平面：

$$\text{minimize}_{\mathbf{w}, b, c, d, \xi_i} J = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(v \left(\frac{1}{2} \|\mathbf{c}\|^2 + d \right) + \frac{1}{N} \sum_{i=1}^N \xi_i \right), \quad (38.1)$$

$$\text{subject to } pos\{y_i(\langle \tilde{W} \cdot \mathbf{x}_i \rangle + \tilde{B}) \geq 1 - \xi_i\} \geq s_i \quad (38.2)$$

$$\xi_i \geq 0 \text{ for all } i=1, \dots, N \quad (38.3)$$

最小化 $\|\mathbf{w}\|^2$ 可以最大化超平面的邊界(margin)，懲罰常數 $C>0$ 控制對於違反模糊可分割限制條件的懲罰分數，差額變數 ξ_i 量測模糊可分割條件被違反的程度，模糊常數 $\nu>0$ 控制對於超平面的模糊程度的懲罰， ν 值越小，得到的超平面越模糊， $\frac{1}{2}\|\mathbf{c}\|^2 + d$ 則是用來量測超平面的模糊度。參數 s_i 代表第 i 個樣本點的模糊程度， s_i 值越小，代表第 i 筆資料向量滿足線性可分割的可能性也越小。我們可以對可能是雜訊的樣本點指派一個較小的 s_i 值，允許這些雜訊樣本點能夠被適當的錯誤分類，以避免過度學習(overfitting)這些雜訊資料，根據定理 3 與 4，上述的模糊數學規劃問題改寫為

$$\underset{\mathbf{w}, \mathbf{c}, b, d, \xi_i}{\text{minimize}} \quad J = \frac{1}{2}\|\mathbf{w}\|^2 + C \left(\nu \left(\frac{1}{2}\|\mathbf{c}\|^2 + d \right) + \frac{1}{N} \sum_{i=1}^N \xi_i \right) \quad (39.1)$$

$$\text{subject to} \quad y_i(\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) + (1 - s_i)(\langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + d) \geq 1 - \xi_i, \quad (39.2)$$

$$\xi_i \geq 0 \text{ for all } i=1, \dots, N. \quad (39.3)$$

使用拉格朗矩乘數(Lagrange multipliers)技巧，我們可以得到上述最佳化問題的對偶問題(dual problem)：

$$\max_{\alpha} \frac{-1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle - \frac{1}{2C\nu} \sum_{i=1}^N \sum_{j=1}^N (1 - s_i)(1 - s_j) \alpha_i \alpha_j \langle |\mathbf{x}_i| \cdot |\mathbf{x}_j| \rangle$$

$$\text{Subject to} \quad \sum_{i=1}^N y_i \alpha_i = 0, \quad \sum_{i=1}^N (1 - s_i) \alpha_i = C\nu, \quad \alpha_i \in [0, \frac{C}{N}] \quad (40)$$

其中 α_i 為非負的拉格朗矩乘數，求解完上述問題後，我們可以得到模糊法向量 $\tilde{\mathbf{W}} = (\mathbf{w}, \mathbf{c})$ ，其中：

$$\mathbf{w} = \sum_{i=1}^N y_i \alpha_i \mathbf{x}_i \quad \text{與} \quad \mathbf{c} = \frac{1}{C\nu} \sum_{i=1}^N \alpha_i (1 - s_i) |\mathbf{x}_i|. \quad (41)$$

模糊偏移量 $\tilde{\mathbf{B}} = (b, d)$ 可以根據 KKT (Karush-Kuhn-Tucker) 條件計算出：

$$b = - \left[\frac{(1-s_j)\langle \mathbf{w} \cdot \mathbf{x}_i \rangle - (1-s_i)\langle \mathbf{w} \cdot \mathbf{x}_j \rangle + (1-s_i)(1-s_j)(\langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle - \langle \mathbf{c} \cdot |\mathbf{x}_j| \rangle) + s_j - s_i}{(1-s_i) + (1-s_j)} \right] \quad (42)$$

$$d = - \left[\frac{\langle \mathbf{w} \cdot \mathbf{x}_i \rangle - \langle \mathbf{w} \cdot \mathbf{x}_j \rangle + (1-s_i)\langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + (1-s_j)\langle \mathbf{c} \cdot |\mathbf{x}_j| \rangle - 2}{(1-s_i) + (1-s_j)} \right] \quad (43)$$

其中 i, j 滿足 $y_i = 1, y_j = -1$ $\alpha_i \in (0, \frac{C}{N})$ 與 $\alpha_j \in (0, \frac{C}{N})$ 。模糊超平面 $Y = \langle \tilde{\mathbf{W}} \cdot \mathbf{x} \rangle + \tilde{\mathbf{B}}$ 的歸屬函數定義為：

$$\mu_Y(y) = 1 - \frac{|y - (\sum_{i=1}^N y_i \alpha_i \langle \mathbf{x}_i \cdot \mathbf{x} \rangle + b)|}{\left(\frac{1}{C\nu} \sum_{i=1}^N (1-s_i) \alpha_i (\langle \mathbf{x}_i \cdot \mathbf{x} \rangle + d) \right)}. \quad (44)$$

給定一個輸入資料點 \mathbf{x}_i ， $Y_i = \langle \tilde{\mathbf{W}} \cdot \mathbf{x}_i \rangle + \tilde{\mathbf{B}}$ 是一個對稱三角模糊數，其中心點為 $\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b$ ，半徑為 $\langle \mathbf{c} \cdot |\mathbf{x}_i| \rangle + d$ 。根據可能性量測的概念，由此模糊超平面切割的模糊邊界定義為：

$$f(\mathbf{x}) = \text{pos}\{y_i((\bar{\mathbf{W}} \cdot \mathbf{x}) + \bar{B}) \geq 0\} = \begin{cases} 1 & \text{if } \langle \mathbf{w} \cdot \mathbf{x} \rangle + b > 0 \text{ and } ((\mathbf{w} \cdot \mathbf{x}) + b) - ((\mathbf{c} \cdot |\mathbf{x}|) + d) > 0 \\ 1 - \frac{(\mathbf{c} \cdot |\mathbf{x}|) + d - ((\mathbf{w} \cdot \mathbf{x}) + b)}{2((\mathbf{c} \cdot |\mathbf{x}|) + d)} & \text{if } \langle \mathbf{w} \cdot \mathbf{x} \rangle + b > 0 \text{ and } ((\mathbf{w} \cdot \mathbf{x}) + b) - ((\mathbf{c} \cdot |\mathbf{x}|) + d) < 0 \\ \frac{(\mathbf{c} \cdot |\mathbf{x}|) + d + ((\mathbf{w} \cdot \mathbf{x}) + b)}{2((\mathbf{c} \cdot |\mathbf{x}|) + d)} & \text{if } \langle \mathbf{w} \cdot \mathbf{x} \rangle + b < 0 \text{ and } ((\mathbf{w} \cdot \mathbf{x}) + b) + ((\mathbf{c} \cdot |\mathbf{x}|) + d) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (45)$$

此模糊邊界對輸入樣本點指派一個值(介於 0 與 1 之間)，來代表此輸入資料隸屬於正類別的模糊程度，回傳的值越大，則代表輸入資料隸屬於正類別的可能性就越大。由於資料中的不確定性被模糊決策函數完美的描述，因此，這提供一個有效的方式來解決現實應用中的模糊與不精確的性質。

四、運用逐層特徵擷取建構的深層模糊支持向量機網路

受到 Li et al. (2019) 在 IEEE Access 期刊上發表的深層支持向量機所啟發，在此小節，我們將導入逐層透過核心函數擷取特徵，以堆疊更深層的模糊支持向量機網路結構。其基本策略如下，首先，使用訓練樣本對第一層的模糊支持向量機進行建構，並且學習數據的淺層特徵。然後，透過「特徵提取公式(feature extraction formula)」用於生成樣本的新表達式，作為下一層的輸入。每進入新的一層，模糊支持向量機就使用在前一層的特徵提取公式所產生的新樣本來進行訓練，並逐層提取和學習資料集的深層特徵；最終，在經過多層的特徵映射後，它將在最後一層輸出分類結果。更明確地說，假設訓練資料集合為 $\{(\mathbf{x}_i, y_i)\}_{i=1, \dots, N}$ ，我們先求解 Eq.(40) 的最佳化問題來建立第一層的模糊支持向量機，當第一層的模糊支持向量機建立好之後(亦即，求解完 Eq.(40) 後)，對應拉格朗日乘數 $\alpha_i > 0$ 的資料向量稱為支持向量，表示為 $\{sv_i\}_{i=1, \dots, Q}$ ， Q 代表支持向量的數目，接著，我們使用特徵提取公式來對訓練資料進行處理，以產生樣本的新表達式，「特徵提取公式」定義如下

$$g(i) = y_{sv_i} \alpha_{sv_i} K_l(sv_i, \mathbf{x}) + b, i = 1, \dots, Q \quad (46)$$

其中， $g(i)$ 表示對資料 \mathbf{x} 提取的第 i 個核心函數特徵的值， sv_i 表示第 i 個支持向量， α_{sv_i} 代表該支持向量對應的拉格朗日乘數， y_{sv_i} 代表該支持向量對應的類別標籤， b 則是偏移量。換句話說，我們使用資料 \mathbf{x} 對支持向量的核心函數激發值來做為特徵，其中， K_l 代表第 l 層所使用的核心函數，我們可以對每一層使用不同的核心函數，來逐層融合不同核心函數的優點，以達到多核心函數學習(multiple kernel learning)的目標。如果第一層的模糊支持向量機得到 Q 個支持向量，則在第一層之後，我們將提取 Q 個特徵，經過特徵提取之後，原來的資料 \mathbf{x} 使用新的特徵，得到新的表達式如下

$$x_i^{2nd} = [y_{sv_1} \alpha_{sv_1} K_l(sv_1, \mathbf{x}) + b, \dots, y_{sv_Q} \alpha_{sv_Q} K_l(sv_Q, \mathbf{x}) + b] \quad (47)$$

圖 3 (a) 顯示特徵提取公式的運作概念，將原本的 N 筆訓練資料用新特徵表示之後，得到 $x_1^{2nd}, x_2^{2nd}, \dots, x_N^{2nd}$ ，這些訓練資料對應的類別標籤 y_1, y_2, \dots, y_N 均保持不

變，接著，那些取好新的特徵的 N 筆訓練樣本將交給模糊支持向量機網路訓練，透過堆疊(stack)的策略來建立第二層的網路結構，我們使用第二層模糊支持向量機學習得到的支持向量，透過相同的特徵提取方式(即 Eq.(46)與(47))，來提取新的特徵，並建立這 N 筆資料的新特徵表示式，即 $x_1^{3rd}, x_2^{3rd}, \dots, x_N^{3rd}$ ，這些資料對應的類別標籤 y_1, y_2, \dots, y_N 均保持不變，這些新取好特徵的 N 筆訓練樣本，將交給模糊支持向量機網路訓練，透過堆疊(stack)的策略來建立第三層的網路結構，如圖 3 (b)所示，透過這種方式，逐層學習新的核心函數特徵表示式，以獲得深層特徵(deep feature)，在最後一層，我們使用深層特徵所表示的 N 筆訓練資料 $x_1^{final}, x_2^{final}, \dots, x_N^{final}$ ，交給模糊支持向量機來學習最終的模糊分類器。

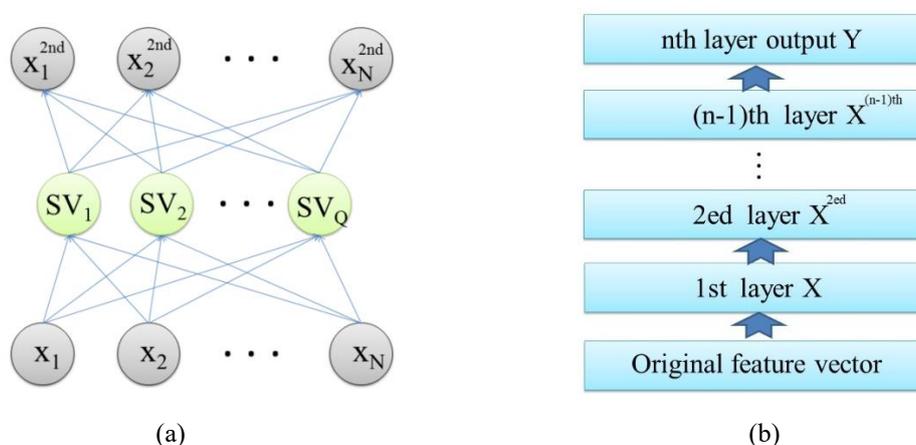


圖 3：(a)透過支持向量的特徵提取公式，與(b)運用逐層特徵擷取建構的深層模糊支持向量機

肆、實驗結果

為了公平的與其他敗血症預警系統進行比較，本研究將使用敗血症公開資料集 PhysioNet/Computing in Cardiology Challenge 2019 (Reyna et al. 2019)來進行實驗比較。該資料集是從三個地理位置不同的美國醫院系統獲得病患數據，包含：貝斯以色列女執事醫療中心(Beth Israel Deaconess Medical Center, 醫院代號 A)、埃默里大學醫院 (Emory University Hospital, 醫院代號 B) 與第三個匿名的醫院系統 (醫院代號 C)。其中，來自醫院系統 A 和 B 的 40,336 名患者的數據和標籤是公開發布並且可以自由下載，而來自醫院系統 C 的資料則未公布，亦沒有提供類別標籤，所以，參考(Lee et al. 2019; Nakhshi et al. 2019; Yu et al. 2019)等其他研究的做法，本研究將以來自醫院系統 A 和 B 的 40,336 名患者的數據建立敗血症早期預警模型，並且評估其預測效能。表格 3 分析醫院 A 和 B 的 40,336 名患者的分布情況，由表格 3 可知，對於醫院 A 和 B，敗血症病患佔總病患數量的比率分別為 8.8% 與 5.7%，有嚴重的數據不平衡的問題。

PhysioNet/Computing in Cardiology Challenge 2019 資料集包含 40 個臨床變量，表 4-6 統計這些變量的遺失值比率、合理範圍、平均值、中位數與標準差。

表 3：醫院 A 與 B 的病患資料分布情況

醫院系統	A	B	總和
病患數目	23,336	20,000	43,336
敗血症病患數目	1,790	1,142	2,932
敗血症病患比率	8.8%	5.7%	7.27%
水平行(時間窗口)的數目	739,663	684,508	1,424,171

表 4：生命體徵的遺失值比率、合理範圍與敘述統計分析

編號	變數名稱	遺失值比率	合理範圍		平均值	中位數	標準差
			最小值	最大值			
1	HR	9.9	10	300	84.6	83.5	17.3
2	O2Sat	13.1	60	100	97.2	98	2.7
3	Temp	66.2	32	42.2	37	37	0.8
4	SBP	14.6	40	280	123.8	121	23.2
5	MAP	12.5	0	300	82.4	80	16.3
6	DBP	31.3	20	130	63.7	62	13.6
7	Resp	15.4	5	60	18.8	18	5
8	EtCO2	96.3	0	150	33	33	8

表 5：實驗室數值的遺失值比率、合理範圍與敘述統計分析

編號	變數名稱	遺失值比率	合理範圍		平均值	中位數	標準差
			最小值	最大值			
9	BaseExcess	94.6	-20	20	-0.7	0	4.2
10	HCO3	95.8	0	50	24.1	24	4.4
11	FiO2	91.7	0	1	0.5	0.5	0.2
12	pH	93.1	6	8	7.4	7.4	0.1
13	PaCO2	94.4	0	200	41	40	9.3
14	SaO2	96.5	0	100	92.7	97	10.9
15	AST	98.4	0	400	64.5	35	73
16	BUN	93.1	0	500	23.9	17	20
17	Alkalinephos	98.4	0	250	82.8	71	43
18	Calcium	94.1	0	20	7.6	8.3	2.4
19	Chloride	95.5	75	145	105.8	106	5.8
20	Creatinine	93.9	0	10	1.4	0.9	1.4
21	Bilirubin direct	99.8	0	50	1.8	0.4	3.7
22	Glucose	82.9	0	1000	136.9	127	51.3
23	Lactate	97.3	0	100	2.6	1.8	2.5
24	Magnesium	93.7	0	10	2.1	2	0.4
25	Phosphate	96.0	0	12	3.5	3.3	1.4
26	Potassium	90.7	1	10	4.1	4.1	0.6
27	Bilirubin total	98.5	0	50	2.1	0.9	4.3
28	TroponinI	99.0	0	200	8	0.3	22.7
29	Hct	91.1	10	70	30.8	30.3	5.5
30	Hgb	92.6	2	22	10.4	10.3	2
31	PTT	97.1	0	250	41.2	32.4	26.2
32	WBC	93.6	0	50	11.2	10.3	5.4
33	Fibrinogen	99.3	0	800	280.2	248	137.5
34	Platelets	94.1	5	1500	196	181	103

表 6：人口統計資料的遺失值比率、合理範圍與敘述統計分析

編號	變數名稱	遺失值比率	合理範圍		平均值	中位數	標準差
			最小值	最大值			
35	Age	0.0	0	150	62	64	16.4
36	Gender	0.0	0	1	0.6	1	0.5
37	Unit1	39.4	0	1	0.5	0	0.5
38	Unit2	39.4	0	1	0.5	1	0.5
39	HospAdmTime	0.0	None	None	-56.1	-6	162.3
30	ICULOS	0.0	1	None	27	21	29

一、資料預處理

由表格 4 至 6 可知，PhysioNet/Computing in Cardiology Challenge 2019 資料集有極嚴重的遺失值問題，其中，實驗室數值的遺失值比率是遠高於生命體徵，主要原因是生命體徵與實驗室數值的量測頻率並不相同，一般來說，生命體徵是每小時量測一次，而實驗室數值則是每天更新一次，由於量測的頻率不同，所以在每小時的時間窗口數據中，實驗室數值的遺失率是大於生命體徵的遺失率。總結而言，來自 A 與 B 兩個醫院系統的公共挑戰數據庫共包括 40,336 名患者，其中 2,932 名 (7.27%) 發展為敗血症，與非敗血症患者相比，敗血症病患的平均記錄時間更長（敗血症病患資料的平均紀錄時間為 58.6 小時，而非敗血症病患資料的平均紀錄時間為 38.9 小時）。在可用的病患變量中，除體溫 (Temp) 的遺失值比率為 66%、舒張壓 (DBP) 的遺失值比率為 31% 和呼氣末二氧化碳 (EtCO₂) 的遺失值比率為 96% 外，每個生命體徵的遺失值比率均小於 16%。對於實驗室值而言，除 Glucose 的遺失值比率為 83% 外，其他測量變量的遺失值比率皆超過 90%。ICU 單元(單元 1 和單元 2)的兩個行政標識符均有近 39% 的遺失值比率，而其他是人口統計的固定變量。在數據分析之後，我們參考(Lee et al. 2019; Nakhshi et al. 2019; Yu et al. 2019)研究的做法，對原始數據進行下列預處理步驟。

第一步：去除遺失值佔 99% 以上的 Bilirubin_direct、Troponin I、Fibrinogen 等變量。

第二步：用前向填充策略估算遺失數據，如果在時間步 $t_p < t$ 有一個先前記錄變量 v 的測量值，我們透過設置 $x_v(t) = x_v(t_p)$ 來執行前向填充，以處理變量 v 在時間步 t 的遺失值。如果沒有以前記錄的測量值，我們則使用該變量 v 的總體平均值來替換該遺失值。

本研究將使用病患的臨床數據來自動識別患者的敗血症發作，並在每個小時的時間窗口內，對敗血症做出陽性或陰性預測。在經過上述預處理步驟後，在訓練過程中將產生超過 100 萬小時的時間窗口數據。然而，在這些數據中，只有大約 1.8% 的時間窗口數據是對應於正類別(敗血症將在 6 小時內發作)。因此，有極嚴重的類別不平衡(class imbalance)的問題，為了解決類別不平衡的問題，我們參考(Lee et al. 2019; Nakhshi et al. 2019; Yu et al. 2019)研究的作法，對大類別內過多的數據實例進行下採樣(down sampling)，也就是說，我們使用所有敗血症發作的數據和只有 5% 沒有敗血症發作的數據進行訓練。

二、預測效能評估準

參考(Lee et al. 2019; Nakhshi et al. 2019; Yu et al. 2019)研究的作法，我們使用 3 折交叉驗證(3-folds cross validation)來評估本研究提出的自調節生成對抗网络和模型對於早期敗血症發作示警的預測效能。此外，參考(Lee et al. 2019; Nakhshi

et al. 2019; Yu et al. 2019)研究的作法，我們使用下述評估準則來衡量模型的預測能力：

準確率 (Accuracy) 為正確預測的數量占總體預測數量的比率，其公式為

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (48)$$

其中，TP (True Positive, 真陽性) 代表正樣本被正確分類的數量；TN (True Negative, 真陰性) 代表負樣本被正確分類的數量；FP (False Positive, 偽陽性) 代表正樣本被錯誤分類的數量；FN (False Negative, 假陰性) 代表負樣本被錯誤分類的數量。基本上，準確率越高越好。

F1-score 是精確率 (Precision) 與召回率 (Recall) 的調和平均數，其公式為

$$F1 - score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (49)$$

其中，精確率 (Precision) = $TP/(TP+FP)$ ，即預測為正的樣本中，是正確預測的比率。召回率 (Recall) = $TP/(TP+FN)$ ，即事實為正的樣本中，是正確預測的比率。

AUROC (Area Under ROC Curve) 代表在 ROC 曲線 (Receiver operating characteristic curve, 接收者操作特徵曲線) 下方的面積，ROC 曲線越靠近右上方越理想，所以，AUROC 的值越大，代表模型的預測效能越高。

AUPRC (Area Under PR Curve) 代表在 PR 曲線 (Precision-Recall Curve, 精確召回曲線) 下方的面積，PR 曲線是以召回率 (Recall) 當 X 軸，以精確率 (Precision) 當 Y 軸。一般而言，精確率與召回率越高，代表模型的預測效能越好，所以 PR 曲線越往右上方靠近越理想。所以，AUPRC 的值越大，代表模型的預測效能越高。

其中，準確率 (Accuracy) 對於資料類別不平衡的狀況，會出現嚴重偏差。以敗血症早期預警系統為例，假如有 100 個病患樣本，其中有 99 位是 Negative (未患敗血症)，只有 1 個 Positive (患敗血症)。假設現在有一個預測模型，針對所有樣本的預測輸出全是 Negative，則此模型的準確率高達 99%。但是，雖然準確率極高，這個敗血症預測模型幾乎是無用的。相反地，AUROC、AUPRC 與 F1-Score 等評估準則較不易受到類別不平衡的影響，所以能夠較合理的衡量模型的預測能力。在調整模型參數時，本研究以最大化 AUROC 為目標。

三、模糊歸屬程度的設定

PhysioNet/Computing in Cardiology Challenge 2019 資料集有極嚴重的類別不平衡與遺失值的問題，而本論文使用的模糊支持向量機可以透過對訓練資料指派適當模糊歸屬程度，來緩解上述問題，令 n_{pos} 與 n_{neg} 表示正類別與負類別的樣本數，則正/負類別的樣本數目的比率 (亦即是， $IR = n_{pos} / n_{neg}$) 為類別不平衡比率 (imbalance ratio, IR)，本論文藉由類別不平衡比率 IR，對每一個訓練樣本指派一個模糊歸屬程度，使得樣本數量眾多的大類別的訓練資料，被指派較小的模糊

歸屬程度，如此可以避免得到一個對大類別具有偏見的預測模型。另外，當資料具有大量的遺失值時，該資料的可靠度較不理想，令 fn 代表資料的特徵數目，而訓練資料 x_i 中的遺失值數量為 m_i ，則訓練資料 x_i 的遺失值比率(missing ratio, MR)的計算方式為 $MR_i = m_i / fn$ ，遺失值比率越低的樣本越可靠，它所指派的模糊歸屬程度也越高，代表在訓練過程中，這些樣本扮演的腳色越重要。本論文將類別不平衡比率與遺失值比率納入考慮，並使用下列公式來計算每一個訓練樣本的模糊歸屬程度：

$$s(x_i) = \begin{cases} 1 & x_i \in \text{minority class} \\ (1 - MR_i) * IR & x_i \in \text{majority class} \end{cases} \quad (50)$$

屬於大類別的訓練資料，公式(50)運用遺失值比率 MR_i 來對遺失值較少的訓練樣本指派較高的模糊歸屬程度，並且運用類別不平衡比率 IR 來對大類別的訓練樣本指派較小的模糊歸屬程度。相反地，屬於小類別的訓練資料，因為樣本的數量已經很缺稀，所以不考慮遺失值比率 MR_i 與類別不平衡比率 IR ，而是將這些樣本的模糊歸屬程度直接指派為 1。

四、與先進的敗血症早期預警模型的預測效能比較

第一個實驗比較本研究提出的自調節生成對抗預測模型與底下的先進的敗血症早期預警模型的預測效能

- **Gultepe 的線性 SVM 模型**：Gultepe et al. (2014) 使用線性支持向量機 (Support Vector Machine, SVM) 來建構敗血症早期預警模型。
- **Yu 的集成架構**：Yu et al. (2019) 提出一個集成學習的架構，它建立三個預測模型，分別為隨機森林 (Random Forests, RF)、XGBoost(XGB) 與 LightGBM(LGB)，再將此三個模型的預測結果集成在一起。
- **Lee 的集成 LSTM 模型**：Lee et al. (2019) 提出一個集成 Transformer 與長短期記憶 (Long short-term memory, LSTM) 的敗血症預測系統。
- **Nakhashi 的集成 RF 架構**：Nakhashi et al. (2019) 提出集成隨機森林 (Random Forests, RF) 的敗血症預測系統。
- **Kok 的 CNN 模型**：Kok et al. (2020) 採用卷積神經網路 (Convolutional Neural Network, CNN) 來建構敗血症早期預警模型。
- **Rafiei 的混和模型**：Rafiei et al. (2021) 混合長短期記憶 (Long short-term memory, LSTM)，卷積層 (convolutional layer) 與全連接層 (fully connected layers, FCL) 來建構一個敗血症早期預警模型。

表 7：各個敗血症早期預警模型的預測效能比較

	AUROC	AUPRC	Accuracy	F1-score
Gultepe 的線性 SVM 模型	0.741	0.053	78.4	0.071
Yu 的集成架構	0.746	0.058	81.2	0.078
Lee 的集成 LSTM 模型	0.771	0.066	82.2	0.089
Nakhashi 的集成 RF 架構	0.736	0.050	76.6	0.067
Kok 的 CNN 模型	0.795	0.084	83.4	0.105
Rafiei 的混和模型	0.807	0.085	84.6	0.119
本研究的預測模型	0.815	0.093	85.0	0.127

表 7 比較本研究提出的自調節生成對抗敗血症早期預測模型與其他先進的敗血症預測模型的預測效能，在表格 7 中顯示的其他研究的實驗結果(Gultepe 的線性 SVM 模型 ... Rafiei 的混和模型)，皆是使用本研究採取的交叉驗證資料重作的結果，而非是直接摘錄自這些研究文獻所發表的實驗結果，所以不用擔心本研究採用的交叉驗證資料分割集的採樣對於公平性的影響。

如表格 7 所示，本論文提出的敗血症預測模型在不同效能評估準則皆優於其他模型，這驗證本論文提出的自調節生成對抗特徵學習模型可以擷取更強健的特徵，同時，本研究提出的深層支持向量機擁有更優異的預測能力。在表格 7 中，有許多先進的敗血症預測模型是採用集成學習(Ensemble learning)的策略，他們將好幾個監督式學習模型的預測結果，以系統化的方式融合在一起，以得到更好的預測效能。事實確實是如此，由於引發敗血症的原因是錯綜複雜，並且表現出許多不同的特徵，只使用單一個簡單的預測模型，很難獲得理想的預測效果。此外，要能夠理想的早期預警敗血症發作，必需要考慮各種不同類型的異質資料來源，其中，人口統計變量是靜態資料，生命體徵與實驗室數值是動態資料，而且，生命體徵與實驗室數值的量測頻率並不相同，生命體徵大多是每小時測量一次，而實驗室數值則是每天更新一次，換句或說，生命體徵顯示短程的時間依賴性，而實驗室數值則是披露長程的時間依賴性。顯而易見地，單一個預測模型不可能同時捕抓以下三種特徵：靜態特徵、短程的時間依賴特徵、與長時間的全局特徵。其中，全連結網路(fully connected network, FCN)擅長分析靜態數據，卷積神經網路(convolutional neural networks, CNN)適合擷取與時間無關的局部模式特徵，雙向長短期記憶體(bi-directional long short-term memory, BiLSTM)可以由「從過去到未來」與「從未來到過去」兩種不同的方向來分析時間依賴性與因果關係特徵，注意力機制(attention mechanism, AM)可以對長時間的全局資料中的重要因素給予較多的關注，而較少關注不重要的特徵，讓準確率可以進一步提升。而模糊支持向量機則可以考慮資料的模糊與不確定性，並且根據統計學習理論打造最佳的分類器。而本研究則是打造一個混合模型(hybrid model)來融合這些模型的優點。實驗結果顯示，本研究提出的混合模型的預測效能，比僅使用單一個深度模型的敗血症預警系統還要優異，而且，本研究提出的混合模型的預測效能，也比僅是將數個監督式模型的預測結果加權整合的集成學習模型還要優異。

五、針對混合模型中的各種組件的消融分析

本論文提出一個混合自調節生成對抗模型、CNN、BiLSTM、AM、模糊支持向量機與多核心學習的深層模糊支持向量機的敗血症早期預警模型，並且結合上述模型的長處。在下一個實驗中，本論文採用消融分析(Ablation Analysis)來驗證混合模型中的各個關鍵組件(component)的重要性。消融分析藉由移除混合模型中某個關鍵組件後，觀察預測效能的變化情況，來衡量被移除的關鍵組件對於預測效能的貢獻程度。如果移除某個關鍵組件後，造成預測效能的下降幅度越大，

代表被移除的組件對於預測效能的貢獻度越大，相反地，若是移除某個組件後的預測效能下降程度並不明顯，代表該關鍵組件對於預測效能的貢獻程度並不顯著，表 8 顯示我們對混合模型各個關鍵組件進行消融分析的結果。

- **移除 CNN 模型：**它是指將卷積神經網路模型 (convolutional neural networks, CNN) 予以刪除，而且將生命體徵與實驗室數值等動態資料直接交給 BiLSTM 來進行分析處理。如表 8 所示，移除 CNN 模型後，AUROC 下降 1.104%，這證明了 CNN 模型的重要性，這是因為 CNN 模型能夠捕抓短期的與時間位置無關的特徵，並且用比較緊密的向量來表示，有利於後續的 BiLSTM 的分析，所以有助於提升預測效能。
- **移除 BiLSTM 模型：**移除雙向長短期記憶體 (bi-directional long short-term memory, BiLSTM) 是指將雙向長短期記憶體改為傳統的「單向」長短期記憶體 (LSTM)，也就是說，我們只找出「從過去到未來」的時間依賴特徵，而不是同時考慮「從過去到未來」與「從未來到過去」兩個不同方向的時間依賴性特徵，從表 8 顯示，移除 BiLSTM 模型後，AUROC 下降 0.122%，證明同時考慮「從過去到未來」與「從未來到過去」兩個不同方向的時間依賴性特徵，有助於提升預測效能。
- **移除 AM 模型：**移除注意力機制 (Attention mechanism, AM) 是指將公式 (19) 至 (21) 的注意力機制移除，並且將 BiLSTM 的輸出結果直接與全連結網路 (FCN, fully connected networks) 的輸出向量串接在一起，並餵給模糊支持向量機來判斷敗血症發作。如表 8 所示，移除 AM 模型後，AUROC 下降 0.368%，證明 AM 模型有助於將焦點放在更能夠幫助判斷敗血症的特徵，從而提高預測效能。
- **移除 SVM 模型：**移除支持向量機 (Support vector machine, SVM) 模型指的是將圖 1 中通道一的「編碼器」學習得到的特徵，並不是交給模糊支持向量機來預測敗血症發作，而是交給通道一的「分類器」來判斷敗血症發作，在此處，通道一的「分類器」是傳統的單層全連結神經網路。如表 8 所示，刪除模糊支持向量機網路後，AUROC 下降 1.349%，這證明模糊支持向量機對於分類正確率有顯著的貢獻，主要原因是傳統的單層全連結神經網路有許多缺點，例如會陷入區域最佳值 (local minimum) 與有過度學習 (over fitting) 的問題，相反地，根據統計學習理論而創建的支持向量機有優異的推理能力，而且 SVM 是透過求解二次最佳化問題 (quadratical programming problem, QPP) 來建立分類模型，所以 SVM 僅有惟一最佳解，並不會陷入區域極值。除此之外，融合模糊理論的模糊支持向量機能夠將訓練樣本的模糊性納入考慮，從而避免雜訊資料的影響，因此可以提高預測效能。
- **移除 DMKL (多核心深層模型, Deep Multiple Kernel Learning)：**它是指將第參之四節的「運用逐層特徵擷取建構的深層模糊支持向量機網路」移除，換句話說，在此處，我們不使用逐層擷取數個核心函數的特徵來堆疊建立深層模糊支持向量機，並且使用傳統的淺層支持向量機來判斷敗血症發作。如

表 8 所示，移除逐層擷取核心函數特徵的深層架構後，AUROC 下降 0.490%，這證明逐層擷取核心函數特徵建立的深層支持向量機能夠融合多個核心函數的優點，進而提升預測效能。

表 8 的消融分析結果證明，本論文提出的混合模型中的各個關鍵組件，都能夠有效地提升敗血症預測效能。

表 8：針對混合模型中不同關鍵組件的消融分析測試

	AUROC	AUPRC	Accuracy	F1-score
移除 CNN	0.806	0.084	83.6	0.107
移除 BiLSTM	0.814	0.091	84.8	0.124
移除 AM	0.812	0.091	84.7	0.121
移除 SVM	0.804	0.087	83.5	0.109
移除 DMKL	0.811	0.090	84.6	0.118
本研究的預測模型	0.815	0.093	85.0	0.127

六、重要模型參數的敏感度分析

在本章節中，我們對本研究提出的混合模型中的重要模型參數進行敏感度分析(sensitive analysis)，來分析這些模型參數的數值對於預測效能有何影響。首先，我們分析調節參數 λ 對於預測效能的影響，如公式(28)所示，模型參數 $\lambda \in [0, \infty)$ 調節分類損失 $L(\hat{y}^c, y)$ 和自調節損失 $L_{diff}(O_E, O_G)$ ，換句話說，當 λ 值越大，則自調節生成對抗模型較注重自調節損失 $L_{diff}(O_E, O_G)$ ，並且更致力於避免模型學習到生成器產生的錯誤虛假特徵，反之，當 λ 值越接近於 0，則自調節生成對抗模型較注重分類損失 $L(\hat{y}^c, y)$ ，並且更致力降低模型的分類錯誤率。表 9 顯示使用不同 λ 值建立模型時，模型的預測效能的變化。如表 9 所示，當 λ 值等於 0 時，預測效能較差(等於移除 GAN)，這是因為當 λ 值等於 0 時，編碼器模型只能從「正面」的方向學習到能夠幫助分類正確的特徵，而不能從「反面」的方向避免學習到會混淆判斷的錯誤特徵，當 λ 值逐漸增加時，預測效能也緩步增加，這是因為編碼器不但能從「正面」的方向學習到能夠幫助分類正確的特徵，也能從「反面」的方向避免學習到會混淆判斷的錯誤特徵，但是，當 λ 值超過 0.0001 時，預測效能又開始逐步降低，這是因為編碼器太過於注重從「反面」的方向避免學習到會混淆判斷的錯誤特徵，而忽略也要學習能夠幫助分類正確的特徵。

表 9：使用不同調節參數 λ 值時的預測效能

調節參數 λ	AUROC	AUPRC	Accuracy	F1-score
0	0.809	0.089	83.9	0.113
0.000001	0.810	0.090	84.2	0.120
0.00001	0.813	0.092	84.7	0.125
0.0001	0.815	0.093	85.0	0.127
0.001	0.813	0.091	84.5	0.124
0.01	0.810	0.089	84.1	0.118
0.1	0.807	0.084	83.5	0.110

在下一個實驗中，我們分析逐層擷取核心函數特徵的深層架構的深度 L 對於預測效能的影響，如圖 3(a)所示，參數 $L \in [1, \infty)$ 控制深層網路的深度，當 L 等於 1 時，即為傳統的淺層模糊支持向量機，表 10 顯示使用不同 L 值建立模型時，模型的預測效能的變化。如表 10 所示，當 L 值等於 1 時，預測效能較差(等於移

除 DMKL)，這是因為當 L 值等於 1 時，我們只是使用傳統的淺層模糊支持向量機，當 L 的值逐漸增加，我們的模型能更學習到更複雜的深度核心函數特徵，所以預測效能也逐步增加，但是，當 L 超過 4 時，預測效能則開始緩慢下降，這是因為深層網路的深度越大，模型要學習的參數也越多，而且太過於複雜的深度核心函數特徵也容易導致過度學習(overfitting)，所以當 L 值為 4 時，預測效能是最理想的。

表 10：使用不同的多層多核心網路的深度 L 值時的預測效能

多層多核心網路的深度 L	AUROC	AUPRC	Accuracy	F1-score
1	0.811	0.090	84.6	0.118
2	0.812	0.090	84.7	0.120
3	0.814	0.092	84.9	0.124
4	0.815	0.093	85.0	0.127
5	0.815	0.092	85.0	0.126
6	0.813	0.090	84.7	0.122

七、對比實驗

在本章節中，我們對本研究提出的模型架構進行對比實驗。首先，本研究在公式(26)中，使用 Frobenius norm 來計算編碼器與生成器輸出編碼的差異，然而，Bousmalis et al. (2016)使用 Frobenius norm 的目的是希望促成這兩種編碼的正交性，而正交性可用向量內積來衡量，假設編碼器 E 和生成器 G 的輸出向量分別為 $o_E \in R^{do}$ 和 $o_G \in R^{do}$ (do 為 o_G 的維度)，此二向量內積平方為 $(\sum_{k=1}^{do} o_{E,k} o_{G,k})^2$ ，如果有 M 個訓練樣本， $o_E \in O_E$ 和 $o_G \in O_G$ ，基於內積平方來衡量這兩種編碼的正交性的自調節損失函數 $L_{dot_diff}(O_E, O_G)$ 的計算如下

$$L_{dot_diff}(O_E, O_G) = \sum_{j=1}^M (\sum_{k=1}^{do} o_{E,k}^j o_{G,k}^j)^2 \quad (51)$$

表格11比較使用公式(26)的基於Frobenius norm的自調節損失函數 $L_{diff}(O_E, O_G)$ 與使用公式(51)的基於內積平方的自調節損失函數 $L_{dot_diff}(O_E, O_G)$ 的預測結果。

表11：使用不同的自調節損失函數的預測效能比較

自調節損失函數	AUROC	AUPRC	Accuracy	F1-score
基於內積平方的自調節損失函數 $L_{dot_diff}(O_E, O_G)$	0.814	0.091	84.7	0.126
本研究的自調節損失函數 $L_{diff}(O_E, O_G)$	0.815	0.093	85.0	0.127

如表11所示，本研究的基於Frobenius norm的自調節損失函數 $L_{diff}(O_E, O_G)$ 的預測效果是略優於基於內積平方的自調節損失函數 $L_{dot_diff}(O_E, O_G)$ ，而在模型學習的收斂速度部分，這兩種自調節損失函數並沒有太大的差異。本論文使用合作網路及深層模糊支持向量機網路，並提出一種自我調節(self-regulated)的生成對抗網路來生成隨機雜訊，以增加特徵的強健性和泛化性。之前的實驗結果數據呈現本研究提出的方法優於其他方法。為了強化本論文提出架構的優點，在接下來的實驗中，我們比較「僅使用合作網路及深層模糊支持向量機網路架構」以及本研究提出的「自調節生成對抗混和模型」的預測效能，以凸顯本論文提出的自我調節(self-regulated)生成對抗網路的價值。表12比較「僅使用合作網路及深層模糊支

持向量機」與本研究的「自調節生成對抗混和模型」的預測效能，其中，僅使用合作網路及深層模糊支持向量機，代表將圖1的雙通道模型中的通道二(生成器與判別器)予以刪除，換句話說，只留下通道一的合作網路(編碼器與分類器)。因此，僅使用合作網路及深層模糊支持向量機，我們的編碼器模型只能從「正面」的方向學習到能夠幫助預測敗血症的特徵，而不能從「反面」的方向避免學習到會混淆判斷的錯誤特徵。如表12所示，僅使用合作網路及深層模糊支持向量機的AUROC下降至0.809，這證明雙通道模型中的通道二內的生成器與判別器互相對抗的GAN模型，對於敗血症預測效能有顯著的貢獻，這是因為生成器與判別器彼此對抗的設計，能夠強化彼此的能力，而且，透過本研究提出的「自我調節(self-regulated)」的架構，能夠幫助「編碼器」避免學習到「生成器」所創建的混淆判斷的虛假特徵，使得編碼器模型可以學習到更加強健的特徵，所以有助於提升預測效能。

表12：僅使用合作網路及深層模糊支持向量機與本研究的混和模型的預測效能比較

模型架構	AUROC	AUPRC	Accuracy	F1-score
僅使用合作網路及深層模糊支持向量機	0.809	0.089	83.9	0.113
本研究的自調節生成對抗混和模型	0.815	0.093	85.0	0.127

伍、結論與建議

敗血症的早期檢測和抗生素治療對於改善敗血症結果至關重要，每延遲治療一小時，死亡率就會增加大約 4-8%。AI 計算方法有望改善敗血症的早期檢測，這些 AI 計算方法可以僅根據提供的臨床數據自動識別患者的敗血症發作風險，並在每小時的時間窗口內對敗血症做出陽性或陰性預測。根據 Sepsis-3 臨床標準，本研究試圖在敗血症發作前至少 6 小時預測敗血症。

引發敗血症的原因是錯綜複雜，必需要考慮各種不同類型的異質資料來源，其中，人口統計變量是靜態資料，生命體徵與實驗室數值是動態資料，而且，生命體徵與實驗室數值的量測頻率並不相同，生命體徵大多是每小時測量一次，而實驗室數值則是每天更新一次，換句或說，生命體徵顯示短程的時間依賴性，而實驗室數值則是披露長程的時間依賴性。顯而易見地，單一個預測模型不可能同時捕抓以下三種特徵：靜態特徵、短程的時間依賴特徵、與長時間的全局特徵。因此，本論文融合全連結神經網路、卷積神經網路、雙向長短期記憶體、注意力機制與模糊支持向量機。其中，全連結網路擅長分析靜態數據、卷積神經網路適合擷取與時間無關的局部模式特徵，雙向長短期記憶體可以從兩種不同的方向(亦即是，從過去到未來與從未來到過去)來分析時間依賴性與因果關係特徵，注意力機制可以對長時間的全局資料中的重要因素給予較多的關注。而模糊支持向量機則可以考慮資料的模糊與不確定性，並且根據統計學習理論打造最佳的分類器。

本研究使用 PhysioNet/Computing in Cardiology Challenge 2019 資料集建立敗血症早期預警模型，該資料集是收集美國的醫院系統內的病患資料，由於不同國

家的醫院環境並不相同，不同國家的人民的體質也可能有所差異，所以本研究建立的敗血症早期預警模型，不見得適用於台灣的敗血症患者，在未來，我們預期蒐集台灣醫院系統的病患資料，來建立針對台灣敗血症病患的早期預警系統。此外，敗血症的早期預測可能會挽救生命，但是，延遲預警或者錯失預警可能會危及生命，而誤報則會消耗醫院資源並削弱對 AI 演算法本身的信任。然而，傳統的評分指標，例如 ROC 曲線下面積 (AUROC) 指標，並不會明確獎勵早期預警，或者懲罰誤報與過度治療。在未來，我們預計設計一種新的基於臨床效用的評分指標，該指標能夠對早期敗血症預測的演算法進行獎勵，對晚期和錯過的敗血症預測以及誤報進行懲罰。並且，我們將開發能夠根據新的評分指標來建立預測模型的學習演算法。最後，雖然循環網路通常是序列數據建模的首選，但 BiLSTM 等循環網路容易過度擬合，而且，PhysioNet/Computing in Cardiology Challenge 2019 資料集擁有大量的遺失值，在向前填充如此眾多的遺失值時，其中許多遺失值被設置為健康人的典型值，使得數據的時間序列性質被破壞，因此失去 BiLSTM 所賦予的優勢，在未來，我們預期開發合適的遺失值處理策略，以保留資料的時間序列性質，並且提升預測效能。

參考文獻

- Apalak, M. & Kiasaleh, K. (2022). Improving sepsis prediction performance using conditional recurrent adversarial networks. *IEEE Access*, 10, 134466-134476.
- Barton, C., Chettipally, U., Zhou, Y., Jiang, Z., Lynn-Palevsky, A., Le, S., Calvert, J. & Das, R. (2019). Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Computers in Biology and Medicine*, 109, 79-84.
- Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D. & Erhan, D. (2016). Domain separation networks, in Lee, D.D., Sugiyama, M., von Luxburg, U., Guyon, I., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems*, 343-351.
- Chen, M., Lu, X., Hu, L., Liu, P., Zhao, W., Yan, H., Tang, L., Zhu, Y., Xiao, Z., Chen, L. & Tan, H. (2017). Development and validation of a mortality risk model for pediatric sepsis. *Medicine*, 96(20), e6923.
- Dark, P., Blackwood, B., Gates, S., McAuley, D., Perkins, G., McMullan, R., Wilson, C., Graham, D., Timms, K. & Warhurst, G. (2015). Accuracy of LightCycler® SeptiFast for the detection and identification of pathogens in the blood of patients with suspected sepsis: A systematic review and meta-analysis. *Intensive Care Medicine*, 41(1), 21-33.
- De Backer D. & Dorman T. (2017). Surviving sepsis guidelines: a continuous move toward better care of patients with sepsis. *JAMA - Journal of the American Medical Association*, 317(8), 807-808.

- Dubois, D. & Prade, H. (1998). Possibility theory: qualitative and quantitative aspects. *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, 1, 169-226.
- Elfeky, S., Golabi, P., Otgonsuren, M., Djurkovic, S., Schmidt, M.E. & Younossi, Z.M. (2017). The epidemiologic characteristics, temporal trends, predictors of death, and discharge disposition in patients with a diagnosis of sepsis: A cross-sectional retrospective cohort study. *Journal of Critical Care*, 39, 48-55.
- El-Rashidy, N., Abuhmed, T., Alarabi, L., El-Bakry, H. M., Abdelrazek, S., Ali, F. & El-Sappagh, S. (2022). Sepsis prediction in intensive care unit based on genetic feature optimization and stacked deep ensemble learning. *Neural Computing and Applications*, 34, 3603-3632.
- Faisal, M., Scally, A., Richardson, D., Beatson, K., Howes, R., Speed, K. & Mohammed M.A. (2018). Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results. *Critical Care Medicine*, 46(4), 612-618.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). Generative adversarial nets, in Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q. (Eds.), *Advances in Neural Information Processing Systems*, 27, 2672-2680.
- 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.
- Graves, A. & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5-6), 602-610.
- Hao, P.-Y., Chiang J.-H. & Chen, Y.-D. (2022). Possibilistic classification by support vector networks. *Neural Networks*, 149, 40-56.
- Hao, P.-Y. (2021). Asymmetric possibility and necessity regression by twin support vector networks. *IEEE Trans. on Fuzzy Systems*, 29(10), 3028-3042.
- Hinton, E. & Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313, 504-507.
- Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *MIT Press*, 9(8), 1735-1780.
- 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 (CVPR'06)*, New York, USA, 284-291.

- Islam, M.M., Yao, X. & Murase, K. (2003). A constructive algorithm for training cooperative neural network ensembles. *IEEE Transactions on Neural Networks* 14 (4), 820–834.
- Kim, J.K., Ahn, W., Park, S., Lee, S.-H. & Kim, L. (2022). Early prediction of sepsis onset using neural architecture search based on genetic algorithms. *International Journal of Environmental Research and Public Health*, 19, 2349.
- Kok, C., Jahmunah, V., Oh, S., Zhou, X., Gururajan, R., Tao, X., Cheong, K., Gururajan, R., Molinari, F. & Acharya, U. (2020). Automated prediction of sepsis using temporal convolutional network. *Computers in Biology and Medicine*, 127, 103957.
- Kumar, S., Tripathy, S., Jyoti, A. & Singh S.G. (2019). Recent advances in biosensors for diagnosis and detection of sepsis: A comprehensive review, *Biosensors and Bioelectronics*, 124-125, 205–215.
- Lagu, T., Rothberg, M.B., Shieh, M.-S., Pekow, P.S., Steingrub, J.S. & Lindenauer P.K. (2012). Hospitalizations, costs, and outcomes of severe sepsis in the United States 2003 to 2007. *Critical Care Medicine* 40(3), 754–761.
- Lee, B., Cho, K., Kwon, O. & Lee, Y. (2019). Improving the performance of a neural network for early prediction of sepsis. *2019 Computing in Cardiology (CinC)*, Singapore, 1-4.
- Li, K., Zhang, R. Li, F. Su, L. Wang, H. & Chen, P. (2019). A new rotation machinery fault diagnosis method based on deep structure and sparse least squares support vector machine. *IEEE Access*, 7, 26571-26580.
- Liu, B. (2000). Dependent-chance programming in fuzzy environments. *Fuzzy Sets and Systems*, 109(1), 97-106.
- Liu, Z., Khojandi, A., Mohammed, A. Li, X., Chinthala, L. K., Davis, R. L. & Kamaleswaranc, R. (2021). HeMA: A hierarchically enriched machine learning approach for managing false alarms in real time: A sepsis prediction case study. *Computers in Biology and Medicine*, 131, 104255.
- Liu, S., Fu, B., Wang, W. Liu, M. & Sun, X. (2022a). Dynamic sepsis prediction for intensive care unit patients using XGBoost-based model with novel time-dependent features. *IEEE Journal of Biomedical and Health Informatics*, 26(8), 4258-4269.
- Liu, Z., Khojandi, A., Li, X., Mohammed, A., Davis, R. L. & Kamaleswaran, R. (2022b). A machine learning-enabled partially observable markov decision process framework for early sepsis prediction. *INFORMS Journal on Computing* 34(4), 2039-2057.

- Martin, G.S., Mannino, D.M., Eaton, S. & Moss, M. (2003). The epidemiology of sepsis in the United States from 1979 through 2000. *New England Journal of Medicine*, 348(16),1546-1554.
- Nakhashi, M., Toffy, A., Achuth, P. V., Palanichamy, L. & Vikas, C. M. (2019). Early prediction of sepsis: using state-of-the-art machine learning techniques on vital sign inputs. *2019 Computing in Cardiology (CinC)*, Singapore, 1-4.
- Nesaragi, N., Patidar, S. & Aggarwal, V. (2021). Tensor learning of pointwise mutual information from EHR data for early prediction of sepsis. *Computers in Biology and Medicine*, 134, 104430.
- Ou, S.-M., Chu, H., Chao, P.-W., Lee, Y.-J., Kuo, S.-C., Chen, T.-J., Tseng, C.-M., Shih, C.-J. & Chen, Y.-T. (2016). Long-term mortality and major adverse cardiovascular events in sepsis survivors a nationwide population-based study. *American Journal of Respiratory and Critical Care Medicine*, 194(2), 209–217.
- 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.
- Reyna, M.A., Josef, C.S., Jeter, R., Shashikumar, S.P., Westover, M.B., Nemati, S., Clifford, G.D. & Sharma, A. (2019). Early prediction of sepsis from clinical data: the PhysioNet/Computing in Cardiology Challenge. *Critical Care Medicine*, 48(2), 210-217.
- Saqib, M., Sha, Y. & Wang, M.D., (2018). Early prediction of sepsis in EMR records using traditional ML techniques and deep learning LSTM networks. *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, USA, 4038-4041.
- Sharma, D. K., Lakhotia, P., Sain, P. & Brahmachari, S. (2022). Early prediction and monitoring of sepsis using sequential long short term memory model. *Expert Systems*, 39(3), e12798.
- Shashikumar, S. P., Josef, C. S., Sharma A. & Nemati, S. (2021). DeepAISE – An interpretable and recurrent neural survival model for early prediction of sepsis. *Artificial Intelligence in Medicine*, 113, 102036.
- Singer, M., Deutschman, C.S., Seymour, C.W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G., Chiche, J.-D., Coopersmith, C., Hotchkiss, R., Levy, M., Marshall, J., Martin, G., Opal, S., Rubenfeld, G., van der Poll, T., Vincent, J.-L. & Angus, D. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA - Journal of the American Medical Association*, 315(8), 801-810.
- Tankaka, H., Uejima, S. & Asai, K. (1982). Linear regression analysis with fuzzy model. *IEEE Transactions on Systems, Man, and Cybernetics*, 12(6), 903-907.

- Taylor, R., Pare, J., Venkatesh, A., Mowafi, H., Melnick, E., Fleischman, W. & Hall, M. (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.
- 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., et al. (2017). Combining biomarkers with EMR data to identify patients in different phases of sepsis. *Scientific Reports*, 7(1), 10800.
- Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-146.
- Vincent, J.-L., Sakr, Y., Sprung, C.L., Ranieri, V.M., Reinhart, K., Gerlach, H., Moreno, R., Carlet, J., Le Gall, J-R. & Payen, D. (2006). Sepsis in European intensive care units: Results of the SOAP study. *Critical Care Medicine*, 34(2), 344–353.
- Wang, G., Zhang, G., Choi, K. & Lu, J. (2019). Deep additive least squares support vector machines for classification with model transfer. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(7), 1527-1540.
- Wang, Z. & Yao, B. (2022). Multi-branching temporal convolutional network for sepsis prediction. *IEEE Journal of Biomedical and Health Informatics*, 26(2), 876-887.
- Xu, Y. & Cohen, S. B. (2018). Stock movement prediction from tweets and historical prices. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, July, 1970–1979.
- Xu, H. Cao, D. & Li, S. (2022). A self-regulated generative adversarial network for stock price movement prediction based on the historical price and tweets. *Knowledge-Based Systems*, 247, 108712.
- Yao, S., Yang, J.-B., Xu, D.-L. & Dark, P. (2021). Probabilistic modeling approach for interpretable inference and prediction with data for sepsis diagnosis. *Expert Systems with Applications*, 183, 115333.
- 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.
- Yu, Q., Huang, X., Li, W., Wang, C., Chen, Y. & Ge, Y. (2019). Using features extracted from vital time series for early prediction of sepsis. *2019 Computing in Cardiology (CinC)*, Singapore, 1-4.

