

出國報告（出國類別：開會）

2024 年
歐洲重症醫學會國際研討會

服務機關：臺中榮民總醫院重症內科

姓名職稱：趙文震 科主任

出國期間：2024年10月04日至2024年10月10日

報告日期：2023年11月08日

摘要

此次重症醫學部重症加護內科趙文震醫師於參加 2024 年歐洲重症醫學會 (ESICM) 國際會議，並發表題為「應用重症資料庫之多任務機器學習模型預測呼吸器脫離」的研究。該研究聚焦於提升 AI 模型的解釋性。透過 SHAP 與 PDP 等視覺化工具呈現模型解釋性，使臨床醫師能夠清楚了解各特徵對預測結果的影響，增進對 AI 的信任度及使用意願，這也是衛福部目前三大中心中「負責任 AI」的核心理念之一。會議期間參加了多場 AI 相關討論，深入探討 AI 模型應用於重症醫學中的挑戰及解決方案。首先，資料來源的偏見是限制 AI 應用的重要因素，如目前常用的 MIMIC IV 資料庫主要來自單一醫院，且多數病患為白人男性，可能造成偏差。為解決此問題，建議整合多中心數據以提升模型的代表性。同時，資料隱私的議題在不同國家具有不同法規需求，如歐盟要求更嚴格的個資保護，而美國則更傾向數據開放，聯邦學習方式是目前相對有效的解決方案。此外，目前 AI 模型在醫院的實際應用率相對低，除了資工技術外，還需評估臨床整合度，特別是醫院智慧化程度，這一點是台中榮總的重大優勢。在敗血症免疫學領域參加了相關講座與海報討論，轉錄組學應用於敗血症病患免疫功能分析的有挑戰也有前景，挑戰在於病患異質性太高，可能要聚焦於敗血症的特殊族群，前景在於目前單細胞轉錄組學已相對較成熟，有機會更精細分析敗血症免疫特徵。藉由參加此次會議深刻體會 AI 與臨床重症醫學整合的必要性，也更確認 AI 的解釋性是醫療 AI 臨床落地應用的關鍵之一，此外亦須持續進行基礎性重症免疫學之研究。

關鍵字： 重症醫學、人工智慧、落地運用、敗血症、轉錄組學

目 次

一、 目的.....	1
二、 過程.....	1-2
三、 心得.....	3-9
四、 建議事項.....	9-10
五、 附錄.....	100

一、 目的

- 重症領域知識更新：參與本次ESICM的國際會議，透過大型演講、中小型研討會、海報發表與討論，可以更有效率的掌握最新研究進展和治療方法。
- 學術交流：ESICM為歐洲最大重症學會年會，透過交流、分享研究成果、討論問題，可以與不同領域的專家交換意見，分享想法和治療經驗。
- 國際曝光：透過展示本院的研究成果，可以讓更多不同國家地區的醫療人員了解並認識中榮，吸引更多的關注和合作機會。

二、 過程

海報展過程:

ESICM 的海報展相對其他學會(如美國胸腔醫學會)有所不同，首先是全部都是電子海報，這也呼應 ESG 的訴求，此外電子海報可讓報告者事先設計重點 Zoom In (我設了四個主要圖表 Zoom in, 如圖一)，這對報告者與評委/現場聽眾溝通相當方便；第二點是全部都是口頭報告，每個海報區都會指派兩位資深學者評論及引導現場討論，每個海報區都有提供各區的耳機，所以不會互相干擾，想聽講的海報或演講在不同區也可很快轉換。[註: 在美國(如美國胸腔醫學會)是一個房間一區，優點是可專心聽講與討論，缺點是不同區通常離很遠，所以就只能選擇少數區的 c-poster 報告]。

我這次報告的題目是應用 MIMIC IV 資料庫多任務機器學習模型預測呼吸器脫離，會議的前兩周論文剛被接受 (Digit Health; PMID: 39381828)，會議期間正在做最後校稿，該文章的焦點是模型解釋性，以 SHapley Additive exPlanations (SHAP)與 partial dependent plot (PDP)這兩個視覺化工具呈現各特徵對結果之影響方向與強度，模型解釋性這也是臨床醫療與資工技術間重要的溝通工具，可以讓跨領域雙方更加理解模型背後的運作邏輯，更重要的是以臨床專業來解釋模型才是臨床醫師在 AI 領域的核心優勢，兩位主持人和現場聽眾都是臨床醫師所以對解釋性這一部分相當有興趣，提問時間我大多數的時間就是逐一說明解釋性背後的原理，兩位主持人與聽眾也理解模型原來可以這樣解釋說明，這樣就能讓臨床醫療人員理解模型才會真的相信模型推論出來的結果，未來才會真的願意使用模型 (圖二)。模型解釋性也是目前所謂負責任人工智慧的基礎之一 (詳見下述)。

Multitask learning to predict successful weaning in critically ill ventilated patients: a retrospective analysis of the MIMIC IV database

Ming-Yen Lin¹, Hsin-You Chi², Wen-Cheng Chao^{3*}

¹ Department of Information Engineering and Computer Science, Feng Chia University, Taichung, Taiwan.

² Department of Critical Care Medicine, Taichung Veterans General Hospital, Taichung, Taiwan.

³ Department of post-Baccalaureate Medicine, College of Medicine, National Chung Hsing University, Taichung, Taiwan.



INTRODUCTION

Weaning is an essential issue in critical care. Multitask learning (MTL) is emerging as a novel approach in artificial intelligence, particularly for handling complex data, and is characterised by shared feature representation, improved generalisation, and resource efficiency. This study used the Medical Information Mart for Intensive Care (MIMIC) IV, a public critical care database with comprehensive ventilator-relevant data, in conjunction with multitask learning techniques to establish a prediction model for successful weaning in critically ill ventilated patients. Our approach aims not only to establish a prediction model with high accuracy but also responsible AI practices, emphasising model interpretability and error analysis.

METHODS

We employed a multitask learning framework with shared bottom network to facilitate common knowledge extraction across all tasks. We explored distinct task combinations of multitask learning. We used Shapley additive explanations (SHAP) plot partial dependence plot (PDP) for model explainability. Area under receiver operating characteristic curve (AUROC), calibration plot and decision curve analysis were used to determine the performance of the model.

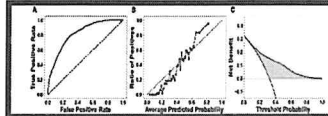


Figure 1. Performance of the proposed model to predict successful weaning. A. ROC plot. B. calibration curve. C. decision curve analysis

RESULTS

We found that 78.5% of 7,758 critically ill patients were successfully weaned. Multitask learning combined with spontaneous breath trial achieved a higher performance to predict successful weaning compared with multitask learning combined with shock and mortality (area under receiver operating characteristic curve, AUROC, 0.820 ± 0.002 vs 0.817 ± 0.001 , $p < 0.001$). We employed not only calibration curve and decision curve analysis to examine the performance of the model but also the SHAP and PDP plots to interpret the model.

CONCLUSIONS

We demonstrate that MTL can effectively enhance predictive accuracy in critical care settings with complex, multifaceted and interrelated data. We established an MTL model with high accuracy to predict successful weaning. Moreover, we conducted the visualised interpretability of the model and error analysis, and these approaches enable the user to realise the model, to know limitations, and to trust the model.

Table 1. Performance of single-task and multitask learning to predict successful weaning

	AUROC	Precision	F1-score	Recall
Single task learning ¹	0.814 ± 0.002	0.673 ± 0.024	0.558 ± 0.037	0.481 ± 0.065
Multi-task learning-1 ²	$0.817 \pm 0.001^*$	0.670 ± 0.042	0.576 ± 0.041	0.513 ± 0.079
Multi-task learning-2 ³	$0.820 \pm 0.002^*$	0.661 ± 0.010	0.582 ± 0.025	0.522 ± 0.043

¹Shock and mortality

²Tasks included shock, mortality, and successful weaning

³Tasks included start of spontaneous breathing trial, successful spontaneous breathing trial, and successful weaning

Table 2. Performance of original test population and sub-group analysis of multitask learning to predict successful weaning

	AUROC	Precision	F1-score	Recall
Original population (n=1917)	0.821	0.660	0.598	0.546
Sub-group (n=1291) ^a	0.827	0.657	0.487	0.387

^a Exclusion of 626 patients whose Pmean lower than 9.5 cmH2O and Enteral feeding higher than 589 ml/day

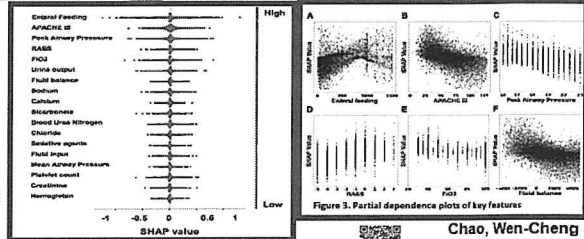


Figure 2. SHAP to illustrate the key features to predict successful weaning.

Figure 3. Partial dependence plots of key features

Chao, Wen-Cheng

cwc081@hotmail.com

圖一、本次發表之海報



圖二、海報報告時與主持人之討論

三、心得

Part I. AI 很熱門，但發展過於快速，進入臨床前可能需要盤整與省思

AI 主題在不同學會包含 ESICM 一樣都很熱門，但和過去幾年不同的點在於今年著重在省思與盤整，這不是壞事，反而是代表已經進入由研究階段進入臨床應用前的評估。巧合的是開會期間剛好也是 AI 大師 Hinton 拿到諾貝爾物理獎的時候，其實多數人不知道，Hinton 2023 刻意離開 Google 回到學校(多倫多大學)就是覺得 AI 進展太快，回到學校才能以純學者身分論述發展 AI 該有的省思。來自 Erasmus University Medical Center 的 Michel E van Genderen 是這個場次主要講者，我晚上回飯店趕快查他最近在重要期刊發表相關文章 (主要有下列 4 篇 1. Extended reality in critically ill patients: "NOT" yet ready for take-off! Intensive Care Med. 2024 Aug;50(8):1365-1367. PMID: 38869674. 2. Federated data access and federated learning: improved data sharing, AI model development, and learning in intensive care. Intensive Care Med. 2024 Jun;50(6):974-977. PMID: 38635044. 3. Charting a new course in healthcare: early-stage AI algorithm registration to enhance trust and transparency. NPJ Digit Med. 2024 May 8;7(1):119. PMID: 38720011. 4. To warrant clinical adoption AI models require a multi-faceted implementation evaluation. NPJ Digit Med. 2024 Mar 6;7(1):58 PMID: 38448743.)。我看了這幾篇論文和本次 ESICM 的討論內容，就重症領域 AI 發展提出的四項主要省思以及可能解決方案綜整如下。

I. 模型資料來源偏見的風險 (data bias)

省思:

其實這個問題一開始就知道，只是走到接近臨床應用才更加重視可能的影響，如 MIMIC 是目前最常用到的重症資料庫，講者統計目前將近 30% 的模型都是以 MIMIC 的資料為主，這是一個很高的數字(背後代表下一個須省思問題，就是各國資料隱私規範不一)，但大家也都知道 MIMIC 是 Beth Israel Deaconess Medical Center 單一醫院的資料庫，將近 80% 為白人資料，多數為男性病患資料。所以研究一直到這一兩年才開始省思 AI 演算法可能加劇健康差異，尤其是在少數族群或特定年齡段的臨床應用中，由於訓練數據集中缺乏這些族群的代表性，導致演算法偏見。

可能解決之道:

在開發 AI 模型時，應要求公開訓練數據的族群特徵，例如年齡、性別和種族，並持續對其進行偏見緩解措施，以確保 AI 系統能夠公平應用於不同患者群體。其實這一點也是我今年預計申請國科會計畫的主要論述之一，因為中榮重症資料庫是單一中心研究，未來朝向與國際上公開之重症資料庫合併使用，這樣不但是多中心研究，也可解決 data bias 的問題，甚至刻意以少數族群(如高齡重症病患)為主軸，因為這才是 research niche/unmet need。

目前已取得之資料庫包含 MIMIC IV, e-ICU, SICdb, 各資料庫之說明如下。

- ▶ 中榮重症資料庫: 台灣，單一中心，個案數 1.6 萬，年度 2015-2022，有長期死亡資料。
- ▶ MIMIC IV: 美國,單一中心,個案數 2.4 萬，年度 2008-2019，2023 開始提供長期死亡資料。
- ▶ e-ICU: 美國,多中心，個案 7.6 萬(不過缺失值多)，年度 2014-2015，無長期死亡資料。
- ▶ SICdb: 德國，單一中心，個案 2.7 萬，年度 2013-2021，約 80%病患有長期死亡資料。

II. 醫療資訊資料隱私問題 (data privacy)

省思:

這個問題在不同國家有不同觀點。如一樣使用 MIMIC 資料庫做研究，美國的研究機構認為”不須”要再另外送 IRB 審查，因為美國以 HIPAA 為基礎，認為在醫療數據去識別後，這些資料就不再屬於醫療個資，因此可以進行自由研究。所以 MIMIC 重症資料庫就被認為是已經不再涉及醫療個資隱私問題，成為一個可以被自由使用的數據資源，為重症醫學研究提供了寶貴的資料支持。但歐盟以 GDPR 為基礎對醫療數據隱私的保護相當嚴格。認為去連結的資料依然有可能透過其他數據交叉引用重新識別出個人身份，因此在使用這些資料時，仍需遵守個資保護原則，並且必須有合理的法律依據與知情同意(Opt-in 或 Opt-out)，不過究竟依據哪一條法律，歐盟就沒再往下制定規範，所以歐洲是嚴格又有灰色空間，有的國家強調關鍵在於資料開放前要有完整去辨識/連結，甚至要應用資工特殊工具將資料模糊化；有的國家則認為需做到須有退出權(Opt-out)。在台灣，則因為健保資料庫釋憲案的影響，趨勢朝向需要保留資料提供者的退出權。

可能解決之道:

如果隱私認定無法突破，聯邦學習就是目前最可行的解決之道，但聯邦學習就是多一層硬體

需求，所以門檻相對高；此外聯邦學習還有另外一個問題就是研究人員通常習慣要看清楚所有資料，才能往下探討新的議題，聯邦學習往往對研究人員相對有一點不能適應/接受。聯邦學習以外的解決之道就是資料隱私要做到什麼程度，以去連結為主或是一定要保留資料提供者的退出權，有趣的是，因為是 ESICM，所以與會者多是歐洲學者，討論通通偏向應該學美國以開放態度減少限制，不然會落後美國與中國發展 AI 的腳步。此外，也有學者提出醫學資料庫的價值是病人與醫療團隊共同付出下的產出，不應通通視為病患個資而予以限制。就這一點類似過去曾與蔡甫昌教授討論健保資料庫釋憲案，蔡教授認為醫界當初應該主動發聲，主張健保資料庫的價值除了病患資料以外也包含全台灣醫療人員投入的智慧結晶，所以不應給予過高的退出權。不過 TAIRB 年會(2024/11/02)和蔡甫昌教授與衛福部資訊處李建璋處長討論，李建璋處長說衛福部資訊處目前是依據釋憲案執行退出權(Opt-out)，短期內趨勢很難改變，不過希望至少可以讓退出權的執行不要太複雜，不然台灣醫療大數據發展會很困難。

III. 醫療人員對模型的信任/透明度 (Trustworthy AI/AI transparency) :

省思:

AI 模型的臨床採用取決於信任，但現有的系統在數據集使用和模型驗證方面缺乏透明度，導致醫療專業人員對於 AI 系統的應用信任度不高，其實就是 black-box issue。

可能解決之道:

其實這一點跟衛福部”負責任 AI 執行中心”的核心概念很像，主要就是 AI 模型要提供解釋性，提升 AI 透明度；此外要有持續性的錯誤分析(error analysis)以持續監測與優化 AI 模型的效能。與會學者還提到，未來可能要像臨床試驗建立強制性 AI 演算法註冊/登錄機制，該註冊/登錄系統應包含數據來源、模型設計及臨床驗證的詳細資訊，以增強倫理監管並促進合規性，從而提高透明度和信任。

IV. 模型臨床實際應用/落地的限制 (Landing issue)

省思:

儘管 AI 技術在臨床決策支持中展現潛力，但實際上只有不到 2% 的 AI 模型能成功應用於臨床環境。主要原因包含模型設計是否能嫁接重症臨床工作流程，更重要的是醫院基礎資訊建設是否足夠與 AI 模型嫁接。

可能解決之道:

在 AI 評估中引入多面向的實施評估方法，不僅僅評估技術指標（如準確性），還應考量臨床接受度與系統整合度，以確保其臨床價值增強 AI 與臨床工作流程的整合。臨床接受度主要是在設計 AI 系統時，應注重其與醫療臨床流程的整合，優化用戶界面和數據輸入流程（舉例如重症跨團隊的資訊整合匯入的最佳時間點應該就是早上 8 點，因為可同時擷取護理、檢驗檢查、生理參數、呼吸器參數、等資料，）優化與重症醫療工作流程的結合，從而促進 AI 系統的採納，並根據 AI 系統往下做臨床決策，這其實就是所謂 Actionable AI 的精神。至於系統整合度則是與醫院資訊/智慧化程度高度相關，這也是中榮入選全球百大智慧醫院的關鍵，智慧化醫院是 AI 的核心基礎建設。與會者討論建議產品開發者可能須開發不同層次的模組，如高準確度但複雜的模組適合高度資訊/智慧化的醫院，但開發者可能也須設計一般準確度但簡單的模組以應用於資訊/智慧化稍差的醫院。

IV. 學界在大型語言模型(Large Language Model)的角色

省思:

過去都是學界和醫院合作研發，做出模型後找業界產品化。但大型語言模型時代來臨，學界其實沒有足夠的資本(包含硬體算力、硬體所需電力)，業界也不願意開源 (open code)，所以研發掌控權在”少數業界”而非不在學界，學界教授和學生做的都是應用型研究而非基礎型開創研究。目前僅有 Meta 的 Llama3 開源，但 Meta 開源的目的也是出於商業目的，因已落後 Open AI，所以開源才能搶市占率，長期不見得會持續開源。所以目前 LLM 研發的趨勢是業界直接找醫院合作，跳過學界。講者和與會討論者都提到這樣的問題會在於業界的思維為短期內開發出產品為導向，很難有先打基礎再求突破的階段性研究思維，所以無法像過去的醫學研究一步一步來，雖然短期可以有產出，但在培育基礎人才和延伸性研究可能就無法兼顧。

可能解決之道:

講者自己覺得很難，可能需要大型研究/國家級機構來統整資源，建立一個以研究應用為目標大型語言模型的 foundation model 供學界使用，其實台灣的 TAIDE 就是類似這樣的背景。但是講者擔心大型語言模型迭代太快，可能還是會跟不上業界，如 GPT 6.0 出來，大型語言模型的生態可能又是另外一個境界。最後結論是大型語言模型的世代來臨，學界要不得不改變，

而且要適應不斷的改變；醫院端其實也一樣，醫院有珍貴的臨床資料，但是如果只和業界合作要很小心可能都是偏向短期性應用型研究產出。最後講者說到，大型語言模型業界獨大的問題是結構性問題，不是只有醫療，所有領域都遇到類似問題，學界不用太悲觀，因為業界競爭激烈，只要不賺錢，可能基礎開創性質的研究就又會回到學界。以我而言，一直保持與逢甲資工林明言合作，林老師每年皆會指派一位碩班學生做醫療 AI 研究，我們每兩周的周三晚上 8:30-9:30 都會固定用線上會議討論進度，不知不覺已經過了四年，這次海報也是我與林老師及碩班學生季星佑三人的研究成果，這篇論文作者也就是我們三個。學界因為要帶學生所以節奏會比較慢，而且學生本來就是會出錯，尤其是清理資料這個階段特別容易出錯，我們開會討論重心都在於仔細分析結果與規畫研究方向，所以我有機會持續在討論過程中理解資工領域觀念性問題。

總結來說，2016 年 AlphaGo 迄今短短 8 年，但 AI 進展快速一路由 CNN，RNN 到現在的 LLM。但在醫療上，一開始都是以資料收集及模型建立為主，這 1-2 年才開始省思上述問題，但會議期間發現其實討論不完，因為問題太多，每個人觀點不同，很難聚焦、結論是上述幾項其實只是目前遇到的基本問題，未來還會有更多問題，如給付問題、
、
。反向思考就是解決這些問題目的在於讓 AI 融入醫療，從而降低醫護人員的負擔，讓醫護人員減少一點整合資料時間，多一點時間與病患及家屬溝通。

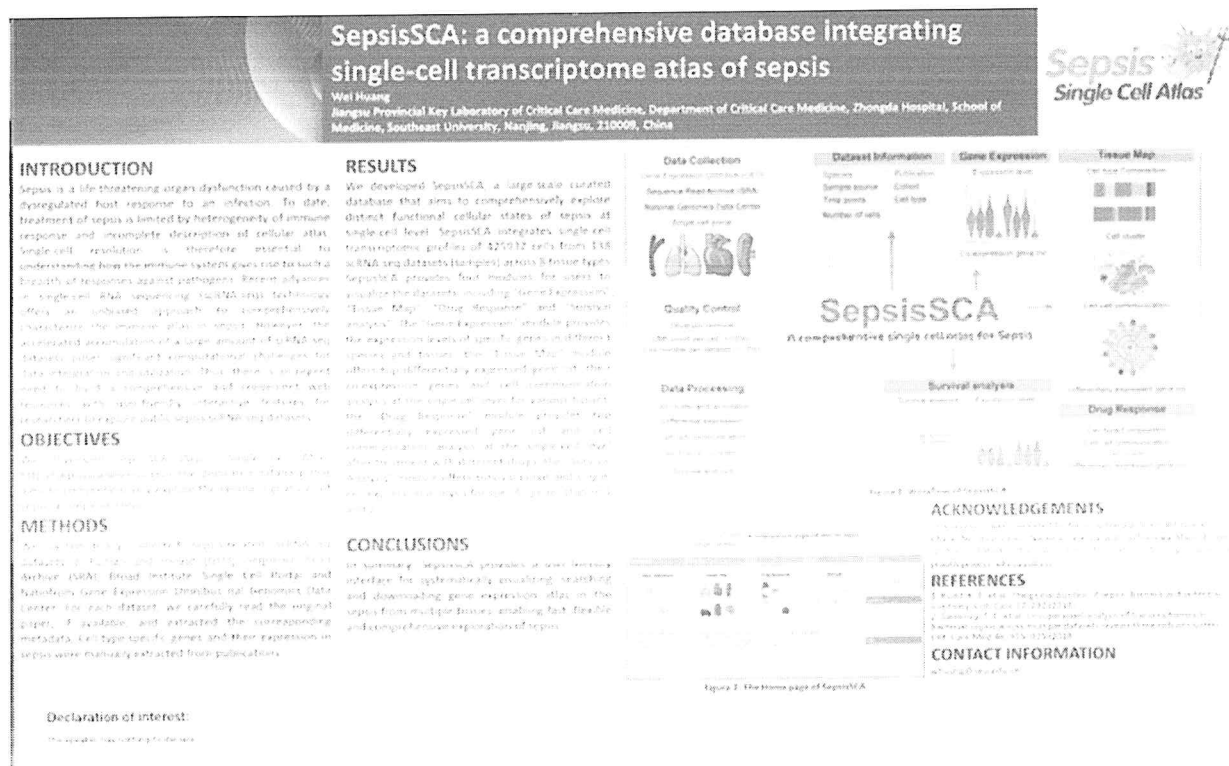
Part II. 敗血症病患的免疫分析很重要，相對冷門，仍需持續投入研究

因為免疫學是我轉譯醫學進修時的基礎，這幾年在國科會及中榮院內計畫支持下也一直以 RNA-Seq 進行敗血症免疫學研究，聚焦於敗血症後(day-8)的 T 細胞免疫不全。這次 ESICM 議程出來我就開始規劃去聽這方面的演講以及海報。說實話，敗血症免疫學領域相對冷門，偏向基礎醫學探討，尚無明確臨床應用。冷門領域有時候就是很冷清，在黑暗中尋找方向，而且往往是找出更多未知的東西，要耐得住寂寞且保持研究熱忱，往前走出一步就能享受突破的樂趣，這就是科學的本質。ESICM 期間我主要多是在兩個中型講堂聽這方面的演講以及在海報區與報告者討論。心得摘錄如下。

轉譯醫學訓練下的習慣就是要從你面前走過的病人留下檢體，未來有新穎工具出來時就可以用不同層次的角度看事情。但是要小心新穎工具有時候並不適合所有的研究，尤其是在敗血症這種高異質性的病人族群。以我最有興趣的 T 細胞免疫來說，3-4 年前很熱門的 T cell repertoire 分析 T cell receptor (TCR)序列今年就幾乎消失，因為我也曾投入 20 多萬做過一輪，所以清楚這個工具不適合使用敗血症，主要是敗血症會引起的是 T 細胞的功能性抑制或耗竭，但 TCR 重組序列主要反映 T 細胞的多樣性，而非其功能狀態。換言之，即使 TCR 多樣性顯示正常，T 細胞可能已因敗血症而呈現功能不全，無法有效參與免疫反應。

相對的，單細胞 RNA 定序 (single cell RNA-Seq) 在敗血症 T 細胞功能研究中具有顯著的優勢。因為可以逐細胞分析 RNA，可以在單細胞層級解析每個 T 細胞的基因表達特徵，從而探討不同 T 細胞亞群的功能性狀態及其在敗血症中的變化。但是單細胞 RNA 定序實在太貴，敗血症異質性太高，如菌種不同、嚴重度不同、共病不同、抽血時間不同、。此外敗血症病患血球大量活化又大量死亡，這又造成單細胞 RNA 定序分析上的困難。這次 ESICM 有位大陸學者的實驗室就把目前有發表的敗血症病患的單細胞 RNA 定序資料全部放到一個平台，稱之為 Single Cell Atlas (SCA)。不過與其討論後就知道其實他們團隊一開始也想找出是否有特殊細胞型態，但是也是遇到敗血症異質性太高的問題，此外每個研究上傳完整的定序資料，但關鍵的臨床資訊常常會有所保留，所以最終只好先把這些資料弄成平台。我們討論了近半小時，共同的想法是敗血症異質性太高所以很難做，單細胞 RNA 定序應該就是目前分析

敗血症免疫學最好的工具，但可能要聚焦在不太會大量死亡的細胞，如 T 細胞。雖然又是找出更多未知的東西，不過這也是科學探索的有趣之處。綜整這次參加 ESICM 的心得，AI 研究很熱門，進展很快，大方向很快就出來；相對敗血症免疫學很冷門，很多時間是在摸索未知，能找到一點小方向就會覺得很有趣。



圖三、敗血症單細胞 RNA 定序資料平台

四、 建議事項

(一) 醫療 AI 參與者須清楚包含解釋性 AI 等負責任 AI 的觀念

說明：多數 AI 應用以高準確性為主，但醫療 AI 很難接受黑盒子效應，所以需解釋性、錯誤分析等所謂負責任 AI 概念才能說服臨床醫療人員相信 AI 與使用 AI，這也是衛福部三大 AI 中心的之一心「負責任 AI 執行中心」的核心概念，建議應提早灌輸醫院對 AI 有興趣同仁此觀念。

(二) 醫療 AI 與業界合作朝產品開發，但也需與學界合作維持研究能量

說明：與業界合作需注意業界有其壓力，偏向快速開發出產品；但 AI 進化很快，仍須與學界保持合作才能跟得上 AI 的進化。建議醫院研究計畫須有一定比例鼓勵與學

界合作 AI 領域研究。

(三) 醫院須以計畫支持不易很快有產出之基礎性質研究

說明：基礎性質研究領域成本高，產出相對少，但還是醫學的重要基礎之一，建議醫院應保持目前以整合型/多年期計畫支持跨基礎與臨床領域的研究。(註：目前研究部的定期演講即是很好的做法，可以讓臨床端更了解醫院基礎性質研究方向以及對內與對外合作對象)。

(四) 醫療 AI 要積極讓年輕醫師參與，因為他們就未來的使用者

說明：AI 熱潮始與 2016 年，目前年輕醫師/醫事人員當年在學校都還沒有機會接觸 AI 領域，但其實年輕人資訊化能力本來就高，而且他們就是未來的使用者，建議於院內推廣 AI 教育訓練應特別注意年輕醫事人員之參與。

五、 附錄

略 (相關圖片皆檢附於上述報告)。

