

分担研究報告書

4. パネルデータを活用したユースケースの検討

研究分担者 飯塚 悠祐 自治医科大学附属さいたま医療センター 講師

研究分担者 橋本 悟 京都府立医科大学 准教授(病院教授)

研究分担者 野村 岳志 東京女子医科大学 教授

研究分担者 長谷川 高志 特定非営利活動法人日本遠隔医療協会 特任上席研究員

研究要旨

<背景>

集中治療領域では、患者の状態悪化を予測する人工知能の開発が進んでいる。近年、ピーチインテリヘルス社の人工知能(AI)を用いて、18~24時間前にSOFA(Sequential Organ Failure Assessment)スコアを予測することが可能となった。このAIは米国やシンガポールの病院の集中治療室電子データを用いて検証されているが、本邦の患者では検証されていない。また集中治療領域では、高い死亡率を有する敗血症の早期発見、早期介入が非常に重要であるが、現行の敗血症診断基準に則って敗血症を高い精度で予測するアルゴリズムは未だ実臨床に投入されていない。ピーチインテリヘルス社は、SOFAスコアの予測に加え、現行の診断基準で定義される敗血症、血管作動薬の必要性、及び全身性炎症反応症候群(SIRS: systemic inflammatory response syndrome)陽性を事前に予測するニューラルネットワークを用いた新しいアルゴリズムを設計した。今回、自治医科大学附属さいたま医療センターではパネルデータを活用したユースケースの検討として、当院のデータベースを活用し、ピーチインテリヘルス社のAIの予測精度を検証することとした。

<研究方法>

単施設の後向き観察研究として、自治医科大学附属さいたま医療センターの集中治療室における匿名化された患者データを使用した。データセットは、2017/7/31から2021/1までの18歳以上の7,283人の入院患者から構成され、必要に応じて日本語から英語へのデータを翻訳し、得られたデータをピーチインテリヘルス社の有するニューラルネットワークによるAIに投入し、重症化の事前予測能を検証した。

<成果>

7,283人の患者のうち、敗血症とSIRSの予測は7,159人(98%)、血管作動薬必要性の予測は7,119人(98%)、予測SOFAスコアは6,258人(86%)について解析された。このデータセットでは、61%(4,468例)が抗生物質を投与され、37%(2,703例)が抗生物質投与及びSOFAスコア2点以上上昇の両者を満たした。予測モデルは、敗血症発症、血管作動薬の必要性、SIRS陽性について、それぞれ1.0、0.98、0.90のROC曲線下面積を達成した。SOFAアルゴリズムの平均絶対誤差は、2時間予測で0.74、12時間予測で0.99であった。

<結論>

ピーチインテリヘルス社のモデルは、SOFAスコアの変化、現行の敗血症診断基準による敗血症発症、血管作動薬必要性、SIRS陽性の予測において、日本の大規模単一施設のデータベースにおいて、優れた精度を示した。この結果は、個々の病院のデータベースに最適化された効果的な予測モデル生成の可能性を示すものであると考えられる。

## A. 研究目的

集中治療重症度ツール開発研究分担施設である自治医大附属さいたま医療センターでは、海外で診療現場に導入済みのAIアルゴリズムを用いた24時間後重症度予測システムの妥当性・有用性に関する実証研究を実施している。

集中治療領域では、患者の状態悪化を予測する人工知能の開発が進んでいる。近年、ピーチインテリヘルス社の人工知能(AI)を用いて、18~24時間前にSOFA(Sequential Organ Failure Assessment)スコアを予測することが可能となった。このAIは米国やアジアの病院の集中治療室電子データを用いて検証されているが、本邦の患者では検証されていない。また集中治療領域では、高い死亡率を有する敗血症の早期発見、早期介入が非常に重要であるが、現行の敗血症診断基準に則って敗血症を高い精度で予測するアルゴリズムは未だ実臨床に投入されていない。(敗血症発症を、人工知能を用いて予測する研究は多いが、過去の敗血症の診断基準を使用しているものが多く[1,2]現行の敗血症診断基準(Sepsis-3基準:感染症によるSOFAスコア2点以上の臓器障害)[3])を使用した研究は少ない。)

ピーチインテリヘルス社は、SOFAスコアの予測に加え、現行の診断基準で定義される敗血症、血管作動薬の必要性、及び全身性炎症反応症候群(SIRS: systemic inflammatory response syndrome)陽性を事前に予測するニューラルネットワークを用いた新しいアルゴリズムを設計した。今回、自治医科大学附属さいたま医療センターではパネルデータを活用したユースケースの検討として、当院のデータベースを活用し、ピーチインテリヘルス社のAIの予測精度を検証することとした。

具体的には、令和3年度においては、令和初年度で確立した院内における自動SOFA算出システム及び令和2年度で整備したシステムを使用し、集中治療室重症部門システムから集中治療室に入室した患者データ(パネルデータ)を抜き出した。そのデータを匿名化された形で海外企業(ピーチインテリヘルス社)が開発したAIアルゴリズムに投入し、当院の患者におけるSOFAスコア、敗血症、血管作動薬の必要性、及び全身性炎症反応症候群(SIRS)陽性の予測精度を検証した。

## B. 研究方法

2017年7月31日(当院集中治療室の重症部門システム改定後)から2021年2月1日に至る期間にICU/CCUまたはEICUへ入室した7283人の患者情報(入退室日時、ベッド番号、入室形式、入室経路、入室理由、経過記録(特殊治療実測、呼吸器実測、観察、処置、デバイス挿入の有無、看護ケア、検査

実施の有無、イベントの有無)、水分出納、血液検査結果、注射・内服情報、生体情報モニタ情報、バイタルサイン測定値等)をCSV形式で抜き出して、匿名化された状態でピーチインテリヘルス社において保有するAIアルゴリズムACEシステム/アプリに投入した。AIアルゴリズムが予測する重症度は、24時間後のSOFAスコア、8時間後の敗血症発症(現抗菌薬使用に加えSOFAスコアが2点以上増悪した場合に敗血症と定義した)、12~24時間後の血管作動薬の必要性、18~24時間後のSIRS発症とであり、24時間後SOFAスコアは、実際のSOFAスコアとの平均絶対誤差(MAE)で、敗血症発症、血管作動薬必要性、SIRS(systemic inflammatory response syndrome)陽性の予測精度については、ROC(receiver operating characteristics)曲線下面積で評価した。なお、ピーチインテリヘルス社の開発した英語環境のACEシステムで使用可能なように、当院のデータは、日本語の項目を英語に変更する等、一部加工した。

### (倫理面への配慮)

本研究は自治医科大学附属さいたま医療センター倫理委員会で審査され認可されている(臨S21-132)。倫理委員会の決定に従い、以下のように倫理面において配慮した。本研究は、既存診療情報を使用した、後ろ向き観察研究のため、研究の参加については、オプトアウト(拒否できる権利を保障)で同意を得たこととした。なおオプトアウト文書については、自治医科大学附属さいたま医療センターICU/CCU、EICU入口に掲示、及び自治医科大学附属さいたま医療センター麻酔科集中治療部HPに掲載した。

## C. 研究結果

7,283人の患者のうち、敗血症とSIRSの予測は7,159人(98%)、血管作動薬必要性の予測は7,119人(98%)、予測SOFAスコアは6,258人(86%)について実施した。(特定の予測に関連するデータが欠落している患者、または滞在期間が短すぎる患者(例えばSOFAの場合、24時間未満)は、予測と対応するイベントとの比較を可能にするために除外された。)

表1に患者背景を示す。

**Table 1. Patient Demographics.**

	Median Age	No. of Patients	% of Total
<b>Adults</b>	75	7,283	100%
<b>SOFA</b>			
Adults	75	6,258	86%
Male	74	4,058	56%
Female	77	2,200	30%
<b>Sepsis, SIRS</b>			
Adults	75	7,159	98%
Male	74	4,624	63%
Female	77	2,535	35%
<b>Vasopressor Need</b>			
Adults	75	7,119	98%
Male	74	4,596	63%
Female	77	2,523	35%

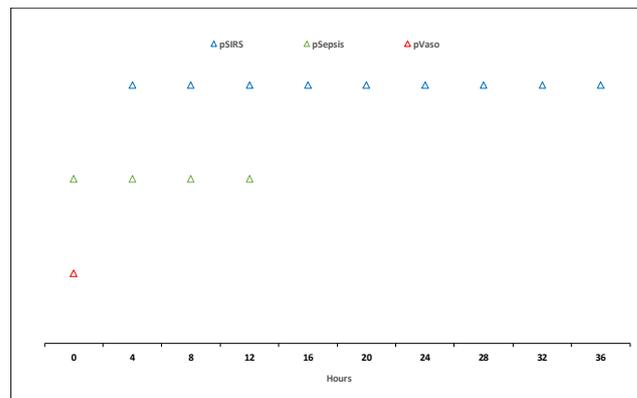
表 2 は、患者が 8 時間以内に敗血症を発症するか、12~24 時間以内に血管作動薬を必要とするか、18~24 時間以内に SIRS が陽性となるか、18~24 時間後の SOFA スコアを予測する 4 つのアルゴリズムの予測精度を示した。予測頻度は、敗血症、血管拡張剤、SIRS については 30 分、SOFA については 2 時間または 12 時間で学習させた。これらの予測モデルは、敗血症発症、血管作動薬必要性、SIRS 陽性について、それぞれ 1.0, 0.98, 0.90 という高い ROC 曲線下面積を達成することができた。また予測 SOFA スコアも 2 時間予測頻度で MAE=0.74、12 時間予測頻度で MAE=0.99 と高い一緻度であった。

**Table 2. Algorithm Performance**

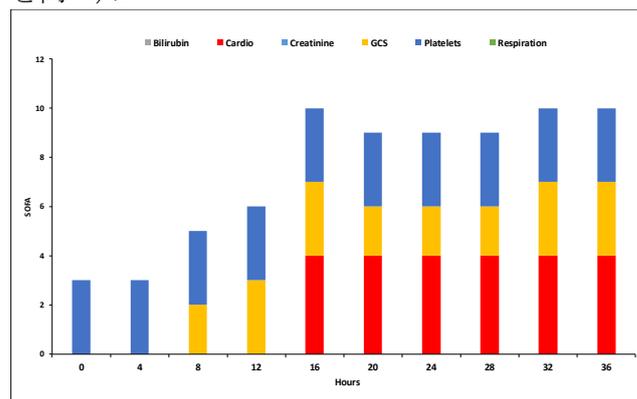
	Proprietary Algorithms				
	Sepsis	Vasopressor	SIRS	SOFA	SOFA
Intervention Window (hours)	8	12 to 24	18 to 24	18 to 24	18 to 24
Prediction Frequency	30 minutes			2 hours	12 hours
Test Samples	214,529	282,981	214,529	52,862	8,800
Performance					
Mean Absolute Error	Not applicable			0.74	0.99
True Positive	5,449	5,697	113,064	Not applicable	
True Negative	208,686	276,107	79,400		
False Positive	394	970	9,830		
False Negative	-	207	12,235		
Specificity	99.8%	99.6%	89.0%		
Sensitivity	100.0%	96.5%	90.2%		
Positive Predictive Value	93.3%	85.5%	92.0%		
Negative Predictive Value	100.0%	99.9%	86.6%		
Accuracy	99.8%	99.6%	89.7%		
AUROC	1.00	0.98	0.90		

図 1 に予測のタイミング (図 1a) と実際の時系列に沿った SOFA スコアの変動 (図 1b) の例を示す。

< 図 1 a 実際のタイムラインに沿った SIRS (pSIRS)、敗血症 (pSepsis)、血管作動薬必要性 (pVaso) の予測 (p は Prediction の略) >



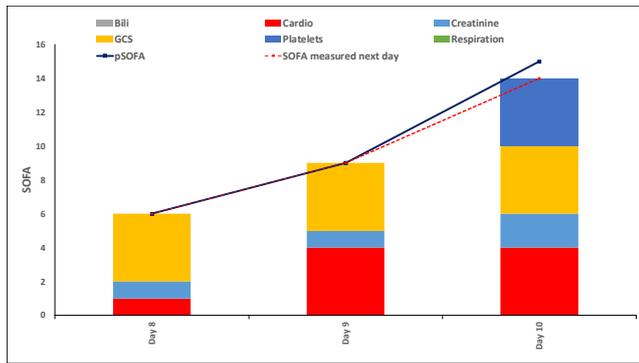
< 図 1 b 実際の SOFA スコアの変化 (時間軸は図 1a と同一) >



この図 (1a 及び 1b) は ICU に入院した 79 歳男性の敗血症予測 (pSepsis)、血管作動薬必要性予測 (pVaso)、SIRS 予測 (pSIRS) を示したものである。集中治療室に滞在期間した 42 日のうちの 36 時間を抽出し 4 時間ごとの予測能を示した。図 1a と図 1b の X 軸は予測時刻に対応し、t=0 は患者の入院 9 日目の午前 10 時 38 分に相当する。図 1b は、個々のサブスコアで計算された実際の SOFA を示す。例えば、pSepsis は t=0、4、8、12 時間のタイムポイントで陽性となったが、これは 8 時間後の実際の ΔSOFA スコアの 2 点以上の増加、具体的には t=0 での 3 から t=8 での 5、t=4 での 3 から t=12 での 6、t=8 での 5 から t=16 での 10、t=12 での 6 から t=20 での 9 に相当する。さらに、SOFA スコアが t=16 から 36 時間まで ±1 以内で安定している場合、pSepsis 予測は陰性となっている。

また図 2 は、予測された SOFA スコアと実際の SOFA スコアの一致を示したものである。

< 図 2: 予測 SOFA スコアと実際の SOFA スコアの一緻率 >



この図は、COVID-19 から急性呼吸不全で ICU に入院した 73 歳男性の予測 SOFA (pSOFA) と実際の SOFA を示したものである。ICU 滞在 29 日間のうち 8-10 日目が示されている。X 軸は予測した時間であり、比較のために 18~24 時間後に測定された実際の SOFA スコアを重ね合わせているが高い一致率をみた。

#### D. 考察

本研究においては、後ろ向き研究において、海外のピーチインテリヘルス社の AI アルゴリズムを使用し、本邦の 1 施設における集中治療患者の重症度予測能を検証した。予測 SOFA スコア、敗血症予測、血管作動薬必要性予測、SIRS 陽性予測はいずれも高い予測能を示した。特に敗血症予測については、事前予測ウィンドウが 4 時間より長くなれば予測精度が落ちることが文献上示唆されているが [4]、8 時間前においても高い精度で予測できたことは特筆に値する。また血管作動薬必要性の予測については、研究が少ないため [5]、この予測アルゴリズムについても臨床上有用であると考えられる。

機械学習 (特にニューラルネットワーク) を用いて敗血症発症を予測した研究は、そのほとんどが MIMIC や eICU データベースなどの大規模なオープンデータベースに依存している [6, 7]。これらのデータベースから生成された高精度の特定の予測モデルは、病院ごとに異なる患者集団や診療パターンを考慮していないため、個別の病院での敗血症の予測精度を保証するものではない。本来は個々の病院に最適化するためには、個々の病院のデータセットを活用すべきだが、しかしながら、個別の病院データは比較的小さなデータセットとなることが多く、機械学習の妨げになる。ピーチインテリヘルス社の AI アルゴリズムは、北米やシンガポールの他国の病院データで事前に学習されており、個々の病院の比較的小さいデータセットを使用した場合の機械学習の精度向上に寄与しており、現に当院 (単施設) における患者の重症度予測精度は非常に高かった。

本研究にはいくつかの Limitation がある。まず、Glasgow Coma Scale (GCS) は SOFA スコアの中樞神経の構成要素であるが、我々のデータベースでは、特に人工呼吸中の患者において鎮静薬の影響を強

く受けている (本来の SOFA スコアの判定法では、鎮静薬の影響がない場合の GCS を予測し評価することが求められるが、今回の研究においては、鎮静され、意識状態が傾眠の場合でも、実際の GCS が反映されている。)。しかしながら、患者に鎮静剤を投与している場合、「真の」意識レベルを判断することは困難であり、本研究では、曖昧さを回避するために鎮静薬の影響を受けた GCS を使用することを許容した。第二に、抗生物質の投与及び SOFA スコアが 2 点以上悪化した場合を敗血症発症と定義したが、このデータセットには術後に手術部位感染予防のためにセファゾリンを継続投与された患者も含まれている。セファゾリンを投与された術後患者においては、SOFA スコアの悪化は敗血症が原因ではない可能性がある。また、今回の敗血症発症予測モデルは、純粋な敗血症の発症だけでなく、敗血症の悪化、例えば、敗血症性ショックへの移行も含まれている。しかしながら実臨床ではどちらも重要であり、敗血症ショックについては、敗血症モデルに加えて SOFA スコア予測や血管作動薬必要性予測も参考になるかもしれない。

世界中でヘルスデータのデジタル化が進む中、このように機械学習を用いて、個々の病院にチューニングした重症度予測モデルの導入が、より一般化していくことが予想される。しかしながら、このようなパネルデータを使用した重症化予測モデルがさらに発展していくためには、この重症化予測モデルが医療従事者の行動変容を促し、患者の予後を本当に改善させるかについての検証が必須である。現時点において、重症化予測モデルが患者予後を改善させたという大規模な前向き研究による結果は得られていない。今後もさらなる研究が必要であろう。

当院においても、令和 3 年度においては、実際に前向き研究 (リアルタイムに SOFA スコアの予測値をピーチインテリヘルス社から得て、実際に介入を行うにあたり最適なスコア変動値を探索する研究 (SOFA スコア 2 点以上増悪が予想される場合に早期に介入した方がよいのか、もしくは 3 点以上の増悪が予想される場合に介入した方がより効果的な予後改善効果が得ることができるのか等、医療従事者の行動変容を促す最適閾値を探索する前向き研究)) を計画していたが、研究を行うにあたっては、当院の電算部との交渉に難渋した。本前向き研究においては、院内の重症部門システムのデータをピーチインテリヘルス社に匿名化された形でインターネットを介して送信、得られたスコアを受信することを考えていたが、所謂「3 省 3 ガイドライン」に準じた情報セキュリティを確保したとしても、やはり院内の電子カルテシステムを外部と接続することについては情報セキュリティ上

の懸念から、院内の了承を得られなかった。このような情報セキュリティに関する問題は、どこの病院においても起こりうる問題であり[8]、外部ソースのAIを使用する場合は、特に重要な課題と考える。しかしながら前向き研究に向け、令和3年度は、AIの解析に必要な重症部門システムのデータを匿名化した状態で簡単に取り出せるプログラムを実装し、一度USB等の外部メディアで抜き出した後、ピーチインテリヘルス社に送信可能とするシステムを院内に構築した。容易にデータを引き出せることは有用であるが、上記には外部メディアで抜き出すというひと手間が必要であり、今後も、院内の情報を外部ソースの人AIにスムーズに投入するためにどのような調整が必要か検討していく。

#### E. 結論

集中治療重症度ツール開発研究分担施設である自治医大附属さいたま医療センターにおいて、ピーチインテリヘルス社の有するAIアルゴリズムを用いたSOFAスコアの変化、現行の敗血症診断基準による敗血症発症、血管作動薬必要性、SIRS陽性の予測は優れた精度を示した。この結果は、病院のデータベースに最適化された効果的な予測モデル生成の可能性を示すものであると考えられる。今後も集中治療領域における更なる人工知能の発展及び、人工知能が医療従事者の行動変容及び患者予後改善効果を有することの検証についての更なる研究が必要である。

#### F. 健康危険情報

なし

#### G. 研究発表

##### 1. 論文発表

現在作成中

##### 2. 学会発表

なし

#### H. 知的財産権の出願・登録状況

なし

##### 1. 特許取得

なし

##### 2. 実用新案登録

なし

##### 3. その他

なし

#### I. 参考文献

- 1 Lauritsen SM, Kalør ME, Kongsgaard EL, Lauritsen KM, Jørgensen MJ, Lange J, Thieson B. Early detection of sepsis utilizing deep learning on electronic health record event sequences. *Artif Intell Med.* 2020 Apr;104:101820. doi: 10.1016/j.artmed.2020.101820.
- 2 Rafiei A, Rezaee A, Hajati F, Gheisari S, Golzan M. SSP: Early prediction of sepsis using fully connected LSTM-CNN model. *Comput Biol Med.* 2021 Jan;128:104110. doi: 10.1016/j.combiomed.2020.104110.
- 3 Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; doi:10.1001/jama.2016.0287.
- 4 Kuo YY, Huang S-T, Chiu H-W. Applying artificial neural network for early detection of sepsis with intentionally preserved highly missing real-world data for simulating clinical situation. *BMC Med Inform Decision Making,* 2021; doi:10.1186/s12911-021-01653-0.
- 5 Rahman A, Chang Y, Dong J, Conroy B, Natarajan A, Kinoshita T, Vicario F, Frassica J, Xu-Wilson M. Early prediction of hemodynamic interventions in the intensive care unit using machine learning. *Crit Care.* 2021 Nov 14;25(1):388. doi: 10.1186/s13054-021-03808-x
- 6 Lucas M, Fleuren, Thomas L. T. Klausch, Charlotte L. Zwager, Linda J. Schoonmade, Tingjie Guo, Luca F. Roggeveen, Eleonora L. Swart, Armand R. J. Girbes, Patrick Thorald, Ari Ercole, Mark Hoogendoorn, Paul W. G. Elbers. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med.* 2020; 46(3): 383-400. Published online 2020 Jan 21. doi: 10.1007/s00134-019-05872-y
- 7 Yan MY, Gustad LT, Nytrø Ø. Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review. *J Am Med Inform Assoc.* 2022 Jan 29;29(3):559-575. doi: 10.1093/jamia/ocab236.

8 高木俊介ら 我が国における遠隔集中治療  
(Tele-ICU) の導入における技術的・社会的課題の  
解決に向けた研究 厚生労働科学研究費補助金  
行政政策研究分野 政策科学総合研究(臨床研究  
等 ICT 基盤構築・人工知能実装研究) 平成 30 年  
度