

分担研究報告書

5. 重症度アルゴリズムの文献的レビュー

研究分担者 大下 慎一郎 広島大学 准教授

研究要旨

**<背景>**人工知能は、人間の認知機能や知的行動を模倣することを目的とした幅広い技術を意味する。その中の機械学習は人工知能のサブフィールドで、コンピュータがプログラムされることなく、経験的データから複雑な関係やパターンを定義できるようにするアルゴリズムのことである。医療領域のデータ量増加に伴い、医療領域における機械学習の重要性が高まってきている。しかし、機械学習の学習手法には様々な種類があり、いずれの学習モデルにもそれぞれ特性があるため、データの種類によって使い分ける必要がある。現時点では、医療領域において、どの機械学習アルゴリズムが適しており、どの程度の臨床学的有用性があるのかが十分には明らかになっていない。

**<研究目的>**本研究の目的は、医療領域（とくに集中治療領域）において、機械学習によって構築された重症度アルゴリズムの有用性を、システマティック・レビュー（網羅的文献調査）の手法で評価することである。

**<研究方法>**3つの医学文献検索エンジン（MEDLINE, Cochrane Central, 医学中央雑誌）を用いて、システマティック・レビューを行った。対象は集中治療領域・救急領域の急性呼吸不全・敗血症・ショック・重症患者とし、介入は人工知能・機械学習・深層学習とした。アウトカムは死亡・重症化・急性腎障害発症とした。6名の調査者が各2名ずつペアになって文献検索を行い、不一致だった文献については議論の上、採否の最終決定を行った。最終的に残った文献内容から、人工知能・機械学習の有用性に関する情報を抽出した。

**<結果>**目的に合致する文献は4,610件が該当した。重複文献、研究デザインの異なる文献、対象の異なる文献等を除外していき、最終的に147文献が選択された。研究数は2017年以降急速に増加し、2021年には1年で58件（40%）の研究が発表された。これまでの研究発表国はアメリカが最多（47件、32%）で、中国（24件、16%）、イギリス（9件、6%）がこれに次いだ。日本は1件（1%）のみだった。診断群分類のような公的ビッグデータを使用した研究は67件（46%）認められた。対象患者数は中央値10,972名（IQR 2,510-33,819）、同定された特徴量は20（IQR 13-55）と研究によって大きな開きがあった。開発ソフトウェアはPython（25件、17%）とR（18件、12%）が多かったが、明記されていない研究も多かった。学習アルゴリズムは、人工ニューラルネットワークが最多で（50件、34%）、ランダムフォレスト（33件、22%）、ロジスティック回帰（26件、18%）、XGBoost（25件、17%）、サポートベクターマシン（23件、16%）、決定木（18件、12%）などがこれに次いだ。研究デザインは、他の重症度指標との比較研究（第3相研究）が最多で（86件、59%）、有用性評価（第2相研究）がこれに次いだ（58件、39%）。機械学習による重症度アルゴリズムが有効であると結論された研究は143件（97%）であった。

**<結論>**機械学習による重症度アルゴリズム構築は、集中治療領域における患者の死亡・重症化予測に有用である可能性が高い。

## A. 研究目的

本研究は、計画書の目的「③ICU 患者に対する AI を用いた重症化予測モデルの構築」に相当する。

人工知能 (Artificial Intelligence: AI) は、人間の認知機能や知的行動を模倣することを目的とした幅広い技術を意味する。その中の機械学習 (Machine Learning: ML) は AI のサブフィールドで、コンピュータがプログラムされることなく、経験的データから複雑な関係やパターンを定義できるようにするアルゴリズムのことである。医療領域のデータ量増加に伴い、医療領域における機械学習の重要性が高まってきている。しかし、機械学習の学習手法には様々な種類があり、いずれの学習モデルにもそれぞれ特性があるため、データの種類によって使い分ける必要がある。機械学習モデルには、大きく分けて教師あり学習・教師なし学習があり、前者では予めラベリングされたデータを用いてアルゴリズムを学習する。教師あり学習には、ランダムフォレスト (Random forest: RF), サポートベクターマシン (Support Vector Machine: SVM), 決定木 (Decision Tree: DT) などの学習モデルが含まれる。一方、教師なし学習ではラベリングされたデータは必要なく、ラベリングされていないデータに内在する関連性を自己解析して学習する仕組みである。いずれの機械学習モデルにおいても、予測誤差率を最適化するために、最適なバイアスと分散のトレードオフを決定する必要がある。つまり、バイアスと分散の選択が不適切だと、アンダーフィッティング、またはオーバーフィッティングという問題が生じうる。このため、両者のバランスが最適になるスイートスポットを見つけることが重要である。

上記のような背景から、現時点では、医療領域において、どの機械学習アルゴリズムが適しており、どの程度の臨床学的有用性があるのかが十分には明らかになっていない。

本研究の目的は、医療領域 (とくに集中治療領域) において、機械学習によって構築された重症度アルゴリズムの有用性を、システマティック・レビュー (網羅的文献調査) の手法で評価することである。

本研究によって、ICU パネルデータを活用した患者重症度の AI 解析を遠隔 ICU に活用するユースケースにおいて、どのような手法を用いれば診断精度を向上させることができるかを調査する。令和 3 年度は過年度の成果を踏まえ、本調査により、ICU・集中治療領域の患者重症度 AI 解析において新規性のある課題整理、有用とされるパネルデータの抽出を行い、AI アルゴリズム構築の精度向上を実現する。

## B. 研究方法

3 つの医学文献検索エンジン (MEDLINE,

Cochrane Central, 医学中央雑誌) を用いて、表 1 の PICOST に従ってシステマティック・レビューを行った。対象 (Patients: P) は集中治療領域・救急領域の急性呼吸不全・敗血症・ショック・重症患者とし、介入 (Intervention: I) は人工知能・機械学習・深層学習とした。アウトカム (Outcome: O) は死亡・重症化・急性腎障害発症とした。研究デザイン (Study Designs: S) は、原著論文 (前向き・後ろ向き研究を含む) とし、症例報告, 学会抄録, 総説, 日本語・英語以外の言語, 動物実験は対象外とした。検索期間 (Timeframe: T) は、2021 年 12 月までに発表されたすべての研究とした。

### <表 1. システマティック・レビューの PICOST>

**臨床的疑問:** ICUにおいて、重症度評価のためにAIがどのような活用をされているのか?  
**目的と方法:** 過去の研究をシステマティック・レビューして、以下を調査する。

対象 (P)	集中治療領域・救急領域の急性呼吸不全・敗血症・ショック・重症患者
介入 (I)	Artificial intelligence (AI), machine learning
アウトカム (O)	死亡, 重症化 (AKI発症)
対象論文	原著論文 (前向き・後ろ向き研究を含む)
除外論文	症例報告, 学会抄録, 総説, 日本語・英語以外の言語, 動物実験

MEDLINE における検索式は以下の内容とし、同様の検索式を Cochrane Central, 医学中央雑誌でも作成した。

### <表 2. 検索式>

	検索式	文献数
#1	"artificial intelligence"[tiab] OR "Artificial Intelligence"[Mesh] OR "machine learning"[tiab] OR "deep learning"[tiab] OR "neural network*"[tiab] OR "Neural Networks, Computer"[Mesh] OR "learner algorithm*"[tiab] OR "reinforcement learning approach*"[tiab] OR "intelligent ICU"[tiab] OR "autonomous monitoring*"[tiab] OR "ensemble learning*"[tiab] OR "computerized algorithm*"[tiab] OR "electronic monitoring*"[tiab] OR "Decision Support Systems, Clinical"[Mesh] OR "continuous risk prediction"[tiab]	214,514

#2	"intensive care*"[tiab] OR "Intensive Care Units"[Mesh] OR "critical care"[tiab] OR "Critical Care"[Mesh] OR "Critical Illness"[Mesh] OR critical[tiab] OR critically[tiab] OR emergency[tiab] OR "Emergency Service, Hospital"[MeSH] OR "Emergency Medical Services"[MeSH] OR "Emergencies"[Mesh] OR "critically ill"[tiab] OR "acutely ill"[tiab] OR "perioperative medicine*"[tiab] OR "perioperative medicine"[Mesh] OR "Perioperative Care"[Mesh] OR postoperative[tiab] OR life-threatening[tiab]	2,178,080
#3	"Severity of Illness Index"[Mesh] OR "severity scoring*"[tiab] OR mortality[tiab] OR "Mortality"[MeSH] OR "mortality risk prediction*"[tiab] OR outcome*[tiab] OR "Critical Care Outcomes"[Mesh] OR onset[tiab] OR incidence[tiab] OR deterioration[tiab]	4,049,971
#4	#1 AND #2 AND #3 AND (english[la] OR japanese[la]) NOT (animals[Mesh] NOT humans[Mesh]) NOT review[pt] NOT case report[pt]	4,543

6名の調査者が各2名ずつペアになって文献検索を行った。一次スクリーニングでは、論文タイトル・抄録の内容をもとにスクリーニングを行い、二次スクリーニングでは全文に目を通して内容を確認して採否判断をした。2名の意見が不一致だった文献については議論の上、採否の最終決定を行った。最終的に残った文献内容から、人工知能・機械学習の有用性に関する情報を抽出し、質的統合を行った。

(倫理面への配慮)

本研究は患者個人を対象とした研究ではなく、患者個人の情報も扱わないため、倫理上の考慮事項には該当しない研究であった。

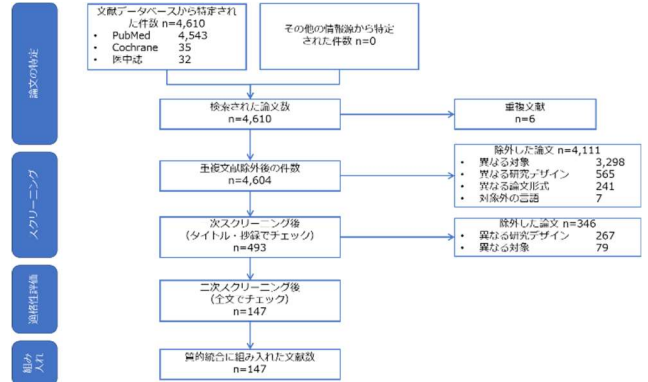
C. 研究結果

1) 文献検索

文献検索の経過を図1に示す。文献検索の結果、MEDLINE 4,543件、Cochrane Central 35件、医学中央雑誌 32件の計4,610件が抽出された。重複文献6件を除外した後、一次スクリーニングを行った。一次スクリーニングで残った493件に対し

て二次スクリーニングを行ったところ、最終的に147件の研究が残った。

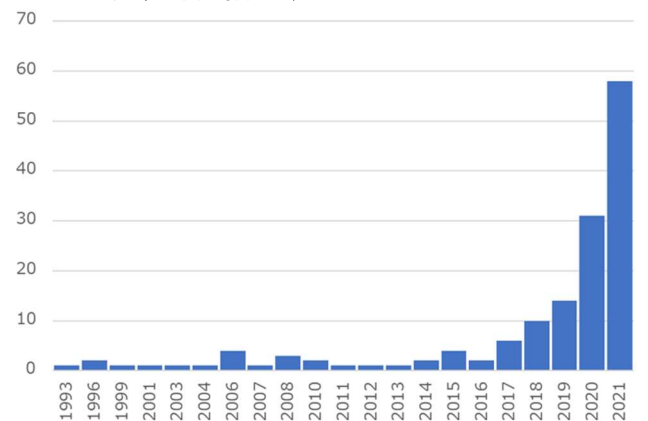
<図 1. 文献検索のフローダイアグラム>



2) 発表年別論文数

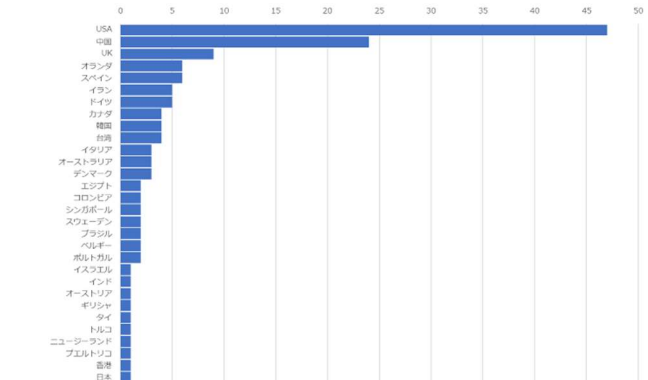
発表年別論文数を図2に示す。2016年までは、危害学習による重症度アルゴリズムに関する研究はほとんど認められなかったが、2017年以降急速に増加し、2021年には1年で58件(40%)の研究が発表されるに至った。

<図 2. 発表年別論文数>



3) 発表国別論文数

<図 3. 発表国別論文数>



発表国別論文数を図3に示す。これまでに研究発表したことのある国数は計30ヶ国だった。発表数はアメリカが最多(47件, 32%)で、中国(24

件, 16%), イギリス (9 件, 6%) がこれに次いだ。日本は 1 件 (1%) のみだった。

#### 4) 対象患者とアウトカム

システムティック・レビューで対象として残った 147 研究における対象患者を表 3 に示す。すべて ICU 患者を対象としていたが、そのうち、16 件 (11%) は敗血症を合併した ICU 患者、7 件 (5%) は新型コロナウイルス肺炎 (COVID-19) を合併した ICU 患者を対象としていた。一方、小児の ICU 患者を対象とした研究も 13 件 (9%) 含まれていた。日本の診断群分類 (Diagnosis Procedure Combination: DPC) のような公的ビッグデータ (Medical Information Mart for Intensive Care III, MIMIC-III; Electronic Intensive Care Unit, eICU, UFHealth 等) を使用した研究は 67 件 (46%) 認められた。

<表 3. 対象患者の背景>

	n	(%)
ICU患者全般	147	(100)
ICU患者+敗血症	16	(11)
ICU患者+COVID-19	7	(5)
ICU患者+急性呼吸不全	5	(3)
ICU患者+その他の疾患	8	(5)
ICU患者+小児	13	(9)
MIMIC-III, eICU, UFHealth, EHR等	67	(46)

対象の 147 研究において検討されたアウトカムの一覧を表 4 に示す。大部分は死亡予測 (111 件, 76%) に対する機械学習アルゴリズムを検討していたが、急性腎障害 (Acute kidney injury: AKI), 敗血症の発症予測や、入院日数・ICU 滞在日数, 重症度・重症化の評価や、再入院の予測における危害学習アルゴリズムを検討している研究も認められた。

<表 4. 検討されたアウトカム>

	n	(%)
死亡	111	(76)
AKI	19	(13)
敗血症	12	(8)
入院/ICU日数	6	(4)
重症度/重症化	5	(3)
再入院	3	(2)
その他	7	(5)

研究の対象患者数は、中央値 10,972 名だったが、四分位範囲 (interquartile range: IQR) は 2,510-33,819 名と極めて広がった (表 5)。これは主

に、施設規模の患者を対象とした研究 (数百~数千名が対象) と、公的ビッグデータを使用した研究 (数万名が対象) が混在していたことによるものであった。同定された特徴量 (アウトカム予測変数として組み込むべき臨床データ) の数は明記されていない研究が大部分だったが、明記されていた研究のみの集計では、中央値で 20 項目だった。

<表 5. 対象サンプル数と特徴量数>

	median	(IQR)
サンプル数	10,972	(2510-33,819)
特徴量数	20	(13-55)

#### 5) 解析に使用されたソフトウェア

機械学習に使用されたソフトウェア名、パッケージ・ライブラリの名称一覧を表 6 に示す。ソフトウェアは Python (25 件, 17%) と R (18 件, 12%) が大部分を占めたが、ゲーム理論のシャープレイ値 (Shapley Value) を機械学習に応用したオープンソース・ライブラリである SHAP を使用している研究も多数認められた。

Python を使用する際は、人工ニューラルネットワーク構築に必要なライブラリである Keras, TensorFlow が高頻度に使用されていた。R では、具体的なライブラリ名が明記されていない研究が多かった。

<表 6. 機械学習に使用されたソフトウェア>

ソフトウェア名 (パッケージ・ライブラリ)	n	(%)
Python (Jupyter)	25	(17)
Keras library	10	(7)
TensorFlow library	9	(6)
Scikit-learn	6	(4)
Adam optimizer	4	(3)
matplotlib	1	(1)
PyCharm	1	(1)
Theano library	1	(1)
R	18	(12)
caret package	1	(1)
mfp library	1	(1)
rms library	1	(1)
deal library	1	(1)
mastino library	1	(1)
SHapley Additive exPlanations (SHAP)	10	(7)
SAS	1	(1)
SPSS	1	(1)

#### 6) 機械学習に使用されたアルゴリズム

機械学習に使用されたアルゴリズム名の一覧を表 7 に示す。人工ニューラルネットワークが最多で (50 件, 34%), ランダムフォレスト (33 件, 22%), ロジスティック回帰 (26 件, 18%), XGBoost (25 件, 17%), サポートベクターマシン (23 件, 16%), 決定木 (18 件, 12%) などがこれに次いだ。

一般公開されているアルゴリズムを使用して機械学習を行っている研究が多かったが、オリジナルのアルゴリズムを開発している研究も認められた (26 件, 18%).

＜表 7. 機械学習のアルゴリズム＞

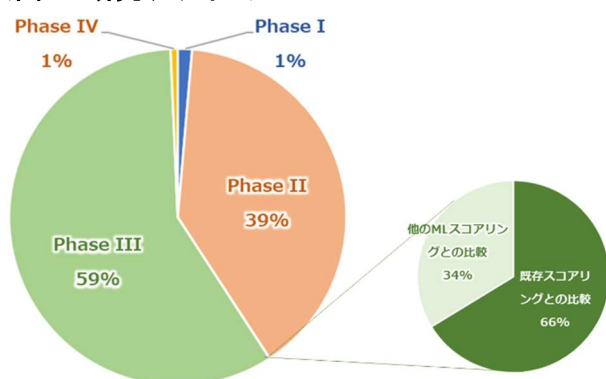
アルゴリズム名	n	(%)
Artificial Neural Network (ANN)	50	(34)
Convolutional neural network (CNN)	19	(13)
Recurrent neural network (RNN)	12	(8)
Unknown	20	(14)
Random Forest (RF)	33	(22)
Logistic Regression (LR)	26	(18)
eXtreme Gradient Boosting (XGBoost)	25	(17)
Support Vector Machines (SVM)	23	(16)
Gadient-boosted decision trees (GBDT)/ Classification trees (CART)	18	(12)
Long-Short term memory (LSTM)	14	(10)
K-Nearest Neighbors (KNN)	10	(7)
Stochastic Gradient Boosting	2	(1)
Light Gradient Boosting Decision Machine (LightGBM)	2	(1)
Random Under-Sampling (RUS)	2	(1)
Latent Semantic Analysis (LSA)	2	(1)
Linear Discriminant Analysis (LDA)	2	(1)
Andadaptive boosting (AdaBoost)	2	(1)
Other	26	(18)
Unknown	9	(6)

### 7) 研究デザイン

研究デザインを、第 1 相研究 (機械学習アルゴリズム構築のための方法論に関する研究)、第 2 相研究 (有用性評価)、第 3 相研究 (他の重症度指標との比較研究)、第 4 相研究 (臨床応用後の治療成績変化に関する研究) と分類した結果を図 4 に示す。その結果を図 4 に示す。第 3 相研究が大部分 (86 件, 59%) を占め、第 2 相研究がこれに次いだ (58 件, 39%)。第 1 相研究、第 4 相研究はごくわずかだった。

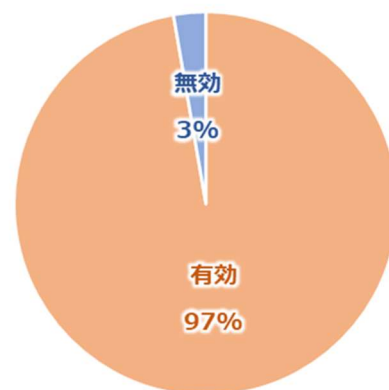
さらに、第 3 相研究は、既存スコアリングとの比較が行われた研究が 57 件 (69%)、他の ML スコアリングとの比較が行われた研究が 29 件 (34%) だった。

＜図 4. 研究デザイン＞



### 8) 機械学習アルゴリズムの有効性

＜図 5. 機械学習アルゴリズムの有効性＞



対象とした 147 研究のうち、機械学習アルゴリズムが有効であると結論していた研究は 143 件 (97%) であった。逆に、機械学習アルゴリズムはまだ十分な有効性がないと結論していた研究は 4 件 (3%) のみであった (図 5)。

### D. 考察

本研究で我々は、機械学習によって構築された重症度アルゴリズムが、ICU 領域の診療において有用であることを、網羅的文献調査の手法を用いて示した。対象とした 147 文献のうち、143 件 (97%) の研究において、機械学習による重症度アルゴリズムが有効であると結論された。患者のアウトカムとしては、死亡を対象としていた研究が最も多く (76%)、急性腎障害、敗血症、入院日数/ICU 滞在日数、重症度/重症化、再入院がこれに次いだ。89 件 (59%) の研究では、既存の臨床スコアリング法または他の機械学習アルゴリズムと比較した優越性を検討していた。

ICU 領域における ML 技術の応用は、近年のコンピューティング技術の向上により急速に拡大している。ML の応用により、複雑で多様な膨大データを、短時間で解析できるようになった [1]。ICU で人工呼吸中の患者の診療をするのに、集中治療医が判断する必要がある変数カテゴリーの数は、236 項目にも上るという報告がある [2]。敗血症による急性呼吸窮迫症候群を例に挙げれば、人工呼吸・動脈血ガス分析・胸部 X 線・循環系・腎機能・薬物・点滴・鎮静・栄養・血栓予防などに含まれる複数の変数カテゴリーを検討した上で、治療方針を決定する必要がある。この中には、真に有用なデータもあれば、そうでないデータも含まれている。臨床現場において意思決定に必要な変数を見つけた上で、膨大な数の変数を処理することを考えると、いかに優れた臨床医だったとしても、その脳に占めるワーキングメモリの容量は大差がないのが現状である。このため、誰でも誤りを犯し、一貫性のない行動をとる可能性がある。この観点から、機械学習による重症度評価アルゴリズムを用いて、電子カルテに保存されたデータを自動分析することは、極めて有望な解決手法と考えられる。

しかし、臨床的に有用な機械学習モデルを構築するためには、データ前処理、特徴量同定、モデル検証の影響を強く上げる。とくにデータ欠損の補完法は、器楽学習が遭遇する最も大きな課題である[3]。一般的には、データをインプットするか削除するかのどちらかを選択する。インプットする方法には、平均値を使用する方法[4]、ガウス関数を使用する方法[5]などがある。しかし、データのインプットは、データ分布を歪曲させ、モデルの精度に影響を与える危険性を含んでいる。どのインプットを使用すべきかは、データの重要性、欠損データの割合、使用する ML モデルによって使い分ける必要がある。

ML モデルは、教師あり学習と教師なし学習に大別される。教師あり学習は、特定の入力(x)が与えられたときに、それに対応する出力(y)を提示する学習法である。教師あり学習では、x と y の大規模なデータセットが前提であるため、新たな未知データに遭遇した際、正確な予測を行うことが期待される。一方、アルゴリズム確立と検証のために、莫大なデータセットと多大な労力を必要とする。教師なし学習では、ルールが不明な大量の未分類データから、コンピュータが規則性を検出することが求められる。大規模データセットの探索的な目的では有用である反面、得られた結果の有用性・正確性を評価するのが困難である側面もある。我々の調査結果では、いずれの ML モデルも ICU 領域において使用されており、いずれも有用性が示された。ML モデルは、特定のデータセットに対して唯一無二の選択肢ではなく、データ特性によってフィッティングの程度が変化するため[6]、適切に使い分けることが重要であろう。

我々の調査結果では、多くの研究が、既存の臨床スコアリング法または他の機械学習アルゴリズムと比較検討を行ったレトロスペクティブ研究だった。この研究手法によって、患者の早期発見・重症化リスクの層別化が可能となるだろう。しかし、より重要なのは、ICU における重症患者の生体変化を、継続的に評価できるインテリジェント・モニターの開発である。このようなモニターを開発し、プロスペクティブな検証を行うことで、自律型 ICU の実現に近づくことが可能となるだろう。

本調査研究にはいくつかの限界点がある。第一に検索式によって見落とされた研究がある可能性である。しかし我々は、必ず含めるべきキー論文を予め調査した上で検索式を立てることにより、見落とされた研究が発生するリスクを最小限に抑える対応をした。第二に ML モデルの性能比較を定量的に行えなかった点である。ML モデルの性能は、使用されたデータ特性、データセットのサイズに依存するため、性能の定量的解析が困難であった。

## E. 結論

機械学習による重症度アルゴリズム構築は、集中治療領域における患者の死亡・重症化予測に有用である可能性が高い。本調査研究によって、患者重症度の AI/ML 解析アルゴリズムを遠隔 ICU に活用するユースケースにおける有用性が示唆された。現時点では、レトロスペクティブ研究が大多数であるため、教師あり学習アルゴリズムを採用しているものが多かった。今後、遠隔 ICU を構築していく上では、教師あり学習のみでなく教師なし学習アルゴリズムを使用した ML モデルの構築が重要性を増していくと考えられる。

## F. 健康危険情報

該当なし

## G. 研究発表

1. 論文発表：現在作成中
2. 学会発表：該当なし

## H. 知的財産権の出願・登録状況

1. 特許取得：該当なし
2. 実用新案登録：該当なし
3. その他：該当なし

## I. 参考文献

1. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017;2:230-243. doi: 10.1136/svn-2017-000101.
2. Morris AH. Human Cognitive Limitations. Broad, Consistent, Clinical Application of Physiological Principles Will Require Decision Support. *Ann Am Thorac Soc* 2018;15:S53-S56. doi: 10.1513/AnnalsATS.201706-449KV.
3. Tang F, Ishwaran H. Random Forest Missing Data Algorithms. *Stat Anal Data Min* 2017;10:363-377. doi: 10.1002/sam.11348.
4. pandas.DataFrame.fillna. pandas 1.4.2 documentation. URL: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.fillna.html> (2022.5.10 accessed)
5. Zhao Y, Udell M. Missing value imputation for mixed data via Gaussian copula, *KDD* 2020. doi.org/10.48550/arXiv.1910.12845
6. Müller AC, Guido S. Introduction to machine learning with Python: a guide for data scientists. Sebastopol, O'Reilly Media, Inc, 2016. ISBN: 9781449369415