

麻酔とテクノロジー

徹底分析
シリーズテクノロジーの
集中治療領域での応用公開データベースの功罪と
ローカルデータの活用

齋藤 大之・内野 滋彦・方山 真朱

人工知能 (AI) による臨床予測は、集中治療のあり方を大きく変える可能性を秘めている。しかし、公開データベースで開発されたモデルは実臨床との乖離が大きく、そのままでは現場で活用できない。この課題を解決し真に信頼できる AI を創出する鍵が、ローカルデータの活用にある。

本稿ではその意義について、自治医科大学附属さいたま医療センター（以下、当院）での実践例をもとに解説する。

集中治療室 (ICU) では、さまざまなデータが日常的に収集されており、AI、特に機械学習モデルの開発に適した環境にある。また後述するとおり、ICU データベースが複数公開されており、これが機械学習研究の発展を後押ししている。

実際の臨床において、早期警戒スコア early warning scores (EWS) *1 として機械学習をもとに開発された eCARTv5 は、従来のルールベースの EWS と比較し精度が高いことが示されている¹⁾。また画像においても深層学習で開発された CheXNet は、胸部 X 線写真において放射線科医よりも高い精度で肺炎を識別することができた²⁾。

一方で、実臨床において AI を活用するまでのハードルは依然として高い。集中治療領域における AI 研究は多数実施されてきたが、前向き研究はごくわずかであり、実臨床への応用はほとんど存在しない³⁾。その一因として、「汎化性能」が不十分であることが挙げられる。これまで行われてきた ICU での AI 研究のうち、外部検証が行われたのは 15% に満たなかった⁴⁾。こ

れは、多くの研究で報告された性能が楽観的すぎる結果である可能性を示唆している。すなわち使用したデータに過剰に適合しており、データ外の実臨床では十分な性能を発揮できない可能性が高い。この問題をより深刻にしている原因として、多くの研究が公開データベースに依存していることにある。

公開データベースの
利点と課題

現在公開されている代表的な ICU データベースは表 1 のとおりである。いずれも研究目的であれば、利用申請を行うことで誰でも利用することができる。

公開データベースの大きな利点は、豊富な症例数にある。これだけの症例数を自施設で収集するには膨大な労力と時間を要するが、公開データベースであれば無償かつ短時間で手に入れることができる。また、複数の研究が同一のデータを用いることで、研究間の性能を客観的に比較することができる。

一方で課題も多く存在する。まずデータベースの多くが単施設に由来している。単施設のデータで開発された AI は、その施設のデータに過剰適合している可能性が高く、ほかの施設のデータでは性能が保証されない。また

- *1 患者の重症度を迅速に判断するためのツールとして、入院患者の急変リスクを評価するために用いられることが多い。
*2 モデルの識別性能を示す指標で、1 に近いほど精度が高い。

表 1 現在公開されている成人 ICU のデータベース

データベース名	国	施設	患者数	データ収集期間
MIMIC- III	米国	単施設	約 4 万例	2001 ~ 2012 年
MIMIC- IV	米国	単施設	約 5 万例	2008 ~ 2019 年
eICU	米国	多施設	約 20 万例	2014 ~ 2015 年
AmsterdamUMCdb	オランダ	単施設	約 2.3 万例	2003 ~ 2016 年
HIRID	スイス	単施設	約 3.4 万例	2008 ~ 2016 年
SICdb	オーストリア	多施設	約 2.7 万例	2013 ~ 2021 年
Zhejiang EHR Critical Care	中国	単施設	約 8 千例	2012 ~ 2022 年

eICU のような多施設データは情報の記録方法が施設によって大きく異なり、品質が保証されていないという問題もある⁵⁾。さらにデータベースのほとんどが欧米圏からのものであり、これらのデータベースを学習したモデルが日本を含むアジア圏やアフリカ圏で十分に適合できるかは不明である。これまでの集中治療領域における AI 研究の 8 割以上が Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) と eICU のデータを用いており、米国の医療システムを強く反映する結果となっている⁶⁾。

公開データベースのおかげで研究へのアクセスは容易になった一方、データベースに強い偏りがあり実臨床を反映していないために、臨床への応用が進まないという現状がある。

ローカルデータの
利点と課題

先述のとおり、公開データベースには実臨床を反映していないという課題がある。この問題を解決する方法は、ローカルデータを用いて実臨床に即したモデルを作成することである。

ローカルデータを用いることの利点は、まずその施設の実情を反映させることができることにある。人種はもちろん、その施設の患者層や社会的背景を反映することができ、施設に特異的な予測パターンをとらえることができる。これは公開データベースでは実現できないことであり、予測性能が向上し、より臨床で活用しやすいモデルとなる。また ICU での治療には、その施設独自のプロトコルや「癖」が存在する。モデルは、この独自プロトコルや

癖が患者のアウトカムに与える影響を学習し、予測へ反映することができる。

ローカルデータの欠点は、施設のデータに適合しているため、ほかの施設で使用した場合は性能が著しく低下する可能性が高いことである。また独自のモデルを作成し、性能を確保するには症例の集積が必要となる。使用できるデータも電子カルテシステムや重症部門管理システムの設定、モニターとの連携、入力された情報によって異なり、必要な情報が欠損している可能性もあり、モデルの構築前に十分な準備が必要となる。そのためローカルデータで独自モデルを構築することは、理想ではあるものの、症例の集積やデータ収集の整備などハードルは高い。

転移学習の
優位性

そこで、公開データベースという膨大な症例数と、ローカルデータという臨床に即した情報を統合することで、少ないローカルデータでも十分な性能を発揮できる可能性がある⁹⁾。これは転移学習と呼ばれる技術で、あるデータセットで学習させたモデルを、別のデータセットを使って微調整し、効率的に高性能なモデルを構築する手法であ

る。汎用モデルを少ないローカルデータでチューニングすることにより、多量のローカルデータを収集したときと同等に近い性能を実現できる。

MIMIC- III を用いて転移学習を行い、院内死亡を予測する研究では、MIMIC- III のみで受信者動作特性曲線下面積 area under the receiver operating characteristic curve (AUROC) *2 0.78、ローカルデータのみで AUROC 0.8 を超えるには症例が 4000 例必要であったが、MIMIC- III のモデルをもとに転移学習を行ったモデルでは、ローカルデータがたった 220 例で AUROC が 0.8 を超えた⁷⁾。このように公開データベースをもとに転移学習を行う方法であれば、ローカルの症例数が少なくても十分な性能を引き出すことができ、臨床に用いるモデルを構築することが期待される。

当院での
AI の活用

最後に当院でのローカルデータを用いた AI 予測、"BEST-AI (Big data-driven Evaluation of Survival and Treatment in Acute Illness)" について紹介する。

2017 年 8 月 ~ 2024 年 7 月に当院 ICU に入室した 1 万 710 例について、

徹底分析
シリーズ

麻酔とテクノロジー

色ベタ + スミ 20% (以下同)
14a ロダンB (以下同)

▼図1 患者ごとに生成される予測画面の一例 (一部改変)

入室後から30分ごとに自動で生成される。左4つが予測、右4つが測定値の推移である。
「病院死亡/ICU死亡/72時間以内の退室」に表示されている割合(例では10.6%)は、退室が成功した場合のその後の再入室予測確率を示す。
同様に「72時間以内の人工気道」に表示されている割合(例では5.8%)は、予測時点または抜管時点での再挿管の予測確率を示す。

1/a MFG BBB
↓
14H
(以下同)

風中 ネム
・基本 1/a MFG BBB
(10a 以下は、ロダンM)
・太くするネム
1/a 太くするB10
(10a 以下は、ロダンBB)
(以下同)

図版は、0.12%
色ベタ+スミ
(以下同)

(±5%以内可!!)
100%
179%

色80%+スミ80%・文は白ヌキ 14a 新3M

コラム

色ベタ+スミ10% AIの「過剰適合」と「汎化性能」 14a 新3M (以下同)

AIモデルの性能を語るうえで欠かせないのが、「過剰適合」と「汎化性能」である。

具体的な臨床シナリオを考えてみる。
あるICUでは「敗血症性ショックでノルアドレナリンが0.5μg/kg/minを超えるまで、バソプレシンは絶対に追加しない」という、非常に厳格な治療方針をもっていたとする。この施設のデータだけで学習したAIは、「バソプレシン開始」が「極めて重篤で、予後不良の強力なサイン」であると学習する。これは、このICUの治療基準が反映された、その施設のデータ上では統計的に正しい「ローカルルール」である。この状態が、データの癖まで学習してしまった「過剰適合」である。

このAIを「ノルアドレナリンが0.2μg/kg/min程度の早期からバソプレシンを併用する」方針の別の施設に導入すると、予測は破綻する。まだ十分に助かる見込みのある患者に対しても「予後不良」と誤った予測を連発してしまう。これは未知のデータに対して正しく機能しない、すなわち「汎化性能」が低い状態である。AIが学習したのは普遍的な医学的因果関係ではなく、特定の施設の治療価値という「癖」であった。

臨床で真に役立つAIとは、特定の施設のローカルルールに「過剰適合」したAIではなく、施設を越えて通用する「普遍的な医学的原則」を学び、高い「汎化性能」を獲得したAIである。

▼図2 ICU全体の予測ランキングの一例 (一部改変)

30分ごとに自動で生成される。上3項目は入退室関連、中3項目は呼吸関連の予測をランキング形式で上位10患者を表示している。
確率が1%未満の場合は表示されない。下2項目は病院・ICU死亡予測確率が24時間前から一定以上上昇した患者を「悪化」として検知し、その上昇幅(ポイント)とともに表示している。

色ベタ

(±10%以内可!!)
170%
179%

計1085個の特徴量を1時間ごとのデータに集約し、計126万6562時間のデータを用いてモデルを作成した。
特徴量にはバイタルや血液ガス、血液検査はもちろん、人工呼吸器の設定や実測値、投与された薬物、看護師の観

察項目、看護記録も含まれる。予測項目は病院・ICU死亡、72時間以内のICU生存退室、退室後の再入室、気管切開、72時間以内の挿管・抜管、72時間以内の再挿管、24時間以内のそのほかの介入・合併症である。予測は

30分ごとに行い、予測結果は重症部門システム(ACSYS)から参照できる(図1)。また退室など一部の項目はICU全体でランキング形式でも参照できる(図2)。予測結果の画面では合わせて計算を要する測定値(スコ

39
e
アや norepinephrine equivalents^{*3}, ROX index^{*4}など)も表示している。性能に関しては2024年8月~2025年2月に入室した症例で検証したところ, AUROCは病院死亡で0.928, ICU死亡で0.960, 気管切開で0.949, ICU生存退室で0.871, 挿管は0.856, 抜管は0.900であった。

生存退室や挿管・抜管を72時間以内としたのは, 週末のために抜管や退室を延期するなど患者によらない要素があり, 24時間以内などにするとその影響を受ける可能性があるためである。

筆者らはこのAI予測を次のような場面で活用している。まずICU退室患者を選定する際, AIの予測と医療者の判断が一致しているか, また候補に漏れないかといった点を確認する。さらに, 候補の中で退室確率が低い患者は本当に退室して問題ないか再検討するきっかけとしている。また抜管を検討している患者では, 抜管の確率が低くないか, 再挿管率が高くないかを検証している。死亡確率や気管切開確率は治療方針の決定をサポートしている。具体的な数値で示されるため, 主科と議論する際に説得力のある資料となっている。特にICU死亡予測確率が一定の割合を一定の日数以上超えている場合(例:9割を超えている日数が7日以上), これまで生存退室できた患者はいないため, 該当する患者では治療撤退を含めた議論において有力な材料となっている。

BEST-AIは方針決定の場面において非常に強力なツールである。今後さらに期待されることとして次のようなものが挙げられる。まず, 現在は診療録などを使用していないため, これら

を予測に組み込むことで数値やカテゴリーデータからでは得られない患者の状態や医療従事者の意図を反映することができると可能性がある⁸⁾。また, 現在は介入の「予測」は表示されているが, さらに精度を高めることや強化学習など別のアプローチを用いることで介入を「提案」することも期待される。

● ● ● ● ● 50%
集中治療領域においてAIの発展はめざましい一方, 偏ったデータによるものが多く, 実臨床への応用はほとんどないことが課題である。ローカルデータを用いることで, 実臨床に沿ったモデルを開発することができる。また転移学習といった技術を用いることで, 少ないローカルデータでも十分な精度を得ることができる。今後さらなる発展が期待される分野であり, その動向に注目である。

13a 見出し MB 31
文献 584 色バツ

1. Edelson DP, Churpek MM, Carey KA, et al. Early warning scores with and without artificial intelligence. JAMA Netw Open 2024 ; 7 : e2438986. 584
2. Rajpurkar P, Irvin J, Zhu K, et al. CheXNet : Radiologist-level pneumonia detection on chest X-rays with deep learning. 2017. (http://arxiv.org/abs/1711.05225) (2025年6月29日閲覧)
3. van de Sande D, van Genderen ME, Huiskens J, et al. Moving from bytes to bedside : a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med 2021 ; 47 : 750-60.
4. Rockenschaub P, Akay EM, Carlisle BG, et al. External validation of AI-based scoring systems in the ICU : a systematic review and meta-analysis. BMC Med Inform Decis Mak 2025 ; 25 : 5.
5. O'Halloran HM, Kwong K, Veldhoen RA, et al. Characterizing the patients,

hospitals, and data quality of the eICU collaborative research database. Crit Care Med 2020 ; 48 : 1737-43.

6. Ebbelohj A, Thunbo MØ, Andersen OE, et al. Transfer learning for non-image data in clinical research : a scoping review. PLOS Digit Health 2022 ; 1 : e0000014.
7. Desautels T, Calvert J, Hoffman J, et al. Using transfer learning for improved mortality prediction in a data-scarce hospital setting. Biomed Inform Insights 2017 ; 9 : 1178222617712994.
8. Marafino BJ, Park M, Davies JM, et al. Validation of prediction models for critical care outcomes using natural language processing of electronic health record data. JAMA Netw Open 2018 ; 1 : e185097.