

〈原著論文〉

機械学習を用いた熱傷患者の入院死亡予測に関する検討

山田 祥子, 椎野 泰和, 立石 寛子, 岡根 亮弘

上野 太輔, 高橋 治郎, 宮本 聡美, 井上 貴博

川崎医科大学救急医学教室

抄録 背景: 熱傷の予後予測指標には様々な指標（従来指標）が提唱されているが、機械学習を用いた予後予測と直接の比較は行われていない。

目的: 熱傷患者の予後予測について従来指標と機械学習モデルとを比較検討する。また、高齢者における予後予測精度を検証する。

対象: 2011年4月から2019年3月までに一般社団法人日本熱傷学会熱傷入院患者レジストリーに登録された20歳以上の患者（JSBI 群）および、川崎医科大学附属病院に2001年1月からの19年間に熱傷で入院した20歳以上の患者（KMS 群）を対象とする。

方法: JSBI 群を、JSBI-train 群と JSBI-test 群にわけ、JSBI-train 群を用いて機械学習モデルを作成し、JSBI-test 群で従来指標および機械学習モデルの予後予測能を、ROC 解析を用いて検討した。また、KMS 群を用いて外部検証を行った。JSBI-test 群を用いて高齢者における従来指標と機械学習モデルの予後予測能を検証した。

結果: 従来指標では Revised baux score（RBS）が最も診断精度が高かったが、機械学習モデルすべてが RBS を上回る精度を示した。年齢層別解析により従来指標、機械学習モデルのいずれにおいても80歳以上の患者でその精度が低下しており、Prognostic burn index（PBI）よりも XGBoost モデルの精度が高かった（ $P = 0.036$ ）。外部検証においても十分な予測能を有していた。

結語: 熱傷の予後予測において機械学習モデルは従来指標を上回る精度を示した。また、高齢者において、本邦で一般的な PBI よりも XGBoost モデルが高い精度を示したがその差は僅かであり、併存疾患や治療撤退などの指標を投入することで、機械学習モデルの精度が上がる可能性がある。

doi:10.11482/KMJ-J202349007（令和5年6月2日受理）

キーワード：熱傷，予後予測，機械学習

背景

熱傷の重症度判定、予後予測指標としては、世界的には熱傷面積（全体表面積に対するパーセンテージ： $\%Total\ body\ surface\ area(\%TBSA)$ ）、 $\%TBSA + 年齢$ （気道熱傷あり：17、なし：0）で計算される Revised baux score（以下 RBS）

が用いられることが多いが¹⁾、本邦では伝統的に熱傷指数（Burn index（BI）＝Ⅲ度熱傷面積＋Ⅱ度熱傷面積 $\times 1/2$ ）と年齢の合計で計算される熱傷予後指数（Prognostic burn index（PBI）＝BI＋年齢）が用いられている²⁾。これら従来から使用されてきた予後予測指標（従来指標）

別刷請求先

山田 祥子

〒701-0192 倉敷市松島577

川崎医科大学救急医学教室

電話：086（462）1111

ファックス：086（464）1044

Eメール：sac_yam_31_42@yahoo.co.jp

は多くの解析ではほぼ同等かつ高精度の予後予測が可能であることが示されているが³⁾、いずれの式も患者の年齢をそのまま予後予測式に代入する点では共通している。しかしながら、年齢が予後に与える影響は非線形である可能性が高く、全年齢層において年齢をそのまま予測式に代入する方法が必ずしも適切ではない可能性がある。重症熱傷の治療には数十回を超える手術が施行される場合もあり、集中治療室の在室日数も極めて長く患者の苦痛は相当のものであるため、正確な予後予測は治療の選択に重大な影響を与えうる。特に高齢者で救命の可能性が低い状況での超侵襲的な治療の適応については臨床現場でも判断に苦慮することが多く、正確な予後予測は極めて重要な判断要因となりうる。

非線形な要因に対する分類については、現在多くの疾病で機械学習モデルによる予後予測の精度が高いことが明らかとなっており、熱傷において年齢が予後に与える影響が非線形であると仮定した場合、従来指標よりも機械学習による予後予測の方がより正確である可能性が考えられる。

しかしながら、熱傷において機械学習モデルの生命予後に対する診断予測精度に関する研究はおこなわれていない。

目 的

従来指標と機械学習モデルによる予後予測の精度を検証する。また、高齢者における機械学習モデルの予後予測精度を検証する。

方 法

1. 対象

一般社団法人日本熱傷学会熱傷入院患者レジストリーは、入院診療を要する熱傷患者を対象として、その重症度、治療状況、転帰を登録し熱傷診療の全体像を把握するとともに、診療情報を分析して診療・教育・研究の質の向上を図ることを目的とした一般社団法人日本熱傷学会が主導で行う全国規模の疫学研究である。今回我々は2011年4月から2019年3月までに一般社

団法人日本熱傷学会熱傷入院患者レジストリー（以下、熱傷レジストリー）に登録された患者13,557例のうち、年齢、性別、受傷起点、Ⅱ度熱傷面積、Ⅲ度熱傷面積、気道熱傷の有無のいずれかが不明な2,227例を除外した。このうち20歳以上の患者9,265例のうち、当院から登録された154例を除いた9,111例をJSBI群としてモデル作成および内部検証として使用した。

また、外部検証として、川崎医科大学附属病院に2001年1月から2019年12月までに熱傷で入院した20歳以上の531例のうち、20歳未満の67例を除いた465例をKMS群とした。なお、外部検証に用いたデータには欠損値を認めなかった。

2. データ準備

JSBI群（ $n = 9,111$ ）をランダムに5,000例のJSBI-train群と4,111例のJSBI-test群にわけそれぞれ（ $n = 5,000$ ）と（ $n = 4,111$ ）とした。これらには、年齢、性別、Ⅱ度熱傷面積、Ⅲ度熱傷面積、気道熱傷の有無、熱傷の原因（火炎、化学熱傷、高温物質、高温液体、爆発、電撃、その他）、および死亡退院が含まれており、これらを用いて機械学習モデルの作成および検証を行なった。

3. 機械学習

機械学習の手法については、過去の報告⁴⁾から分類予測として広く使用されているRandom forest, Support vector machine, Artificial neural networkに近年注目されているExtreme Gradient Boosting (XGBoost)を用いて解析を行なった。いずれの解析も、JSBI-train群を用いてモデルを作成し、JSBI-test群を用いてその精度を検証した。また比較検討のため、従来指標として%TBSA, PBI, RBSについて検証データにおける予後予測精度の解析を行なった。予後予測精度についてはROC解析を用いて解析を行なった。作成された機械学習による予後予測モデルを、KMS群を用いて外部検証を行なった。

Subgroup解析として、10歳ごとの区分に分

けたRBS,PBI および機械学習モデルの診断精度をJSBI-test 群を用いて解析し，高齢者における診断精度についての検討を行った。

解析にはRandom forest (randomForest パッケージ⁵⁾)，Support vector machine (e1071 パッケージ⁶⁾)，Artificial neural network (NeuralNetTools パッケージ⁷⁾)，XGBoost (xgboost パッケージ⁸⁾)を用いた。パッケージの選択については，過去の研究やパッケージのアップデート状況などから選択した。アルゴリズムごとに複数のパッケージでの比較は行っていない。

本研究は川崎医科大学倫理委員会の承認（承認番号 5540-00）を得て実施している。

統計解析

群間の検定は，正規分布数値はStudent's t-test（両側検定），非正規分布数値はMann-Whitney's U-test（両側検定）を使用した。測定結果は，正規分布数値は平均値±標準偏差で示し，非正規分布値は中央値（第1四分位数－第3四分位数）で示した。比率の検定には，Pearson カイ二乗検定あるいはFisherの直接確率法を使用した。機械学習を含め，すべての計算はR（Version 4.1.2）⁹⁾を用いて行なった。

結 果

JSBI 群とKMS 群の背景を表1示す。KMS 群の方で男性が多く，%TBSA, BI, PBI が高く，死亡率も高かった。JSBI-test 群を対象とした，%TBSA, BI, PBI, RBS の死亡に対する診断精度（図1）を示す。いずれも，高い診断精度を有しており，有意差は認めないがRBS が最も診断精度が高かった。このうち最も診断精度の高かったRBSと本邦で広く使われているPBIの各年齢層別（10歳ごと）の診断精度を示す（図2a, b）。RBS, PBI のいずれも年齢が上がるにつれて診断精度が落ちており，特に80代で顕著な低下が認められた。JSBI-train 群を使用して作成した機械学習モデルによるJSBI-test 群を対象とした診断精度を示す（図3）。有意差は認めないものの，機械学習モデルに4種類はすべてRBSを上回り，極めて高い診断精度を示した。機械学習モデルのJSBI-test に対する一致率を示す（図4）。いずれも94%以上の一致率を示し，極めて優秀な予後予測能力を有していることが示された。RBSと機械学習モデルによる年齢層別の診断精度をAUC（95%CI）にて示す（表2）。RBS，機械学習モデルのいずれにおいても80代で診断精度の低下が認められた。RBSの診断精度も十分なため有意差は

表1 患者背景

	JSBI 群 (n=9, 111)	KMS 群 (n=465)	P 値
年齢（中央値（四分位範囲））	62 (43-76)	62 (43-77)	0.577
性別・男性	5,525 (60.6%)	326 (70.1%)	<0.001
Ⅱ度熱傷面積	7.0 (9.7)	9.5 (11.6)	<0.001
Ⅲ度熱傷面積	6.1 (15.6)	13.8 (24.0)	<0.001
気道熱傷あり	2,573 (28.2%)	134 (28.8%)	0.788
受傷原因			
火炎	4,131 (45.3%)	291 (62.6%)	<0.001
化学熱傷	368 (4.0%)	22 (4.7%)	0.461
高温物質	774 (8.5%)	12 (2.6%)	<0.001
高温液体	2,472 (27.1%)	91 (19.6%)	<0.001
爆発	505 (5.5%)	30 (6.5%)	0.405
電撃	154 (1.7%)	11 (2.4%)	0.275
その他	646 (7.1%)	8 (1.7%)	<0.001
% Total body surface area	13.10 (18.78)	23.63 (25.80)	<0.001
Burn index	9.61 (16.54)	18.59 (24.11)	<0.001
Prognostic Burn Index	68.79 (26.59)	78.30 (32.34)	<0.001
Revised Baux Score	77.07 (30.15)	88.23 (35.52)	<0.001
死亡退院	894 (9.8%)	115 (24.7%)	<0.001

性別，気道熱傷，原因は実数（割合），Ⅱ度熱傷面積，Ⅲ度熱傷面積，% Total body surface area, Burn index, Prognostic Burn index, Revised Baux Score は平均（標準偏差）で示す。

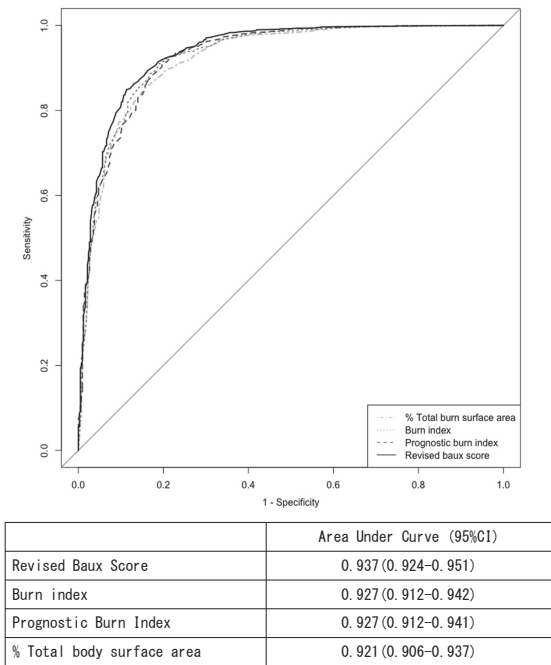
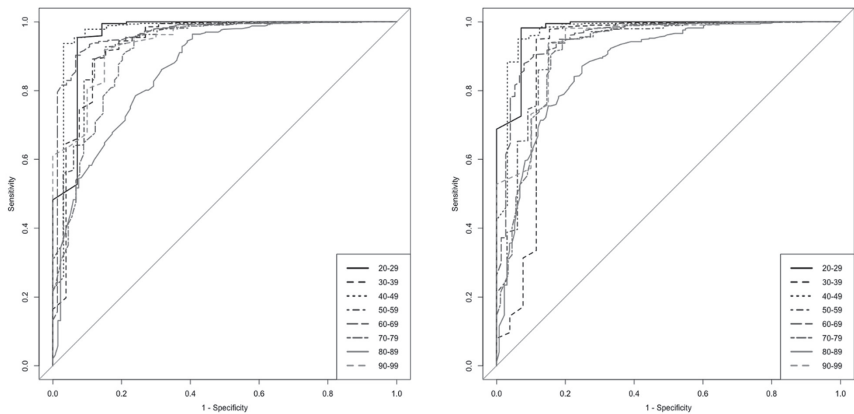


図1 従来指標による全入院死亡に対する予測精度（ROC 解析：JSBI-test 群）



年齢（歳）	Area Under Curve (95%CI)	
	Revised Baux Score	Prognositc Burn Index
20 - 29	0.961 (0.892-1)	0.977 (0.892-1)
30 - 39	0.933 (0.866-1)	0.901 (0.866-1)
40 - 49	0.969 (0.923-1)	0.973 (0.923-1)
50 - 59	0.937 (0.887-0.988)	0.924 (0.887-0.988)
60 - 69	0.963 (0.939-0.988)	0.954 (0.939-0.988)
70 - 79	0.91 (0.871-0.949)	0.914 (0.871-0.949)
80 - 89	0.863 (0.826-0.9)	0.88 (0.826-0.9)
90 - 99	0.943 (0.891-0.995)	0.934 (0.891-0.995)

図2 年齢層別の Revised Baux Score（a: 左）と Prognostic Burn Index（b: 右）の全入院死亡に対する予測精度（ROC 解析：JSBI-test 群）

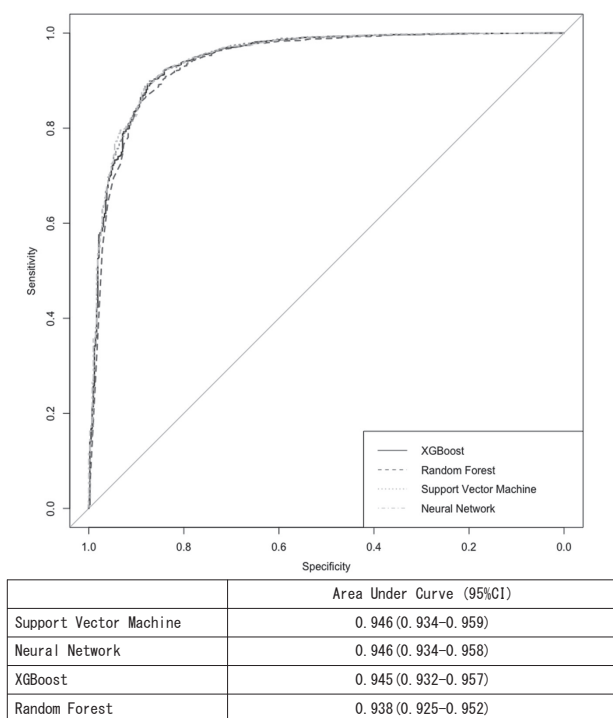


図3 機械学習モデルによる全入院死亡に対する予測精度（ROC 解析：JSBI-test 群）

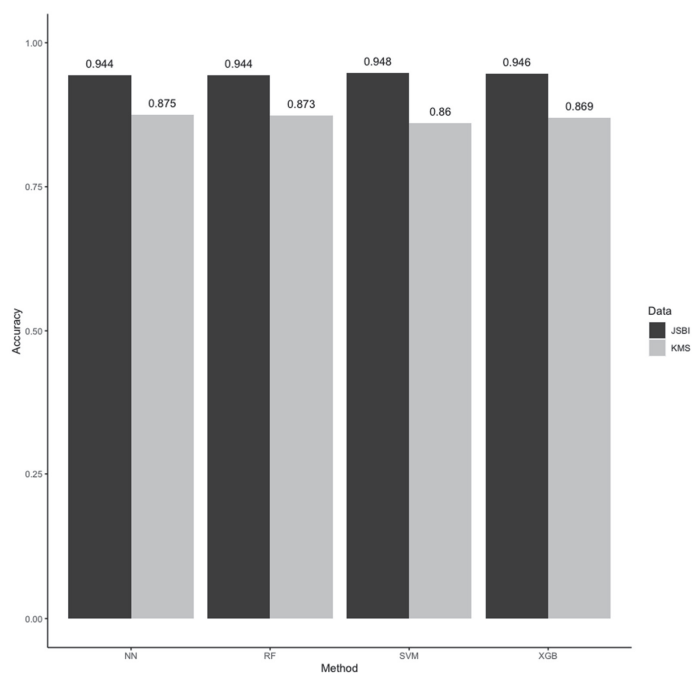


図4 機械学習モデルによる内部検証（JSBI-test 群）と外部検証（KMS 群）の一致率

表2 年齢層別の従来指標および機械学習モデルの全入院死亡に対する予測精度 (JSBI-test 群)

年齢 (歳)	Area Under Curve (95%CI)				
	Revised Baux Score	Random Forest	Support Vector Machine	Neural Network	XGBoost
20 - 29	0.961 (0.892 - 1)	0.987 (0.892 - 1)	0.938 (0.822 - 1)	0.991 (0.98 - 1)	0.93 (0.799 - 1)
30 - 39	0.933 (0.866 - 1)	0.938 (0.866 - 1)	0.929 (0.861 - 0.996)	0.938 (0.881 - 0.995)	0.932 (0.868 - 0.996)
40 - 49	0.969 (0.923 - 1)	0.967 (0.923 - 1)	0.978 (0.949 - 1)	0.977 (0.95 - 1)	0.975 (0.942 - 1)
50 - 59	0.937 (0.887 - 0.988)	0.929 (0.887 - 0.988)	0.937 (0.886 - 0.987)	0.935 (0.885 - 0.985)	0.926 (0.866 - 0.986)
60 - 69	0.963 (0.939 - 0.988)	0.971 (0.939 - 0.988)	0.965 (0.945 - 0.986)	0.965 (0.945 - 0.986)	0.958 (0.933 - 0.982)
70 - 79	0.91 (0.871 - 0.949)	0.908 (0.871 - 0.949)	0.928 (0.896 - 0.961)	0.927 (0.895 - 0.96)	0.925 (0.891 - 0.959)
80 - 89	0.863 (0.826 - 0.9)	0.883 (0.826 - 0.9)	0.892 (0.861 - 0.924)	0.893 (0.862 - 0.924)	0.894 (0.863 - 0.925)
90 - 99	0.943 (0.891 - 0.995)	0.938 (0.891 - 0.995)	0.963 (0.932 - 0.995)	0.964 (0.931 - 0.998)	0.953 (0.9 - 1)

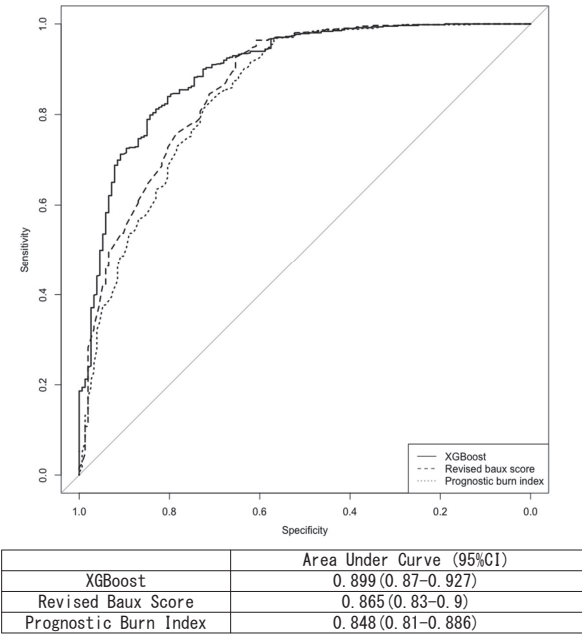


図5 JSBI-test 群の80歳以上に対する XGBoost, RBS, PBI の診断精度

認められなかったが、この年齢層においても機械学習モデルに4種類はすべてRBSを上回った。JSBI-test 群の80歳以上に対する XGBoost, RBS, PBI の診断制度を示す (図5)。RBS と XGB の間には有意差はないが (p = 0.153) PBI と XGB の間で有意差を認めた (p = 0.03611)。

外部検証

続いて KMS 群における AI の外部検証結果を示す (図6)。外部検証の結果も有意差はないが RBS,PBI よりも機械学習モデルの方が高い精度を示した。外部検証の一致率を示す (図4)。全てのモデルで85%を超える一致率を示したが、内部検証に比較して低下していた。

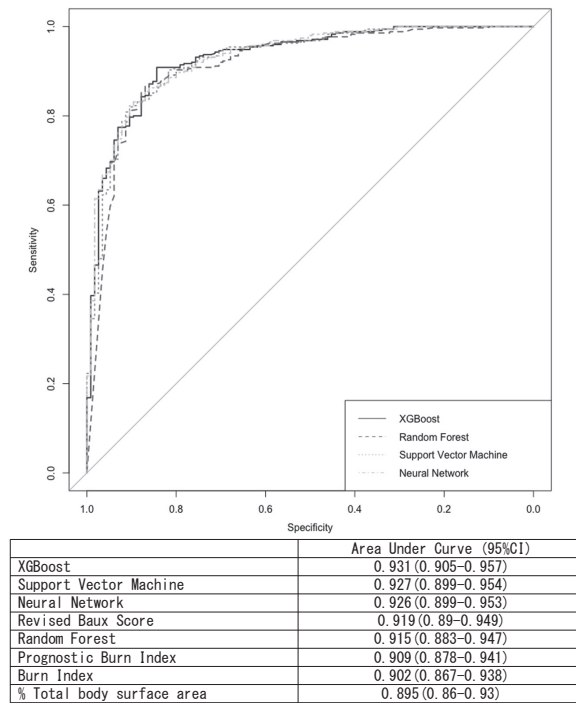


図6 KMS 群における外部検証結果

考 察

今回我々は、熱傷患者における従来指標と機械学習の予後予測精度を比較することを目的として本研究を行なった。従来指標については歴史も深く、数多くの研究が行われほぼ同等かつ高い予後予測精度を持っていることが報告されているが^{3, 10-15)}、今回の我々の研究においてもほぼ同様の結果であった。一方で、高齢者において従来指標ではその予後予測精度の著しい低下がみられた。これは、多大な苦痛を伴う熱傷治療において、不適切に治療撤退が選択される可能性に繋がり、今後大きな問題となりうると考えられた。

機械学習による熱傷患者の予後予測については、いくつかの研究が行われている。Frye らは1,585人の熱傷患者に対し、気道熱傷の有無、年齢、%TBSA を指標に用いて artificial neural network により救急外来での死亡予測を行っている。その死亡予測能は極めて高く、その精度は98%であったと報告しているが、救急外来で

の死亡に限定した単施設の研究であり、検証に使われたデータは159人と決して多くない¹⁶⁾。

2002年には Estahbanati らが2,096人の熱傷患者のデータを用いて、Frye らが用いた気道熱傷の有無、年齢、%TBSA に、性別を因子に加えた上で artificial neural network を用いて予後予測を行った結果、その精度は90%であったと報告している¹⁷⁾。また、artificial neural network に加えて Decision tree, Naïve Bayes classifier, support vector machine を用いた研究もあり、それぞれ高い精度を示したとしているが^{18, 19)}、単施設で対象症例が180人、99人とあまりに小規模であり、外的妥当性に疑問がある。

これらの結果から、先行研究でその有用性が示されている artificial neural network, Random forest, Support Vector Machine に加えて近年その性能の高さが注目されている XGBoost を用いて解析を行なった。

次に、機械学習の各アルゴリズムについて考察する。

Random forest はツリー状に条件分岐を繰り返すことで分類を予測するモデルを作成する決定木法を多数行い、統合し解析するものである。決定木法だけでは過学習しやすいという弱点があるが、Random forest を用いることでその欠点を補うことができる。アウトカムに対して意義のある変数が、意義のない変数よりも極端に少ない場合に良好な結果を残せないと言われている²⁰⁾。今回我々は熱傷レジストリーデータを用いたため、熱傷の予後に影響が少ない因子は含まれておらず、内部検証で一致率94.4%、当院のデータを用いた外部検証で87.3%と高い一致率を示した。一方で、のちに述べる XGBoost, Support Vector Machine, Artificial neural network と比較すると内部検証データ、外部検証データのいずれにおいても有意差はないものの AUC が低く、計算が簡易とはいえ、今回使用した機械学習法の中での優位性はないものと考えられた。

Support vector machine はデータを分類するにあたり、その境界と近傍のデータの距離が可能な限り大きくなるように境界を定める方法である。基本的に2クラス分類がその特徴であり、今回のような生死の判定には向いているといえる。また、未知のデータに対しても比較的精度が高く予測ができる方法であるが、モデルの作成においてチューニングが重要である。また、データが大きくなった場合、計算量が大きくなり簡便性にかける²¹⁾。今回、我々は Bootstrapping によるチューニングを行なった。その結果、内部検証データでは94.8%、外部検証データでも86.0%と高い予測精度を示した。

Artificial neural network は上述の通り熱傷の予後予測として過去に研究がなされており、非常に高い精度を示したと報告されている^{16, 17)}。今回の我々の研究における一致率は内部検証データで94.4%、外部検証データで87.5%であった。Frye は内部検証データを用いた一致率を98%と報告しており、これは我々の94.4%よりの高い数字であるが、救急外来での死亡予測に限定したものであり、入院後の様々な交絡因子

が含まれない状況での解析のため、診断精度が高くなったものと思われる。一方、Estahbanati らの研究では、内部検証データを用いた一致率を90%と報告している。また、彼らの研究では40歳以下の患者が過半数で、80歳以上の患者は急性熱傷1,082人のうち4%しか含まれておらず、本邦の状況とは乖離している。今回の熱傷レジストリーの対象データにおいても18.5% (1,685人 / 9,111人) が80歳以上であり、直接の比較は難しいと考えられる。

XGBoost は勾配ブースティングともよばれ、決定木を作成する際に前に構築された結果を利用して、正しく分類されなかったデータに対する予測を向上させていく仕組みである²²⁾。本アルゴリズムではパラメータの設定により、その精度が変化しやすいが、今回我々は、ParamHelpers²³⁾を用いてパラメータのチューニングをランダムサーチで行い、最終的に nround = 494, max depth = 1, eta = 0.335, lambda = 1.12 となり、過剰な過学習にはなっていないと推測される。XGBoost は Random forest と同じく決定木法であり、前の決定木の誤差を引き継ぎモデルを作成するため、一般的には Random forest よりも精度が高くなりやすいが、今回の我々の研究でもその AUC は内部検証、外部検証のいずれでも有意差はないものの Random forest を上回っており、今後は XGBoost に類似する Light Gradient Boosting Machine や Category Boosting などが有力な候補となる可能性がある。

本研究では機械学習モデルは、従来指標である RBS や PBI を上回ることができたが、その差はわずかであった。これは熱傷自体が、外的要因による比較的シンプルな病態であることが原因であると考えられる。つまり、熱傷の重症度は、その温度、接触時間、接触面積により規定されるのであり、基本的には熱傷面積と熱傷深達度がその重症度を決定している。生命予後はその重症度に、患者背景や合併症を加味したものであると考えられ、今後、多くのパラメータを機械学習モデルに投入することで、さらに高い精度を得られる可能性がある。

一方で高齢者の熱傷における従来指標での予後予測精度には課題が残る。過去の報告では、予後を区別または予測する明確なカットオフ年齢はないが、65歳を越え年齢が上がるにつれて死亡リスクが増加することから、高齢者熱傷の定義として60歳以上とすることが提案されている²⁴⁾。また、高齢者熱傷の予後予測指標としては今回検討した%TBSA、Ⅲ度熱傷面積、気道熱傷、性別だけでなく、合併症（慢性閉塞性肺疾患、糖尿病、心臓病）が提唱されている²⁵⁻²⁸⁾。

今回の我々の検討では、年齢層別解析にて精度が落ちることが明らかになった80歳以上の患者において、XGBoost モデルは本邦で広く用いられている PBI よりも有意に高い精度を示した（AUC: $p = 0.031$ ）が、その他の年代よりも従来指標、機械学習モデルのいずれも精度の低下が見られた。これは上記の既往症の有無が従来指標、機械学習モデルのいずれにも含まれていないことがその要因と考えられ、チャールソン併存疾患指数（Charlson Comorbidity Index）²⁹⁾を含めた解析を行うことでモデルが改善する可能性がある。また、高齢者熱傷患者においてしばしば経験する「積極的治療の辞退」の影響も無視できない。「積極的治療の辞退」についての情報はこれまでの熱傷レジストリーに含まれておらず、2023年度に改訂される見込みである熱傷レジストリーによるデータ集積が待たれる。

本研究の限界と今後の展望を示す。今回の研究は10年以上前のデータを含むものであり、現時点での生命予後を正確に示しているわけではないことに注意が必要である。熱傷診療も他の疾患と同じく、技術の進歩や経験の積み重ねにより救命率は向上していると考えられ、今回作成した機械学習モデルによる予後予測は真実よりも悪く推定される可能性が残っており、今回の研究結果が積極的治療撤退を促すものになってはいけないと考える。しかしながら熱傷治療に伴う苦痛は大きく、医療費も高額であり患者・家族の身体的・精神的苦痛や経済的問題、熱傷治療に関する医療資源の有効活用など考慮する

と、機械学習を用いてさらに多くの因子を総合的に判断し、より精度の高い予後予測に取り組むことには意義があると考ええる。今後、機械学習を日常診療に応用し、基礎疾患や日々の臨床検査、熱傷創部の情報など情報量をさらに多くすることで、熱傷の治療経験が少ない医療施設であっても適切に判断ができるシステムを構築できる可能性があり、救命率の向上、治療の効率化、後遺障害の軽減、入院期間の減少や医療費削減、治療方針の決定につなげられると期待できる。

結 語

熱傷レジストリーを用いた機械学習による熱傷患者の予後予測は、従来から本邦で広く用いられている PBI と同等の予後予測能を有していた。また、80歳以上の高齢者において XGBoost を用いた機械学習モデルは PBI よりも優れた予後予測能を示した。今後、合併症、積極的治療の辞退などの情報を含めた多くの情報を機械学習モデルに投入することで、従来指標では予測精度が下がる可能性がある高齢者熱傷の予後予測の精度をさらに改善することができる可能性がある。

引用文献

- 1) Osler T, Glance LG, Hosmer DW. Simplified Estimates of the Probability of Death After Burn Injuries: Extending and Updating the Baux Score. *Journal of Trauma: Injury, Infection & Critical Care*. 2010; 68: 690-697. doi: 10.1097/TA.0b013e3181c453b3.
- 2) Kaita Y, Tarui T, Tanaka Y, Suzuki J, Yoshikawa K, Yamaguchi Y. Reevaluation for prognostic value of prognostic burn index in severe burn patients. *Acute Medicine & Surgery*. 2020; 7. doi: 10.1002/ams2.499.
- 3) Yoshimura Y, Saitoh D, Yamada K, Nakamura T, Terayama T, Ikeuchi H, Sasaki J, Nemoto M. Comparison of prognostic models for burn patients: A retrospective nationwide registry study. *Burns*. 2020; 46: 1746-1755. doi: 10.1016/j.burns.2020.10.008.
- 4) Liu NT, Salinas J. Machine learning in burn care and research: A systematic review of the literature. *Burns*. 2015; 41: 1636-1641. doi: 10.1016/j.burns.2015.07.001.
- 5) Liaw A, Wiener M. Classification and Regression by

- randomForest. R News. 2002; 2: 18-22.
- 6) Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. 2022, R package version 1.7-12., <https://CRAN.R-project.org/package=e1071>.
 - 7) Beck MW. NeuralNetTools: Visualization and Analysis Tools for Neural Networks. Journal of Statistical Software. 2018; 85: 1-20. doi: 10.18637/jss.v085.i11.
 - 8) Chen T, He T, Benesty M, *et al.* xgboost: Extreme Gradient Boosting. 2022, R package version 1.6.0.1, <https://CRAN.R-project.org/package=xgboost>.
 - 9) R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.
 - 10) Tagami T, Matsui H, Fushimi K, Yasunaga H. Validation of the prognostic burn index: a nationwide retrospective study. Burns. 2015; 41: 1169-75. doi: 10.1016/j.burns.2015.02.017.
 - 11) Blot S, Brusselsaers N, Monstrey S, *et al.* Development and validation of a model for prediction of mortality in patients with acute burn injury. British Journal of Surgery. 2008; 96: 111-117. doi: 10.1002/bjs.6329.
 - 12) McGwin G, George RL, Cross JM, Rue LW. Improving the ability to predict mortality among burn patients. Burns. 2008; 34: 320-327. doi: 10.1016/j.burns.2007.06.003.
 - 13) Ryan CM, Schoenfeld DA, Thorpe WP, Sheridan RL, Cassem EH, Tompkins RG. Objective Estimates of the Probability of Death from Burn Injuries. The New England Journal of Medicine. 1998; 338: 362-366. doi: 10.1056/nejm199802053380604.
 - 14) Sheppard NN, Hemington-Gorse S, Shelley OP, Philp B, Dziewulski P. Prognostic scoring systems in burns: A review. Burns. 2011; 37: 1288-1295. doi: 10.1016/j.burns.2011.07.017.
 - 15) Moore EC, Pilcher D, Bailey M, Stephens H, Cleland H. The Burns Evaluation and Mortality Study (BEAMS): Predicting deaths in Australian and New Zealand burn patients admitted to intensive care with burns. Journal of Trauma-Injury Infection and Critical Care. 2013; 75: 298-303. doi: 10.1097/ta.0b013e318295409d.
 - 16) Frye KE, Izenberg SD, Izenberg SD, Williams MD, Luterma A. Simulated biologic intelligence used to predict length of stay and survival of burns. Journal of Burn Care & Rehabilitation. 1996; 17: 540-546. doi: 10.1097/00004630-199611000-00011.
 - 17) Estahbanati HK, Bouduhi N. Role of artificial neural networks in prediction of survival of burn patients-a new approach. Burns. 2002; 28: 579-586. doi: 10.1016/s0305-4179(02)00045-1.
 - 18) Jiménez F, Sánchez G, Juárez JM. Multi-objective evolutionary algorithms for fuzzy classification in survival prediction. Artificial Intelligence in Medicine. 2014; 60: 197-219. doi: 10.1016/j.artmed.2013.12.006.
 - 19) Patil BM, Joshi RC, Toshniwal D, Biradar S. A New Approach: Role of Data Mining in Prediction of Survival of Burn Patients. Journal of Medical Systems. 2011; 35: 1531-1542. doi: 10.1007/s10916-010-9430-2.
 - 20) Breiman L. Random Forests. Machine Learning. 2001; 45: 5-32. doi: 10.1023/A:1010933404324.
 - 21) Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995; 20: 273-297. doi: 10.1007/BF00994018.
 - 22) Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016; Accessed: Mar. 12, 2023. <https://dl.acm.org/doi/10.1145/2939672.2939785>.
 - 23) Bischl B, Lang M, Richter J, Bossek J, Horn D, Kerschke P. ParamHelpers: Helpers for Parameters in Black-Box Optimization, Tuning and Machine Learning. 2022.
 - 24) Jeschke MG, Pinto R, Costford SR, Amini-Nik S. Threshold age and burn size associated with poor outcomes in the elderly after burn injury. Burns. 2016; 42: 276-281. doi: 10.1016/j.burns.2015.12.008.
 - 25) Pham TN, Kramer CB, Wang J, Rivara FP, Heimbach DM, Gibran NS, Klein MB. Epidemiology and Outcomes of Older Adults With Burn Injury: An Analysis of the National Burn Repository. Journal of Burn Care & Research. 2009; 30: 30-36. doi: 10.1097/bcr.0b013e3181921efc.
 - 26) Baux S, Mimoun M, Saade H, Lioret N, Esteve M, Nolland XB, Bertiere MN. Burns in the elderly. Burns. 1989; 15: 239-240. doi: 10.1016/0305-4179(89)90039-9.
 - 27) Ho W, Ying SY, Ying S, Chan HH. A study of burn injuries in the elderly in a regional burn centre. Burns. 2001; 27: 382-385. doi: 10.1016/s0305-4179(00)00146-7.
 - 28) Mahar PD, Wasiak J, Bailey M, Cleland H. Clinical factors affecting mortality in elderly burn patients admitted to a burns service. Burns. 2008; 34: 629-636. doi: 10.1016/j.burns.2007.09.006.
 - 29) Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of chronic diseases. 1987; 40: 373-383. doi: 10.1016/0021-9681(87)90171-8.

〈Regular Article〉

Machine Learning Models to Predict In-hospital Mortality in Burn Patients

Sachiko YAMADA, Yasukazu SHIINO, Hiroko TATEISHI, Takahiro OKANE
Daisuke UENO, Jiro TAKAHASHI, Satomi MIYAMOTO, Takahiro INOUE

Department of Acute Medicine, Kawasaki Medical School

ABSTRACT BACKGROUND: Various conventional indices have been proposed as prognostic indicators for burns, but they have not been directly compared with machine learning-based prognostic models.

OBJECTIVES: We compared conventional indices and machine learning models with hospitalization death as the primary outcome. Furthermore, we examined the prediction accuracy of the indices and models in older adults.

PARTICIPANTS: Patients aged 20 years or older who were registered in the Burn Inpatient Registry of the Japan Burn Injury Society between April 2011 and March 2019 (JSBI group) and patients aged 20 years or older who were admitted to Kawasaki Medical School Hospital for burn injuries between January 2001, and December 2019 (KMS group).

METHODS: The JSBI group was divided into JSBI-train and JSBI-test groups, and a prognostic model was created via machine learning (Random Forest, Support Vector Machine, Artificial Neural Network, XGBoost) using the JSBI-train group. The prognostic ability of the conventional indices and machine learning models in older adults was examined in the JSBI-test group using receiver operating characteristic analysis. External validation was conducted using the KMS group.

RESULTS: Among the conventional indices, the revised Baux score (RBS) showed the highest accuracy in predicting death during hospitalization; however, all the machine learning models showed higher accuracy than the RBS. The accuracy of the XGBoost model was higher than that of the prognostic burn index (PBI) in a subgroup analysis of patients aged 80 years or older ($P = 0.036$). The external validation analysis revealed that the machine learning model adequately predicted in-hospital death, although its accuracy was lower than that of the internal validation model.

DISCUSSION: The accuracy in the KMS group was lower than that in the JSBI-test group, possibly because the patients in the KMS group had more severe burns than those in the JSBI-test group. Moreover, the influence of comorbidities and treatment discontinuation in older adults cannot be ruled out.

CONCLUSION: The machine learning models were more accurate than the conventional indices in predicting the prognosis of burns. In older patients with burns, the XGBoost model was more accurate than the PBI, which is commonly used in Japan, but the difference

was small. In the future, the accuracy of machine learning models should be improved by considering factors such as comorbidities and treatment discontinuation.

(Accepted on June 2, 2023)

Key words : Burn injury, Machine learning, Mortality prediction

Corresponding author

Sachiko Yamada

Department of Acute Medicine, Kawasaki Medical
School, 577 Matsushima, Kurashiki, 701-0192, Japan

Phone : 81 86 462 1111

Fax : 81 86 464 1044

E-mail : sac_yam_31_42@yahoo.co.jp