

一般口演 | 医療データ解析

一般口演1

医療データ解析

2019年11月22日(金) 09:00 ~ 11:00 C会場 (国際会議場 2階国際会議室)

[2-C-1-04] SS-MIX2標準化ストレージを用いた入院後の死亡退院リスク予測モデルの開発

○関 倫久¹、河添 悦昌¹、大江 和彦¹ (1. 東京大学医学部附属病院)

キーワード : SS-MIX2, Machine Learning, Outcome Prediction, In-hospital mortality

【背景】入院時の患者重症度評価において患者の死亡リスクを評価することは、重症度に応じた入院管理にとって重要である。現行では集中治療室・ハイケアユニットの施設基準評価のために重症度、医療・看護必要度の評価が行われているが、主に医療者側の行為のみが反映され、患者側自体の要因を反映仕切れていない現状がある。入院時の限定された患者要因項目から機械的に死亡リスクを予測でき、患者のリスクを層別化することができれば臨床的に有用と考えられる。

【目的】入院時の患者要因項目を用いて死亡退院リスクの予測モデルを作成する。

【方法】東京大学医学部附属病院における2009年から2017年の匿名化された SS-MIX2標準化ストレージデータを本研究に用いた。2009年-2015年の症例(125283回の入院イベント)の入院時の採血19項目、性別、年齢の計21項目を用いて死亡退院を予測するモデルを作成し、2016年-2017年の症例(35588回の入院イベント)をテストデータとした。機械学習手法として、ロジスティック回帰モデル、ランダムフォレストモデル、ニューラルネットワークモデルをそれぞれ作成し、AUC (Area Under the Curve) を用いて死亡退院に対する予測性能を評価した。

【結果】入院時採血、性別、年齢を用いた死亡退院予測において、ロジスティック回帰モデル、ランダムフォレストモデル、ニューラルネットワークモデルでそれぞれ AUCは0.913 [95%信頼区間(CI) 0.908-0.918]、0.922 [CI 0.917-0.927]、0.919 [CI 0.914-0.924]であった。

【結論】入院時の採血19項目、性別、年齢の計21項目のみを用いて患者の死亡退院を予測するモデルを開発し、良好な予測性能が得られた。作成したモデルは適切な重症度に応じた入院管理を行う上で有用と考えられる。

SS-MIX2 標準化ストレージを用いた入院後の死亡退院リスク予測モデルの開発

関 倫久^{*1}、河添 悦昌^{*2}、大江 和彦^{*1,3}

*1 東京大学医学部附属病院企画情報運営部

*2 東京大学大学院医学系研究科医療 AI 開発学講座

*3 東京大学大学院医学系研究科社会医学専攻医療情報学分野

Machine Learning-Based Prediction of In-Hospital Mortality Using Data in SS-MIX2 Standardized Storage

Tomohisa Seki^{*1}, Yoshimasa Kawazoe^{*2}, Kazuhiko Ohe^{*1,3}

*1 Department of Healthcare Information Management, The University of Tokyo Hospital,

*2 Artificial Intelligence in Healthcare, Graduate School of Medicine, The University of Tokyo,

*3 Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo

[Background] Risk assessment for in-hospital mortality of patients at the time of hospitalization is necessary to determine the scale of required medical resources for each patient according to the severity of their condition. [Objective] To generate an accurate prediction model of in-hospital mortality using machine learning techniques. [Methods] Admission cases with patients 18 years of age or older admitted to the University of Tokyo Hospital between January 1, 2009 and December 26, 2017 in the SS-MIX2 standardized storage were used in this study; the data were divided into a training data set (n = 125,283) and a test data set (n = 35,588) according to the time of admission. To generate the prediction model of in-hospital mortality, 21 variables (age, sex, and 19 laboratory test items) were taken into account using logistic regression, random forest, and deep neural network analyses. [Results] Using the test data set, the prediction models of in-hospital mortality showed C statistic values of 0.913 [95% CI, 0.908–0.919], 0.922 [95% CI, 0.917–0.927], and 0.916 [95% CI, 0.911–0.921], respectively, in the case of logistic regression, random forest, and deep neural network, respectively. [Conclusions] Machine learning-based prediction of in-hospital mortality using admission laboratory data demonstrated an outstanding prediction capability.

Keywords: In-hospital mortality, Machine Learning, Outcome prediction, SS-MIX2 standardized storage

1. 緒論

患者の入院時における死亡リスクの評価は臨床現場において大きな重要性を有している。患者の治療に必要な医療資源のレベルの評価、適切な医療レベルを提供可能な病室の選択、高度医療が必要な施設へ転送を行うかの判断など、多くの場面において死亡リスクの評価は有用となり得る。これまで入院患者の院内死亡の予測は、重症例に対する死亡リスク評価を目的とし、治療方針選択時の補助的利用や臨床研究の対象選択、施設間治療成績の比較などを目的として広く研究されてきた。¹⁾²⁾しかし APACHE スコア³⁾⁴⁾⁵⁾⁶⁾、SAPS スコア⁷⁾⁸⁾⁹⁾、SOFA スコア¹⁰⁾などの院内死亡のリスクを予測するモデルは、主に ICU に入室した集学的治療を要する重症例を評価する目的で開発されており、軽症例まで含めた全入院患者集団から高度な医療資源を要さない軽症例と、集学的治療を要する重症例を見分けるという目的に対しては、単純に流用できる訳ではない。必要な医療レベルの評価を行い、軽症例と重症例を見分けることを目的とした予測モデルを作成するためには、予測モデルの作成過程で軽症から重症まで全症例を含めた集団を対象とする必要がある。

また、いずれの情報を用いて患者予後を予測するかという点に関し、臨床の現場においてその入力を簡便化させるという目的から言えば、診療室で短時間に動的に変化し続けるような項目より、一定の時間内であれば静的な情報として得られる情報の方が望ましい。さらに予測モデルを開発するにあたって、データとしてアクセスが可能な患者アウトカムと紐付けできる状態で、既存の電子診療録内に保存されている情報を利用すれば、新たなレジストリの構築などを行わずに容易にモデル化することが可能である。これらの点から、多くの

入院患者において入院時に日常的に行われており、その検査結果が電子診療録に記録されている採血検査結果に着目した。採血結果の多くは入院時の短時間のエピソードで急激に変動することは少なく、いずれの値を予測に用いるかの選択に臨床医が迷うケースも少ないと考えられる。また、ロジスティック回帰を含めた既存手法を用いた予測モデリングでは、採血検査が患者の院内死亡の予測に関して有用となることが既に報告されている。¹¹⁾近年では非線形分離が可能な機械学習技術が医療分野の予測モデル作成に応用されており¹²⁾¹³⁾、入院時の採血検査結果を含めた患者情報を用いて院内死亡を予測する非線形機械学習モデルを作成することで、結果として高い予測性能の達成が期待できると考えられた。

2. 目的

本研究は、電子診療録に記録された入院実施時点での年齢、性別、採血検査結果から、機械学習技術を用いて院内死亡を高い予測性能で予測するモデルを作成することを目的とする。

3. 方法

3.1 対象とデータ抽出

2009年1月1日から2017年12月26日までの間に東京大学医学部附属病院に入院となった入院例(80,729名、173,578回の入院)を、匿名化した SS-MIX2 標準化ストレージから抽出した。SS-MIX2 に格納されている HL7 version 2.5 (ISO 27931:2009)メッセージ¹⁴⁾の中から入院イベントを SS-MIX2 標準化ストレージデータ種別の ADT-22(入院実施)を元に取り出し、それに該当する日付群を取り出したのちに対応するデータ種別 ADT-52(退院実施)を有する日付を取り出

して最終的に入院実施時刻と退院実施時刻で接合して入院時、退院時と定義した。入院実施のみ、退院実施のみが存在し、矛盾なく入院実施と退院実施を接合できない入退院イベントは抽出対象から除外した。データ抽出後に入院日によって入られておらず判読不能であった 1 ケースを対象から除外した。データ分割後、入院時に患者年齢が 18 歳未満であった入院ケースを除外し、入院時年齢 18 歳以上のみの入院ケースを対象とした(学習用データ: $n = 125,283$, テスト用データ: $n = 35,588$)。予測に用いる入院時の採血検査結果は、入院実施オーダの前後 6 時間の中で、検体採取時刻が入院実施オーダの時刻に最も近いものを採用した。データの変数として当初 1946 項目が抽出されたが、全入院ケースのうち機械学習に用いることが不可能な項目(匿名化 ID など)及び 20% 以上のケースで欠損する変数は対象から除外し、最終的に 23 項目(入院期間、退院時状態、年齢、性別、19 項目の採血項目)を研究対象とした(図 1)。なお、本研究の実施は東京大学大学院医学系研究科・医学部倫理委員会の承認を得ている(承認番号 10791)。

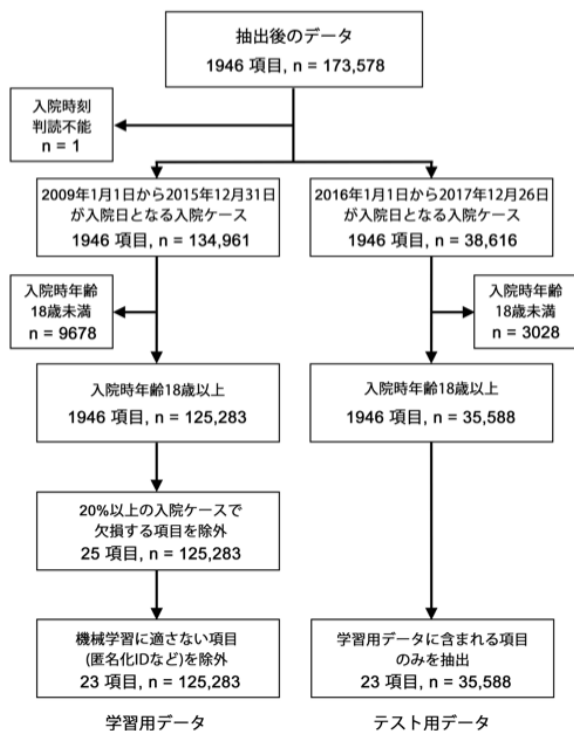


図 1 学習用データとテスト用データの作成

SS-MIX2 標準化ストレージから抽出されたデータから入院時年齢が 18 歳以上の入院ケースを抽出し、さらに機械学習の予測に用いる項目を抽出した。

3.2 欠損値補完

抽出によって得られたデータは、学習データでは 125,283 ケース中の 35,844 ケース、テスト用データでは 35,588 ケース中の 8,443 ケースが欠損値を有していた。学習データ、テスト用データそれぞれに対して R (version 3.3.2) の micemd パッケージ (version 1.2.0)^{15) 16)} を用いて多重代入による欠損値補完を行なった。欠損値はそれぞれ 10 回代入値を計算し、結果として得られた学習用データ、テスト用データそれぞれ 10 セットずつのデータをモデリングと検証に用いた。

データを学習用データとテスト用データに分割した(学習用データ: $n = 134,961$, 2009 年 1 月 1 日から 2015 年 12 月 31 日、テスト用データ: $n = 38,616$, 2016 年 1 月 1 日から 2017 年 12 月 26 日)。データ分割時に、入院時期の項目が正しく記

3.3 機械学習手法

本研究では教師あり機械学習の手法として、ロジスティック回帰¹⁷⁾、ランダムフォレスト¹⁸⁾、ニューラルネットワーク¹⁹⁾を用いて学習データでモデルの学習を行い、それぞれ院内死亡予測モデルの作成を行った。モデルに用いる説明変数は 21 項目(年齢、性別、19 項目の採血項目)とし、予測ターゲットとなる患者アウトカムは退院時状態(生存または死亡)とした。学習にあたっては、SMOTE (Synthetic Minority Over-sampling Technique)アルゴリズム²⁰⁾を用いて、院内死亡例をデータ数の 20%まで増幅したのちに、学習時の不正解時の誤差がデータ内の頻度に合わせて均等となるよう重み付けを行なった。モデルのハイパーパラメータの調整にあたっては、欠損値補完した全ての学習データを用い、学習データ内で 10 分割交差検証を行って C 統計量の平均値を比較し、最も優れた結果が得られたパラメータを採用することで調整を行なった。交差検証にあたっては、多重代入後のデータ間で同一の入院のケースが交差検証における学習データと検証データをまたぐことのないようにデータの分割を行なった。

ロジスティック回帰では、正則化項の係数($10^{-2} \sim 10^4$)及び正則化手法(L1 または L2)についてグリッドサーチを用いてパラメータの検証を行なった。

ランダムフォレストでは、決定木数を 500 に固定した上で、各決定木の深さを 1 から 20 まで、決定木において分類に利用する説明変数の数を 1 から 20 までグリッドサーチを用いて検証し、パラメータの調整を行なった。

ニューラルネットワークでは、それぞれが 128 個のパーセプトロンを有する 3 層のニューラルネットワークを用い、最適化手法として確率的勾配降下法を用いた。各パーセプトロンのドロップアウト比率は同一として 0 から 0.9 まで、学習率は 10^{-2} から 10^{-3} まで、バッチサイズは 128 から 2048 までの値でグリッドサーチを用いて検証し、パラメータの調整を行なった。

学習を終えた予測モデルは、全てのテストデータを用いて検証を行い、それぞれのモデルごとに C 統計量を算出した。C 統計量の算出にあたっては、交差検証で作成された 10 パターンのモデルによる出力値の平均値を用いて計算を行った。それぞれの機械学習モデルの出力値のキャリブレーションは、交差検証時の検証データを用いた isotonic 回帰によって行った。

4 結果

データ抽出テストデータを用いた検証の結果、学習用データ及びテスト用データにおける各説明変数とアウトカムの統計値を表 1 に示す。院内死亡例の比率は学習用データにおいて 1.86%、テスト用データにおいて 1.56%であった。

機械学習を用いて 21 項目(年齢、性別、19 項目の採血項目)を用いて院内死亡を予測した際の C 統計量は、ロジスティック回帰、ランダムフォレスト(図 2)、ニューラルネットワークにおいてそれぞれ 0.913 [95% CI, 0.908-0.919]、0.922[95% CI, 0.917-0.927]、0.916 [95% CI, 0.911-0.921]であった。

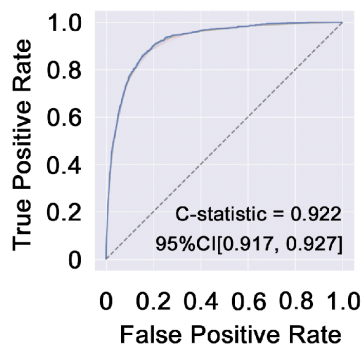


図2 ランダムフォレストによる院内死亡予測

テストデータで検証時のROC曲線及びC統計量を図示した。細い線は各交差検証モデルの結果、青色の太線が統合された平均の結果を示す。

C統計量において最も優れていたランダムフォレストモデルを用い、ジニ不純度の減少量の比較を利用してそれぞれの説明変数の重要度の順位づけを行い、交差検証時の結果全てにおける逆順位和を用いて各説明変数の重要度を可視化した結果、院内死亡の予測に最も重要と評価された説明変数はLDH、最も重要度が低いと評価された項目は性別であった(図3)。

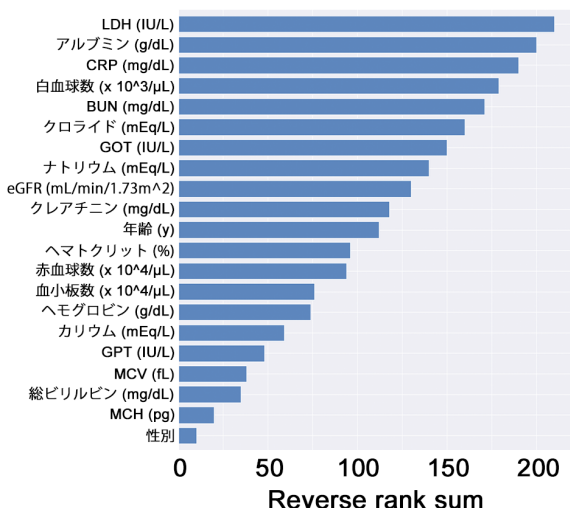


図3 各説明変数の重要度

ジニ不純度減少の比較によりそれぞれの説明変数の重要度の順位づけを行い、交差検証時の結果全てを用いて逆順位和を算出し可視化した。

また、テストデータに対してモデルが出力した院内死亡予測確率を層別化し、十分位での分割サブグループを用いた Hosmer-Lemeshow 検定による適合度の評価²¹⁾では $p = 0.00022$ (< 0.05)となり、モデルのデータ適合性が低い領域が存在することが検出された(図4左)ものの、実際の臨床利用を想定して院内死亡予測確率30%以下を層別化した範囲では、予測確率の上昇によって実際の院内死亡率も上昇傾向にあることが確認され、作成した機械学習モデルが院内死亡リスクの層別化に応用可能性を有することが示唆された(図4右)。

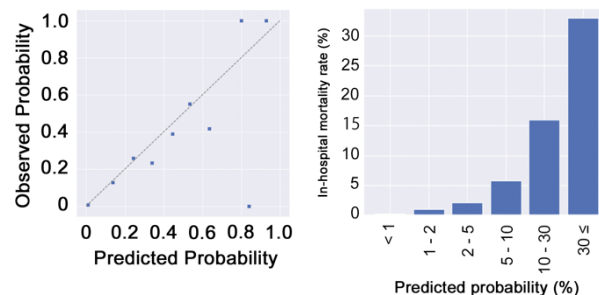


図4 院内死亡率と院内死亡予測確率の比較

テストデータによる検証時のランダムフォレストモデルの出力値を十分位で分割したサブグループを用いたキャリブレーションプロットを左図に示した。また、院内死亡予測確率30%以下を層別化しグループごとの院内死亡率を算出した(右図)。

5 考察

本研究では、東京大学医学部附属病院のSS-MIX2標準化ストレージに保存されたデータを用い、入院時の21項目(年齢、性別、19項目の採血項目)を用いて院内死亡率を高い精度で予測する機械学習モデルの作成に成功した。これまでも採血検査結果を用いたロジスティック回帰による院内死亡予測モデルは報告されている¹¹⁾²²⁾²³⁾²⁴⁾ものの、本研究で開発されたモデルは既存の報告を上回る予測性能を達成した。本研究で使用した10万件を超える十分なデータ量と非線形分類に対応可能な機械学習手法を用いたことが、高い予測性能の達成に寄与した可能性が考えられた。

しかし、本研究は現時点ではいくつかの限界を有しており、まずその一つとして特定の確率値領域における適合度が不十分である可能性が挙げられる。Hosmer-Lemeshow 検定による適合度の評価²¹⁾ではモデルのデータ適合性が低い領域が存在することが検出され、キャリブレーションプロット(図4右)では高い確率値において適合度が低い領域が存在することが示唆された。本モデルを実臨床に応用するとすれば、このモデルの利用者は高い確率値において適合度が低いことを留意しなくてはならない。このようなデータ適合性の低い領域が存在する理由については、本研究で用いたデータが極端な不均衡データであったことによる影響が考えられ、さらにテストデータが偏ったデータであった可能性と、本モデルの作成手法ではそもそも高い確率値領域が不安定となる可能性が挙げられる。一方で、実際の患者の入院時リスク評価にあたっては、図4の右図で示したような層別化範囲での院内死亡予測が実用に則していると考えられ、少なくともその範囲では予測確率に沿った観察確率が得られていることから、本モデルは一定の有用性が期待できると考えられた。

さらに本研究の重要な限界として単施設のデータを用いた研究であることが挙げられる。本研究で得られた予測性能は同一施設内のデータに対して得られた結果であり、得られた高い予測性能が他の施設のデータに対して再現されるかどうかは今後検証が必要である。本研究は単施設の大学病院のデータを利用して予測モデリングを行なっているため、大学病院に集まる一般的ではないような重症例に当てはまる特徴に対して大きな重要度が置かれた内部的アルゴリズムで予測を行なっている可能性がある。本モデルは地域一般病院での日常診療における症例頻度と異なるデータかつ由来のモデルであることが本研究の限界として非常に重要である。また、本研究で用いたデータは救命ではなく緩和ケアなどを目的と

表1 各説明変数とアウトカム

	学習用データ (n=125, 283)		テスト用データ (n=35, 588)	
	平均値 (SD)	0	平均値 (SD)	0
性別 (男性 %)	51.82	0	51.47	0
年齢 (year, mean (SD))	61.49 (16.44)	0	61.22 (17.05)	0
入院期間 (day, mean (SD))	14.68 (23.30)	0	13.79 (17.07)	0
白血球数 (g/dL, mean (SD))	6.82 (5.85)	5.91	7.27 (6.65)	5.47
赤血球数 (x10 ⁴ /μL, mean (SD))	400.34 (73.74)	5.91	409.20 (75.07)	5.47
ヘモグロビン値 (g/dL, mean (SD))	12.12 (2.20)	5.91	12.26 (2.19)	5.47
ヘマトクリット (%, mean (SD))	37.10 (6.20)	5.91	37.67 (6.22)	5.47
平均赤血球容積 (MCV) (fL, mean (SD))	93.16 (6.80)	5.91	92.61 (7.10)	5.47
平均赤血球ヘモグロビン 量 (MCH) (pg, mean (SD))	30.41 (2.65)	5.91	30.11 (2.73)	5.47
血小板数 (x10 ⁴ /μL, mean (SD))	22.59 (9.41)	5.92	23.38 (9.62)	5.47
アルブミン (g/dL, mean (SD))	3.76 (0.58)	10.72	3.81 (0.57)	8.55
総ビリルビン (mg/dL, mean (SD))	0.85 (1.35)	13.60	0.83 (1.33)	11.95
GOT (IU/L, mean (SD))	35.88 (94.75)	8.25	35.80 (124.20)	6.99
GPT (IU/L, mean (SD))	29.58 (77.87)	7.66	29.11 (86.44)	6.72
LDH (IU/L, mean (SD))	251.30 (211.22)	13.49	248.83 (196.17)	11.53
ナトリウム (mEq/L, mean (SD))	138.97 (3.58)	10.04	139.19 (3.46)	8.64
カリウム (mEq/L, mean (SD))	4.22 (0.51)	9.79	4.22 (0.51)	8.44
クロライド (mEq/L, mean (SD))	104.17 (3.98)	15.96	103.26 (3.93)	14.02
血中尿素窒素 (BUN) (mg/dL, mean (SD))	18.33 (12.44)	7.88	17.89 (12.12)	7.02
クレアチニン (mg/dL, mean (SD))	0.99 (1.17)	7.69	1.01 (1.20)	6.76
eGFR (mL/min/1.73m ² , mean (SD))	74.28 (35.64)	7.69	73.94 (40.75)	6.78
CRP (mg/dL, mean (SD))	1.55 (4.15)	8.57	1.58 (4.20)	7.54
院内死亡 (死亡 %)	1.86	0	1.56	0

学習用データ、テスト用データのカラムの右側には欠損値補完前の欠損率(%)を記載した。連続値は平均と標準偏差(SD)を、2分類項目は一方の存在比率を記載した。

して入院した症例を含んでおり、そのような症例はそもそも重症度評価の対象外となるため、救命が必要な患者の院内死亡予測モデルを作成するという研究の学習データとすることは本来ふさわしくない。本モデルはそのような救命目的ではない入院の院内死亡の特徴を学習し分類に利用している可能性があり、本モデルを臨床で実用するとすれば、利用者は本モデルが緩和ケア目的で入院した症例を含むデータで作成されたモデルであることを留意する必要がある。

また、本研究は院内死亡の予測に際して疾患名、診断名を利用していないため、現状ではそれぞれの疾患群、診断群

の間において予測性能が変動する可能性が存在する。また、各説明変数の重要度の検証結果で重要と判定された説明変数が、それぞれの疾患群、診断群内に限定して評価した場合に本研究の結果と同様に院内死亡予測に重要性を有するかどうかを判断することは、本研究結果の範囲内からは不可能である。本研究では、モデルによる院内死亡予測に際してLDHが最も大きな重要性を有するとの結果が得られた。LDHは実際に入院患者において予後を予測するマーカーとなると報告がなされており²⁵⁾、作成されたモデルが既存の報告の結果と矛盾のない内部アルゴリズムに沿って予測を行なっていると考えられる一方で、当然ながら本研究結果はLDHが予後予測因子となるかどうかをこれまで検証されていない疾患群、診断群に対してLDHが院内死亡予測に有用であることを示唆するものではない。以上を踏まえた上で考えれば、本研究で作成された予測モデルは、既に死亡予測のモデルが確立されている特定の疾患群、診断群においては、既存の予測モデルによるリスク評価が優先されるべきである。そのため、本機械学習モデルが実用できる可能性のある対象範囲は、そのような既存の予測モデルが存在しないケースや使用できないケースにおける補助的な利用に限定されるべきであると考えられる。

6 結論

本研究により、電子診療録に記録された入院実施時点での年齢、性別、及び19項目の採血検査結果から、機械学習技術を用いて院内死亡を予測するモデルを開発し、高い予測性能を達成した。本機械学習モデルは入院時の患者の院内死亡予測率の評価に有用となる可能性が考えられた。

参考文献

- 1) Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *Bmj*. 2009;338:b375.
- 2) Vincent JL, Moreno R. Clinical review: scoring systems in the critically ill. *Crit Care*. 2010;14(2):207.
- 3) Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13(10):818-29.
- 4) Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991;100(6):1619-36.
- 5) Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med*. 1981;9(8):591-7.
- 6) Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med*. 2006;34(5):1297-310.
- 7) Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* : the journal of the American Medical Association. 1993;270(24):2957-63.
- 8) Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. A simplified acute physiology score for ICU patients. *Crit Care Med*. 1984;12(11):975-7.
- 9) Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3-From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive Care*

- Med. 2005;31(10):1345-55.
- 10) Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 1996;22(7):707-10.
 - 11) Pine M, Jones B, Lou YB. Laboratory values improve predictions of hospital mortality. *Int J Qual Health Care.* 1998;10(6):491-501.
 - 12) Deo RC. Machine Learning in Medicine. *Circulation.* 2015;132(20):1920-30.
 - 13) Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. (1533-4406 (Electronic)).
 - 14) Campbell SL. HL7 (Health Level 7)--the future becomes a reality. *Healthc Inform.* 1990;7(5):24, 6.
 - 15) Audigier V, White IR, Jolani S, Debray TPA, Quartagno M, Carpenter J, et al. Multiple Imputation for Multilevel Data with Continuous and Binary Variables. *Statistical Science.* 2018;33(2):160-83.
 - 16) Jolani S, Debray TP, Koffijberg H, van Buuren S, Moons KG. Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. *Stat Med.* 2015;34(11):1841-63.
 - 17) Cox DR. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society Series B (Methodological).* 1958;20(2):215-42.
 - 18) Breiman L. Random Forests. *Machine Learning.* 2001;45(1):5-32.
 - 19) Battiti R. First- and Second-Order Methods for Learning: Between Steepest Descent and Newton's Method. *Neural computation.* 1992;4(2):141-66.
 - 20) Chawla N, Bowyer K, O. Hall L, Philip Kegelmeyer W. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res (JAIR).* 2002;16:321-57.
 - 21) Meurer WJ, Tolles J. Logistic Regression Diagnostics: Understanding How Well a Model Predicts Outcomes. *JAMA.* 2017;317(10):1068-9.
 - 22) Froom P, Shimoni Z. Prediction of hospital mortality rates by admission laboratory tests. *Clin Chem.* 2006;52(2):325-8.
 - 23) Redfern OC, Pimentel MAF, Prytherch D, Meredith P, Clifton DA, Tarassenko L, et al. Predicting in-hospital mortality and unanticipated admissions to the intensive care unit using routinely collected blood tests and vital signs: Development and validation of a multivariable model. *Resuscitation.* 2018;133:75-81.
 - 24) Zhang Z, Goyal H, Lange T, Hong Y. Healthcare processes of laboratory tests for the prediction of mortality in the intensive care unit: a retrospective study based on electronic healthcare records in the USA. *BMJ Open.* 2019;9(6):e028101.
 - 25) Erez A, Shental O, Tchebiner JZ, Laufer-Perl M, Wasserman A, Sella T, et al. Diagnostic and prognostic value of very high serum lactate dehydrogenase in admitted medical patients. *Isr Med Assoc J.* 2014;16(7):439-43.