

分担研究報告書

3. 遠隔 ICU データを用いた集中治療重症度予測 AI モデルの開発

研究代表者 高木俊介 横浜市立大学 准教授

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 自治医科大学附属さいたま医療センター 学内講師

研究要旨

<背景>ICU 患者のバイタルサインの AI で予測することで、急変の予兆を事前に察知するアルゴリズムを構築することが本研究の目的である。

<方法> 昭和大学に導入されている遠隔 ICU システムから血圧、心拍数、血中酸素飽和度、呼吸数の各バイタルサインのデータを抽出した。データの前処理を行った後に、あと 6 時間および 3 時間後の数値を予測するモデルを作成するためにニューラルネットワークで学習を行った。

<結果>

昭和大学病院 ICU で蓄積されたデータ 2019 年 2 月～ 2021 年 2 月 の 24 ヶ月分を用いた。患者数は 5856 人、のべ患者数は 6322 人であった。6 時間後の上記 4 バイタルを予測する AI モデルの構築を行った。過去 24 時間分、または 12 時間分の各バイタル値を入力し、6 時間後、または 3 時間後の値を予測する時系列回帰モデルとしてアプローチした。性能評価は平均絶対値誤差を指標とし、その誤差は、最も予測難易度の高かった血圧において 5.5～6.0、心拍数で 3.3～5.0、呼吸数で 1.3～2.2、血中酸素飽和度で 0.6～0.8 であった。6 時間後予測と比較し、3 時間後予測の方が概して性能が高かった。入力時間長の比較ではポジティブな結果が得られ、24 時間と比較して 12 時間において、心拍数の 3 時間予測で軽度の性能低下を認めた。

<結論>遠隔 ICU システムで記録された 24 ヶ月分のデータに基づき、6 時間後および 3 時間後のバイタル値を予測する AI モデルを構築した。モデルの学習曲線からさらなる改善の可能性が示唆された一方、入力時間長は 24 時間ではなくその半分の 12 時間分でも十分な性能を目指せることが判明した。

A. 研究目的

昭和大学に導入されている日本初の遠隔 ICU システム(以下、eICU)にて用いられている豊富な重症系診療データを用いた人工知能を使った重症度予測モデルの構築をおよびその臨床応用である。令和 2 年度に、6 ヶ月分のデータを用いて、過去 24 時間分の血圧の値から未来 3 時間後の値を予測する 1 次元 CNN(Convolutional neural network)を採用した AI モデルを構築し、1 サンプルに対する予測結果とそのペアである 3 時間後の実測値との乖離が平均して 2～6mmHg と極めて良好な結果が得られた。

学習の都度のゆらぎの安定化といった課題が見られたが、データ数を増やしていくにつれバリエーションが増え安定化すると見込まれる。

令和 3 年度は、バイタルサインのスコープを広げ、2 年分のデータベースを元に、時系列モデリングで

一般に用いられる Long short-term memory、および画像認識で用いられる Convolutional neural network を時系列データに特化させた 1 次元 CNN を用いて、最終的に、6 時間後の血圧、心拍数、呼吸数、酸素飽和度を予測する AI モデルを構築する。

B. 研究方法

抽出したデータに対して前処理を行なった。前処理は欠損値除去、線形補間、および指数移動平均(係数 0.05)によるスムージングを行った。(詳細は別紙参照。)

次に、学習用と評価用のデータセットの構築について述べる。学習と評価のデータセット分割は、データ前処理を適用する前に、入退室単位(入退室の記録単位のため同一患者による重複はあり得る。)にランダムに 19 : 1 に分割し、95%を AI モ

デルの学習用、残る 5%をその評価用とした。学習に用いた患者の記録から採取したサンプルが、学習と評価の両方に含まれることを避ける意図がある。前処理は入退室単位で行い、学習用と評価用で同質の処理を適用している。

前処理を行った後に、ABP (M), HR, SpO2, RR のつの 4 バイタルサインに対して、6 時間先と 3 時間先を予測する AI モデルを作った。またその入力として過去 24 間分を用いるモデルと 12 時間分を用いるモデルを作り、1 つのバイタルサインに対し 4 つのモデルを作り性能を評価した。もし予測性能が同等であるならば、3 時間先よりも 6 時間先を予測出来たほうが望ましい。また同様に、24 時間分の入力を必要とするモデルよりも 12 時間分の入力で済むモデルのほうが、入室後半日後から未来予測が可能であるという点と、単純にモデルの軽量化つまり学習に必要なパラメータ数の削減という点から、望ましく、それぞれのモデルの性能を比較する。

準備 AI モデルにはニューラルネットワークを用い、1 次元 Convolutional Neural Networks (1dCNN) と呼ばれるアーキテクチャを採用した。CNN は一般に画像処理の分野で広く用いられるが、それを時系列データ向けに派生させたモデルが 1dCNN であり、音声認識および自然言語処理の分野で応用の実績が知られている。画像は縦と横の 2 軸の方向をもつため、画像では一般に 2 次元 CNN が適用されるが、時系列データであるバイタルサインは時間軸の 1 方向のみとなるため、1dCNN 用いる。今回用いたアーキテクチャを述べる。1dCNN はフィルターサイズ 32, 出力の特徴量 map を 64 とした。1dCNN に続きバッチ正規化、LeakyReLU 関数による活性化、プーリングサイズおよびストライドサイズ 2 とした max-pooling, drop 率 0.5 のドロップアウトの適用までを 1 つのセットとし、これを 2 セット適用し、最後に full コネクト層で予測長分の値を出力する。ニューラルネットワークの学習は一般に、loss 関数と呼ばれる指標 (loss) をタスク・目的ごとに設計し、その loss 値を最小化する最適化プロセスが '学習' に相当する。今回我々は loss 関数を $|y_{true} - y_{pred}|$ として設計し、予測対象となる y_{true} とモデルが出力した予測値 y_{pred} の差の絶対値とした。 y_{true} の値は前処理適用後のデータである。学習は、Adam と呼ばれる確率的勾配降下法を用いて、GPU デバイス (NVIDIA Quadro P5000) を用いた並列演算を 1000 エポック、すなわち学習用データセットを 1000 周させた。評価にも学習時の loss を指標として用い、loss 値で性能を報告する。

(倫理面への配慮)

後方視的な研究であり、対象者に対する健康被害

のリスクは認めない。個人情報扱うため、昭和大学における倫理委員会の承認を得て研究を行なっている。

C. 研究結果

昭和大学病院 ICU で蓄積されたデータ 2019 年 2 月～ 2021 年 2 月の 24 ヶ月分を用いた。患者数は 5856、のべ患者数は 6322 であった。

4 つのバイタルサインの予測値と実測値の誤差を下記の表で示す。

モデル(入力長-予測長)	ABP(M)	HR	SpO2	RR
24-6	6.7 - 8.2	5 - 6	1.0 - 1.3	1.8 - 1.95
24-3	5.5 - 6.0	3.3 - 5.0	0.8 - 1.6	1.3 - 2.2
12-6	7.2 - 7.7	4.8 - 6.8	0.8 - 1.0	1.8 - 2.2
12-3	5.5 - 7.0	4.0 - 6.5	0.6 - 0.8	1.4 - 2.2

D. 考察

6 時間後予測と比較し、3 時間後予測の方が概して性能が高く、これは想定のとおり結果となった。入力時間長の比較ではポジティブな結果が得られ、24 時間と比較して 12 時間において性能低下は、HR の 3 時間予測で若干見られたものの、著しい劣化は見られなかった。一方、SpO2 を除く 3 つのバイタルサインでは、評価データの loss 値の収束が弱く、後半エポックにおいても値の振れ幅が大きかった。またこれら 3 つのバイタルサインに共通して、学習データの loss が前半エポックで緩やかな上昇が見られた。原因の特定には至っていないが、loss 値がそれ以降減少せずに停滞することから過学習への関与が考えられ、モデルアーキテクチャもしくは学習方法において改善の必要性が示唆されている。予測の様子を図 10-13 に示す。4 バイタルとも 1000 エポック学習後のモデルを用い、評価データに対する予測 (橙) と正解 (青) をプロットした。HR と SpO2 で、実測値よりも全体的に低めを予測する傾向があった。これも原因の特定には至っていないが、経験的に、CNN を用いた予測においてみられる傾向であり、改善の余地を示唆している。

E. 結論

遠隔 ICU システムで記録された 24 ヶ月分のデータに基づき、6 時間後および 3 時間後のバイタル値を予測する AI モデルを構築した。モデルの学習曲線からさらなる改善の可能性が示唆された一方、入力時間長は 24 時間ではなくその半分の 12 時間分でも十分な性能を目指せることが判明した。

F. 健康危険情報

なし。

G. 研究発表

1. 論文発表

なし。

2. 学会発表

なし。

H. 知的財産権の出願・登録状況

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. Lucas M. Fleuren, Thomas L. et.al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* (2020) 46:383–400
2. Alistair E. W. Johnson, Mohammad M. et.al. Machine Learning and Decision Support in Critical Care. Vol. 104, No. 2, February 2016 | *Proceedings of the IEEE*.
3. Joon-myung Kwon, Youngnam LeeID, Yeha Lee,et.al. Validation of deep-learning-based triage and acuity score using a large national dataset. *PLOS ONE*. October 15, 2018