

令和2年度厚生労働行政推進調査事業費補助金政策科学総合研究事業
(政策科学推進研究事業)

医療・介護のデータの利活用の推進のための、NDB・介護DBの連結可能性および活用可能性の評価に関する研究

分担研究報告書

レセプトデータを用いて死亡を予測する深層学習モデルの開発

研究分担者 康永秀生 東京大学大学院医学系研究科臨床疫学・経済学教授

研究要旨

レセプトデータを用いて死亡を予測する深層学習モデル(deep learning model)を開発とバリデーションを行った。46,665,942人のDPC入院患者を対象とし、レセプト情報から取得できる情報として、年齢・性別、病名、過去の入院歴、初日の診療行為内容を予測変数として用い、アウトカム変数は在院死亡とした。その結果、モデルのarea under curve (AUC)は0.954(95%信頼区間:0.9537-0.9547)と高値を示した。急性心筋梗塞・心不全・脳卒中・肺炎の4疾患について、既存の重症度指標を用いたモデルと、今回開発した深層学習モデルを比較した場合、いずれの疾患についても後者の方がAUCはより高値であり、モデルの適合度もより良好であった。本研究結果から、レセプトに含まれる膨大な診療行為情報を網羅的に深層学習モデルに投入することにより、在院死亡を非常に精密に予測できることが分かった。こうした手法は、今後整備される医療・介護連結データにも適用可能である。さらに様々なクリニカルクエスチョンに対しても本法を適用し、患者の転帰の予測やリスク調整に応用可能であると考えられる。

研究分担者 松居宏樹 東京大学大学院医学系研究科臨床疫学・経済学助教

A. 研究目的

本研究班は全体として、「医療・介護の連結データの利活用を具体的にイメージして実現させる」ことを目的として、レセプト情報等データベース(NDB)、および要介護認定情報・介護レセプト等情報(介護DB)のデータ連結の可能性を研究するとともに、両者が仮にデータベースとして連結された際に、保健医療分野におけるどう

いった課題やテーマに対して強みを発揮できるのかを評価するものである。

分担研究者らは、特に後者の部分に貢献するために、上記のような大規模なデータベースの特性を踏まえたうえで、それに合致するテーマとして、今年度はレセプトデータを用いて死亡を予測する深層学習モデル(deep learning model)を開発とバリデーションを行ったので報告する。

レセプトデータは近年臨床研究にも利活用が進められている。しかしレセプトデータには、病名情報や診療行為

履歴情報が豊富に含まれるものの、検査値情報等の臨床的情報は少ない。疾患の重症度に関するデータが不足しているため、リスク調整が十分にできないことが多い。

これまで海外では、レセプトの病名情報を用いたリスク調整指標が開発されてきた。Charlson comorbidity index がその代表例である。また、我々の先行研究において、病名情報に医師の専門的知見に基づきいくつかの診療行為履歴情報を追加して、より精度の高いリスク調整指標を開発した(BMC Health Services Research 2015;15:261)。しかしこれら既存の指標は、少数の変数を用いたロジスティック回帰に基づくものがある。それに加え、診療領域の専門的知見に基づいてモデルに投入する変数を決定するため、専門領域をまたいだ精緻なモデルの構築は難しかった。

近年の機械学習の臨床分野への応用は、専門分野をまたぐ精緻な予後予測モデルの構築を可能とした。しかしながら、臨床研究におけるリスク層別化に予測モデルを応用するためには、モデル適合度の評価が必要なにもかかわらず、多くのモデルはそのモデル適合度が報告されていない。(PLoS One. 2020 Jun 12;15(6):e0234722, ESC Heart Fail. 2020.)。

そこで今回我々は、深層学習モデルを用いて、レセプトデータに含まれる各個人の病名および診療行為履歴情報を網羅的にモデルに投入し、より高い精度で在院死亡を予測するモデルを開発し、予測精度と適合度からバリデーションを実施した。

B. 研究方法

対象

DPC データ調査研究班の DPC レセプ

トデータにおいて 2010 年 7 月-2017 年 3 月の期間に退院した全患者データ (n=46,665,942) を用いた。

レセプト情報から取得できる情報として、年齢・性別、病名、過去の入院歴、初日の診療行為内容を予測変数として用いた。アウトカム変数は在院死亡とした。

メインモデル

対象患者はランダムに 95% (derivation cohort) と 5% (validation cohort) に分割した。validation cohort からは入院当日に退院した患者は除外した。

深層学習モデルとして、4 層の全結合中間層を有する多層ニューラルネットワークを構築した。入力層は 49297 ノード、各中間層は 1000 ノードで 20% の drop out を設定した。実装は Python 上で行い、学習には TensorFlow を用いた。

既存の重症度指標を用いたモデル

メインモデルとは別に、急性心筋梗塞・心不全・脳卒中・肺炎の 4 疾患について、既存の重症度指標を用いたモデルも作成した。

年齢・性別、過去の入院歴に加えて、DPC データの様式 1 にある下記の重症度指標をそれぞれモデルに投入した。(初日の診療行為内容はモデルに投入しなかった)

急性心筋梗塞；Killip 分類

心不全；NYHA 分類

脳卒中 (stroke)；Japan Coma Scale, Barthel Index

肺炎；A-DROP

予測精度の比較

メインモデルを急性心筋梗塞・心不全・脳卒中・肺炎の 4 疾患に適用した場合の予測精度と、各疾患の既存の重

症度指標を用いたモデルの予測精度を比較した。

予測精度の計算には、Delong の方法による Area under curve (AUC) を用いた。また、モデルの適合度 (goodness of fit) を評価するためにキャリブレーション曲線(calibration curve)を作図した。

C. 研究結果

各モデルの症例数を図 1 に示す。全体の院内死亡率は 4.3%であった。

全症例によるメインモデルの AUC は 0.954 (95% 信頼区間: 0.9537-0.9547)であった。

メインモデルと既存の重症度指標を用いたモデルの AUC の比較を表 1 に示す。4 疾患のいずれにおいても、メインモデルの方が AUC の値はより高値であった。

図 2 は心筋梗塞に関するキャリブレーション曲線を示す。メインモデルでは予測死亡と実測死亡がほぼ一致しており、モデルの適合度が極めて良好である。一方、既存の重症度指標を用いたモデルでは曲線が 45 度線の下方に偏位しており適合度はあまりよくないことがわかる。

D. 考察

DPC レセプトデータの初日の治療内容情報を用いて在院死亡を予測する深層学習モデルを構築した。AUC は 0.954 と高値であった。

レセプトデータの初日の治療内容情報を用いた深層学習モデルと、既存の疾患別重症度指標を用いたモデルを比較した場合、前者の方が在院死亡の予測精度は高かった。

臨床においてよく用いられる重症度指標よりも、初日の診療行為内容の情報に基づいて深層学習を用いる方が、

在院死亡をより精度よく予測できることが明らかとなった。

E. 結論

今回の結果から、レセプトに含まれる膨大な診療行為情報を網羅的に深層学習モデルに投入することにより、在院死亡を非常に精密に予測できることが分かった。

こうした手法は、今後整備される医療・介護連結データにも適用可能である。さらに様々なクリニカルクエスションに対しても本法を適用し、患者の転帰の予測やリスク調整に応用可能であると考えられる。

F. 健康危険情報

なし

G. 研究発表

1. 論文発表

(英文原著論文投稿中)

2. 学会発表

なし

H. 知的財産権の出願・登録状況

なし

図 1. 各モデルの症例数

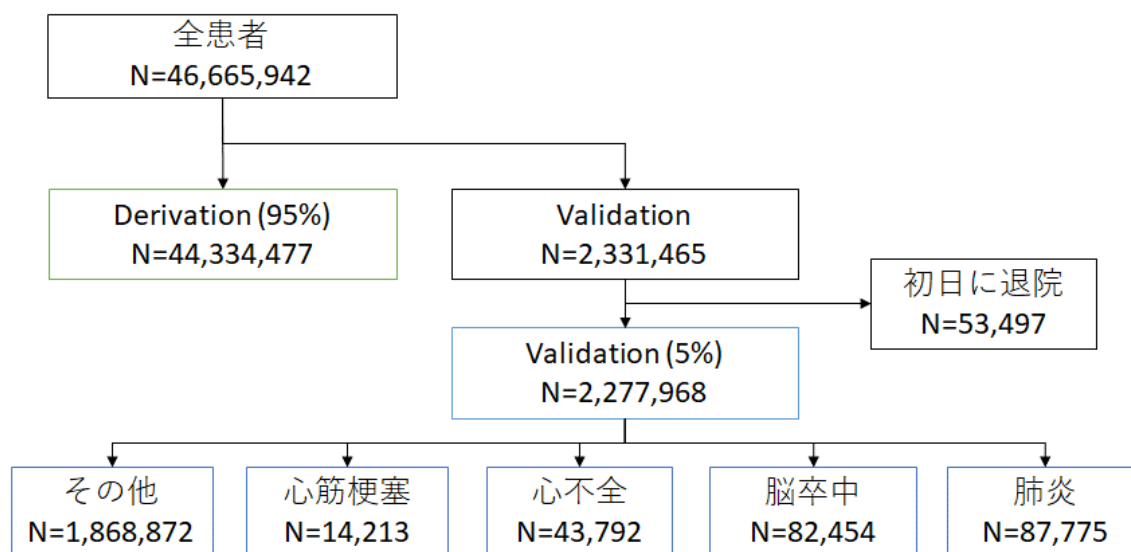


表 1. メインモデルと既存の重症度指標を用いたモデルの AUC の比較

Population	n	メインモデルの AUC (95%信頼区間)	既存の重症度指標を 用いたモデルの AUC (95%信頼区間)
AMI	14,213	0.944 (0.938-0.950)	0.876 (0.866-0.887)
Heart failure	43,792	0.831 (0.825-0.837)	0.745 (0.738-0.753)
Stroke	82,454	0.921 (0.918-0.925)	0.894 (0.890-0.898)
Pneumonia	87,775	0.918 (0.915-0.920)	0.863 (0.859-0.867)

図2. 心筋梗塞のモデルにおけるキャリブレーション曲線

X 軸: 予測死亡, Y 軸: 実測死亡.

実戦: メインモデル, 点線: 既存の重症度指標を用いたモデル

