

救急外来データベースでの機械学習による、 入院予測システムの構築

Admission predicting system using pre-arrival information in emergency departments

岩井 聡¹ 井口 竜太² 園生 智弘¹ 中島 勸¹ 矢作 直樹¹

Satoshi Iwai¹, Ryota Inokuchi¹, Tomohiro Sonoo¹, Susumu Nakajima¹ and Naoki Yahagi¹

¹ 東京大学医学部附属病院 救急部・集中治療部

¹Department of Emergency and Critical Care Medicine, The University of Tokyo Hospital

²JR 東京総合病院

²JR Tokyo General Hospital

Abstract: In emergency departments in Japan, physicians prepare for patients using information from emergency medical technicians before their arrival. But those information were not recorded on previous electronic health record system. We constructed a database to record pre-arrival information and developed experimental prediction models to assess whether the patients need hospital admission or not. The prediction accuracy of the models was superior to that of human predictions. We focused on the cases which physicians correctly predicted the outcome, and the models had almost perfect concordance. Our database and prediction model may support clinical decision making in emergency department.

まえがき

2・3次救急は、救急隊からの一本の電話から始まる。救急隊からの患者情報には、患者氏名や生年月日、救急隊を呼ぶに至った経緯（現病歴という）や血圧・脈拍・呼吸数などの生理学的データ（バイタルサインという）等が含まれており、病院の医師はこれらの情報を元に、重症度を想定し診察する人数や資材の調整を救急車到着前に行う。

しかし救急隊の情報が、カルテに記載されることはあまりなく、今までデータベースとして残ることは少なかった。救急外来での検査結果や主訴による入院・帰宅の予測を、ロジスティック回帰にて行った研究や[1]、画像検査の有無等まで含めたベイズネットワークによる解析[2]などが先行研究としてあり、救急隊の情報に基づく入院の要否についての予測の研究は、調べる範囲ではなかった。救急外来・集中治療室での死亡を予測したロジスティック回帰によるモデルは、人間のトリアージより、優れていたという研究があり[3]、救急隊情報を使った体系的なトリアージ方法を確立することで、必要な資材や人数をより適切に調整することができると考えられる。

今回、救急外来に特化したデータベースシステムを独自に作成、運用し、救急隊からの情報の蓄積を

開始した。これら救急隊からの情報を使い、機械学習モデルを用いた入院の要否についての予測と、人間の考える予測の精度を比較検討した。

本論

目的

救急医療にとどまらず、全ての医療は人間の判定による方針決定がなされている。しかし、人間の判断は、一個人の経験症例および知識を総合して決める、客観性に乏しいものである。診断支援システムを使用することによって、この判断を平均化し、医療の質の底上げを図ることができる可能性がある。

患者の救急搬送では、病院によってできる処置が異なるため、搬送先病院を適切に決定することが重要である。前述の診断支援システムは、搬送先を決定する支援をすることで、救急隊のワークロードを減らす可能性がある。

今回、救急隊からの情報をデータベースに蓄積し、これらを使用して患者の入院の要否を機械学習モデルによって判定できるか、検証した。

方法

東大救急部に Filemaker® で作成したデータベース蓄積システムを導入し、運用した。

(図1 a:救急外来システム インターフェイス)



(図1 b:患者一覧画面)



4月15日から8月15日までの4ヶ月間に蓄積された、救急隊から得た患者情報(救急隊現場到着時の患者の血圧・脈拍・呼吸数・体温・血中酸素飽和度・意識レベル・年齢・性別・主訴)および、その患者の当院での転帰(帰宅・入院・転院・死亡など)を使用した。バイタル入力に完全なデータの中で、データベース運用開始から1400件までのデータを抽出した。

入力を救急隊からの患者情報、出力を入院の要否とし、機械学習にかけた。主訴データについては、事前に設定した主訴リストの中から選択して入力するようなインターフェイスを作成し、解析する際にはそれらの中からデータベースに入力された回数の多い上位30個について、それぞれ有りを1、無しを0として入力データとした。複数の主訴が選択されていた場合は、「1/(その症例での主訴の選択個数)」を、それぞれの主訴の値として採用した。性別は男性を1、女性を-1とし、年齢、呼吸数、心拍、収縮期血圧、拡張期血圧、体温、血中酸素飽和度については実測値を、意識レベルは意識障害有りを1、無しを0として入力データとした。

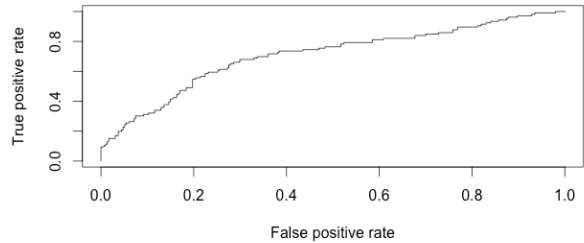
アルゴリズムには、多重ロジスティック回帰分析およびニューラル・ネットワークを使用した。

収集されたデータから1000件を学習データとし、1)残りの400件を判定させた。2)同400件のうち、医師が入院・帰宅の予測をし、正しく予測できたものを、人間が判定した場合に明らかな予測ができる症例セットと仮定して、これらに絞って判定をさせた。

結果

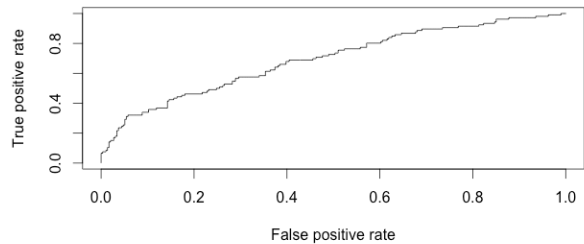
1) 多重ロジスティック回帰による帰宅可能かどうかの予測では、Precision 0.45 Recall 0.68 AUROCは0.71であった。

(図2:多重ロジスティック回帰でのROCカーブ)



ニューラルネットワークでの予測モデルでは、隠れ層は1層で800ニューロン、出力層には入院または帰宅の2ニューロンを使用した。Precision 0.63 Recall 0.31 AUROCは0.69であった。

(図3:ニューラルネットワークでのROCカーブ)



(表1:それぞれのモデルによる予測結果)

400件	入院と予測		帰宅と予測		Prec	Rec	正答率	AUROC
	TP	FP	TN	FN				
多重ロジスティック回帰	72	88	206	34	0.45	0.68	69.5	0.71
ニューラルネットワーク	33	19	275	73	0.63	0.31	77	0.69

医師2人による入院・帰宅の判断を400件について施行したところ、それぞれの医師による判定は、Precision 0.39 Recall 0.40, Precision 0.39 Recall 0.58となった。

(表2:医師による判定と、モデルとの比較)

	正解	不正解	Prec	Rec	正答率
多重ロジスティック回帰	278	122	0.45	0.68	69.5
ニューラルネットワーク	308	92	0.63	0.31	77
医師1	271	129	0.39	0.4	67.75
医師2	259	141	0.39	0.58	64.75

医師の中で正答率の高かった医師1と、機械学習の中で正答率の高かったニューラルネットワークによる結果をカイ二乗検定で比較したところ、有意にニューラルネットワークのほうが正答率が高かった(p<0.01)

2) また、医師2人の意見が一致し、なおかつそれが正しく予測できた件数は184件あった。これらに対し、先の学習データ1000件を学習した予

測モデルに判定させたところ、多重ロジスティック回帰分析によるモデルでは Precision 0.90 Recall 0.82 AUROC は 0.98 であった。また、ニューラルネットワークでの予測モデルでは、Precision 1.00 Recall 0.36 AUROC は 0.90 であった。

(表3：医師が判定し、一致かつ正しい予測をした対象のみの予測結果)

医師一致回答184件	入院と予測		帰宅と予測		Pre	Rec	正答率
	TP	FP	TN	FN			
医師	11	0	183	0			
多重ロジスティック回帰	9	1	172	2	0.9	0.82	98.37
ニューラルネットワーク	4	0	173	7	1	0.36	96.2

考察

今回、機械学習を用いて入院の要否の予測システムを作成した。機械学習による予測は、人間のそれよりも精度で上回っており、人間が正確に予測できる症例についてはほぼ正確に判定した。

救急隊からの患者情報だけで予測する本モデルでは、多重ロジスティック回帰では Precision 0.45 Recall 0.68、ニューラルネットワークでは Precision 0.63 Recall 0.31 と、あまりよい結果が出なかった。精度改善法として、情報の精度を上げる方法が考えられる。本研究では、データベース蓄積の際に主訴を選択式にした。本研究に用いたデータの中で、10回以上使用された主訴は35件あった。今回上位30件に絞って入力データとしたが、これらのうち、入院・帰宅に関与する割合の高いものに絞って予測をさせると、より精度が上がる可能性がある。

機械学習による精度は、人間の精度を上回ったが、判定した医師の習熟度が成績に寄与している可能性は否定できない。

しかし一方で、医師によって正しく予測できた184件に絞って予測させると、機械学習モデルは非常に正確に判定している。適切な予測が存在する範囲では、機械学習は医師による予測とほぼ遜色ない結果を残せると考えられる。また、本研究での情報のみでは、医師でも正確に予測できない症例があると考えられる。つまり、今回蓄積されたデータには、正確な予測に必要な情報が欠損している可能性がある。例えば、同じ疾患でも追加情報の有無によって転帰が大きく異なる状態も考えられる。(例：痙攣発作などでは、初回の痙攣発作と、怠薬による痙攣発作では対応が異なる) 今後さらに精度を上げていくためには、各主訴に対して、追加情報を入力するチェック項目を設けるなどの方法が考えられる。

今回機械学習モデルが判定に失敗した例を見ると、医師でも判定に迷うような例が多々あった。

(表4：判定に失敗した例)

	年齢	呼吸数	心拍	sBP	dBP	体温	SpO2	性別	意識障害	主訴
帰宅と予測したが実際は入院した例	52	24	102	152	100	36.2	99	男性	なし	嘔気・嘔吐
	26	18	78	136	72	36.9	100	男性	なし	発熱・しびれ
	57	18	96	112	62	37.8	98	女性	なし	腹痛
	17	18	84	118	60	37	94	男性	あり	発熱・浮遊性めまい
入院と予測したが、実際は帰宅した例	15	18	102	126	56	36.7	95	男性	あり	発熱・浮遊性めまい
	72	18	96	200	116	36.8	97	男性	なし	発熱・頭痛
	91	18	66	170	70	35.7	97	女性	あり	胸痛
	91	24	90	170	60	37.4	96	女性	あり	全身倦怠感
	71	18	60	124	62	36.3	94	男性	なし	頭部外傷・発熱・失神
	87	18	102	150	80	36.6	95	女性	なし	発熱・上肢痛

実際の救急の現場では、救急隊からの情報と、実際の患者の状況がずれていることが多々ある。このため、例えば救急隊情報から軽症だったが、来院時には重症化しており、入院が必要であった症例では、救急隊情報からの予測は正しいものの、False Negative と判定されているであろう。また逆に、救急隊情報では重症であったが、救急外来での治療によって、入院が必要でなかった症例では、False Positive と判定されているであろう。このような誤差があるため、予測的中率は完全に100%には達しないと考えられる。

しかし、同じ条件下で機械学習および医師を比較した際には、機械学習のほうが精度が優れていた。人間の感覚に頼った予測より、機械学習による予測のほうが精度が高いということになる。大量のデータを処理することができる機械学習に、適切な入力をすることによって、人間より正確な判定ができる可能性を示している。予後予測モデル作成に機械学習を使用する試みが近年増えており[4-6]、今後正確な診断支援システムを構築するためにも、必要な情報を蓄積するデータベースが求められている。

謝辞

本研究にあたって、システム導入に尽力して下さった救急部の先生方や救急外来医師による入力、運用方法の変更にご理解を頂いた看護師ならびに事務の方々には、大変感謝しております。この場をお借りして御礼申し上げます。

参考文献

- [1] Kim SW, Li JY, Hakendorf P, Teubner DJ, Ben-Tovim DI, Thompson CH: Predicting admission of patients by their presentation to the emergency department, Emerg Med Australas, 26, 361-7,(2014)
- [2] Dexheimer JW, Leegon J, Aronsky D: Predicting hospital admission at triage in an emergency department, AMIA Annu Symp Proc, 937,(2007)

- [3] Coslovsky M, Takala J, Exadaktylos AK, Martinolli L, Merz TM: A clinical prediction model to identify patients at high risk of death in the emergency department, *Intensive Care Med*, 41, 1029-36,(2015)
- [4] Ong ME, Lee Ng CH, Goh K, Liu N, Koh ZX, Shahidah N, et al.: Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score, *Crit Care*, 16, R108,(2012)
- [5] Ayaru L, Ypsilantis PP, Nanapragasam A, Choi RC, Thillanathan A, Min-Ho L, et al.: Prediction of Outcome in Acute Lower Gastrointestinal Bleeding Using Gradient Boosting, *PLoS One*, 10, e0132485,(2015)
- [6] Green M, Bjork J, Forberg J, Ekelund U, Edenbrandt L, Ohlsson M: Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room, *Artif Intell Med*, 38, 305-18,(2006)