

一般口演 | 医療データ解析

## 一般口演17 データマイニング

2021年11月20日(土) 16:30 ~ 18:00 F会場 (2号館2階224)

### [3-F-3-06] AIを用いた定量的かつ動的な転倒転落リスクの解析

\*園部 真也<sup>1,2</sup>、門脇 水樹<sup>3,4</sup>、石川 哲朗<sup>5,6</sup>、高屋 英知<sup>1,5</sup>、菅原 寛子<sup>7</sup>、中川 敦寛<sup>2,4</sup>、田畑 雅央<sup>7</sup>、植田 琢也<sup>1,5</sup>、川上 英良<sup>5,6</sup>、富永 悌二<sup>2</sup> (1. 東北大学病院 AI Lab, 2. 東北大学病院 脳神経外科, 3. University of California San Diego, 4. 東北大学病院 臨床研究推進センター バイオデザイン部門, 5. 東北大学 大学院医学系研究科 保健学専攻 生体応用技術科学領域 画像診断学分野, 6. 理化学研究所 情報統合本部 先端データサイエンスプロジェクト, 7. 東北大学病院 医療安全推進室)

\*Shinya Sonobe<sup>1,2</sup>, Mizuki Kadowaki<sup>3,4</sup>, Tetsuo Ishikawa<sup>5,6</sup>, Eichi Takaya<sup>1,5</sup>, Hiroko Sugawara<sup>7</sup>, Atsuhiko Nakagawa<sup>2,4</sup>, Masao Tabata<sup>7</sup>, Takuya Ueda<sup>1,5</sup>, Eiryu Kawakami<sup>5,6</sup>, Teiji Tominaga<sup>2</sup> (1. 東北大学病院 AI Lab, 2. 東北大学病院 脳神経外科, 3. University of California San Diego, 4. 東北大学病院 臨床研究推進センター バイオデザイン部門, 5. 東北大学 大学院医学系研究科 保健学専攻 生体応用技術科学領域 画像診断学分野, 6. 理化学研究所 情報統合本部 先端データサイエンスプロジェクト, 7. 東北大学病院 医療安全推進室)

キーワード : fall risk, incident report, artificial intelligence, unsupervised learning

【目的】 入院中における転倒転落のリスク因子として、年齢、低血圧、認知機能障害の有無、神経症状の有無、排尿機能低下の有無、転倒転落歴の有無、低栄養、特定の薬物の内服、喫煙、心疾患が報告されており、転倒転落リスクの評価に用いられている。しかしこれらの情報は日々の変動が少ないため、どのような症例で転倒転落リスクが高いかを示すには向いているが、どのような状況で転倒転落リスクが高いかを示すには向いていない。我々はこの度、より定量的かつ動的なリスク因子を抽出するために、AIを用いて解析した。

【方法】 インシデントレポートで転倒転落が報告された肺癌症例188例を転倒群とし、インシデントレポートで転倒転落が報告されていない肺癌症例228例を非転倒群とした。脈拍数、体温、白血球数、血小板数、ヘモグロビン値、総蛋白量値、LDH値、CRP値、D-dimer値の値をもとに、クラスタリングの手法を用いて転倒群を分類した。得られたサブグループにおいて、25パーセンタイル値ないし75パーセンタイル値が母集団の平均値を超えて偏倚している特徴を抽出し、これらの特徴を持つ症例が転倒群と非転倒群にどれだけ存在するか比較した。

【結果】 A、B、C、D、E、の5つのサブグループが検出された。AとBは際立った特徴を持たない、Cは白血球数と血小板数が高い、Dは総蛋白量とヘモグロビン値が低くD-dimer値が高い、Eは総蛋白量とヘモグロビン値が低く脈拍数が高い、という結果であった。それぞれの特徴を持つ症例の割合は、Cが転倒群：37.8%で非転倒群：22.7% ( $p<0.01$ )、Dが転倒群：22.9%で非転倒群：4.7% ( $p<0.01$ )、Eが転倒群：28.2%で非転倒群：14.2% ( $p<0.01$ ) であった。

【考察】 AIは複数の因子からなる複合要因の解析に強く、過去に報告の無い複合的なリスク因子を検出できる。

# AIを用いた定量的かつ動的な転倒転落リスクの解析

園部 真也<sup>\*1,2</sup>、門脇 水樹<sup>\*3,4</sup>、石川 哲朗<sup>\*5,6</sup>、高屋 英知<sup>\*1,5</sup>、菅原 寛子<sup>\*7</sup>、  
中川 敦寛<sup>\*2,4</sup>、田畑 雅央<sup>\*7</sup>、植田 琢也<sup>\*1,5</sup>、川上 英良<sup>\*5,6</sup>、富永 悌二<sup>\*2</sup>

\*1 東北大学病院 AI Lab、\*2 東北大学病院 脳神経外科、

\*3 University of California San Diego、\*4 東北大学病院 臨床研究推進センター バイオデザイン部門、

\*5 東北大学 大学院医学系研究科 保健学専攻 生体応用技術科学領域 画像診断学分野、

\*6 理化学研究所 情報統合本部 先端データサイエンスプロジェクト、\*7 東北大学病院 医療安全推進室

## AI analysis for quantitative and dynamic fall risk

Shinya Sonobe<sup>\*1,2</sup>, Mizuki Kadowaki<sup>\*3,4</sup>, Tetsuo Ishikawa<sup>\*5,6</sup>, Eichi Takaya<sup>\*1,5</sup>, Hiroko Sugawara<sup>\*7</sup>  
Atsuhiko Nakagawa<sup>\*2,4</sup>, Masao Tabata<sup>\*7</sup>, Takuya Ueda<sup>\*1,5</sup>, Eiryu Kawakami<sup>\*5,6</sup>, Teiji Tominaga<sup>\*2</sup>

\*1 AI Lab, Tohoku University Hospital, \*2 Department of Neurosurgery, Tohoku University Hospital,

\*3 University of California San Diego, \*4 CRIETO, Tohoku University Hospital

\*5 Department of Clinical Imaging, Tohoku University Graduate School of Medicine,

\*6 Advanced Data Science Project, Information R&D and Strategy Headquarters, RIKEN,

\*7 Patient Safety Management Office, Tohoku University Hospital

**Background and Purpose:** Various fall risk factors have been reported. Since most of the information is static, it is suitable for assessing which cases have a high risk of falls, but not for assessing what situations have a high risk of falls. We searched for dynamic risk factors by machine learning analysis.

**Materials and Methods:** We analyzed 188 lung cancer cases with falls and 211 lung cancer cases without falls. The fall group was clustered using the pulse rate, body temperature, white blood cell count, platelet count, hemoglobin value, total protein amount value, LDH value, CRP value, and D-dimer value. The proportion of cases with the characteristics seen in these subgroups was compared between the fall group and the non-fall group.

**Results:** Five subgroups were detected. Two had no specific features, one had high white blood cell and platelet counts, one had low total protein and low hemoglobin and high D-dimer levels, and one had low total protein and low hemoglobin levels and high pulse rate. There was a significant difference in the proportion of cases with these characteristics between the fall group and the non-fall group.

**Discussion:** Complex risk factors that have not been reported will be detected using machine learning.

**Keywords:** fall risk, incident report, artificial intelligence, unsupervised learning

## 1. 緒論

入院中における転倒転落のリスク因子として、年齢、低血圧、認知機能障害の有無、神経症状の有無、排尿機能低下の有無、転倒転落歴の有無、低栄養、特定の薬物の内服、喫煙、心疾患が報告されており、転倒転落リスクの評価に用いられている。しかし、これらの情報は日々の変動が少ないため、どのような症例で転倒転落リスクが高いかを示すのには向いているが、どのような状況で転倒転落リスクが高いかを示すのには向いていない。またこれらの情報のうち大部分は定量性が低いという課題がある。かねてより統計学的手法を用いて定量的で動的なリスク因子の探索が行われてきたが、因子がそれぞれ単独で有する影響力は高くないため、統計学的有意差を示すまでには至っていない。

## 2. 目的

転倒転落における定量的かつ動的なリスク因子を、機械学習を用いた解析により探索した。

## 3. 方法

2011年04月01日から2021年03月31日までの期間に肺癌の診断で東北大学病院呼吸器内科へ入院し、インシデントレポートで転倒転落が報告された全症例を転倒群として登録した。2020年04月01日から2021年03月31日までの期間に肺癌の診断で東北大学病院呼吸器内科へ入院し転倒転落歴の無い全症例を非転倒群として登録した。

## 3.1 クラスタリング

転倒群と非転倒群においてそれぞれ後方視的に変数を収集した。変数は、脈拍数、体温、白血球数、血小板数、ヘモグロビン値、総蛋白量値、LDH値、CRP値、D-dimer値を採用した。なお転倒群においては転倒から遡り最初に記録のある時点の値を各項目とも採用し、非転倒群においては入院期間中で記録のある時点からランダムに時点を選択して同時点の値を各項目とも採用した。これらの値をもとに、凝集的階層的クラスタリングの手法を用いて転倒群を分類した。クラスタリングのサブグループ数はコンセンサスクラスタリングの手法を用いて決定した。

## 3.2 特徴抽出

得られたサブグループごとに、各サブグループに特徴的な計測値の偏りを抽出した。具体的には、非転倒群における変数ごとの平均値および標準偏差を用いて転倒群の症例における偏差値を算出し、各グループにおける25パーセンタイル値が50を上回らないし75パーセンタイル値が50を下回る変数を特徴と定義した。

## 3.3 群間比較

抽出された特徴を部分的に有する症例が転倒群と非転倒群にどれだけの割合で存在するかを比較し、有意水準を5%として統計学的有意差の有無を判定した。

## 4. 結果

転倒群は 188 例、非転倒群は 211 例であった。

### 4.1 クラスタリング

A、B、C、D、E、の 5 サブグループに分類された (図 1)。

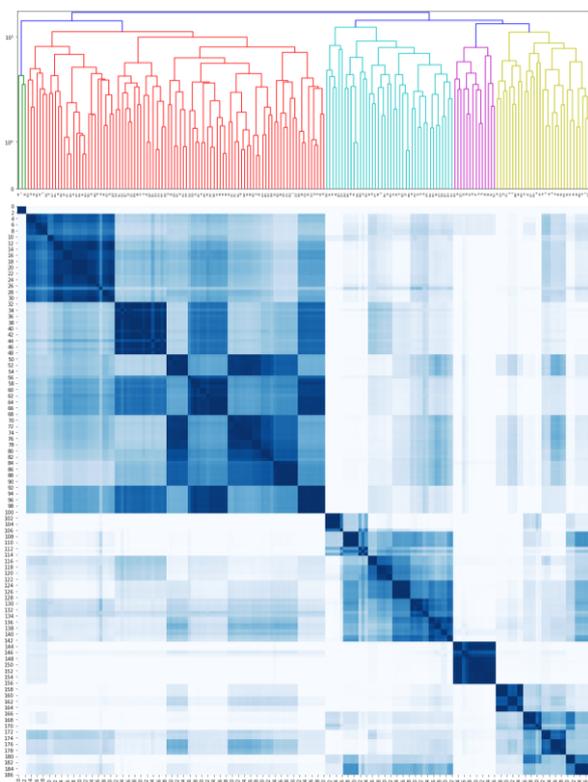


図 1 凝集的階層的クラスタリング

脈拍数、体温、白血球数、血小板数、ヘモグロビン値、総蛋白量値、LDH 値、CRP 値、D-dimer 値の値から凝集的階層的クラスタリングの手法で転倒群を分類した。

### 4.2 特徴抽出

いずれの群においても収縮期血圧の 75 パーセンタイル値が 50 を下回っており、これは特徴量として抽出しなかった。

A 群は、高い心拍数、高い総蛋白量、高い白血球数を特徴として認めた。B 群は、低い CRP 値を特徴として認めた。C 群は、高い血小板数と、高い白血球数と、高い体温と、低い総蛋白量値と、低いヘモグロビン値と、高い LDH 値を特徴として認めた。D 群は、低い総蛋白量値と、低いヘモグロビン値と、高い D-dimer 値と、高い体温を特徴として認めた。E 群は、高い脈拍数と、低い総蛋白量値と、低いヘモグロビン値と、高い CRP 値を特徴として認めた (図 2)。

### 4.3 群間比較

抽出された特徴をもとに、臨床の見地から、経験的に C 群と D 群と E 群が何らかの特徴的な状態を示している可能性が高いと判断した。これらの三群の特徴を参考に、状態Ⅰ：血小板数が高く白血球数が高い、状態Ⅱ：総蛋白量値が低くヘモグロビン値が低く D-dimer 値が高い、状態Ⅲ：心拍数が高く総蛋白量が低くヘモグロビン値が低い、の三つの状態を定義した。

状態Ⅰの症例は、転倒群では 37.8%に認められ、非転倒群では 22.7%に認められ、有意 ( $p < 0.01$ ) な群間差を認めた。状態Ⅱの症例は、転倒群では 22.9%に認められ、非転倒群では 4.7%に認められ、有意 ( $p < 0.01$ ) な群間差を認めた。状態Ⅲの症例は、転倒群では 28.2%に認められ、非転倒群では 14.2%に認められ、有意 ( $p < 0.01$ ) な群間差を認めた。

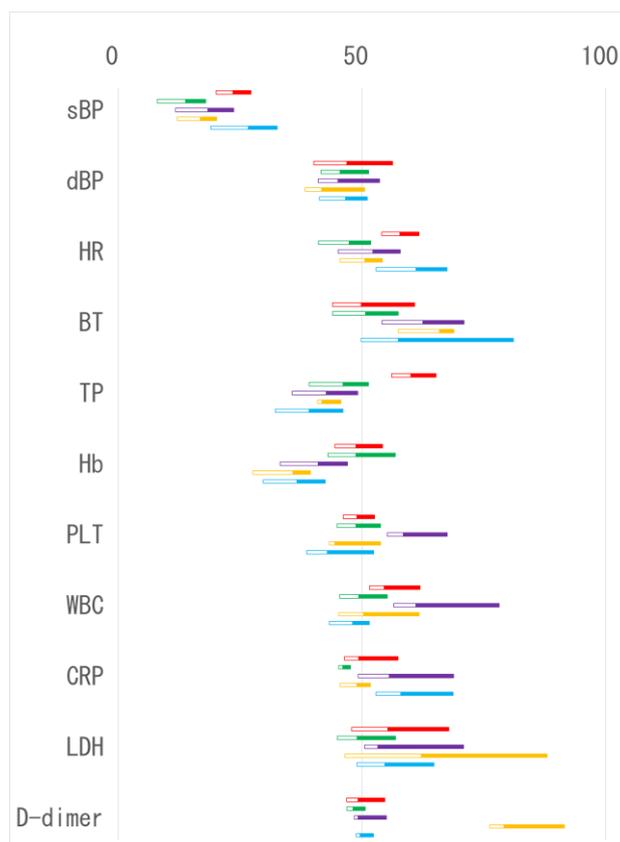


図 2 特徴抽出: サブグループの計測値分布

各サブグループにおける計測値の偏りを示す。

赤は A 群、緑は B 群、紫は C 群、黄は D 群、青は E 群。棒グラフの中空部分は 25 パーセンタイル値から中央値を、塗り潰された部分は中央値から 75 パーセンタイル値である。

## 5. 考察

機械学習によるクラスタリングで抽出された、複数の因子の複合による特徴は、統計学的有意差をもって転倒転落と関連していた。これらの特徴を用いて定量的活動的に転倒転落のリスク評価を行える可能性があることが示された。

これらの特徴は、多変量解析をはじめとする従来の統計学的手法では抽出できていない。その原因として、単一の因子それぞれは強い影響力を持たないこと、統計学的手法では複数の因子による複合的な特徴を捉えにくいこと、転倒転落の原因は偶発的な要因も多分に含まれるため例え強力な予測因子としての特徴があるとしても確実に結果を判定できるには限らないこと、などが挙げられる。本研究で用いた手法は、教師なし機械学習の一種であり<sup>1)</sup>、アルゴリズムの性質上、これらの課題に対応し得る側面を有する。

本研究では、サンプル数の少なさを考慮し、特徴として採用する変数を数種類に制限した。サンプル数を増やして解析することで、さらなる特徴の抽出につながると期待される。

## 6. 結論

機械学習を用いた解析は、過去に報告されていない複合的なリスク因子を検出できる可能性がある。

## 参考文献

- 1) Stefano Monti, Pablo Tamayo, Jill Mesirov, Todd Golub. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine Learning 2003, 52 : 91-118.