

ポスター

## ポスター9

### 医療データ解析・評価

2017年11月22日(水) 13:30 ~ 14:30 L会場（ポスター会場1）（12F ホワイエ）

## [3-L-3-PP9-3] 血液検査データに対する判別分析を用いた予測因子候補の抽出

永田 夏海<sup>1</sup>, 松井 藤五郎<sup>1,2</sup>, 平手 裕市<sup>1</sup> (1.中部大学 生命健康科学部 臨床工学科, 2.中部大学 工学部 情報工学科)

### 1.はじめに

本研究では、医学的知識を用いずに血液検査データの全検査項目を対象として判別分析を行うことによって、医学的研究における予測因子の候補を抽出する方法について検討する。

### 2.方法

本研究は名古屋掖済会病院倫理審査委員会の承認を得ている（No. 2016-030）。

本研究では、アウトカムの例として死亡を対象に、死亡直近の検査結果を1、そうでない結果を0として従属変数とした。独立変数は、Mann-WhitneyのU検定にて帰無仮説が棄却された41個の検査項目を用いた。ステップワイズ法を用いてさらに変数を選択し、選択された変数を用いて判別関数を求めた。

### 3.結果

選択された41個の検査項目を独立変数の候補としてステップワイズ法を行った結果、pH, AGAP (Anion GAP), LD/AST (Lactate Dehydrogenase / Aspartate Aminotransferase), 総蛋白, Lymphocyte, Eosinophilの6項目を独立変数として用いた判別関数が得られた。

交差確認の結果、得られた判別関数の精度は90.1%と高いものであった。

### 4.考察

ステップワイズ法によって選択された6項目の中で、2番目のステップにおいてpHとの組み合わせでAGAPが採用されたことは、pH低値となるアシドーシスの鑑別診断でAGAPが有用となる臨床的意義に一致しており、興味深いものである。

### 5.結論

本論文では、医学的研究における予測因子の候補を抽出するために、血液検査データに対して判別分析を行う方法について検討した。アウトカムの例として死亡を対象に、検査結果を死亡直近の結果とそれ以外の結果に分け、Mann-WhitneyのU検定により帰無仮説が棄却された検査項目を独立変数の候補とし、ステップワイズ法を用いた判別分析によって独立変数を選択した。

その結果、死亡直近を90.1%の精度で判別できる判別関数を得ることができた。この判別関数には6つの検査項目が独立変数として用いられ、これらの中には臨床的意義に一致したものも含まれていた。このことから、医学的知識を用いずにデータマイニングを行うことによって、医学的研究における予測因子の候補を抽出できることが示唆された。

# 血液検査データに対する判別分析を用いた予測因子候補の抽出

永田夏海<sup>\*1</sup>、松井藤五郎<sup>\*1\*2</sup>、  
平手裕市<sup>\*1</sup>

\*1 中部大学 生命健康科学部 臨床工学科、

\*2 中部大学 工学部 情報工学科

## Extracting Predictor Candidates Using Discriminant Analysis of Blood Test Results

Natsumi Nagata<sup>\*1</sup>, Tohgoroh Matsui<sup>\*1\*2</sup>, Yuichi Hirate<sup>\*1</sup>

\*1 Department of Clinical Engineering, College of Life and Health Sciences, Chubu University,

\*2 Department of Computer Science, College of Engineering, Chubu University

In this paper, we propose a method using discriminant analysis of blood test results in order to extract candidates of predictive factors in medical research. As an example of outcome, in this paper, we divided the test results into two groups: results closest to death and results that have been tested before those. We used Mann-Whitney's U test and selected tests whose null hypotheses were rejected as a candidate for independent variables. We then carried out stepwise discriminant analysis. As a result, we could obtain a discriminant function that can discriminate near death with accuracy of 90.1%. The positive predictive value was 85.2% and the sensitivity as 62.5%. Six test items that includes pH, AGAP, LD/AST, total protein, Lymphocyte, and Eosinophil were used as independent variables in this discriminant function, and some of them coincided with clinical significance. We suggest that candidates of predictors in medical research can be extracted by performing data mining without using medical knowledge.

**Keywords:** Extracting Predictor Candidates, Pattern Recognition, Discriminant Analysis, Blood Test

### 1. 緒論

一般的に、医学的研究に取り組む際、例えば「術前に栄養状態の悪かった症例は術後合併症がおこりやすい」というように医学的な知識に基づいて仮説が立てられている。これまで、血液検査データに対して判別分析を行うことによって臓器がん診断の支援を試みる研究[1]や健康状態の把握と将来予測を行う研究[2]などが行われているが、これらの研究では、医学的知識に基づいて検査項目を選択している。

現在、多くの病院が電子カルテを導入している。その中に蓄積されている検査データや診療データといった膨大なデータを分析することによって、医学的な因果関係を予測できると期待されるが、現状多くの病院ではこれらのデータを活用できていない。

そこで本研究では、電子カルテに保存された血液検査データに焦点を当て、医学的研究へ活用するために血液検査の結果を分類する方法を開発する。

### 2. 目的

本論文では、医学的知識を用いずに血液検査データの全検査項目を対象として判別分析を行うことによって、医学的研究における予測因子の候補を抽出する方法について検討する。

### 3. 方法

#### 3.1 血液検査データ

本研究では、名古屋掖済会病院にて2016年1月1日から2016年4月18日までに死亡した患者359名の、死亡日の1年前から死亡日までに行われた血液検査の結果を対象とした。患者の氏名は匿名化され検査値ごとのレコードとして保存されており、全部で260,965件ある。本研究における死

亡日直近の判断基準は、最も死亡日に近い検査結果を死亡直近とした。

#### 3.2 データのクリーニング

各患者について検査日ごとに検査項目名が同じ検査値を一つの属性としてまとめ、検査値ごとのレコードを患者および検査日ごとのレコードとして4,408件にまとめた。検査項目は621個あった。このとき、観測できなかったことを示す --- や /、不等号記号を含む数値を空欄に置き換えた。また、欠損率の高い項目を削除した。

#### 3.3 Mann-Whitney の U 検定

死亡直近の検査結果を1、そうでない結果を0とした変数を用意し、これを従属変数(目的変数)とした。621個の検査項目から欠損率が90%以上の項目、性別のような名義尺度の項目、乳びや溶血のような定性的な項目を除外し、独立変数(説明変数)の候補とした。ノンパラメトリック検定の一つである Mann-Whitney の U 検定を用いて検定を行い、有意水準0.1%で帰無仮説が棄却された検査項目を独立変数として選択した。

#### 3.4 判別分析

最後に、死亡直近であるグループとそうでないグループの事前確率をそれぞれのグループサイズから計算し、判別に有用な独立変数一つずつ採用するステップワイズ法を用いて判別分析を行い、その結果について考察した。

### 4. 結果

クリーニングの結果、621個あった検査項目のうち77個が Mann-Whitney の U 検定の対象となった。U 検定により有意水準0.1%で帰無仮説が棄却された41個の検査項目を独立変数の候補としてステップワイズ法を行った結果、pH、AGAP

(Anion gap), LD/AST (Lactate Dehydrogenase/Aspartate Aminotransferase), 総蛋白, Lymphocyte, Eosinophil の 6 項目を独立変数として用いた判別関数が得られた。標準化された判別関数係数を表 1 に、標準化されていない判別関数係数を表 2 に示す。グループの重心は、死亡直近のグループが 0.486, そうでないグループが -1.920 であった。

交差確認の結果を表 3 に示す。得られた判別関数の分類精度は 90.1%, 陽性的中率は 85.2%, 感度は 62.5% であった。

全 4,409 件のレコードのうち判別関数に用いられた 6 つの項目が一つも欠損していないレコードは 587 件あり, そのうち死亡直近のケースが 467 件, 死亡直近でないケースが 120 件であった。

得られた判別関数の標準化された係数のうち pH と AGAP の係数の絶対値が大きいことから, この 2 つの項目を用いて散布図を作成した (図 1)。× が死亡直近のグループ, ○ が死亡直近でないグループを表している。図中の直線は, pH と AGAP を独立変数として求めた判別関数を表している。

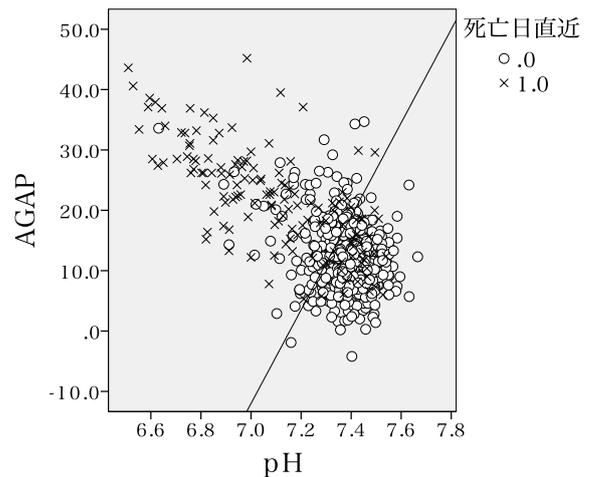


図 1 pH と BE の散布図

表 1 標準化された判別関数係数

	関数
	1
Lymphocyte	-.212
Eosinophil	.149
LD/AST	.170
総蛋白	.259
AGAP	-.384
pH	.618

表 2 標準化されていない判別関数係数

	関数
	1
Lymphocyte	-.163
Eosinophil	.720
LD/AST	.028
総蛋白	.251
AGAP	-.070
pH	4.267
(定数)	-31.918

表 3 交差確認の結果

		死亡直近	予測グループ番号		合計
			.0	1.0	
元のデータ	度数	.0	455	12	467
		1.0	44	76	120
	%	.0	97.4	2.6	100.0
		1.0	36.7	63.3	100.0
交差確認済み	度数	.0	454	13	467
		1.0	45	75	120
	%	.0	97.2	2.8	100.0
		1.0	37.5	62.5	100.0

## 5. 考察

得られた判別関数は, 感度が 62.5% とやや低いものの, 陽性的中率は 85.2% で十分に高いものであった。

ステップワイズ法による変数選択では, pH, AGAP, 総蛋白, LD/AST, Lymphocyte, Eosinophil の 6 項目が選択されたが, 最初のステップでは pH の他に BE (Base Excess), AGAP, K, HCO<sub>3</sub>, ABG K の F 値が 100 より大きく, 重要度が高かった。このことより, これらは単独では生命予後予測に有用な因子であると考えられる。この中で, 2 番目のステップにおいて pH との組み合わせで AGAP が採用されたことは, pH 低値となるアシドーシスの鑑別診断で AGAP が有用となる臨床的意義に一致しており, 興味深いものである。

## 6. 結論

本研究では, 医学的研究における予測因子の候補を抽出するために, 血液検査データに対して判別分析を行う方法について検討した。アウトカムの例として死亡を対象に, 検査結果を死亡直近の結果とそれ以外の結果に分け, Mann-Whitney の U 検定により帰無仮説が棄却された検査項目を独立変数の候補とし, ステップワイズ法を用いた判別分析によって独立変数を選択した。

その結果, 感度 62.5%, 陽性的中率 85.2%, 分類精度 90.1% で判別できる判別関数を得ることができた。この判別関数には, pH, AGAP, LD/AST, 総蛋白, Lymphocyte, Eosinophil の 6 つの検査項目が独立変数として用いられ, これらの中には臨床的意義に一致したものも含まれていた。このことから, 医学的知識を用いずにデータマイニングを行うことによって, 医学的研究における予測因子の候補を抽出できることが示唆された。

本論文では, 死亡直近をアウトカムとして用いたが, 疾患の発症や再発等をアウトカムとしても同じように分析することができると考えられる。ただし, 事例を十分に用意することが課題となるだろう。

## 参考文献

- 1) 田中達規 他: データマイニングを利用した血液検査項目からの膵がん診断支援の検討, 信学技報, 111(234):13-17 (2011).
- 2) 高橋為生: 人間ドックの一次予防および二次予防のための判定・教育支援システムの開発に関する研究, 健康医学, 18(1):49-56 (2003).