

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 8 日現在

機関番号：15501

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500206

研究課題名(和文) ブール代数的パターン認識によるゲノムデータと臨床データを統合した肝癌再発予測

研究課題名(英文) A study on the prediction of the recurrence of hepatocellular carcinoma (HCC) by a novel Boolean classifier with use of the combination of genomic data and clinical data

研究代表者

浜本 義彦 (HAMAMOTO, YOSHIHIKO)

山口大学・医学(系)研究科(研究院)・教授

研究者番号：90198820

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：本研究では血液検査により肝癌の再発予測システムの開発を目的とする。このため、ゲノムデータと臨床データを統合して2値パターンとし、2値パターンを識別する新しいブール識別器を開発した。実験では肝癌を手術で完全に切除した患者を訓練サンプル40とテストサンプル33として用いた。特徴には、ゲノムデータとしてSPINTとSRDというメチル化遺伝子を、また臨床データとして肝癌の代表的な腫瘍マーカーであるAFPとPIVKAを用いた。ブール識別器を用いて、テストサンプルの血液検査で感度64%、特異度65%、Youden-index0.29という高精度で肝癌の再発を予測できたことから、臨床応用の見通しを得た。

研究成果の概要(英文)：The purpose of this study is to develop a novel blood test system for predicting the recurrence of hepatocellular carcinoma (HCC) associated with hepatitis C virus (HCV) infection. In order to develop the test system, we established the combination of genomic data and clinical data, and the design of a novel Boolean classifier for binary patterns. We used 40 training samples and 33 test samples, who underwent complete surgical resection of HCC. Regarding features used for a classifier, 2 methylated genes (SPINT2 and SRD5A2), traditional tumor maker AFP and PIVKA were used. We developed the blood test system with use of the Boolean classifier for predicting the recurrence of HCC, which achieved high abilities (sensitivity of 64%, specificity of 65% and Youden-index of 0.29) for 33 test samples. This suggests that the proposed method might lead to clinical applications.

研究分野：情報工学

キーワード：癌 パターン認識 診断

1. 研究開始当初の背景

肝癌の難治性は、その再発の高さにある。たとえ手術で全ての癌を摘出して1年後には30%近くの再発が生じている。癌再発の診断において手術後の癌はCT検査や超音波検査では見えず、診断が困難となっている。しかし、この見えない癌が再発するか否かは、患者には重大な関心事である。

これまで取り組んできた肝癌の再発予測では、まず癌の組織から得られた遺伝子発現データを用いて再発の有無を予測した。このアプローチでは93%の高精度で再発を予測することができたが、何らかの手段で患者から癌の組織を抽出しなければならないという欠点がある。これは患者にとって大きな負担である。そこで簡易な手段として、患者からの血液検査で癌の再発を予測しようと考えた。

では血液検査による肝癌の再発予測の現状を調べてみると、肝癌の代表的検査としてAFPとPIVKAがあるが、特異度はある一定のレベルがあるものの、感度が低く、癌の見逃しがあり、十分な機能を果たしているとはいえない。血液検査による肝癌の診断では、再発予測よりも難易度が低いものの、臨床データであるAFPとPIVKAとゲノムデータであるメチル化遺伝子のデータを組合すことにより高精度化に成功している。これは、癌の多様性を考えると、単一のデータではなく、ゲノムデータと臨床データの統合による予測法の有効性を示唆している。この統合という立場からの開発促進は、平成24年度科学技術重要施策アクションプランでも指摘されている。

2. 研究の目的

本研究では、ゲノムデータと臨床データを統合して、手術等で癌を取り除いた患者を対象にした肝癌の再発予測システムの開発を目的とする。

仮に再発しないと正しく予測されると、つらい抗癌剤治療を受ける必要がなくなり、また被爆の心配のあるCT検査も不要となり、更に医療費の高騰を抑制することができる。一方、再発と予測されると先制医療を効果的に行うことができる。いずれにしても再発の予測を簡易な手段で、しかも高精度に行うことができれば、患者はこの上ない恩恵を受けることができる。

3. 研究の方法

ゲノムデータと臨床データの統合に際して、臨床データは検査会社によって、正常の範囲が異なるという問題がある。重要なのは、臨床データの値そのものではなく、そのデータが正常か否かの2値である。臨床データの

2値化により時間を遡って過去に蓄積されたデータや、検査会社の垣根を越えたデータを統一的に取り扱うことができ、更にゲノムデータとの統合も容易となる。このことから、ブール代数に基づいて2値パターンを対象としたパターン認識理論を構築する。

再発の予測問題を定式化すると、患者からのデータは2値で、それらを成分する2値パターンベクトル(以下、2値パターン)で患者を記述し、特徴選択、識別、誤予測の分析のアルゴリズムを確立し、再発の有無を高精度に予測するシステム(図1参照)を構築する。

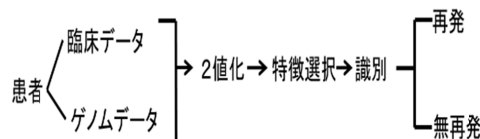


図1 癌の臨床データとゲノムデータの統合による再発予測

(1) 特徴選択

いま n 個のデータの組があり、それぞれが0と1のいずれかの2値を取るものとした、 n 次元の2値パターンが与えられたとする。訓練サンプル上で、相異なるクラスに属するパターンの全ての組合せに対してそれぞれのパターン間の距離(一致すれば0、そうでなければ1)を求め、これの総和を特徴評価関数とする。このとき、特徴評価関数の値が大きいデータを特徴として選択する。

(2) ブール識別器の設計

まず、選択された特徴上で再発クラスの2値パターンと無再発クラスの2値パターンで表された訓練サンプルを、0と1の組合せで分類する。同一分類内で多数を占める訓練サンプルのクラスを、その分類のクラスとする。ブール代数に基づきクラス c_1 を識別する識別関数 f_1 の真理値表を作成する。この真理値表から主加法標準形により、再発であれば1、そうでなければ0とする再発の識別関数 f_1 を、また同様に無再発の識別関数 f_2 を定める。

ここでブール代数によれば、識別関数は簡略化され、変数を削減することができる。これは、一種の特徴選択であり、また再発の有無をいくつかの検査データの組合せで定義できる。この組合せは一つの診断基準であり、これが統計的にデータから導出されたことは帰納的推論の立場から興味あることである。

一般に、臨床データ、ゲノムデータをあわせると n の値が数十となり、本アプローチは網羅的な統計的推論により医学の分野で今まで気づかなかった新しい診断基準を見出す可能性を期待させる。

4. 研究成果

提案手法の有効性を評価するため、山口大学医学部から提供された再発の患者35サンプルと無再発の患者38サンプルを用いて実験を行った。そのうち訓練サンプルとして再発20症例、無再発20症例をランダムに抽出し、残りをテストサンプルとして再発は15症例、無再発は18症例を用いた。特徴には、ゲノムデータとしてSPINTとSRDというメチル化遺伝子を、また臨床データとして肝臓癌の代表的な腫瘍マーカーであるAFPとPIVKAを用いた。

手術で全ての肝臓を完全に切除するため、CT検査やエコー検査では術後の肝臓を捉えることはできない。この診えない癌を血液検査だけで再発を予測することは極めて難しい。参考のため、上記のマーカーAFPとPIVKAを単独で用いた予測について検討した。臨床ではAFP、PIVKAのCut off値をそれぞれ20、40と設定して肝臓の再発の有無を予測する。評価指標に用いる Youden-Index は、感度+特異度-1で定義され、この値が高い程優れていると評価される。上記の再発35症例と無再発38症例に対し、AFP、PIVKAのYouden-Indexの値はそれぞれ0.07、0.02となり、既存の腫瘍マーカー単独では十分な予測ができないことが分かる。

ゲノムデータと臨床データを組合せた2値パターンに対する識別器の識別性能を比較した。比較対象の識別器には2値パターンを対象とするBayes識別器とChow識別器を用いた。識別性能の指標には、識別率、感度、特異度、及びYouden-Indexを採用した。

比較結果を表1に示す。いずれの指標においても提案手法であるブール識別器が優れ、肝臓再発予測において本研究の臨床応用への見通しを得た。

表1 識別器の比較

識別器	ブール識別器	Chow識別器	Bayes識別器
識別率	0.64	0.58	0.48
感度	0.64	0.55	0.46
特異度	0.65	0.60	0.49
Youden-Index	0.29	0.15	-0.05

最後に、AFP、PIVKA、SPINT、SRDの4次元2値パターン(4-dimensional binary pattern) $x = [x_1, x_2, x_3, x_4]$ に対する再発の識別関数 f_1 は、

$$f_1 = \bar{x}_1 \bar{x}_2 \bar{x}_3 x_4 + \bar{x}_1 x_2 \bar{x}_3 \bar{x}_4 + x_1 \bar{x}_2 x_3 + x_1 \bar{x}_2 \bar{x}_4 + \bar{x}_2 x_3 \bar{x}_4$$

となった。ここで $x_1 = \text{AFP}$ 、 $x_2 = \text{PIVKA}$ 、 $x_3 = \text{SPINT}$ 、 $x_4 = \text{SRD}$ である。一方、無再発の識別関数 f_2 は

$$f_2 = \bar{x}_1 \bar{x}_2 \bar{x}_3 \bar{x}_4 + \bar{x}_1 \bar{x}_2 x_3 x_4 + \bar{x}_1 x_2 x_3 \bar{x}_4 + x_1 \bar{x}_3 x_4$$

となった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 1件)

Y. Tokumitsu, T. Tamesa, S. Matsukuma, N. Hashimoto, Y. Maeda, Y. Tokuhisa, K. Sakamoto, T. Ueno, S. Hazama, H. Ogihara, Y. Fujita, Y. Hamamoto, M. Oka and N. Iizuka, An accurate prognostic staging system for hepatocellular carcinoma patients after curative hepatectomy International Journal of Oncology, 査読有 DOI 10.3892/ijo.2014.2798, pp.944-952, 2015.

[学会発表](計 4件)

H. Ogihara, Y. Fujita, N. Iizuka, M. Oka and Y. Hamamoto, Comparative study of classifiers for prediction of recurrence of liver cancer using binary patterns, 27th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, Kaohsiung (Taiwan), 4 June 2014.

浜本義彦、統計的パターン認識による肝臓再発予測、第28回信号処理シンポジウム、海峡メッセ下関(山口県下関市)、2013年11月21日。

H. Ogihara, Y. Fujita, N. Iizuka, M. Oka and Y. Hamamoto, Classification based on Boolean algebra and its application to the prediction of recurrence of liver cancer, Workshop on Recent Advances in Computer Vision and Pattern Recognition, ロワジュールホテル沖縄(沖縄県沖縄市), 5 November, 2013.

荻原宏是、那須大気、藤田悠介、飯塚徳男、岡正朗、浜本義彦、ブール代数に基づく識別とその肝臓診断への応用、情報処理学会第75回全国大会、東北大学(宮城県仙台市)、2013年3月6日。

[図書](計 0件)

[産業財産権]

出願状況(計 1件)

名称: 情報処理装置と情報処理プログラム並びに情報処理方法

発明者: 浜本義彦 飯塚徳男

権利者: 国立大学法人 山口大学

種類: 特許

番号: 特願 2014-147690

出願年月日: 平成 26 年 7 月 18 日

国内外の別: 国内

取得状況(計 1件)

名称：情報処理装置と情報処理プログラム並びに情報処理方法
発明者：浜本義彦 飯塚徳男
権利者：国立大学法人 山口大学
種類：特許
番号：特許第 5669162 号
出願年月日：平成 26 年 7 月 18 日
取得年月日：平成 26 年 12 月 26 日
国内外の別：国内

〔その他〕

ホームページ等

<http://www.ir.csse.yamaguchi-u.ac.jp/>

6. 研究組織

(1) 研究代表者

浜本 義彦 (HAMAMOTO YOSHIHIKO)
山口大学・大学院医学系研究科・教授
研究者番号：90198820

(2) 研究分担者

飯塚 徳男 (IIZUKA NORIO)
山口大学・医学部付属病院・准教授
研究者番号：80332807

藤田 悠介 (FUJITA YUSUKE)
山口大学・大学院医学系研究科・准教授
研究者番号：40509527

(3) 連携研究者

なし