

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7157941号  
(P7157941)

(45)発行日 令和4年10月21日(2022.10.21)

(24)登録日 令和4年10月13日(2022.10.13)

(51)国際特許分類 F I  
 G 0 1 N 33/49 (2006.01) G 0 1 N 33/49 Y  
 G 1 6 H 50/20 (2018.01) G 1 6 H 50/20

請求項の数 10 (全17頁)

(21)出願番号	特願2020-133346(P2020-133346)	(73)特許権者	520295041 佐藤 憲一 宮城県仙台市泉区旭丘堤1丁目17-2
(22)出願日	令和2年8月5日(2020.8.5)		
(65)公開番号	特開2022-29824(P2022-29824A)	(73)特許権者	520295672 野地 満 東京都世田谷区上馬2丁目20-9
(43)公開日	令和4年2月18日(2022.2.18)	(74)代理人	100120581 弁理士 市原 政喜
審査請求日	令和3年2月6日(2021.2.6)	(72)発明者	佐藤 憲一 宮城県仙台市泉区旭丘堤1丁目17-2
		(72)発明者	及川 香代 宮城県仙台市泉区紫山1丁目8-2
		(72)発明者	野地 満

最終頁に続く

(54)【発明の名称】 癌罹患判定方法、装置、およびプログラム

(57)【特許請求の範囲】

【請求項1】

血液検査情報を含む健康診断情報に基づいて癌罹患の可能性を判定する癌罹患判定方法であって、

複数の個人の既知の健康診断情報のうちHbA1c(ヘモグロビンA1c)、GTP(グルタミルトランスペプチダーゼ)、Alb(アルブミン)、TC(総コレステロール)、TG(中性脂肪;トリグリセリド)、GLU(グルコース)、CRP(C反応性蛋白)およびBMI(ボディマス指数)と、当該個人の癌罹患に関する癌罹患情報とに基づきコンピュータ解析を行って、判別式の係数および判定閾値を取得する取得ステップと、

対象個人の健康診断情報から、前記取得した係数による判別式および判定閾値を用いて該対象個人の癌罹患の可能性を判定する判定ステップとを備えることを特徴とする癌罹患判定方法。

【請求項2】

前記血液検査情報は、血液の所定のインデックスを含むことを特徴とする請求項1に記載の癌罹患判定方法。

【請求項3】

前記健康診断情報は、心拍数(HR)および脈拍数(PR)の少なくとも1つを含むことを特徴とする請求項1または2に記載の癌罹患判定方法。

【請求項4】

前記コンピュータ解析は、パターン認識手法による解析であることを特徴とする請求項 1 ないし 3 のいずれかに記載の癌罹患判定方法。

【請求項 5】

前記癌罹患情報は、罹患した癌のステージ情報を含み、  
前記取得ステップは、前記ステージ情報について前記係数および前記判定閾値を取得し、  
前記判定ステップは、前記対象個人の癌罹患の可能性を前記ステージ情報について判定することを特徴とする請求項 1 ないし 4 のいずれかに記載の癌罹患判定方法。

【請求項 6】

血液検査情報を含む健康診断情報に基づいて癌罹患の可能性を判定する癌罹患判定方法であって、

複数の個人の既知の健康診断情報と、当該個人の癌罹患に関する癌罹患情報とに基づきコンピュータ解析を行って、判別式の係数および判定閾値を取得する取得ステップと、  
対象個人の健康診断情報から、前記取得した係数による判別式および判定閾値を用いて該対象個人の癌罹患の可能性を判定する判定ステップと、

前記癌罹患の可能性について判定された対象個人の前記健康診断情報のうち、癌以外の疾病に関連する健康診断情報を用いて前記癌以外の疾病の可能性を排除して、該対象個人の癌罹患の可能性をさらに判定する別疾患鑑別ステップと

を備えることを特徴とする癌罹患判定方法。

【請求項 7】

前記癌以外の疾病は、糖尿病を含むことを特徴とする請求項 6 に記載の癌罹患判定方法。

【請求項 8】

前記癌罹患情報は、罹患した癌の癌種情報を含み、  
前記取得ステップは、前記癌種情報について前記係数および前記判定閾値を取得し、  
前記判定ステップは、該対象個人の癌罹患の可能性を前記癌種について判定することを特徴とする請求項 1 ないし 7 のいずれかに記載の癌罹患判定方法。

【請求項 9】

血液検査情報を含む健康診断情報に基づいて癌罹患の可能性を判定する癌罹患判定方法をコンピュータに実行させるプログラムであって、該癌罹患判定方法は、

複数の個人の既知の健康診断情報のうち H b A 1 c (ヘモグロビン A 1 c)、G T P (グルタミルトランスぺプチダーゼ)、A 1 b (アルブミン)、T C (総コレステロール)、T G (中性脂肪；トリグリセリド)、G L U (グルコース)、C R P (C 反応性蛋白) および B M I (ボディマス指数) と、当該個人の癌罹患に関する癌罹患情報とに基づきコンピュータ解析を行って、判別式の係数および判定閾値を取得する取得ステップと、

対象個人の健康診断情報から、前記取得した係数による判別式および判定閾値を用いて該対象個人の癌罹患の可能性を判定する判定ステップと

を備えることを特徴とするプログラム。

【請求項 10】

血液検査情報を含む健康診断情報に基づいて癌罹患の可能性を判定する癌罹患判定装置であって、

複数の個人の既知の健康診断情報のうち H b A 1 c (ヘモグロビン A 1 c)、G T P (グルタミルトランスぺプチダーゼ)、A 1 b (アルブミン)、T C (総コレステロール)、T G (中性脂肪；トリグリセリド)、G L U (グルコース)、C R P (C 反応性蛋白) および B M I (ボディマス指数) と、当該個人の癌罹患に関する癌罹患情報とに基づきコンピュータ解析を行って、判別式の係数および判定閾値を取得する取得手段と、

対象個人の健康診断情報から、前記取得した係数による判別式および判定閾値を用いて該対象個人の癌罹患の可能性を判定する判定手段と

を備えることを特徴とする癌罹患判定装置。

【発明の詳細な説明】

【技術分野】

10

20

30

40

50

## 【 0 0 0 1 】

本発明は癌罹患判定方法、装置、およびプログラムに関し、より具体的には、一般的に入手容易な指標を用いて癌罹患の可能性を判定することが可能な癌罹患判定方法、装置、およびプログラムに関する。

## 【背景技術】

## 【 0 0 0 2 】

癌治療は早期発見および早期治療が最も治癒、生存率を高めるとされる。したがって、早期に精度が高い癌診断を行うことが望まれている。高精度という観点からすると、検体を直接採取するような侵襲性の高い検査を行うことが确实ではあるが、早期発見を目的として初期の癌を発見するために侵襲性の高い検査を採用することは、対象者の身体に過度な負担を与えることとなり実用的ではない。そこで、早期発見には、血液・血清検査を主とする検査パラメータのように一般的な健康診断等の検査で入手可能な指標を用いることが有効と考えられている。

10

## 【 0 0 0 3 】

一方、癌罹患の判定には「腫瘍マーカー」という指標が臨床的には使用されているが、上述のような侵襲性の高い検査とは異なり、あくまで癌特異的ではない分子・物質が、正常部位で産生されるよりも産生量が多いというように癌細胞の存在を間接的にとらえるものであって、単一のマーカーで精度の高い診断を行うことは困難である。また、癌に特化した指標のため採取にあたっては信頼性の課題もある。信頼性が保証された検査指標としては、やはり、対象者が網羅的で、大量に長年の蓄積があり、通常健康診断等で使用されている一般的な血液検査データを用いることができれば有効である。血液検査は、自動測定器が完備され、標準化も進んでおり、その精度も保証されて、臨床現場で日常的に使用されてもいるので、安価、容易に使用することができる。通常血液検査においては、各血液指標ごとに対象者個人の健康状態、身体癌罹患可能性などが提示されているが、複数の指標を所定のアルゴリズムでデータ解析することにより、例えば甲状腺機能異常を予測、診断支援する実効性のある有効な提案もなされている。

20

## 【 0 0 0 4 】

また、血液中のアミノ酸の濃度のうち胃癌の状態と関連するアミノ酸の濃度を利用して胃癌の状態を評価することができる胃癌の評価方法、ならびに胃癌評価装置を提供することを目的として、評価対象から採取した血液から、アミノ酸の濃度値に関するアミノ酸濃度データを測定し、測定した評価対象のアミノ酸濃度データに含まれる A s n , C y s , H i s , M e t , O r n , P h e , T r p , P r o , L y s , L e u , G l u , A r g , A l a , T h r , T y r のうち少なくとも1つの濃度値に基づいて、評価対象につき胃癌の状態を評価する技術が提案されている。

30

## 【先行技術文献】

## 【特許文献】

## 【 0 0 0 5 】

【文献】特許第5976987号

## 【非特許文献】

## 【 0 0 0 6 】

【文献】佐藤憲一ほか著、「人間ドック受診時の基本的検査データを用いた甲状腺機能異常の診断支援 - クラウドスクリーニングサービスの展開と有用性 - 」、人間ドック34巻第5号(2020年3月発行)

40

## 【発明の概要】

## 【発明が解決しようとする課題】

## 【 0 0 0 7 】

しかし、従来、特許文献1に開示されている指標式で胃癌発症の有無の評価を行って一定の精度が得られていると考えられるが、アミノ酸測定自体が一般的ではなく安価、容易に実現することができないという問題がある。また、甲状腺機能の異常といった特定の疾病の判定は関連する特定の検査項目に絞ったモデルの導出が可能だが、癌罹患はそのよ

50

うに特定の検査項目に絞ることができないという問題がある。

【0008】

このため本発明では、健康診断等の信頼性のある検査データを利用することにより、癌の状態を精度良く評価することができる罹患判定方法、装置、およびプログラムを提供することを目的とする。

【課題を解決するための手段】

【0009】

請求項1に記載の発明は、血液検査情報を含む健康診断情報に基づいて癌罹患の可能性を判定する癌罹患判定方法であって、複数の個人の既知の健康診断情報のうちHbA1c（ヘモグロビンA1c）、GTP（グルタミルトランスペプチダーゼ）、Alb（アルブミン）、TC（総コレステロール）、TG（中性脂肪；トリグリセリド）、GLU（グルコース）、CRP（C反応性蛋白）およびBMI（ボディマス指数）と、個人の癌罹患に関する癌罹患情報とに基づきコンピュータ解析を行って、判別式の係数および判定閾値を取得する取得ステップと、対象個人の健康診断情報から、取得した係数による判別式および判定閾値を用いて対象個人の癌罹患の可能性を判定する判定ステップとを備えることを特徴とする。

10

【0010】

請求項2に記載の発明は、請求項1に記載の癌罹患判定方法において、血液検査情報は、血液の所定のインデックスを含むことを特徴とする。

【0011】

請求項3に記載の発明は、請求項1または2に記載の癌罹患判定方法において、健康診断情報は、心拍数（HR）および脈拍（PR）の少なくとも1つを含むことを特徴とする。

20

【0012】

請求項4に記載の発明は、請求項1ないし3のいずれかに記載の癌罹患判定方法において、コンピュータ解析は、パターン認識手法による解析または多変量解析であることを特徴とする。

【0013】

請求項5に記載の発明は、請求項1ないし4のいずれかに記載の癌罹患判定方法において、癌罹患情報は、罹患した癌のステージ情報を含み、取得ステップは、ステージ情報について係数および判定閾値を取得し、判定ステップは、対象個人の癌罹患の可能性をステージ情報について判定することを特徴とする。

30

【0014】

請求項6に記載の発明は、血液検査情報を含む健康診断情報に基づいて癌罹患の可能性を判定する癌罹患判定方法であって、複数の個人の既知の健康診断情報と、個人の癌罹患に関する癌罹患情報とに基づきコンピュータ解析を行って、判別式の係数および判定閾値を取得する取得ステップと、対象個人の健康診断情報から、取得した係数による判別式および判定閾値を用いて対象個人の癌罹患の可能性を判定する判定ステップと、癌罹患の可能性について判定された対象個人の健康診断情報のうち、癌以外の疾病に関連する健康診断情報を用いて癌以外の疾病の可能性を排除して、対象個人の癌罹患の可能性をさらに判定する別疾患鑑別ステップをさらに備えることを特徴とする。

40

【0016】

請求項7に記載の発明は、請求項6に記載の癌罹患判定方法において、癌以外の疾病は、糖尿病を含むことを特徴とする。

【0017】

請求項8に記載の発明は、請求項1ないし7のいずれかに記載の癌罹患判定方法において、癌罹患情報は、罹患した癌の癌種情報を含み、取得ステップは、癌種情報について係数および前記判定閾値を取得し、判定ステップは、対象個人の癌罹患の可能性を癌種について判定することを特徴とする。

【0018】

請求項9に記載の発明は、血液検査情報を含む健康診断情報に基づいて癌罹患の可能性

50



図である。

【図 17】本発明の一実施形態の一実施例において胃癌患者群と食道癌患者群とを区別する手法を説明するための図である。

【発明を実施するための形態】

【0022】

以下、本発明の罹患判定方法、装置、およびプログラムについて図面を参照して実施形態を説明する。なお、異なる図面でも、同一の処理、構成を示すときは同一の符号を用いる。

(本願発明の原理)

本願発明では、胃癌の早期発見を目的とし、侵襲性の高い検査や、特定の腫瘍マーカーを用いることなく、一般的な健康診断で取得することが可能な、血液検査を中心とする検査データを利用して、より精度の高い癌罹患判定を実現することを目的とする。このような一般的な指標を複数組み合わせることで罹患状況や身体の状態を予測、診断する方法としては、例えば甲状腺機能異常を予測、診断する技術が知られている。この技術によると、甲状腺異常により生じるホルモン産生の過剰または不足による影響を、ホルモン量を直接測定することに替えて、血液検査指標で代用すべく開発されたものであり、総コレステロール、クレアチニン、アルカリフォスファターゼ等を組み合わせることでチロキシン量を推定し甲状腺機能異常を判定するものである。具体的には、実際に健康診断等で取得された総コレステロール、クレアチニン、アルカリフォスファターゼ等の値と、既知の甲状腺機能の状況との相関関係を種々のコンピュータ解析にかけ、判定アルゴリズムを生成して、対象個人のデータに判定アルゴリズムを適用し、異常の有無を判定する。

【0023】

しかし一般に、同様の手法を癌診断に適用することは困難である。すなわち、癌は基本的に遺伝子変異や染色体の傷に伴う細胞の異常の積み重ねが長年月かけて変化をきたしたものであるため、急性期の疾病とは異なり慢性疾患の要素が主たるものであり、また癌罹患状態は単一の腫瘍ではなく、種々の異なる遺伝子変異が混合した状態の集合体を有するものであるため、一般的な指標を複数組み合わせることで罹患状況や身体の状態を予測、診断するのは困難である。以上の通り、例えば、甲状腺機能異常へのアプローチと同様なアプローチを癌の罹患判定に適用して高精度の判定結果を得ようとしても、同様の精度を実現するのは困難であるが、一方で、癌の確定的診断の前に早期に疑わしい対象者を見出すというスクリーニングの観点では、偽陽性の混在については一定程度存在するとしても、偽陰性をできる限り排除することができれば、十分実用とすることができる。

【0024】

本発明者は、甲状腺機能異常で用いた指標よりも多くの指標を採用し、さらに様々なコンピュータ解析手法を取り入れることにより、一般的な健康診断等で取得できる指標によっても、これを組み合わせることにより胃癌の罹患可能性を偽陽性、偽陰性を含みながらかなりの確度で判定することが可能であるとの知見を得た。図4を参照して、本願発明の原理を説明する。図4を参照すると、一般健康診断データを取得して必要な前処理を行い(S401)、前処理済みのデータに基づいて複数のパターン認識手法を用いることにより一次スクリーニングの予測スコアを算出する(S402)。この段階でも、一定以上の偽陽性、偽陰性を示しており、実用化の可能性があるが、さらにここで胃癌の罹患の可能性の高い対象者についてさらに鑑別スコアを算出し(S403)、最終的に胃癌の罹患状態を評価する(S404)ことにより良好な評価結果を得ることができる。

【0025】

以上のような原理により胃癌の罹患可能性を判定することができることを本発明者は見出したが、本原理は胃癌以外にも様々な部位の癌に適用することができることも確認された。したがって、以下の本実施形態においては胃癌に関し、本願発明を説明するが、本願発明は胃癌に限定されることなく、本技術分野で知られたように解析手法を調整することによって、様々な癌種に適用することができる。なお、本実施形態のスクリーニングを行っても、特定の癌の罹患であることを最終判断するためには、更なる検査や専門医の確認が必

10

20

30

40

50

要なことは言うまでもない。

(システム構成)

図1は、本発明の一実施形態の全体のシステム構成図である。本実施形態では、図1に示すように、本システムの各種処理を実行する情報処理装置である癌罹患判定サーバ101には、本システムで使用する利用者情報、健康診断情報などを含むマスターデータ群102および判別解析データなどを格納する癌罹患判定データベース103が接続されている。また、癌罹患判定サーバ201は、各種クライアント端末111、121とネットワーク104を介して接続し、各種クライアント端末からの要求等により、各種データベースにアクセスして癌罹患判定を実行する。

【0026】

本実施形態では、図3の機能ブロック図に示すように1つの癌罹患判定サーバ101で、各種処理、例えばデータ管理モジュール301によるデータ管理処理、判別式取得302による判別式取得処理、他疾病管理モジュール303による他疾病管理処理、癌罹患判定モジュール304による癌罹患判定処理などを実行するが、これに限られず、さらに多くの処理を実行することができる。ここで、少なくとも以上4つの処理は癌罹患判定サーバ101で実行するように記載したが、これとは別に複数のサーバを用意して以上の各種処理の機能を分担させることもできる。

【0027】

また、マスターデータ群102には、ユーザの各グループを管理するためのグループマスターおよびユーザマスター、健康診断情報を管理するためのマスターなどを含むが、これに限られず各種のデータをマスターとして整理して格納することができる。

【0028】

クライアント端末111、121は、例えば、各種ユーザがシステム管理、マスター管理等の管理業務に使用したり、癌罹患判定に使用したりするが、ネットワーク104を介して癌罹患判定サーバ101と通信できるものであればパソコンに限られず、いずれかの端末装置を使用することができ、タブレット等の移動端末を使用することもできる。また、例えばタブレット端末とする場合、本実施形態でネットワーク104との接続は、例えば携帯電話の回線や、Wi-Fi、BLUETOOTH(登録商標)等の無線ネットワークにより行うことができる。

【0029】

本実施形態では、以上の構成により、癌罹患判定処理を実行するが、具体的には、予め複数の個人の健康診断情報と、その個人の罹患状況を示す情報を取得し、各種のコンピュータ解析を実行して判別式と判別閾値を取得する。対象個人の健康診断情報に判別式を適用することにより判別値を算出し、判別閾値と比較することにより、罹患の有無あるいは罹患の可能性を判定する。以下、各機能の処理を説明するが、各データの保守、管理処理は以下の説明で各データの構成を理解できれば本技術分野で知られたいずれかのデータの処理手法を使用して実行できるので、詳述しない。

【0030】

図3は、本実施形態のシステムの機能ブロック図であり、図5は、本実施形態の癌罹患判定処理の一例を示すフローチャートである。以下に詳述するように、癌罹患判定サーバ101の判別式取得モジュール302および癌罹患判定モジュール304は、図5に示すフローチャートに沿って健康診断データの入力処理を行い、コンピュータ解析を行って判定モデルを構築し、モデルに沿って対象個人から採取されたデータに判別式を適用して判別値を算出、処理は終了する(ステップ501~506)。

【0031】

(本実施形態で使用する指標および解析手法)

本実施形態では、対象者の負担にならずに取得が可能な一般的な健康診断の血液検査による手法などを使用することができ、主にRBC(赤血球)、WBC(白血球)、Ht(ヘマトクリット)、Hb(ヘモグロビン)、PLT(血小板)、HbA1c(ヘモグロビンA1c)、ALP(アルカリフォスファターゼ)、ALT(アラニンアミノトランスフ

10

20

30

40

50

エラーゼ)、AST(アスパラギン酸アミノトランスフェラーゼ)、GTP(グルタミルトランスペプチダーゼ)、LDH(乳酸脱水素酵素)、AMY(アミラーゼ)、TP(総蛋白)、A/G(アルブミン/グロブリン比)、Alb(アルブミン)、BUN(尿素窒素)、Cr(クレアチニン)、Bil(ビリルビン)、TC(総コレステロール)、HDL(HDLコレステロール)、TG(中性脂肪;トリグリセリド)、LDL(LDLコレステロール)、UA(尿酸)、GLU(グルコース)、RF(リュウマチ因子)、CRP(C反応性蛋白)、SG(尿比重)、BMI(ボディマス指数)を使用することができる。以上の検査にさらにHR(心拍数)またはPR(脈拍数)、MONO(単球)、LYMPH(リンパ球)、NEUT(好中球)、EOSINO(好酸球)、BASO(好塩基球)、Na(ナトリウム)、K(カリウム)、Ca(カルシウム)、P(リン)を加えることができるが、これに限れずにより少ない、あるいはより多い本技術分野で知られた指標を用いることができる。

10

## 【0032】

本実施形態では以上の検査指標を用い、解析のアルゴリズムとしては、教師無し学習として自己組織化マップ(SOM)手法、教師あり学習として、ベイズ正則型ニューラルネットワーク(BRNN)手法やサポートベクトルマシン(SVM)、その他利用可能と考えられるパターン認識手法を用いて判別式を取得することができる。

## 【0033】

## [第1実施例]

図6および7に示す複数の個人に対する健康診断データを用いて、種々の解析手法を用い判別式を構築する。

20

## 【0034】

図8は、本発明の一実施形態の胃癌群と健常対照群の健康診断データを示す図である。すなわち、分類によるRBCの一元配置分析801、分類によるHtの一元配置分析802、分類によるPLTの一元配置分析803、分類によるWBCの一元配置分析804、分類によるHbの一元配置分析805および分類によるHbA1cの一元配置分析806を示す。図8に示すように、胃癌群と非胃癌群(健常対照)のサンプルデータを採用して、パターン認識手法(自己組織化マップ、ベイズ正則型ニューラルネットワーク、サポートベクトルマシン)、多変量解析、統計ソフトSPSS、JMPなどを用いて予測モデルを構築した。使用した検査項目は、RBC(赤血球)、WBC(白血球)、Ht(ヘマトクリット)、Hb(ヘモグロビン)、PLT(血小板)、HbA1c(ヘモグロビンA1c)、ALP(アルカリフォスファターゼ)、ALT(アラニンアミノトランスフェラーゼ)、AST(アスパラギン酸アミノトランスフェラーゼ)、GTP(グルタミルトランスペプチダーゼ)、LDH(乳酸脱水素酵素)、AMY(アミラーゼ)、TP(総蛋白)、A/G(アルブミン/グロブリン比)、Alb(アルブミン)、BUN(尿素窒素)、Cr(クレアチニン)、Bil(ビリルビン)、TC(総コレステロール)、HDL(HDLコレステロール)、TG(中性脂肪;トリグリセリド)、LDL(LDLコレステロール)、UA(尿酸)、GLU(グルコース)、RF(リュウマチ因子)、CRP(C反応性蛋白)、SG(尿比重)、BMI(ボディマス指数)、HR(心拍数)またはPR(脈拍数)、MONO(単球)、LYMPH(リンパ球)、NEUT(好中球)、EOSINO(好酸球)、BASO(好塩基球)、Na(ナトリウム)、K(カリウム)、Ca(カルシウム)、P(リン)であり、教師なし学習アルゴリズムのSOM手法や、教師ありの学習アルゴリズムのBRNN手法などで解析した結果、感度、特異度とも優れた予測モデルを構築することができた。最終の予測モデルは、それらの複数の予測モデルを併用して最適化を図り構築した。

30

40

## 【0035】

具体的には、教師ありの学習アルゴリズムであるBRNN手法により、図6および7に示すような実際の健康診断のデータをパターン認識手法により解析してモデルを生成し、生成されたモデルにより指標の組み合わせに対応する胃癌と判断される確率を算出することにより、ターゲットの実際の診断結果を入力することにより予測率を出力することがで

50

きる。ここで、ベイズの先見的結合荷重確率が結合荷重の正則化項と自然に対応しており、自動関与度決定 (ARD) により判定に大きく影響した一連の入力変数を評価することができる。また、例えば、最適なカットオフ値を設定して胃癌罹患の可能性の有無を判定することもできる。

#### 【0036】

計算条件を変化させて得られた最大の感度および特異度はそれぞれ、85.7%および98.2%であった。図9は、BRNN予測計算のROC曲線を示す。図9のLOO感度902に示すようにLOOでは、カットオフ値=0.47でAUC=0.948、同じくCV感度901に示すようにCVでは、カットオフ値=0.5でAUC=1.0であった。同様に、最適な計算条件の下でSOMを用いた投射位置の自動判定の良好な予測モデルも構築できた。変数に欠損のある場合は、後者の予測モデルしか対応できない。最終の予測モデルはそれらの複数の予測モデルを併用し最適化を図って構築した。

10

#### 【0037】

本実施例で採用したモデルについてさらに癌の進行度との関係で検討すると、より進行したステージの胃癌ほどの中率は高い。図10は、胃癌のステージごとの予想率の分布を示す図である。図10に示すように、ステージの進行と予測率とがよく相関していることが理解できる。また、IA、IBといった早期癌でも6~7割が予測可能である。

#### 【0038】

さらに、図示しないが、特に分化型管状腺癌 (tub) はやや精度が高く、未分化型の低分化腺癌 (por)、印環細胞癌 (sig)、粘液癌 (muc) もほぼ50%は予測可能である。そこで、上述の学習データとして胃癌群をさらに罹患した癌のステージごとに分け、ステージ情報を含めて学習させることにより、癌のステージも加味したモデルを生成する。上述の通りステージにより予測率は変動する可能性はあるが、このようなモデルを用いることによりターゲットとなる患者が癌に罹患している可能性に加え、どのようなステージにいるか (すなわち、胃癌のステージ情報) も判定することが可能となる。

20

#### 【0039】

##### [第2実施例]

本実施例では、上述の第1実施例と同様の学習データを使用し、1次スクリーニングとして先ずSOM手法によりクラスター分類を行うことにより、胃癌群の2~3のサブクラスターに分離する。すなわち、SOM手法により得られた学習モデルに、ターゲットの健康診断データを投影して胃癌群のクラスターに投影されるか否かで、胃癌罹患の可能性を判定することができる。具体的には、自己組織化ネットワークの入力層のニューロン数は基本的に検査項目数とし、出力層には解像度を考慮して学習サンプル数より十分多いニューロンを二次元的に並べるが、発火するニューロン数は学習サンプル数となる。発火するニューロンの配置は、学習初期はランダムな状態であるが、競合学習が進むにつれて基本的に検査項目の値の組の類似度が高いもの同士ほど近接して配置されていき、徐々にデータ構造を反映したものに变化していく。個々の学習サンプルに対応する発火ニューロンの最適な配置図が自己組織化マップ (SOM) となる。

30

#### 【0040】

ここで、このように複数の胃癌群のサブクラスターに分離した場合、各サブクラスター内のデータは異なるパターンを示す。図11にそのようにして得られたSOMの一例を示す。図11に示すように、胃癌群(1)1001および胃癌群(2)1002に示すように明確にサブクラスターに分離されており、各サブクラスターに含まれるサンプルの検査項目の値を比較するといずれのサブクラスターでも、高値あるいは低値を示す検査項目があるだけでなく、あるサブクラスターでは高値を示すが、他のサブクラスターではより低い値あるいは低値を示す検査項目がある一方、逆にあるサブクラスターでは低値を示すが、他のサブクラスターではより高い値あるいは高値を示す検査項目が存在することとなる。以上の特徴を利用し、2~3の胃癌群のサブクラスターと健常対照群とをさらにBRNNなどを用いて教師あり学習分類を行って予測モデルを構築することにより、感度を数ポイント上昇させることができ、特異度の向上も得られた。

40

50

【 0 0 4 1 】

[ 第 3 実施例 ]

上述の 2 つの実施例では基本的に胃癌患者と健常者との分類を行ったが、本実施例では糖尿病患者も含めた分類を行う。すなわち、一次スクリーニングでは胃癌群に多くの糖尿病患者が紛れ込むことが判明しており、糖尿病患者を排除することにより、胃癌患者の検出精度が高くなることが期待される。本実施例では、パターン認識手法により上記実施例でも使用した検査項目を使用して胃癌患者群と糖尿病患者群とを分類する。具体的に図 1 2 を参照して説明する。図 1 2 は、このようにして得られた胃癌の糖尿病との鑑別の R O C 曲線を示す図である。上述のいずれかの実施例による 1 回目のスクリーニングの後、癌罹患と判定される群からさらに糖尿病に関連する指標を用いて糖尿病罹患者を排除することにより、より高精度に判定を行うことができる。計算条件を変化させて得られた最大の感度および特異度はそれぞれ 8 8 . 1 % および 9 8 . 1 % であり、曲線 1 2 0 2 で示すように A R D \_ o n で A U C = 0 . 9 5 4、曲線 1 2 0 1 で示すように A R D \_ o f f で A U C = 0 . 9 3 8 であった。

10

【 0 0 4 2 】

糖尿病との鑑別で特に重要な検査項目は、H b A 1 c、A / G、G L U、L D H、C r などである。また、胃癌群の中でさらに H b A 1 c、空腹時血糖値で判別するステップを組み込むことも有効である。ただし、これらの値は医療従事者が診れば明白な評価因子である

[ 第 4 実施例 ]

上述の第 1 および第 2 実施例では基本的に胃癌患者と健常者の分類を行ったが、本実施例では一般的な癌以外の疾患のデータも含めて分類を行う。すなわち、胃癌とその他の疾患とを区別する鑑別スクリーニングを導入する。図 1 3 は各検査項目についての胃癌患者群および疾患患者群 1 ~ 1 8 の一元配置分析を示す図である。図 1 3 には、疾患による検索項目 1 の一元配置分析 1 3 0 1、疾患による検索項目 2 の一元配置分析 1 3 0 2、疾患による検索項目 3 の一元配置分析 1 3 0 3、疾患による検索項目 4 の一元配置分析 1 3 0 4、疾患による検索項目 5 の一元配置分析 1 3 0 5、疾患による検索項目 6 の一元配置分析 1 3 0 6、疾患による検索項目 7 の一元配置分析 1 3 0 7、疾患による検索項目 8 の一元配置分析 1 3 0 8 および疾患による検索項目 9 の一元配置分析 1 3 0 9 を示す。具体的には図 1 3 に示すように、本願発明で処理の対象となる一般的な健康診断で得られる検査指標は、同時に癌以外の疾患に関連する情報を含んでいると考えられるので、癌以外の疾患 1 ~ 1 8 の各疾患患者群と胃癌患者群との鑑別を、パターン認識手法を用いて行う。具体的には、上述のいずれかの実施例による 1 回目のスクリーニングの後、癌罹患と判定される群からさらに各疾患に関連する指標を用いて各疾患の患者を排除することにより、より高精度に判定を行うことができる。その結果、疾患により精度は異なるが、いずれも 8 0 % ~ 9 8 % といった高い精度で鑑別が行えることが分かった。図 1 4 は、本実施例の鑑別スクリーニングの R O C 曲線を示す図である。鑑別の具体的な手法については、本技術分野で知られたいずれの手法も用いることができる。

20

30

【 0 0 4 3 】

以上の通り、本実施例の胃癌とその他の疾患とを区別する鑑別スクリーニングを導入することにより偽陽性を大幅に減らすことができた。すなわち、ターゲットとなる患者が特定の疾病に罹患していることが事前に判明していれば、本実施例の手法を用いることで、さらに高精度で癌罹患を予測することができる。

40

【 0 0 4 4 】

図 1 5 は、本実施例による胃癌予測における強弱を示す表である。図 1 5 において、は B R N N などの教師あり学習モデルと、S O M などの教師なし学習モデルとのいずれでも強く予測された高予測率群、同様に O は中等度予測率群、は低予測率群を示す。予測不可の x は偽陰性である。図 1 5 に示すように、進行度によらず一定の偽陰性の患者は存在するが、早期癌でも多くが中等度予測可能であることが理解できる。

【 0 0 4 5 】

50

本実施例と第1実施例を用いて、特定の四半期の実際のサンプルデータ(2034名分)によるスクリーニング調査結果を図16に示す。図16は、本発明の一実施形態の胃癌予測におけるサンプルデータの一例の予測結果を示す箱ひげ図である。図16の調査結果に基づき、感度等を計算すると、感度100%(または75%)、特異度99.3%、PPV21.1%(または16.7%)、NPV100%(または99.95%)と、良好な結果が得られた。

【0046】

[第5実施例]

上述の実施例では基本的に癌の中でも主に胃癌患者の分類を行ったが、本実施例では胃癌以外の癌との判別を行う。具体的には、胃癌と食道癌とはもともと異なる組織由来の癌であるため、一定程度区別することが可能である。上述の実施例に加えて食道癌患者のデータも加えた学習データを使用して、鑑別計算を行った結果、図17に示すように胃癌患者群と食道癌患者群とを区別することができた。両群を区別する有益な検査パラメータとしては、特にRBC(赤血球)、ALT(アラニンアミノトランスフェラーゼ)、TP(総蛋白)、A/G(アルブミン/グロブリン比)、TC(総コレステロール)、TG(中性脂肪;トリグリセリド)、GLU(グルコース)、BMI(ボディマス指数)などを挙げるができる。

10

【0047】

[第6実施例]

上述の各実施例では様々な手法で予測モデルを構築することを説明してきたが、これらの予測モデルは基本的に各患者の個別の検査データを学習サンプルまたはテストサンプルとするものであり、測定時期という時間的概念はない。しかし、一般に、疾病は時間の経過とともに症状や状態が変化し、それに伴い関連する検査項目やその値の大小の傾向も変化する可能性がある。本実施例では、各患者の所定の期間を置いて測定された測定値を考慮することにより、より高精度で予測モデルを構築するものである。すなわち、前回受診時から今回受診時にかけての各検査値の変動率を例えば1年あたりの変動率に換算した年間変動速度Vを、 $V(\%/年) = ((\text{今回検査値} - \text{前回検査値}) / \text{前回検査値}) \times (365日 / \text{受信間隔})$ とする。検査値だけでなく、その年間変動速度も変数として使用しパターン認識手法を用いた予測モデルを構築した。

20

【0048】

例えば、上述の各予測モデル構築の際の変数としてRBC、WBC、Ht、Hb、PLT、HbA1c、ALP、ALT、AST、LDH、GTP、AMY、TP、A/G、Alb、BUN、Cr、Bil、TC、HDL、TG、LDL、UA、GLU、RF、CRP、SG、BMI、HR、MONO、LYMPH、NEUT、EOSINO、BASO、Na、K、Ca、Pに加え、これらの年間変動速度V\_RBC、V\_WBC、V\_Ht、V\_Hb、V\_PLT、V\_HbA1c、V\_ALP、V\_ALT、V\_AST、V\_LDH、V\_GTP、V\_AMY、V\_TP、V\_A/G、V\_Alb、V\_BUN、V\_Cr、V\_Bil、V\_TC、V\_HDL、V\_TG、V\_LDL、V\_UA、V\_GLU、V\_RF、V\_CRP、V\_SG、V\_BMI、V\_HR、V\_MONO、V\_LYMPH、V\_NEUT、V\_EOSINO、V\_BASO、V\_Na、K、V\_Ca、V\_Pを変数として用いて予測モデルの構築を行った。その結果、通常測定値のみを使用して構築した予測モデルよりも精度の向上が診られえた。本実施例は、本願発明のような通常定期的に測定される健康診断などの測定値を用いて予測モデルを構築するような態様では、各患者の測定を定期的に入手することができることから有効である。

30

40

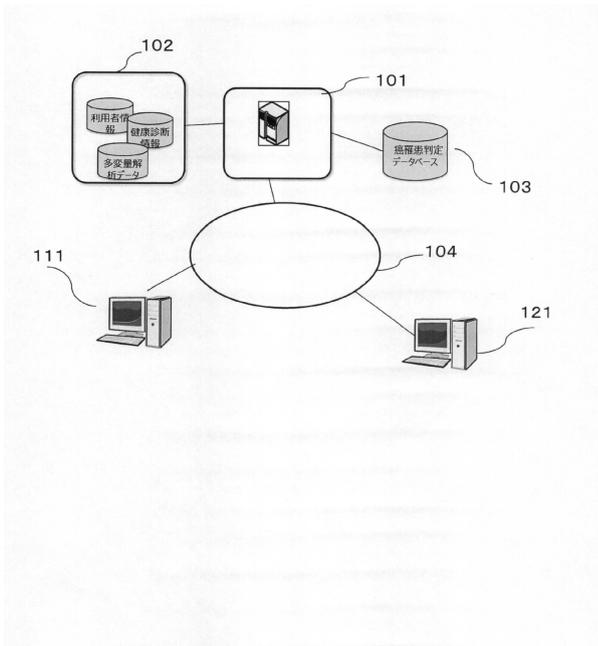
【0049】

以上、一般的な健康診断で得られる検査指標により予測モデルを構築して胃癌について主に予測可能であることが分かったが、以上の点は胃癌に限らず広くその他の癌にも適用できるものであり、本願発明は胃癌および上述の各実施例自体に限定されるものではない。

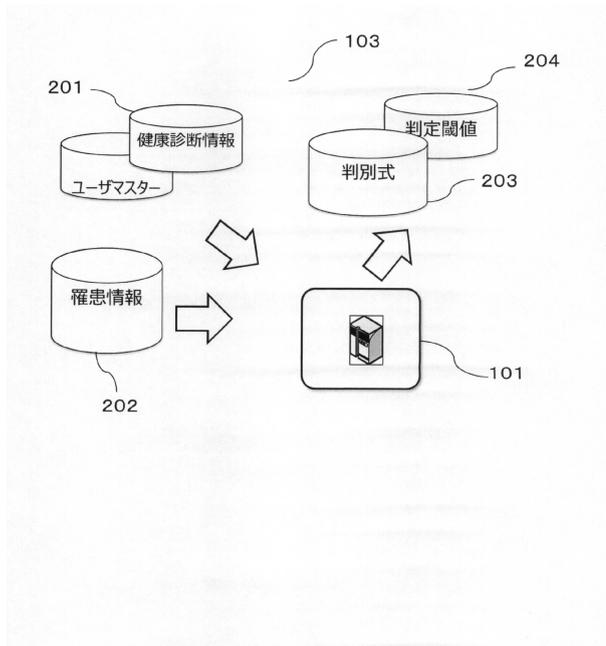
50

【図面】

【図 1】



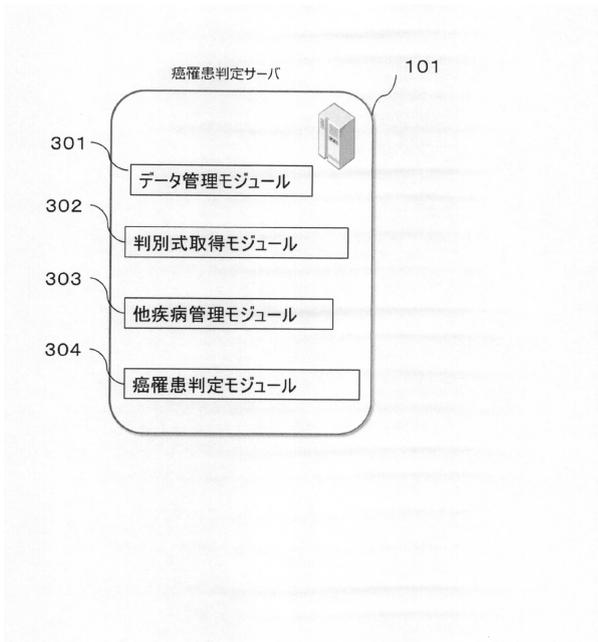
【図 2】



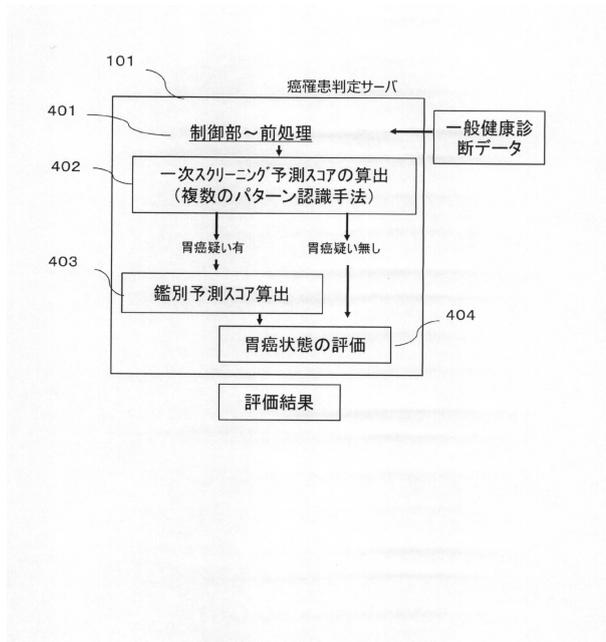
10

20

【図 3】



【図 4】

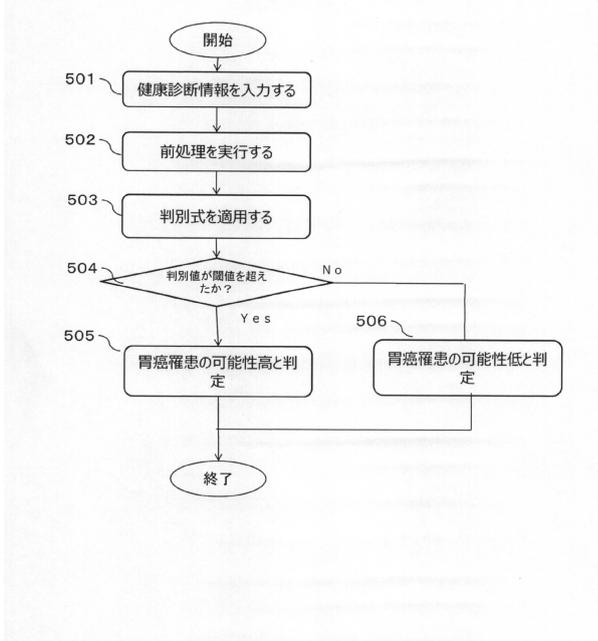


30

40

50

【図5】



【図6】

受検者ID	性別	受診日	RBC	HbC	Ht	Hb	PLT	SG	BMI	癌罹患
0000	M	...	00	xx	△△	◇◇	000	...	xxx	胃癌
0000	M	...	00	xx	△△	◇◇	000	...	xxx	未
0000	F	...	00	xx	△△	◇◇	000	...	xxx	未
...	...	...	...	...	...	...	...	...	...	...
0000	F	...	00	xx	△△	◇◇	000	...	xxx	胃癌
0000	M	...	00	xx	△△	◇◇	000	...	xxx	未

10

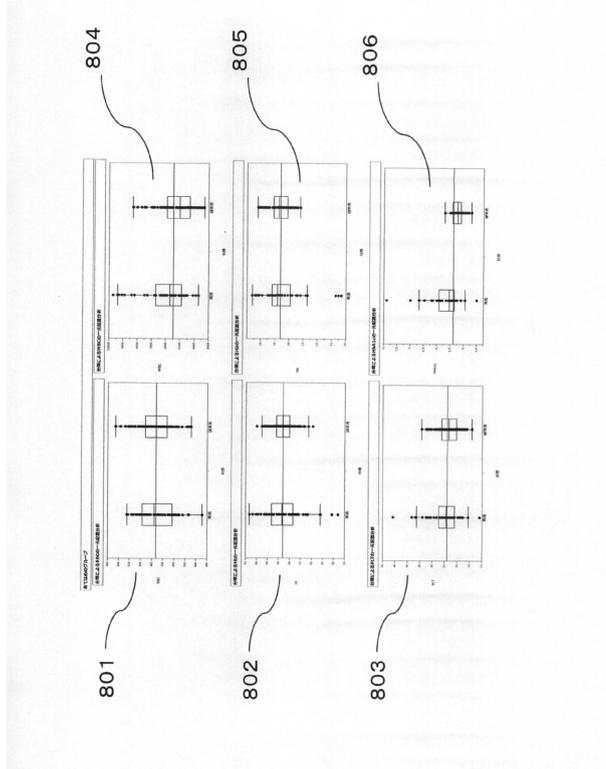
20

【図7】

受検者ID	性別	受診日	RBC	HbC	Ht	Hb	PLT	SG	BMI	癌罹患	癌ステージ
0000	M	...	00	xx	△△	◇◇	000	...	xxx	胃癌	IA
0000	M	...	00	xx	△△	◇◇	000	...	xxx	未	未
0000	F	...	00	xx	△△	◇◇	000	...	xxx	未	未
...	...	...	...	...	...	...	...	...	...	...	...
0000	F	...	00	xx	△△	◇◇	000	...	xxx	胃癌	III
0000	M	...	00	xx	△△	◇◇	000	...	xxx	未	未

30

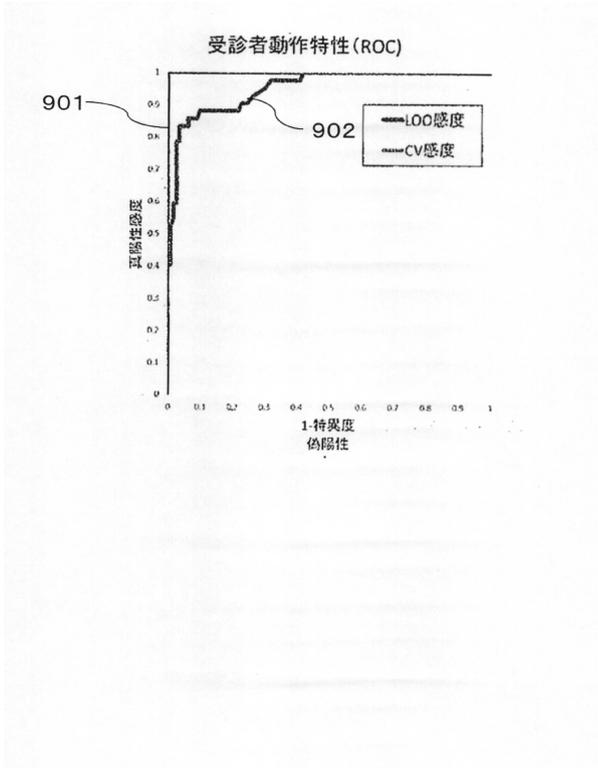
【図8】



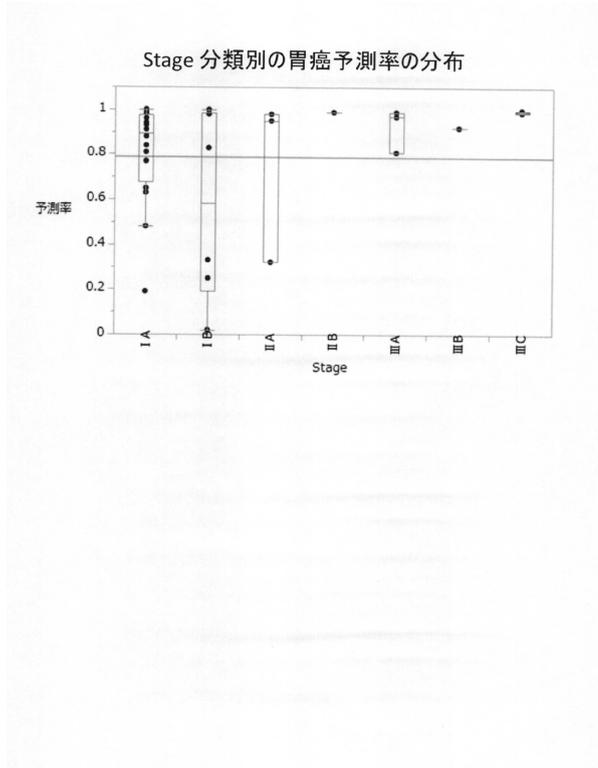
40

50

【 図 9 】



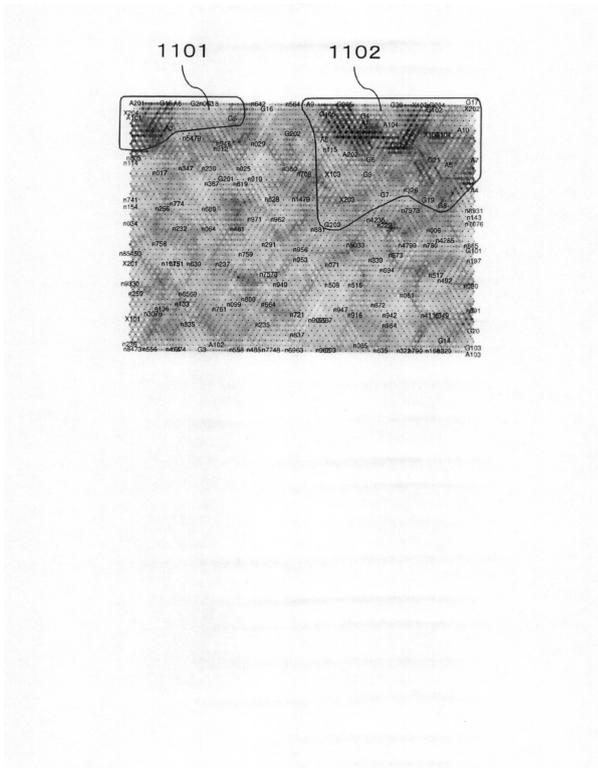
【 図 1 0 】



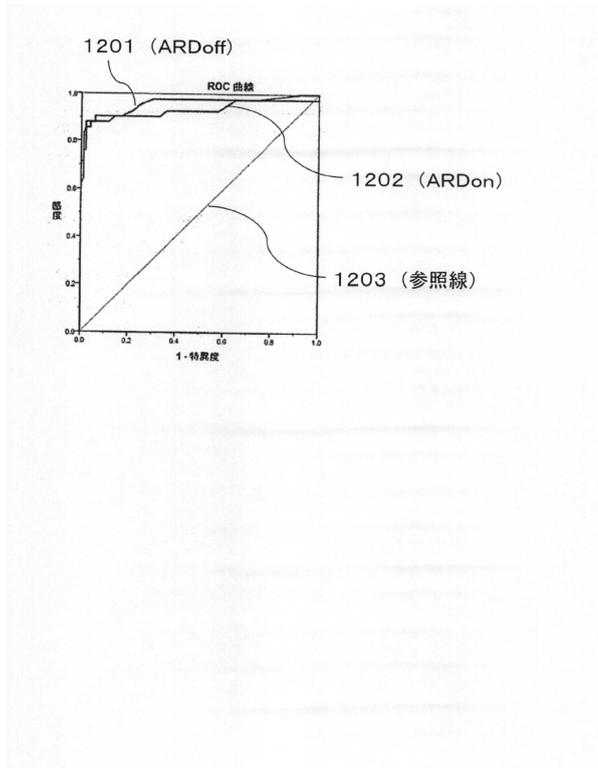
10

20

【 図 1 1 】



【 図 1 2 】

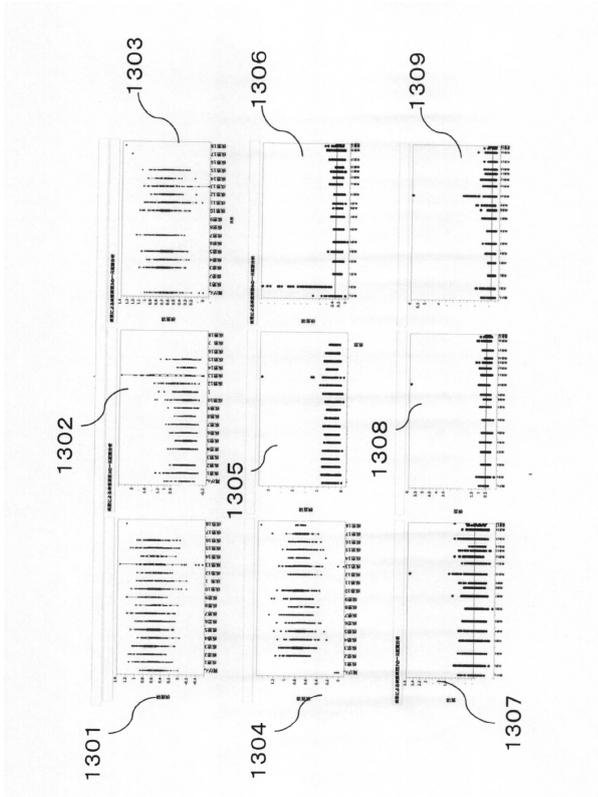


30

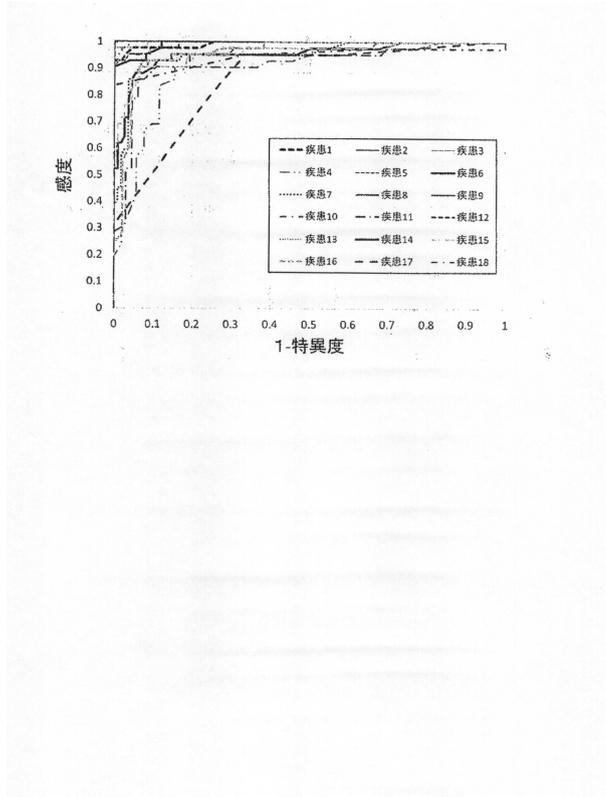
40

50

【図 13】



【図 14】



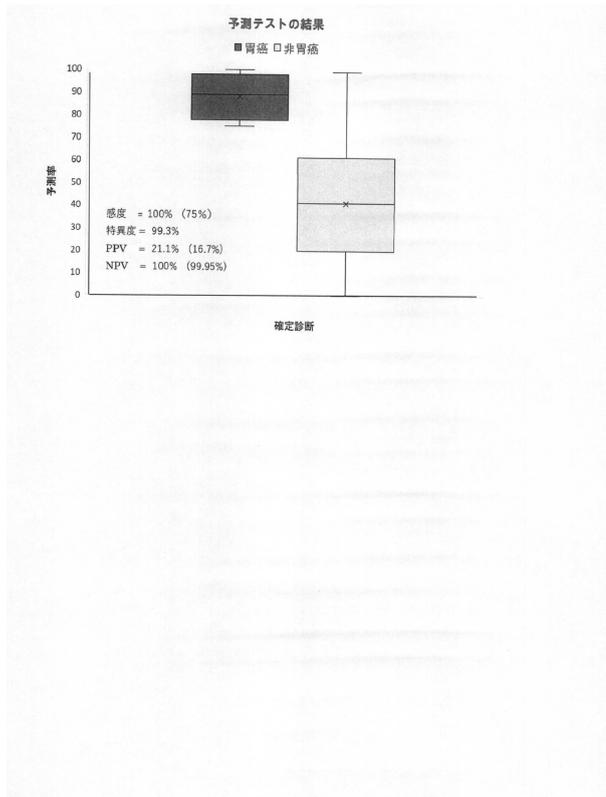
10

20

【図 15】

予測強度	早期胃癌	進行胃癌	不明	全体
	人数 (%)	人数 (%)	人数 (%)	人数 (%)
◎	9.8	15.7	3.9	29.4
○	23.5	5.9	3.9	33.3
△	13.7	3.9	2.0	19.6
×	7.8	3.9	5.9	17.6
	54.9	29.4	15.7	100.0

【図 16】

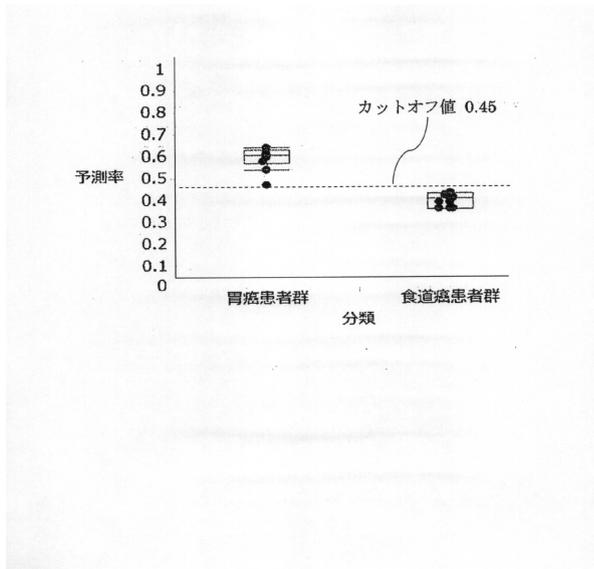


30

40

50

【 図 17 】



10

20

30

40

50

## フロントページの続き

東京都世田谷区上馬2丁目20-9

審査官 白形 優依

- (56)参考文献 特開2019-060749(JP,A)  
特開2007-052774(JP,A)  
特開2008-052511(JP,A)  
特開2019-105451(JP,A)  
国際公開第2018/186434(WO,A1)  
特開2015-108515(JP,A)  
特表2011-501845(JP,A)  
韓国登録特許第1951727(KR,B1)  
特表2012-531581(JP,A)  
国際公開第2017/115816(WO,A1)  
国際公開第2016/181912(WO,A1)  
国際公開第2016/121695(WO,A1)  
国際公開第2016/013597(WO,A1)  
特開2008-224526(JP,A)
- (58)調査した分野 (Int.Cl., DB名)  
G01N 33/48 - 33/98  
G06Q 50/22  
JSTPlus/JMEDPlus/JST7580(JDreamII)