

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2009-247688  
(P2009-247688A)

(43) 公開日 平成21年10月29日(2009.10.29)

(51) Int.Cl. F I テーマコード (参考)  
**A 6 1 B 5/00 (2006.01)** A 6 1 B 5/00 G 4 C 1 1 7  
**G 0 6 Q 50/00 (2006.01)** G 0 6 F 17/60 1 2 6 H

審査請求 未請求 請求項の数 21 O L (全 23 頁)

(21) 出願番号	特願2008-100587 (P2008-100587)	(71) 出願人	504174180 国立大学法人高知大学 高知県高知市曙町二丁目5番1号
(22) 出願日	平成20年4月8日(2008.4.8)	(74) 代理人	100083806 弁理士 三好 秀和
		(74) 代理人	100101247 弁理士 高橋 俊一
		(74) 代理人	100109380 弁理士 小西 恵
		(72) 発明者	倉本 秋 高知県南国市岡豊町小蓮 国立大学法人高知大学医学部内
		(72) 発明者	島山 豊 高知県南国市岡豊町小蓮 国立大学法人高知大学医学部内

最終頁に続く

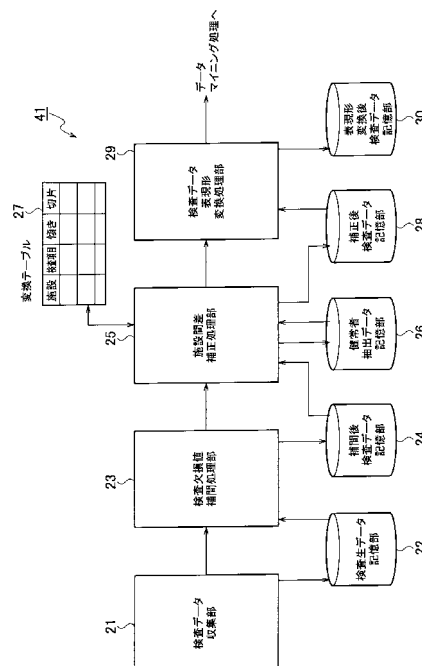
(54) 【発明の名称】 臨床検査データ解析支援装置、臨床検査データ解析支援方法及びそのプログラム

(57) 【要約】

【課題】 各施設間で臨床検査データに生ずる施設間誤差を効率的かつ高精度に補正する。

【解決手段】 入力された前記臨床検査データから、健常者の臨床検査データと推定される臨床検査データのみを抽出して、第1の基準検査データとして第1の記憶装置に格納する健常者検査データ抽出部(21)と、第1の基準検査データを読み出し、正規化された第2の基準検査データに変換する正規化処理部(25)と、第2の基準検査データ群の正規分布パターン曲線上において、標準偏差に基づき決定される近似直線を求め、その傾き及び切片を補正用パラメータとして導出する補正用パラメータ導出部(25)と、入力された臨床検査データの項目値と一致する検査施設及び臨床検査項目について定義された前記補正用パラメータで規定される一次関数を、前記臨床検査データに対して適用して、前記臨床検査データを補正する補正処理部(25)とを具備する。

【選択図】 図2



## 【特許請求の範囲】

## 【請求項 1】

同一施設で分析された被検者を母集団とする、1つの臨床検査項目についての臨床検査データを入力する入力部と、

入力された前記臨床検査データから、健常者の臨床検査データと推定される臨床検査データのみを抽出して、第1の基準検査データとして第1の記憶装置に格納する健常者検査データ抽出部と、

前記第1の記憶装置から前記第1の基準検査データを読み出し、その分布型が正規分布に近似する第2の基準検査データに変換する正規化処理部と、

前記第2の基準検査データ群の正規分布パターン曲線上において、該正規分布の標準偏差に基づき決定される近似直線を求め、該近似直線の傾き及び切片を補正用パラメータとして導出し、該補正用パラメータを変換テーブルに記憶する補正用パラメータ導出部と、

入力部により入力された前記臨床検査データの項目値と一致する施設及び臨床検査項目について定義された前記補正用パラメータを、前記変換テーブルを参照して取得し、前記臨床検査データに対して前記補正用パラメータで規定される一次関数を適用することにより、前記臨床検査データを補正する補正処理部と、

補正された臨床検査データを第2の記憶装置に格納する補正後データ格納部とを具備する

ことを特徴とする臨床検査データ解析支援装置。

## 【請求項 2】

前記補正用パラメータ導出部は、前記近似直線を、前記第2の基準検査データ群の正規分布パターン曲線上で、 $x = \mu \pm 2$  である偏位点（ここで、 $\mu$  は平均値、 $\sigma$  は標準偏差）における接線として求める

ことを特徴とする請求項 1 に記載の臨床検査データ解析支援装置。

## 【請求項 3】

前記健常者検査データ抽出部は、

入力された前記臨床検査データのうち、被検者の年齢を値とする、第1の閾値と、該第1の閾値より高い第2の閾値との間にある値の年齢値を有する被検者のみを母集団とする臨床検査データを抽出する第1のフィルタリング処理部を具備する

ことを特徴とする請求項 1 又は 2 に記載の臨床検査データ解析支援装置。

## 【請求項 4】

前記健常者検査データ抽出部は、

同一被検者について、入力された前記臨床検査データの臨床検査項目との間で所定の閾値以上の相関性を有する他の臨床検査項目が、異常値を示した場合には、当該被検者の臨床検査データを、前記第1の基準検査データとして抽出すべき臨床検査データから除外する第2のフィルタリング処理部を具備する

ことを特徴とする請求項 1 ないし 3 のいずれか記載の臨床検査データ解析支援装置。

## 【請求項 5】

前記健常者検査データ抽出部は、

各臨床検査項目をそれぞれ軸とする  $n$  次元（ $n \geq 1$  の整数）の確率分布上で、該確率分布の等確率楕円外にプロットされる臨床検査データを除外し、

該処理を、入力された前記臨床検査データの臨床検査項目との間で所定の閾値以上の相関性を有する全臨床検査項目間について、除外されるべきデータがなくなるまで繰り返した後得られた臨床検査データを前記第1の基準検査データとして抽出する第3のフィルタリング処理部を具備する

ことを特徴とする請求項 1 ないし 4 のいずれか記載の臨床検査データ解析支援装置。

## 【請求項 6】

上記臨床検査データ解析支援装置は、さらに、

前記入力部により入力された臨床検査データに欠損臨床検査項目が存在する場合には、該欠損臨床検査項目の検査値を補間により生成する補間処理部を具備する

10

20

30

40

50

ことを特徴とする請求項 1 ないし 5 のいずれか記載の臨床検査データ解析支援装置。

【請求項 7】

上記臨床検査データ解析支援装置は、さらに、

前記入力部により入力された臨床検査データの検査値のデータ型を判定し、文字型で記述された検査値を、数値型の連続値に変換する表現形変換処理部を具備する

ことを特徴とする請求項 1 ないし 6 のいずれか記載の臨床検査データ解析支援装置。

【請求項 8】

入力部と、健常者検査データ抽出部と、正規化処理部と、補正用パラメータ導出部と、補正処理部と、補正後データ格納部とを備える臨床検査データ解析支援装置により実行される臨床検査データ解析支援方法であって、

前記入力部が、同一施設で分析された被検者を母集団とする、1つの臨床検査項目についての臨床検査データを入力するステップと、

前記健常者検査データ抽出部が、入力された前記臨床検査データから、健常者の臨床検査データと推定される臨床検査データのみを抽出して、第1の基準検査データとして第1の記憶装置に格納するステップと、

前記正規化処理部が、前記第1の記憶装置から前記第1の基準検査データを読み出し、その分布型が正規分布に近似する第2の基準検査データに変換するステップと、

前記補正用パラメータ導出部が、前記第2の基準検査データ群の正規分布パターン曲線上において、該正規分布の標準偏差に基づき決定される近似直線を求め、該近似直線の傾き及び切片を補正用パラメータとして導出し、該補正用パラメータを変換テーブルに記憶するステップと、

前記補正処理部が、入力部により入力された前記臨床検査データの項目値と一致する施設及び臨床検査項目について定義された前記補正用パラメータを、前記変換テーブルを参照して取得し、前記臨床検査データに対して前記補正用パラメータで規定される一次関数を適用することにより、前記臨床検査データを補正するステップと、

前記補正後データ格納部が、補正された臨床検査データを第2の記憶装置に格納するステップとを含む

ことを特徴とする方法。

【請求項 9】

前記補正用パラメータを導出するステップにおいて、前記近似直線を、前記第2の基準検査データ群の正規分布パターン曲線上で、 $x = \mu \pm 2$  である偏位点（ここで、 $\mu$  は平均値、 $\sigma$  は標準偏差）における接線として求める

ことを特徴とする請求項 8 に記載の方法。

【請求項 10】

前記健常者検査データを抽出するステップは、さらに、

入力された前記臨床検査データのうち、被検者の年齢を値とする、第1の閾値と、該第1の閾値より高い第2の閾値との間にある値の年齢値を有する被検者のみを母集団とする臨床検査データを抽出するステップを含む

ことを特徴とする請求項 8 又は 9 に記載の方法。

【請求項 11】

前記健常者検査データを抽出するステップは、さらに、

同一被検者について、入力された前記臨床検査データの臨床検査項目との間で所定の閾値以上の相関性を有する他の臨床検査項目が、異常値を示した場合には、当該被検者の臨床検査データを、前記第1の基準検査データとして抽出すべき臨床検査データから除外するステップを含む

ことを特徴とする請求項 8 ないし 10 のいずれか記載の方法。

【請求項 12】

前記健常者検査データを抽出するステップは、さらに、

各臨床検査項目をそれぞれ軸とする  $n$  次元（ $n \geq 1$  の整数）の確率分布上で、該確率分布の等確率楕円外にプロットされる臨床検査データを除外し、

10

20

30

40

50

該処理を、入力された前記臨床検査データの臨床検査項目との間で所定の閾値以上の相関性を有する全臨床検査項目間について、除外されるべきデータがなくなるまで繰り返した後得られた臨床検査データを前記第1の基準検査データとして抽出するステップを含むことを特徴とする請求項8ないし11のいずれか記載の方法。

【請求項13】

上記方法は、さらに、

前記入力部により入力された臨床検査データに欠損臨床検査項目が存在する場合には、該欠損臨床検査項目の検査値を補間により生成するステップを含むことを特徴とする請求項8ないし12のいずれか記載の方法。

【請求項14】

上記方法は、さらに、

前記入力部により入力された臨床検査データの検査値のデータ型を判定し、文字型で記述された検査値を、数値型の連続値に変換するステップを含むことを特徴とする請求項8ないし13のいずれか記載の方法。

【請求項15】

臨床検査データ解析支援処理をコンピュータに実行させるための臨床検査データ解析支援プログラムであって、該プログラムは、前記コンピュータに、

同一施設で分析された被検者を母集団とする、1つの臨床検査項目についての臨床検査データを入力する入力処理と、

入力された前記臨床検査データから、健常者の臨床検査データと推定される臨床検査データのみを抽出して、第1の基準検査データとして第1の記憶装置に格納する健常者検査データ抽出処理と、

前記第1の記憶装置から前記第1の基準検査データを読み出し、その分布型が正規分布に近似する第2の基準検査データに変換する正規化処理と、

前記第2の基準検査データ群の正規分布パターン曲線上において、該正規分布の標準偏差に基づき決定される近似直線を求め、該近似直線の傾き及び切片を補正用パラメータとして導出し、該補正用パラメータを変換テーブルに記憶する補正用パラメータ導出処理と、

入力部により入力された前記臨床検査データの項目値と一致する施設及び臨床検査項目について定義された前記補正用パラメータを、前記変換テーブルを参照して取得し、前記臨床検査データに対して前記補正用パラメータで規定される一次関数を適用することにより、前記臨床検査データを補正する補正処理と、

補正された臨床検査データを第2の記憶装置に格納する補正後データ格納処理とを含む処理を実行させるためのものであることを特徴とするプログラム。

【請求項16】

前記補正用パラメータ導出処理は、前記近似直線を、前記第2の基準検査データ群の正規分布パターン曲線上で、 $x = \mu \pm 2$  である偏位点（ここで、 $\mu$  は平均値、 $\sigma$  は標準偏差）における接線として求める

ことを特徴とする請求項15に記載のプログラム。

【請求項17】

前記健常者検査データ抽出処理は、

入力された前記臨床検査データのうち、被検者の年齢を値とする、第1の閾値と、該第1の閾値より高い第2の閾値との間にある値の年齢値を有する被検者のみを母集団とする臨床検査データを抽出する第1のフィルタリング処理を含む

ことを特徴とする請求項15又は16に記載のプログラム。

【請求項18】

前記健常者検査データ抽出処理は、

同一被検者について、入力された前記臨床検査データの臨床検査項目との間で所定の閾値以上の相関性を有する他の臨床検査項目が、異常値を示した場合には、当該被検者の臨床検査データを、前記第1の基準検査データとして抽出すべき臨床検査データから除外す

10

20

30

40

50

る第2のフィルタリング処理を含む

ことを特徴とする請求項15ないし17のいずれか記載のプログラム。

【請求項19】

前記健常者検査データ抽出処理は、

各臨床検査項目をそれぞれ軸とする $n$ 次元( $n \geq 1$ の整数)の確率分布上で、該確率分布の等確率楕円外にプロットされる臨床検査データを除外し、

該処理を、入力された前記臨床検査データの臨床検査項目との間で所定の閾値以上の相関性を有する全臨床検査項目間について、除外されるべきデータがなくなるまで繰り返した後得られた臨床検査データを前記第1の基準検査データとして抽出する第3のフィルタリング処理を含む

10

ことを特徴とする請求項15ないし18のいずれか記載のプログラム。

【請求項20】

上記プログラムは、さらに、

前記入力処理により入力された臨床検査データに欠損臨床検査項目が存在する場合には、該欠損臨床検査項目の検査値を補間により生成する補間処理を含む

ことを特徴とする請求項15ないし19のいずれか記載のプログラム。

【請求項21】

上記プログラムは、さらに、

前記入力処理により入力された臨床検査データの検査値のデータ型を判定し、文字型で記述された検査値を、数値型の連続値に変換する表現形変換処理を含む

20

ことを特徴とする請求項15ないし20のいずれか記載のプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、臨床検査データ解析支援装置、臨床検査データ解析支援方法及びそのプログラムに関する。より詳細には、特に、複数の病院、検査施設、分析施設ないし在宅サイトで収集された臨床検査データを統合し、集積して解析する統合医療システムにおいて、統合医療データベースに格納されるべき複数の病院、検査施設ないし在宅サイトでそれぞれ収集される臨床検査データを効率的かつ高精度に統合医療データを解析する医療データマイニングエンジンに供給するための技術に関する。

30

【背景技術】

【0002】

医療の経済効率を低下させることなく医療の質的向上を実現するためのプラットフォームとして、Evidence Based Medicine (EBM) システムの構築が図られている。このEBMシステムによれば、統合医療データベースに格納された医療データの解析により動的に診断根拠を導出し、各医療機関や在宅において、医師、医療スタッフや栄養管理士等が、客観的診断根拠に基づく診断、治療、栄養指導等を行なうことが可能となる。

【0003】

例えば、特許文献1は、本願発明者らの一部を含む発明者らによる特許出願に係り、臨床検査で取得され、中央に収集された複数の分析結果項目について、ニューラルネットや自己組織化マップ等の手法を用いて集積された多数患者の過去の臨床検査結果に基づき、予め定義された複数の出現パターンのうち最も近似する出現パターンに対する近似度を算出し、熟練した医師の判断によらなくとも、再検査が必要な検査結果であるか否かを自動的に判断することのできる臨床検査分析装置を開示する。

40

【0004】

一方、特許文献2は、臨床検査で取得され、中央に収集された膨大な医療データを網羅的に解析して、医療データ相互間の相関性を示す相関ルール及びその頻度データを抽出するデータマイニング処理により、複数の分析項目間の関係を可視的に表示する医療データ解析システムを開示する。

50

【特許文献1】特開2003-114231号公報

【特許文献2】特開2004-185547号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

ところで、上記の統合医療データを解析する医療データマイニングエンジンにより用いられる統合医療データベースには、複数の病院、検査施設、分析施設、ないし在宅サイト（本明細書及び特許請求の範囲において、特に断らない限り、これらを総称して単に「施設」という。）でそれぞれ収集された臨床検査データが集約される。しかしながら、従来、施設ごとに取得された臨床検査データに施設間誤差が生じ、これにより、統合的な医療データマイニングエンジンに提供される統合医療データの精度が損なわれ、結果として導出されるべき診断根拠の信頼性を低下させていた。すなわち、臨床検査実施機関である複数の施設では、相互に画一的でない条件下で臨床検査データが取得、分析されるため、例えばMedical Electronics (ME)測定機器、試薬、分析法、分析用標準液等の施設間での相違に起因して、施設ごとに取得された臨床検査データに誤差が生ずることは避け得ない。

10

【0006】

また、例えば、受けるべき検査を受けていない、体重、腹囲等の測定を行なわない、問診に回答しない、等の種々の理由から、ある被検者について臨床検査データの欠損は生じ得るが、欠損が生じた臨床検査項目について後から実データである検査値を取得することはできない。欠損値が存在するとルール導出のための論理演算においてエラーとして処理されてしまうため、欠損値を含む臨床検査データについては、後続の医療データマイニングエンジン及び知識処理エンジンで診断ルールを導出する基礎とすることができなかつた。

20

【0007】

さらに、各施設で取得されるデータの表現形は、標準化ないし統一化されておらず、例えば文字型、数値型と多様なデータ型であり得るため、医療データマイニングエンジンに供給されても、効率的なクラスタリング、分類等の処理を阻害していた。

【0008】

臨床検査データは、一般に、データ量が膨大である上、各施設で分析される被検者の母集団が異なるため施設間で同一サンプルによる対比ができず、また、時系列データであるためある臨床検査項目について再度同じ被検者から同じ条件で検査データの再取得ができず、このため施設間誤差を有効に解消することができない。

30

【0009】

ここで、例えば、基準となるマーカー物質を試料として分析を行なって臨床検査データを得る手法や、健康診断検査項目について外部から提供される精度管理データにより補正する手法や、分析法ないし標準液等を標準化することにより誤差を解消するための手法も提案されている。しかしながら、第1の手法においては、臨床検査試料の分析段階でマーカー物質を分析させなければならないため、過去において既に取得された臨床検査データに対して行なうことができず、第2の手法においては、外部提供の精度管理データによる補正はまた健康診断項目以外の臨床検査項目についての有効な精度管理データは未だ提供されておらず、第3の手法においては、分析法ないし標準液等を標準化する手法によっても標準化されない条件が作用するため、臨床検査データについての微妙な施設間誤差を解消することは依然なし得なかつた。

40

【0010】

本発明は、上記課題に鑑みてされたものであり、その目的は、各施設間で臨床検査データに生ずる施設間誤差を効率的かつ高精度に補正することで、複数の施設で収集される臨床検査データを、医療データマイニングエンジンが本来想定する正しいデータにデータクレンジングして、1つの統合医療データベースに統合可能とすることのできる臨床検査データ解析支援装置、臨床検査データ解析支援方法及びそのプログラムを提供することにあ

50

る。

【0011】

また、本発明の他の目的は、臨床検査データ中に生じた検査値欠損を有効に補間して、検査値欠損がある被検者の臨床検査データであっても、医療データマイニングにおけるクラスタリング、分類等の処理の基礎とすることができる臨床検査データ解析支援装置、臨床検査データ解析支援方法及びプログラムを提供することにある。

【0012】

また、本発明の他の目的は、各施設間で相違する臨床検査データのデータ形式の相違を吸収して、医療データマイニングにおけるクラスタリング、分類等の処理の基礎とすることができる臨床検査データ解析支援装置、臨床検査データ解析支援方法及びプログラムを提供することにある。

10

【課題を解決するための手段】

【0013】

本願発明者らは、検査・分析条件が施設間で相違しない理想状態においては、各施設間で取得される臨床検査データは本来均一性を示すべきものとの知見に基づき、本願発明に想到した。

【0014】

本発明の原理は、同一施設で分析された被検者を母集団とする臨床検査データから、健常者の臨床検査データと推定される臨床検査データのみを抽出して、施設間誤差補正用の基準臨床検査データとし、この基準臨床検査データから補正用パラメータを導出するものである。好適には、この基準臨床検査データは、正規化された後、補正用パラメータ導出の基礎とされる。各施設から収集される臨床検査データは、該臨床検査データの項目値と一致する検査施設及び臨床検査項目について定義された補正用パラメータを適用することにより補正される。

20

【0015】

なお、本明細書及び特許請求の範囲において、「臨床検査データ」とは、医師が患者の病気の診断及び治療に直接利用するデータに限定されるものではなく、およそ病気の診断、治療、予防並びに栄養指導等に利用され得るあらゆる被検者のデータを意味するものであり、被検者から採取された検体を分析する検体検査により得られるデータと、被検者を直接検査する生理機能検査により得られるデータの双方を含む。

30

【0016】

本発明のある特徴によれば、同一施設で分析された被検者を母集団とする、1つの臨床検査項目についての臨床検査データを入力する入力部と、入力された前記臨床検査データから、健常者の臨床検査データと推定される臨床検査データのみを抽出して、第1の基準検査データとして第1の記憶装置に格納する健常者検査データ抽出部と、前記第1の記憶装置から前記第1の基準検査データを読み出し、その分布型が正規分布に近似する第2の基準検査データに変換する正規化処理部と、前記第2の基準検査データ群の正規分布パターン曲線上において、該正規分布の標準偏差に基づき決定される近似直線を求め、該近似直線の傾き及び切片を補正用パラメータとして導出し、該補正用パラメータを変換テーブルに記憶する補正用パラメータ導出部と、入力部により入力された前記臨床検査データの項目値と一致する検査施設及び臨床検査項目について定義された前記補正用パラメータを、前記変換テーブルを参照して取得し、前記臨床検査データに対して前記補正用パラメータで規定される一次関数を適用することにより、前記臨床検査データを補正する補正処理部と、補正された臨床検査データを第2の記憶装置に格納する補正後データ格納部とを具備するとを特徴とする臨床検査データ解析支援装置が提供される。

40

【0017】

前記補正用パラメータ導出部は、前記近似直線を、前記第2の基準検査データ群の正規分布パターン曲線上で、 $x = \mu \pm 2$  である偏位点（ここで、 $\mu$ は平均値、 $\sigma$ は標準偏差）における接線として求めてよい。

【0018】

50

前記健常者検査データ抽出部は、入力された前記臨床検査データのうち、被検者の年齢を値とする、第1の閾値と、該第1の閾値より高い第2の閾値との間にある値の年齢値を有する被検者のみを母集団とする臨床検査データを抽出する第1のフィルタリング処理部を具備してよい。

【0019】

前記健常者検査データ抽出部は、同一被検者について、入力された前記臨床検査データの臨床検査項目との間で所定の閾値以上の相関性を有する他の臨床検査項目が、異常値を示した場合には、当該被検者の臨床検査データを、前記第1の基準検査データとして抽出すべき臨床検査データから除外する第2のフィルタリング処理部を具備してよい。

【0020】

前記健常者検査データ抽出部は、各臨床検査項目をそれぞれ軸とする $n$ 次元( $n-1$ の整数)の確率分布上で、該確率分布の等確率楕円外にプロットされる臨床検査データを除外し、該処理を、入力された前記臨床検査データの臨床検査項目との間で所定の閾値以上の相関性を有する全臨床検査項目間について、除外されるべきデータがなくなるまで繰り返した後得られた臨床検査データを前記第1の基準検査データとして抽出する第3のフィルタリング処理部を具備してよい。

【0021】

上記臨床検査データ解析支援装置は、さらに、前記入力部により入力された臨床検査データに欠損臨床検査項目が存在する場合には、該欠損臨床検査項目の検査値を補間により生成する補間処理部を具備してよい。

【0022】

上記臨床検査データ解析支援装置は、さらに、前記入力部により入力された臨床検査データの検査値のデータ型を判定し、文字型で記述された検査値を、数値型の連続値に変換する表現形変換処理部を具備してよい。

【0023】

本発明のある特徴によれば、入力部と、健常者検査データ抽出部と、正規化処理部と、補正用パラメータ導出部と、補正処理部と、補正後データ格納部とを備える臨床検査データ解析支援装置により実行される臨床検査データ解析支援方法であって、前記入力部が、同一施設で分析された被検者を母集団とする、1つの臨床検査項目についての臨床検査データを入力するステップと、前記健常者検査データ抽出部が、入力された前記臨床検査データから、健常者の臨床検査データと推定される臨床検査データのみを抽出して、第1の基準検査データとして第1の記憶装置に格納するステップと、前記正規化処理部が、前記第1の記憶装置から前記第1の基準検査データを読み出し、その分布型が正規分布に近似する第2の基準検査データに変換するステップと、前記補正用パラメータ導出部が、前記第2の基準検査データ群の正規分布パターン曲線上において、該正規分布の標準偏差に基づき決定される近似直線を求め、該近似直線の傾き及び切片を補正用パラメータとして導出し、該補正用パラメータを変換テーブルに記憶するステップと、前記補正処理部が、入力部により入力された前記臨床検査データの項目値と一致する施設及び臨床検査項目について定義された前記補正用パラメータを、前記変換テーブルを参照して取得し、前記臨床検査データに対して前記補正用パラメータで規定される一次関数を適用することにより、前記臨床検査データを補正するステップと、前記補正後データ格納部が、補正された臨床検査データを第2の記憶装置に格納するステップとを含むことを特徴とする方法が提供される。

【0024】

本発明の他の特徴によれば、臨床検査データ解析支援処理をコンピュータに実行させるための臨床検査データ解析支援プログラムであって、該プログラムは、前記コンピュータに、同一施設で分析された被検者を母集団とする、1つの臨床検査項目についての臨床検査データを入力する入力処理と、入力された前記臨床検査データから、健常者の臨床検査データと推定される臨床検査データのみを抽出して、第1の基準検査データとして第1の記憶装置に格納する健常者検査データ抽出処理と、前記第1の記憶装置から前記第1の基

10

20

30

40

50



準検査データを読み出し、その分布型が正規分布に近似する第2の基準検査データに変換する正規化処理と、前記第2の基準検査データ群の正規分布パターン曲線上において、該正規分布の標準偏差に基づき決定される近似直線を求め、該近似直線の傾き及び切片を補正用パラメータとして導出し、該補正用パラメータを変換テーブルに記憶する補正用パラメータ導出処理と、入力部により入力された前記臨床検査データの項目値と一致する施設及び臨床検査項目について定義された前記補正用パラメータを、前記変換テーブルを参照して取得し、前記臨床検査データに対して前記補正用パラメータで規定される一次関数を適用することにより、前記臨床検査データを補正する補正処理と、補正された臨床検査データを第2の記憶装置に格納する補正後データ格納処理とを含む処理を実行させるためのものであることを特徴とするプログラムが提供される。

10

【発明の効果】

【0025】

本発明によれば、各施設間で臨床検査データに生ずる施設間誤差を効率的かつ高精度に補正することで、複数の施設で収集される臨床検査データを、医療データマイニングエンジンが本来想定する正しいデータにデータクレンジングして、1つの統合医療データベースに統合可能とすることが可能となる。

【0026】

また、臨床検査データ中に生じた検査値欠損を有効に補間して、検査値欠損がある被検者の臨床検査データであっても、医療データマイニングにおけるクラスタリング、分類等の処理の基礎とすることが可能となる。

20

【0027】

また、各施設間で相違する臨床検査データのデータ形式の相違を吸収して、医療データマイニングにおけるクラスタリング、分類等の処理の基礎とすることが可能となる。

【0028】

従って、本発明に係る臨床検査データ解析支援装置、臨床検査データ解析支援方法及びそのプログラムによれば、臨床検査データに不可避免的に生ずる施設間誤差をデータクレンジングにより解消して、高精度での医療データマイニングが実現され、EBMの向上に資する。

【発明を実施するための最良の形態】

【0029】

以下、図面を参照して、本発明の実施の形態を説明する。

30

【0030】

<本実施形態に係るネットワークシステムの構成>

図1は、本実施形態に係る臨床検査データ解析支援装置41がデータクレンジング後の臨床検査データを出力する統合医療データベース4Aを構成要素とするユビキタス統合医療情報システムのネットワークシステムの一構成例を示す。ユビキタス統合医療情報システムは、病院ないし施設1A, 1B・・・と、在宅サイト2A, 2B・・・と、ネットワーク3と、統合医療サーバ4とを具備する。

【0031】

各病院ないし施設1A, 1B・・・、各在宅サイト2A, 2B・・・はそれぞれ、例えばインターネット等のネットワーク3を介して、統合医療データベース管理装置4及び知識処理装置5とに接続される。

40

【0032】

各病院ないし施設1A, 1B・・・は、それぞれ、Medical Electronics (ME) 計測器ME-A1, ME-A2, ME-B1, ME-B2、及び医療データベース1A3、1B3とを備える。各ME計測器ME-A1, ME-A2, ME-B1, ME-B2から入力される臨床検査データは、いったん対応する医療データベース1A3、1B3に入力、蓄積される。好適には、この医療データベース1A3、1B3には、臨床検査データの他、さらに医療スタッフ等により診断データ及び診断結果データ等(以下、「アウトカムデータ」という。)が入力されてよい。各施設の医療データベース1A

50

3、1 B 3 に集積された臨床検査データないしアウトカムデータ等は、ネットワーク 3 を介して統合医療サーバ 4 により適宜のタイミングで収集され、該サーバ 4 により管理される統合医療データベース 4 A に集積される。

【0033】

各在宅サイト 2 A , 2 B . . . は、それぞれ、例えば体重計、血圧計等の健康関連計測器 C M E - A 1 , C M E - B 1 を備える。各在宅サイト 2 A , 2 B . . . は、さらに、例えばネットワークに接続されるパーソナルコンピュータ P C - 2 A , P C - 2 B により管理されるクライアントデータベース 2 A 3、2 B 3 を備えてよい。パーソナルコンピュータ P C - 2 A , P C - 2 B は、例えば在宅訪問を行なう栄養指導員が携行してもよく、或いは被検者本人が携行するものでもよい。健康関連計測器 C M E - A 1 , C M E - B 1 から入力される臨床検査データは、いったん対応するクライアントデータベース 2 A 3、2 B 3 に入力、蓄積された上所定のトリガーにより統合医療サーバ 4 にバッチ的に送信されてもよく、代替的に、パーソナルコンピュータ P C - 2 A , P C - 2 B により、適時統合医療サーバ 4 に送信され、統合医療データベース 4 A に集積されてよい。在宅サイトにおいては、常時インターネット接続が得られるとは限らないが、この場合、好適には、クライアントデータベース 2 A 3、2 B 3 と統合医療サーバ 4 とは、携行されるパーソナルコンピュータ P C - 2 A , P C - 2 B がインターネット接続を得られた際に、パーソナルコンピュータ P C - 2 A , P C - 2 B に接続されるクライアントデータベース 2 A 3、2 B 3 と統合医療サーバ 4 との間でデータベース同期処理が実行され得る。

10

【0034】

統合医療サーバ 4 は、本発明に係る臨床検査データ解析支援装置 4 1 と、データマイニング装置 4 2 と、知識処理装置 4 3 とを具備し、その外部記憶装置上に、統合医療データベース 4 A と、アウトカムデータベース 4 B と、知識データベース 4 C とを管理する。

20

【0035】

この統合医療サーバ 4 が備える本発明に係る臨床検査データ解析支援装置 4 1 は、臨床検査データが医療データベース 1 A 3、1 B 3 から受信されてから統合医療データベース 4 A に集積されるまでの間に、本発明に係る臨床検査データ解析支援処理を実行する。代替的に、臨床検査データ解析支援装置は、臨床検査データが統合医療データベース 4 A に集積された後、非同期的に、統合医療データベース 4 A から各施設で収集された臨床検査データを適宜読み出し、本発明に係る臨床検査データ解析支援処理を実行してもよい。さらに代替的に、本発明に係る臨床検査データ解析支援処理の全部又は一部は、統合医療サーバ 4 側ではなく、各病院ないし施設 1 A , 1 B . . . において実行されてもよい。

30

【0036】

統合医療データベース 4 A に集積され、本発明に係る臨床検査データ解析支援処理により各施設間誤差が解消され統合された臨床検査データは、データマイニング装置 4 2 に受け渡される。このデータマイニング装置 4 2 は、受け渡された臨床検査データ及びアウトカムデータに対してクラスタリング、分類等の各種解析処理を実行する。データマイニング装置 4 2 が出力するマイニング後の医療データは、知識処理装置 4 3 に受け渡され、知識処理装置 4 3 は、例えば公知の決定木生成アルゴリズム等を利用して、診断ルール、栄養指導ルール等の知識を生成し、知識データベース 4 C にこれらの知識を記憶する。生成された診断ルール、栄養指導ルール等は、ネットワーク 3 を介して、各病院ないし施設 1 A , 1 B . . .、各在宅サイト 2 A , 2 B . . . に診断ルールや健康アドバイスとしてフィードバックされ得る。なお、統合医療サーバ 4 において、データベースに格納される各種データは、統合疫学データベースを構成し、適宜、可視化され、検証され得るが、かかる機能は本発明の主題ではないため、詳細の説明は省略する。

40

【0037】

< 本実施形態に係る臨床管理データ解析支援装置の機能構成 >

図 2 は、本実施形態に係る臨床管理データ解析支援装置の機能構成の一例を示す。臨床管理データ解析支援装置 4 は、検査データ収集部 2 1 と、検査欠損値補間処理部 2 3 と、施設間誤差補正処理部 2 5 と、検査データ表現形変換処理部 2 9 とを具備する。代替的に

50

、臨床管理データ解析支援装置 4 において、検査欠損値補間処理部 2 3、検査データ表現形変換処理部 2 9 とは省略され得る。また、図 2 に示される検査欠損値補間処理部 2 3 と、施設間誤差補正処理部 2 5 と、検査データ表現形変換処理部 2 9 とが実行する各処理の順序は、図示された順序に限定されず、検査欠損値補間処理部 2 3、施設間誤差補正処理部 2 5、検査データ表現形変換処理部 2 9 の実行する各処理順序は、任意に変更され得る。

【 0 0 3 8 】

検査データ収集部 2 1 は、ネットワーク 3 を介して各病院ないし施設 1 A , 1 B . . . 、各在宅サイト 2 A , 2 B . . . から送信される臨床検査データを受信し、受信された臨床検査データを外部記憶装置等の記憶装置にデータベースとして構成される検査生データ記憶部 2 2 に格納する。

10

【 0 0 3 9 】

図 3 は、各施設及び在宅サイトからそれぞれ統合医療サーバ 4 に送信され、統合医療データベース 4 A 中の検査生データ記憶部 2 2 に格納される臨床検査データの構造の非限定的一例を示す。臨床検査データは、好適には、施設 A , 施設 B , 施設 C 等施設 I D ごとに 1 つのテーブルないしサブテーブルを備え、各テーブルのレコードはそれぞれ、例えば、被検者 I D、被検者性別、被検者年齢、検査日等の被検者属性項目と、図 3 の非限定的一例においては、体重、BMI 値、最大血圧、GT ( GTP ) , T c h o ( 総コレステロール )、H D L C ( H D L コレステロール ) , L D L C ( L D L コレステロール ) , T G ( 中性脂肪 ) , G L U ( 血糖 ) , H b A 1 c ( グリコヘモグロビン ) , 腹囲等の臨床検査項目とを備える。なお、図 3 に示すテーブルレイアウトは、理解の容易のため冗長性を維持したまま図示されているものに過ぎず、公知の技術を用いて、図示される項目間を階層化し、関係付け、或いはその他冗長性を排除することは適宜なし得る。また、臨床検査項目が、図 3 に図示されたものに限定されないことも容易に理解され得る。

20

【 0 0 4 0 】

検査欠損値補間処理部 2 3 は、検査生データ記憶部 2 2 から同一検査施設で分析された被検者を母集団とする臨床検査データを読み出し、図 3 ないし図 5 を参照して後述される検査欠損値補間処理を実行し、補間後の臨床検査データを外部記憶装置等の記憶装置にデータベースとして構成される補間後検査データ 2 4 に格納する。

【 0 0 4 1 】

施設間誤差補正処理部 2 5 は、補間後検査データ記憶部 2 4 から同一検査施設で分析された被検者を母集団とする補間後の臨床検査データを読み出して、補正する際の基準データとなる健常者抽出データを生成して健常者抽出データとして健常者抽出データ記憶部 2 6 に一次的ないし恒常的に記憶し、さらに変換テーブル 2 7 を参照して、図 6 ないし図 9 を参照して後述される施設間差補正処理を実行し、補正後の臨床検査データを補正後検査データ記憶部 2 8 に出力する。

30

【 0 0 4 2 】

検査データ表現形変換処理部 2 9 は、補正後検査データ 2 9 を読み出して、図 1 0 を参照して後述される表現形変換処理を実行し、表現形変換後検査データとして表現形変換後検査データ記憶部 3 0 に出力する。

40

【 0 0 4 3 】

検査欠損値が補間され、施設間誤差が補正され、さらにデータ表現形が統一された臨床検査データは、後続の臨床検査データマイニング処理に受け渡される。

【 0 0 4 4 】

< 検査欠損値補間処理詳細 >

本実施形態に係る検査欠損値補間処理部 2 3 は、所定の複数の臨床検査項目についてそれぞれ検査値が取得されているべき 1 つの被検者の臨床検査データについて、1 つ又は複数の検査項目について欠損値がある場合に、この欠損値を補間する。

【 0 0 4 5 】

図 4 A は、本実施形態に係る検査欠損値補間処理部 2 3 により参照される O R 演算につ

50

いての真理値表の一例を示し、図4Bは、本実施形態に係る検査欠損値補間処理部23により参照されるAND演算についての真理値表の一例を示す。好適には、いずれの真理値表も、本実施形態に係る臨床検査データ解析支援装置4が備える外部記憶装置或いは一時記憶装置に、例えばテーブルとして構成されてよい。

【0046】

例えば、診断ルールとして、空腹時血糖  $\geq 100$  (mg/dl) OR HbA1c  $\geq 5.2$  (%) OR 糖尿病治療を受けている；  
 ならば、血糖リスク+1、と判断するものとする。この場合、例えば空腹時血糖  $\geq 100$  (mg/dl)の値があり、糖尿病治療を受けている=Yの値があり、一方HbA1cの検査値が欠損していたとすると、本実施形態に係る検査欠損値補間処理部23は、図3のOR演算についての真理値表を参照して、HbA1cの検査値について、 $m \geq 5.2$  (%)であるmの値を生成する。

10

【0047】

図5は、本実施形態に係る検査欠損値補間処理部23により参照される臨床検査項目間の相関を行列表に示すマトリックスの一例である。図5を参照すると、例えば、LDL-CとTchoとの間の相関係数は大きく、高い正の相関性が相互に認められることが理解される。ここで、例えばLDL-Cの臨床検査値が欠損していたとすると、本実施形態に係る検査欠損値補間処理部23は、1つの被検者(同一施設、同一検査時)の臨床検査データを、検査生データ記憶部22から検索し、検索された臨床生データについて実データが得られたTchoの値を参照して、欠損する検査項目値の基準値に、図5に示される項目間の相関性に従って規定される所定の係数を乗じて、同一被検者について欠損するLDL-Cの臨床検査値を生成する。好適には、検査結果が得られている複数検査項目を説明因子として用いる公知の重回帰分析の手法を利用して、欠損値を推定してもよい。代替的に、或いはこれらに追加して、時系列的要因を考慮して、同じ被検者(かつ同一施設)の異なる検査時の臨床検査データ中で、欠損している検査値に相当する実データを得て、この実データにより、欠損値を補間してもよく、さらに、異なる検査時の臨床検査データ配列を同時に重回帰分析エンジンに与えて、補間すべき欠損値を生成してもよい。

20

【0048】

本実施形態に係る検査欠損値補間処理によれば、欠損値を含む臨床検査データについて、欠損値を補間して検査値を生成するので、臨床検査において欠損値が存在する臨床検査データについても、欠損値が補間された臨床検査データに基づき遜色なく診断ルールを導出する基礎とすることができる。

30

【0049】

<施設間差補正処理詳細>

図6ないし図9を参照して、本実施形態に係る施設間差補正処理部25が実行する施設間差補正処理を説明する。

【0050】

図6は、本実施形態に係る施設間差補正処理部25が実行する施設間差補正処理の処理手順の非限定的一例を示すフローチャートである。

【0051】

施設間差補正処理部25は、まず1つの施設、例えば施設Aについて補間後検査データ記憶部24に格納されている臨床検査データを、まず1つの検査項目に関連するフィールドについて、施設識別子をキーとして読み込む(ステップS61)。代替的に、施設間差補正処理部25は、検査生データ記憶部22から臨床検査データを読み込んでよい。

40

【0052】

読み込まれた施設Aについての臨床検査データは、当該施設Aで取得、分析されたすべての被検者を母集団とする、1つの検査項目についての臨床検査データを含む。施設間差補正処理部25は、読み込まれた臨床検査データから、健常者を母集団とする臨床検査データであると推定される臨床検査データのみを抽出し、健常者抽出データとして健常者抽出データ記憶部26に出力する(ステップS63)。

50

## 【 0 0 5 3 】

図 6 中のステップ S 6 3 の処理詳細を示す図 7 を参照して、より詳細には、施設間差補正処理部 2 5 は、以下の複数のフィルタリング処理を少なくとも 1 つ実行することにより、健常者を母集団とする臨床検査データであると推定される臨床検査データのみを抽出する。まず、男性被検者と女性被検者とでは、臨床検査項目の許容正常値範囲が異なるため、男性被検者を母集団とする群の臨床検査データと女性被検者を母集団とする群の臨床検査データとを別々に抽出する（ステップ S 6 3 1）。

## 【 0 0 5 4 】

## ( a ) 年齢層によるフィルタリング

一般に、疾病或いは定期健康診断等に起因して臨床検査を受検する被検者は、特定年齢層にピークを持つ傾向を持つ。例えば、若年層に属する被検者は、少数であってサンプルのばらつきが大きいことが推定され、他方、老年層に属する被検者の多くは何らかの疾病を持つため、検査異常値が出現する頻度が高いものと推定される。このため、本実施形態に係る施設間差補正処理部 2 5 は、臨床検査の被検者が多数存在する年齢層であって健常者の割合が高いものと推定される年齢層を抽出する。例えば図 8 に示す例によれば 4 0 歳を下限閾値、5 0 歳を上限閾値として設定することにより、4 0 歳から 5 0 歳までの年齢層の臨床検査データのみをフィルタリング処理により抽出し、健常者と推定される母集団の臨床検査データの候補とする（ステップ S 6 3 3）。

10

## 【 0 0 5 5 】

## ( b ) 検査異常値を示した被検者のサンプルを除くフィルタリング

ある被検者が複数の臨床検査項目について臨床検査を受検したものとする。この場合、他の検査項目で異常値を示した被検者は、この他の検査項目が当該検査項目と高い相関性を示すものであれば、当該検査項目についても異常値を示す確率が高いものと推定される。このため、本実施形態に係る施設間差補正処理部 2 5 は、臨床検査項目間の相関に基づき、ある検査項目で異常値を示した被検者については、該検査項目と他の検査項目とが高い相関性を示すものであれば、他の検査項目について、当該被検者の臨床検査データを除外する（ステップ S 6 3 5）。好適には、この項目間相関に基づくフィルタリング処理は、図 5 の臨床検査項目間の相関マトリックスを利用した回帰分析により実行することができる。

20

## 【 0 0 5 6 】

代替的に、項目間相関に基づき異常値を除外する方法として、例えば、図 5 のマトリックスのそれぞれに対応する 2 次元の相関図平面を 6 4 等分のメッシュで区切り、歪んだパターンそのものの領域マップを作ることにより異常値を抽出する方法として、公知の「出現実績ゾーン法」（特許文献 4、特許文献 5）を利用することができる。

30

【特許文献 4】特開平 1 1 - 0 4 5 3 0 2

【特許文献 5】特開 2 0 0 2 - 0 9 0 3 7 0 さらに、図 8 の分布図 8 0 3 に示すように、例えば年齢層ごとに規定される許容正常値範囲内に入らないサンプルを除外することも当然になし得る。

## 【 0 0 5 7 】

図 7 のステップ S 6 3 1 からステップ S 6 3 5 の処理が終了後、検査項目ごとに、健常者と推定されるサンプルのみを母集団とする臨床検査データが抽出されて健常者抽出データ記憶部 2 6 に格納され、健常者のみについての臨床検査データ分布パターンが生成され得る（ステップ S 6 3、S 6 3 7）。

40

## 【 0 0 5 8 】

図 6 に戻り、ステップ S 6 5 において、施設間差補正処理部 2 5 は、健常者抽出データ記憶部 2 6 から、1 つの検査項目について健常者のものと推定される母集団の臨床検査データを読み出し、この健常者の臨床検査データを正規化する（ステップ S 6 5）。この正規化には、例えばべき乗変換を用いることができる。

## 【 0 0 5 9 】

臨床検査項目の検査値の分布型は一様ではなく、正規分布、対数正規分布、両者の中間

50

にある平方根正規分布、3乗根正規分布等があるが、一般に、変換元データ群  $x(p, i)$  に対して、べき乗値  $k(p)$  と変換原点  $a(p)$  とを指定して、べき乗変換を行なうと、下記式1のとおり、その変換後のデータ群  $X(p, i)$  の分布が正規分布を近似するものとなることが知られている。

【0060】

[数1]

$k(p) = 0$  のとき、一般べき乗変換

$$X(p, i) = (x(p, i) - a(p))^{k(p) - 1} / k(p)$$

$k(p) = 0$  のとき、対数変換

$$X(p, i) = \log(x(p, i) - a) \quad (\text{式1})$$

10

特許文献3及び非特許文献1は、このべき乗変換によるデータ配列の正規化手法を開示する。また、べき乗変換パラメータの最適値を、例えばBox-Cox変換方式によりシンプレックス法で導出する手法は、例えば、<http://aoki2.si.gunma-u.ac.jp/R/Box-Cox-transformation2.html> に関示されている。

【特許文献3】特開2007-199787号公報

【非特許文献1】Ichihara, K. and Kawai, T: Determination of reference intervals for 13 plasma proteins based on IFCC international reference preparation (CRM470) and NCCLS proposed guideline (C28-P, 1992): trial to select reference individuals by results of screening tests and application of maximal likelihood method. J Clin Lab Anal. 10(2): 110-7, 1996.

20

【0061】

図8を参照して、施設間差補正処理部25は、補間後検査データ記憶部24から、例えば臨床検査項目HbA1cについての施設Aにおける全被検者の臨床検査データを読み出す。補間後検査データ記憶部24に格納され、読み出された臨床検査データは、年齢をX軸、検査値をY軸として2次元上プロットしたものが801として図示され、その分布パターンを検査値をX軸として図示したものが802として図示されている。

30

【0062】

図8においては、一例として、施設間差補正処理部25が、40歳から50歳までの年齢層を抽出したものと図示されている。施設間差補正処理部25は、この他にも上記(a)及び(b)のフィルタリング処理を適宜実行することにより、サンプル数が絞られた、健常者と推定される母集団のみについての臨床検査データを抽出し、健常者抽出データ記憶部26に格納する。図8における804は、802のパターンに対してフィルタリング処理を適用することにより抽出された健常者と推定される母集団のみについての臨床検査データの分布パターン、すなわちノイズ除去後の分布パターンを図示するものである。図8における805は、ノイズ除去後の分布パターン804に対して、上記の例えばべき乗変換を用いた正規化処理を適用することにより変換された健常者と推定される母集団のみについての臨床検査データの正規化分布パターンを図示するものである。

40

【0063】

図8の健常者と推定される母集団のみについての臨床検査データの正規化分布パターンが得られると、図6に戻り、ある臨床検査項目について、正規化された健常者の正規分布パターンを基準分布パターンとして、補正用パラメータを導出する(ステップS67、ステップS69)。

【0064】

一般に、正規分布パターンの平均値を  $\mu$ 、標準偏差を  $\sigma$  とすると、X軸上、平均値  $\mu \pm 2\sigma$  の範囲に、サンプルの95.44%が確率的に含まれることが知られている。正規分布パターン805は、確率密度関数が示すカーブであり、 $x = \mu \pm 2\sigma$  のとき正規分布パタ

50

ーン上変曲点を持つ。図 8 を参照して、施設間差補正処理部 2 5 は、例えば、正規分布パターン 8 0 5 上の  $x = \mu - 2\sigma$  の偏位点における近似直線 8 0 5 a を求め（図 6、ステップ S 6 7）、求められた 1 次関数である直線の傾き  $a$ （一次係数）及び切片  $b$ （定数項）を、施設間誤差の補正用パラメータとして導出する（ステップ S 6 9）。一例として、近似直線 8 0 5 a は、例えば  $\mu - 2\sigma$  の偏位点における分布パターンカーブへの接線として得ることができるが、これに限定されず、例えば  $x = \mu \pm n\sigma$ （ $1 \leq n \leq 3$ ）の間の所定の偏位点において接線が求められてもよい。こうして求められた一次関数の傾き  $a$  は、正規化分布パターンの所定幅内についてのばらつきを示す指標となる。

#### 【0065】

この補正用パラメータ、すなわち施設 A について、ある臨床検査項目についての基準分布パターン 8 0 5 の中央（平均値中心）から例えば 2 標準偏差だけ偏位した偏位点における近似直線の傾き、及び切片は、施設間誤差補正用の変換テーブル 2 7 に登録される（ステップ S 6 7）。

10

#### 【0066】

このようにして施設ごと、かつ臨床検査項目ごとに、健常者の臨床検査データに基づいて補正用パラメータが導出された後、施設間誤差補正処理部 2 5 は、補間後検査データ記憶部 2 4 から、施設ごと、かつ臨床検査項目ごとに、全被検者を母集団とする補間処理後の臨床検査データを読み込み、読み込まれた施設識別子及び臨床検査項目識別子と一致する識別子を有する施設及び検査項目について定義された傾き  $a$  及び切片  $b$  を補正用パラメータとして適用することによって補間処理後の臨床検査データが有する施設間誤差を補正し（ステップ S 7 3）、補正された臨床検査データを、補正後検査データ記憶部 2 8 に格納する。

20

#### 【0067】

図 9 を参照して、より詳細には、例えば、病院 A（9 1）内の分析施設 A で分析された検査項目 1（9 1 1）についての臨床検査データに対しては、同じ分析施設識別子 = A 及び検査項目識別子 = 1 である補正パラメータとして、変換テーブル 2 7 から、1 カラム目の傾き = 0.9、切片 = 1、が読み出される。施設間誤差補正処理部 2 5 は、臨床検査データの全サンプルについて、下記式 2 において  $a = 0.9$ 、 $b = 1$  を代入し、補正前の臨床検査データ  $X$  から、補正後の臨床検査データ  $Y$  を得る。

#### 【0068】

$$Y = aX + b \quad (\text{式 2})$$

30

図 6 に戻り、ステップ S 6 1 からステップ S 7 3 の処理は、全臨床検査項目について、さらに全施設について、繰り返し実行される。

#### 【0069】

< 検査データ表現形変換処理詳細 >

各施設で収集される同一の臨床検査項目について得られた検査値は、各施設間で異なる表現形で記述されている。本実施形態に係る検査データ表現形変換処理部 2 9 は、これら相違する表現形で記述された検査値を、統一された表現形式へと変換する。好適には、検査データ表現形変換処理部 2 9 は、臨床検査値のデータ型を判定し、臨床検査データの文字型で記述された検査値を、数値型で記述された連続値へと変換する。数値型で記述された連続値の検査値は、検索条件に一致することが容易に判定でき、また他の検査値との間で大小関係が容易に把握できる。このため、統合された臨床検査データのデータマイニング処理がシームレスに実現される。

40

#### 【0070】

図 1 0 は、本実施形態に係る検査データ表現形変換処理部 2 9 が参照する連続値変換テーブルの一例を示す。臨床検査結果値は、表現形フィールド 1 0 A の表現形で記述されている場合、連続値フィールド 1 0 B の対応する連続値に変換される。好適には、表現形変換後検査データ記憶部 3 0 に格納される表現形変換後の臨床検査データは、検査結果として、収集時に記述されていた文字型での結果値を格納する文字型フィールドと、連続値への変換後の数値型での結果値を格納する数値型フィールドとの双方を有する。

50

## 【 0 0 7 1 】

< 施設間差補正処理における健常者臨床検査データ抽出処理の変形例 >

図 1 1 ないし図 1 3 を参照して、図 6 のステップ S 6 3 に示される、施設間差補正処理部 2 5 が行なう健常者のものと推定される検査データのみを検査項目ごと抽出する処理の変形例を説明する。

## 【 0 0 7 2 】

本変形例は、図 7 に替えて、図 6 のステップ S 6 3 の詳細処理として、図 1 1 に示す処理手順を実行する。

## 【 0 0 7 3 】

まず、男女別に臨床検査データを抽出し（ステップ S 6 3 1）、特定年齢層の臨床検査データのみを抽出する処理（ステップ S 6 3 3）は、図 7 に示される上記実施形態における処理と同様である。

## 【 0 0 7 4 】

次に、図 1 1 のステップ S 1 1 3 1 において、ステップ S 6 3 3 により抽出された特定年齢層の臨床検査データは、例えばべき乗変換等の手法を用いて、正規分布に変換される（ステップ S 1 1 3 1）。ステップ S 1 1 3 3 において、正規分布化された特定年齢層の臨床検査データ中、例えば  $x = \mu \pm 2$  の 2 標準偏差以内に入らない臨床検査データが、異常値を示すものとして除外される（ステップ S 1 1 3 3）。

## 【 0 0 7 5 】

次に、ステップ S 1 1 3 5 において、相関性のある  $n$  項目間（ $n - 1$  の整数）、例えば 2 項目間、での  $n$  次元等偏差楕円内に入らない臨床検査データが、異常値を示すものとして除外される（ステップ S 1 1 3 5）。なお、ステップ S 1 1 3 5 の処理を単項目（1 次元）について行なうこともできる（ $n = 1$ ）。

## 【 0 0 7 6 】

図 1 2 は、2 つの臨床検査項目、例えば ALT（GPT）と AST（GOT）間での 2 次元正規分布（確率分布）を示し、図 1 2 中の楕円は、その長軸が回帰直線を、その中心が平均値を、それぞれ示す。例えば、図 1 2 の楕円は、2 次元正規分布における 2 標準偏差（2）楕円を示すため、楕円内にはサンプルの 95.44% が確率的に含まれることになる。ステップ S 1 1 3 5 において、この楕円内に入らない臨床検査データが除外される。なお、データ除外の閾値を確定する楕円は、2 に限定されることなく、例えば  $n$ （ $1 < n < 3$ ）等から任意に規定された楕円を用いてよい。

## 【 0 0 7 7 】

上記ステップ S 1 1 3 1 から S 1 1 3 5 までの処理を、与えられた全臨床検査項目間について、除外すべき臨床検査データがなくなるまで、繰り返し処理する（ステップ S 1 1 3 7）。すなわち、ステップ S 1 1 3 1 から S 1 1 3 5 までの処理は、入力された前記臨床検査データの臨床検査項目との間で所定の閾値以上の相関性を有する全ての他の臨床検査項目と、当該臨床検査項目との間で、除外すべき臨床検査データがなくなるまで、繰り返される。

## 【 0 0 7 8 】

図 1 3 は、図 1 2 に示される 2 次元正規分布に対して、ステップ S 1 1 3 1 から S 1 1 3 5 までの処理を繰り返して得られた、異常値が十分に除外された臨床検査データのみを母集団とする、臨床検査項目 ALT（GPT）と AST（GOT）との間の 2 次元分布パターンを示す。本変形例においては、この図 1 3 に例示されるステップ S 1 1 3 1 から S 1 1 3 5 までの処理を繰り返して得られた、異常値が十分に除外された臨床検査データのみを母集団とする臨床検査データを、健常者のものと推定して、図 6 のステップ S 6 5 に出力する。

## 【 0 0 7 9 】

図 6 のステップ S 6 5 に戻り、健常者のものと推定される臨床検査データは、再度例えばべき乗変換等の手法により、正規分布化される（ステップ S 6 5）。

## 【 0 0 8 0 】

10

20

30

40

50



以降、図6におけるステップS67ないしステップS73の処理は、上記実施形態における処理と同様である。

【0081】

<本実施形態に係る臨床検査データ解析支援装置のハードウェア構成>

図14は、本実施形態による臨床検査データ解析支援装置44のハードウェア構成の非限定的一例を示すブロック図である。図14に示されるコンピュータ装置110である臨床検査データ解析支援装置44において、CPU111は、ROM114および/またはハードディスクドライブ116に格納されたプログラムに従い、RAM115を一次記憶用ワークメモリとして利用して、システム全体を制御する。さらに、CPU111は、マウス112aまたはキーボード112を介して入力される利用者の指示に従い、ハードディスクドライブ116に格納されたプログラムに基づき、本実施形態に係る臨床検査データ解析処理を実行する。ディスプレイインタフェース113には、CRTやLCDなどのディスプレイが接続され、CPU111が実行する臨床検査データ解析処理の入力待ち受け画面、処理経過や処理結果、各種画像などが表示される。リムーバブルメディアドライブ117は、主に、リムーバブルメディアからハードディスクドライブ116へファイルを書き込んだり、ハードディスクドライブ116から読み出したファイルをリムーバブルメディアへ書き込む場合に利用される。リムーバブルメディアとしては、フロッピディスク(FD)、CD-ROM、CD-R、CD-R/W、DVD-ROM、DVD-R、DVD-R/W、DVD-RAMやMO、あるいはメモリカード、CFカード、スマートメディア、SDカード、メモリスティックなどが利用可能である。

10

20

【0082】

プリンタインタフェース118には、レーザビームプリンタやインクジェットプリンタなどのプリンタが接続される。ネットワークインタフェース119は、コンピュータ装置をネットワークへ接続するためのインターフェースである。

【0083】

なお、上記各実施形態に係る臨床検査データ解析装置44における入力部は、マウス112aあるいはキーボード112に限定されることなく、任意のポインティングデバイス、例えばトラックボール、トラックパッド、タブレットなどを適宜用いることができる。携帯情報端末を本実施形態に係る臨床検査データ解析装置44への送信装置として用いる場合には、入力部をボタンやモードダイヤル等で構成してもよい。

30

【0084】

また、図14に示した上記各実施形態に係る臨床検査データ解析装置のハードウェア構成は一例に過ぎず、その他の任意のハードウェア構成を用いることができることはいうまでもない。

【0085】

殊に、上記各実施形態に係る臨床検査データ解析処理の全部又は一部は、上記コンピュータ端末装置100あるいはPDA等の携帯情報端末装置等によって実現されてもよく、コンピュータ端末装置等とサーバー装置とをBluetooth(登録商標)等の無線、あるいはインターネット(TCP/IP)、公共電話網(PSTN)、統合サービス・デジタル網(ISDN)等の有線通信回線で相互接続した、インターネットあるいは任意の周知のローカル・エリア・ネットワーク(LAN)またはワイド・エリア・ネットワーク(WAN)からなるネットワークシステムによって臨床検査データ解析処理が実現されてもよい。

40

【0086】

以上のとおり、本実施形態によれば、各施設間で臨床検査データに生ずる施設間誤差を効率的かつ高精度に補正することで、複数の施設で収集される臨床検査データを、医療データマイニングエンジンが本来想定する正しいデータにデータクレンジングして、1つの統合医療データベースに統合可能とすることが可能となる。

【0087】

また、臨床検査データ中に生じた検査値欠損を有効に補間して、検査値欠損がある被検

50

者の臨床検査データであっても、医療データマイニングにおけるクラスタリング、分類等の処理の基礎とすることが可能となる。

【0088】

また、各施設間で相違する臨床検査データのデータ形式の相違を吸収して、医療データマイニングにおけるクラスタリング、分類等の処理の基礎とすることが可能となる。

【0089】

本発明の範囲は、図示され記載された例示的な実施形態に限定されるものではなく、本発明が目的とするものと均等な効果をもたらすすべての実施形態をも含む。さらに、本発明の範囲は、請求項1により画される発明の特徴の組み合わせに限定されるものではなく、すべての開示されたそれぞれの特徴のうち特定の特徴のあらゆる所望する組み合わせによって画されうる。

10

【0090】

例えば、施設間で被検者の母集団に有意な地域差等が観察される場合には、これを考慮して、後続のデータマイニング及び知識処理を行なってもよい。

【図面の簡単な説明】

【0091】

【図1】本発明の一実施形態に係る臨床検査データ解析支援装置を含むユビキタス医療サービスシステムのネットワーク構成の一例を示すブロック図である。

【図2】本発明の一実施形態に係る臨床検査データ解析支援装置の機能構成の一例を示す機能ブロック図である。

20

【図3】検査生データ記憶部22に格納される臨床検査データのデータレイアウトの一例を示す模式図である。

【図4】本発明の一実施形態に係る臨床検査データ解析支援装置の欠損値補間処理部2が行なう欠損値補間処理が参照するOR演算およびAND演算の審理値表の一例である。

【図5】臨床検査項目間の相関行列の一例を示す模式図である。

【図6】本発明の一実施形態に係る臨床検査データ解析支援装置の施設間誤差補正処理部23が行なう補正処理の処理手順の一例を示すフローチャートである。

【図7】図6のステップS63により実行される処理の詳細処理手順の一例を示すフローチャートである。

【図8】補正処理部3が各施設間の臨床検査データの補正を行なうために算出する健常者臨床検査データに基づく補正用換算係数算出の手順を説明する模式図である。

30

【図9】各施設ないし病院における臨床検査項目と、変換テーブル項目との対応を示す模式図である。

【図10】本発明の一実施形態に係る臨床検査データ属性値の連続値への変換用テーブルの項目の一例を示す模式図である。

【図11】本実施形態の変形例による、健常者臨床検査データ抽出処理の詳細処理手順の一例を示すフローチャートである。

【図12】本実施形態の変形例による異常値抽出処理を適用する前の2項目間での2次元正規分布の一例を示す図である。

【図13】図12の2次元正規分布に対して、本実施形態の変形例による異常値抽出処理を適用して得られる、健常者のものと推定される臨床検査データについてのみの、2項目間での2次元正規分布の一例を示す図である。

40

【図14】本発明の各実施形態に係る臨床検査データ解析支援装置のハードウェア構成の一例を示す図である。

【符号の説明】

【0092】

検査データ収集部 21

検査生データ記憶部 22

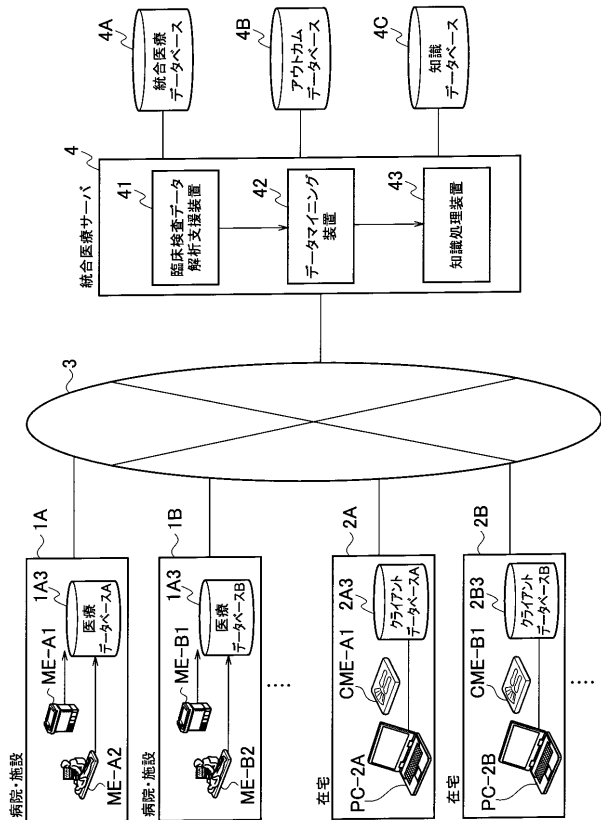
検査欠損値補間処理部 23

補間後検査データ記憶部 24

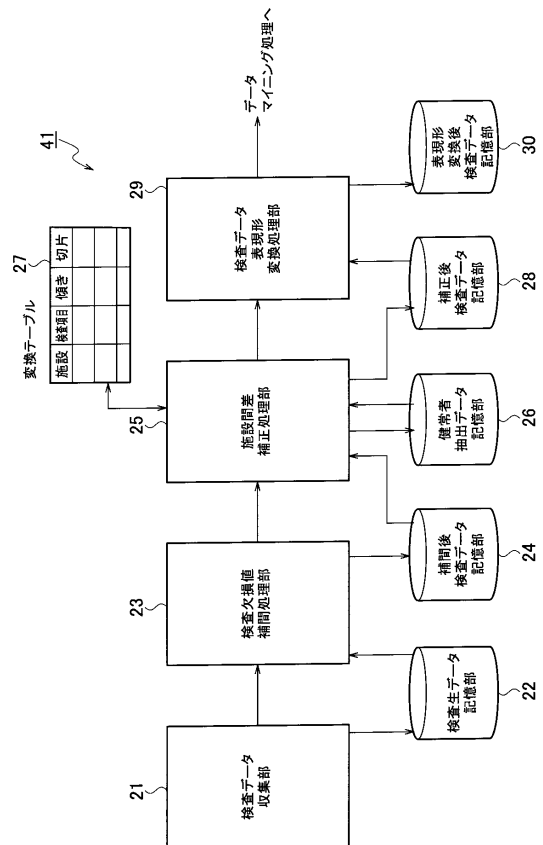
50

- 施設間差補正処理部 25
- 健常者抽出データ記憶部 26
- 変換テーブル 27
- 補正後検査データ記憶部 28
- 検査データ表現形変換処理部 29
- 表現形変換後検査データ記憶部 30

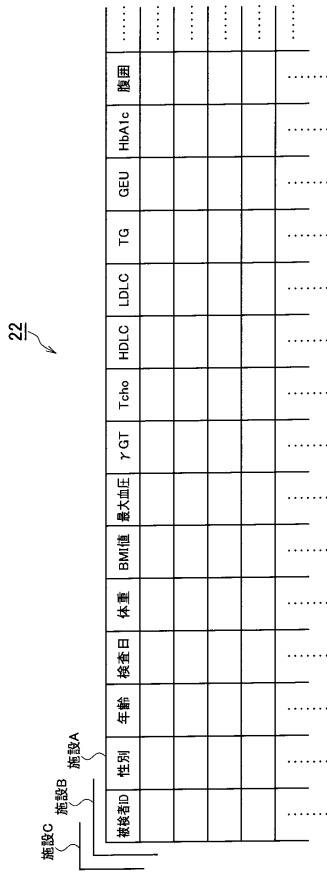
【 図 1 】



【 図 2 】



【 図 3 】



【 図 4 】

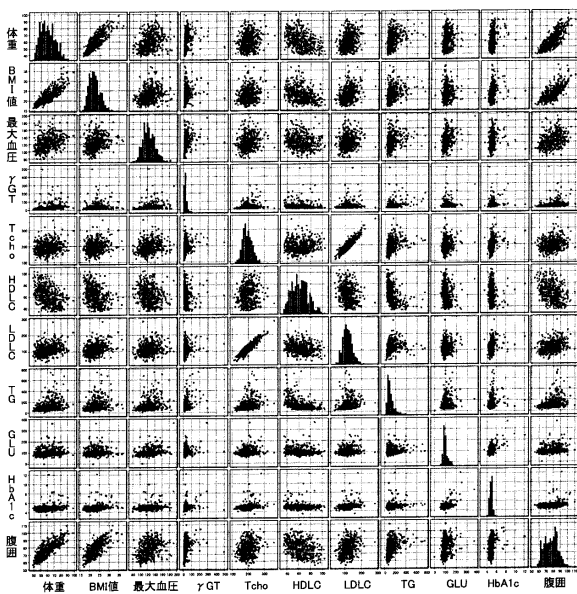
(A)

	401	402	403
A	B	OR	
0	0	0	
0	1	1	
0	null	0	
1	0	1	
1	1	1	
1	null	1	
null	0	0	
null	1	1	
null	null	null	

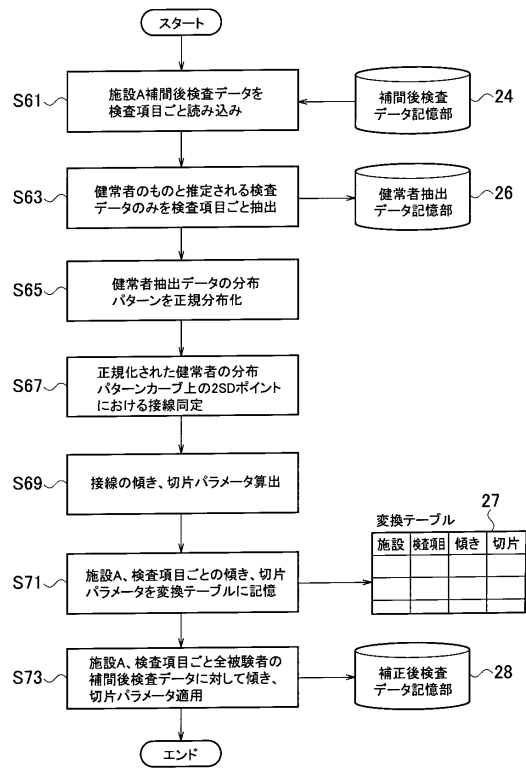
(B)

	404	405	406
A	B	AND	
0	0	0	
0	1	0	
0	null	0	
1	0	0	
1	1	1	
1	null	1	
null	0	0	
null	1	1	
null	null	null	

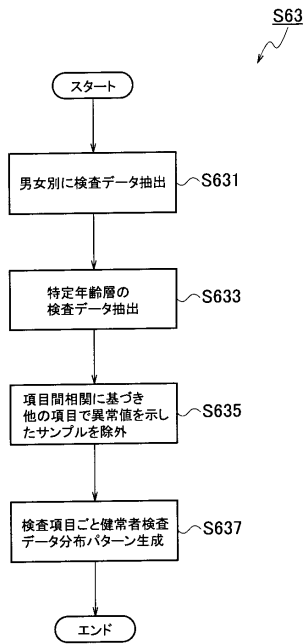
【 図 5 】



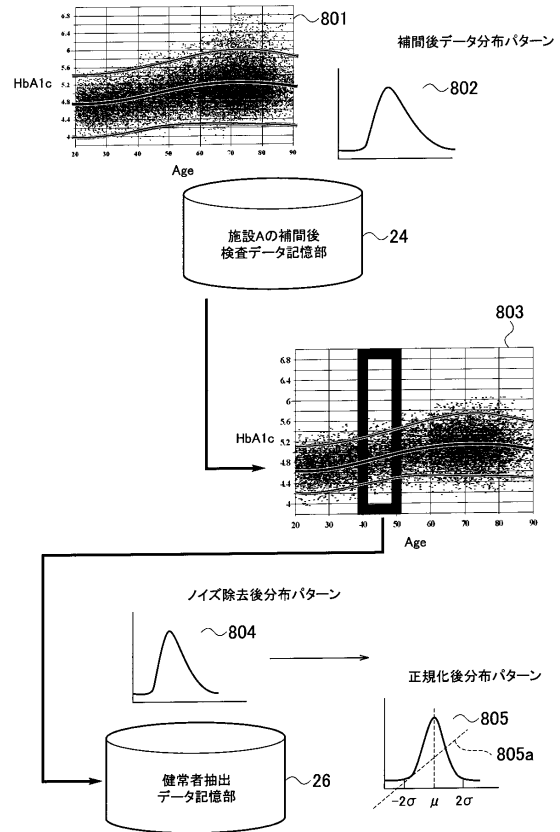
【 図 6 】



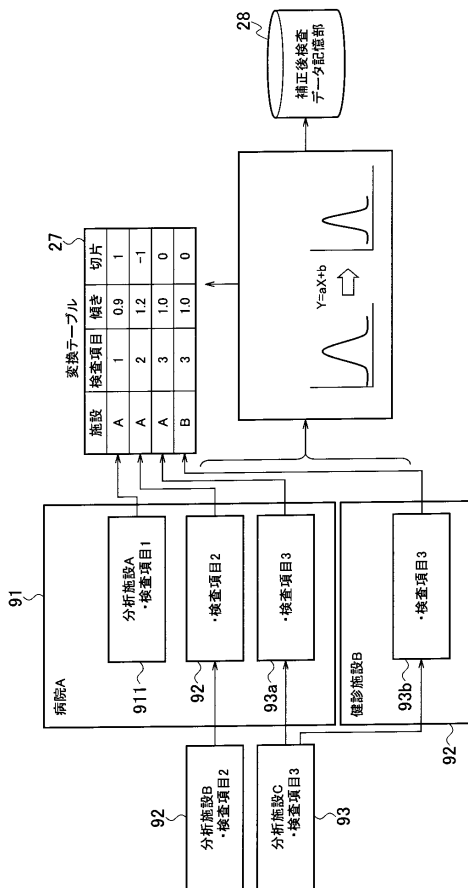
【 図 7 】



【 図 8 】



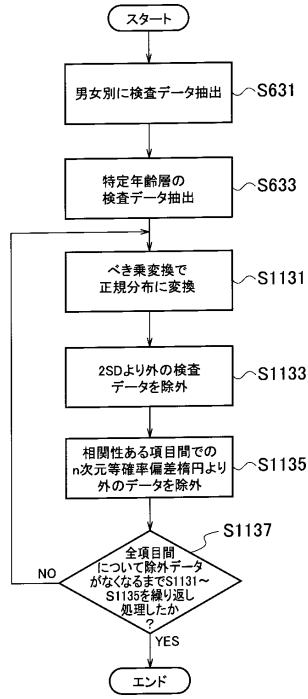
【 図 9 】



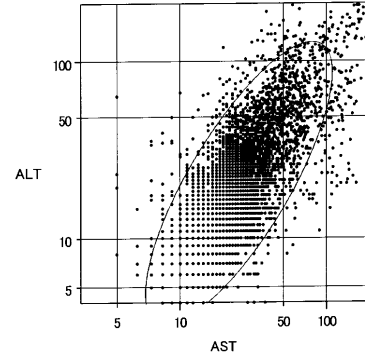
【 図 10 】

10A	10B
表現形	連続値
>100	200
≥100	200
<100	50
≤100	50
-	-1
±	0.5
+	1
1+	1
2+	2
3+	3
+1	1
+2	2
+3	3
陽性	1
陰性	-1
[空白]	null

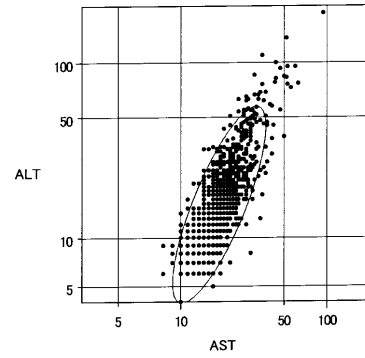
【 図 1 1 】



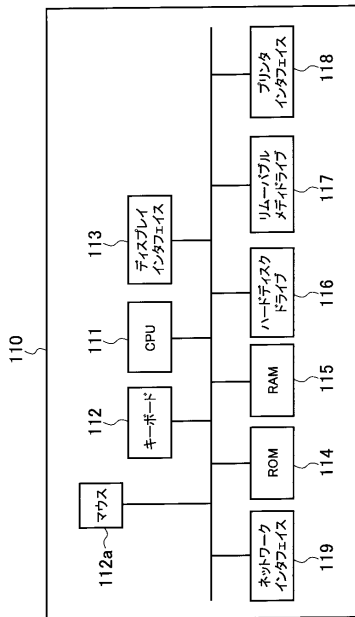
【 図 1 2 】



【 図 1 3 】



【 図 1 4 】



---

フロントページの続き

(72)発明者 片岡 浩巳

高知県南国市岡豊町小蓮 国立大学法人高知大学医学部内

(72)発明者 相良 祐輔

高知県高知市曙町二丁目5番1号 国立大学法人高知大学内

(72)発明者 曾根原 登

東京都国分寺市東恋ヶ窪2-14-16

Fターム(参考) 4C117 XA07 XB09 XF01 XF14 XF16 XF17 XF19 XF22 XF27 XG01  
XG02 XH16 XJ03 XJ05 XJ17 XJ23 XJ34 XL01 XL11 XM01  
XM05 XQ07