

明 細 書

発明の名称：

多項目同時測定データを活用した高精度診断システム、高精度診断方法及びプログラム

技術分野

[0001] 本発明は、多項目同時測定データを活用し、既知の変異と、未知の変異と、を検出し可視化する高精度診断方法及びプログラムに関する。

背景技術

[0002] 例えば、塩基配列の変異、つまり遺伝子変異、を検出するには、全ゲノムシーケンス (Whole Genome Sequencing: WGS) を用いてゲノムを解析する方法が最も包括的な手法である。しかしながら、WGSは高度な解析機器、技術、高価な試薬を必要とするため、簡便で迅速に遺伝子の変異を定量的に検出し、診断する技術が求められている。

[0003] 簡便で迅速に変異箇所を定量的に検出する技術の一つとして、例えば、DNAマイクロアレイを用いた多項目同時遺伝子関連検査が挙げられる。多項目同時遺伝子関連検査により、既知の遺伝子変異の有無が変異箇所ごとに多項目同時測定データとして検出することが可能である。

[0004] 従来、前述した多項目同時測定データから変異箇所の有無が検出され、既知の典型的な変異パターンと合致するか否かで、既知の変異か未知の変異かが診断されている。

[0005] 例えば、特許文献1では、リード配列 (シーケンシング対象塩基配列の配列情報) の各々を参照配列と比較して1つ1つのリード配列から塩基の変異を検出し、その結果を解析することで、変異のパターン及びその頻度を算出する解析方法が開示されている。

[0006] また、特許文献2では、遺伝子変異の傾向を調べるために、配列のより広い領域の変異を検出し、それらの変異パターンを解析して、広い配列領域内での遺伝子変異の傾向を解析することを優先するという考え方が開示されて

いる。

- [0007] さらに、非特許文献1では、ウイルスの遺伝子から進化距離をp距離で求め、多次元尺度構成法を用いてユークリッド空間に布置する方法が開示されている。

先行技術文献

特許文献

- [0008] 特許文献1：特開2018-130114

特許文献2：特開2021-126125

非特許文献

- [0009] 非特許文献1：特集学生の研究活動報告—国内学会大会・国際会議参加記22, [online], [令和4年5月20日検索], インターネット, <<https://www.rikou.ryukoku.ac.jp/journal/journal69/RJ69S-19.pdf>>

発明の概要

発明が解決しようとする課題

- [0010] しかしながら、特許文献1、特許文献2、非特許文献1、に記載の技術では、既知の遺伝子変異を判別と、未知の遺伝子変異を検出と、遺伝子変異の差異の進化距離を可視化することは可能であるが、

既知の変異と、未知の変異とを分類して可視化した上で、未知の変異が既知のどの変異に近いかを高精度で示し、未知の変異であるという解析結果の判断根拠を視覚的に示すことはできない。

- [0011] 本発明は、多項目同時測定データを活用し、既知の変異と、未知の変異と、を高精度に分類し、分類した結果を解析して可視化した上で判断根拠を視覚的に示す高精度診断システムを提供する。

課題を解決するための手段

- [0012] 本発明では、以下のような解決手段を提供する。

- [0013] 第1の特徴に係る発明は、多項目同時測定データを活用し、既知の試料と、未知の試料と、を検出し可視化する高精度判定システムであって、

未知の試料の多項目同時測定データと、既知の試料の多項目同時測定データと、を少なくとも取得する取得部と、

取得した前記未知の試料の多項目同時測定データと、前記既知の試料の多項目同時測定データと、の近似性を解析して定量化し、既知の試料と傾向が異なる未知の試料を分類し、分類データを生成する分類部と、

生成された前記分類データを次元圧縮して２次元散布図を判定データとして生成する判定部と、

前記未知の多項目同時測定データと、前記既知の多項目同時測定データと、前記分類データと、前記判定データと、を少なくとも格納する格納部と、

格納した当該各種データを少なくとも提供する提供部と、を備えることを特徴とする高精度判定システムを提供する。

[0014] 第１の特徴に係る発明によれば、多項目同時測定データを活用し、既知の試料と、未知の試料と、を検出し可視化するために、未知の試料の多項目同時測定データと、既知の試料の多項目同時測定データと、を少なくとも取得し、取得した前記未知の試料の多項目同時測定データと、前記既知の試料の多項目同時測定データと、の近似性を解析して定量化し、既知の試料と傾向が異なる未知の試料を分類し、分類データを生成し、生成された前記分類データを次元圧縮して２次元散布図を判定データとして生成し、未知の多項目同時測定データと、前記既知の多項目同時測定データと、前記分類データと、前記判定データと、を少なくとも格納し、格納した当該各種データを少なくとも提供する。

[0015] 第１の特徴に係る発明は、高精度診断システムのカテゴリであるが、高精度診断方法、及びプログラムであっても同様の作用、効果を奏する。

[0016] 第２の特徴に係る発明は、第１の特徴に係る発明である高精度診断システムであって、

前記多項目同時測定データおよび前記既存データは塩基配列に係るデータであって、前記多項目同時測定データを活用し、既知の遺伝子変異と、未知の遺伝子変異と、を検出し可視化する高精度診断システムであって、

前記多項目同時測定データと、既知の塩基配列データと、を少なくとも取得する取得部と、

取得した前記多項目同時測定データと、前記塩基配列データと、からk近傍法を用いて分類し、局所外れ値因子法を用いて既知の遺伝子変異と傾向が異なる未知の遺伝子変異を検出し、多項目変異部位パターンデータを生成する学習モデルを作成する学習モデル作成部と、

前記学習モデルに基づいて、生成された前記多項目変異部位パターンデータと、前記学習モデルの学習データと、から多次元尺度構成法を用いて2次元散布図を診断データとして生成する診断部と、

前記多項目同時測定データと、前記既存データと、前記多項目変異部位パターンデータと、前記診断データと、を少なくとも格納する格納部と、

格納した当該各種データを少なくとも提供する提供部と、
を備える高精度診断システムを提供する。

[0017] 第2の特徴に係る発明によれば、第1の特徴に係る発明であって、前記分類部は、k近傍法（*k-Nearest Neighbor*, *kNN*）を用いて取得した前記未知の試料の多項目同時測定データと、前記既知の試料の多項目同時測定データと、の近似性を解析して定量化し、局所外れ値因子法（*Local Outlier Factor*, *LOF*）を用いて前記既知の試料と前記傾向が異なる未知の試料を分類し、前記判定部は、多次元尺度構成法（*MDS: Multi-Dimensional Scaling*）を用いて、前記分類データから2次元散布図を判定データとして生成する。

[0018] 第3の特徴に係る発明は、第1の特徴または第2の特徴に係る発明であって、前記多項目同時測定データは、前記未知の試料および前記既知の試料から得た核酸配列に係るデータである。

[0019] 第4の特徴に係る発明は、第1の特徴または第2の特徴に係る発明である高精度診断システムであって、既知の新型コロナウイルスと、未知の新型コロナウイルスと、を判定する高精度診断システムを提供する。

[0020] 第4の特徴に係る発明によれば、第1の特徴または第2の特徴に係る発明

である高精度診断システムにおいて、既知の新型コロナ変異ウイルスと、未知の新型コロナ変異ウイルスと、を診断する。

[0021] 第5の特徴に係る発明は、第3の特徴に係る発明である高精度診断システムであって、前記未知の試料および前記既知の試料は、がん細胞である高精度診断システムを提供する。

[0022] 第6の特徴に係る発明は、前記多項目同時測定データは、特定タンパク質の存在量を測定したデータであって、病態を判定する高精度診断システムを提供する。

[0023] 第7の特徴に係る発明は、第1の特徴に係る発明であって、定量化した一つ乃至複数の項目を、一つ乃至複数のパターンとして認識して分類精度を向上する学習モデルを作成する学習モデル作成部と、

前記学習モデルに基づいて、前記分類データからパターンデータを生成するパターンデータ生成部と、を更に備え、

前記格納部は、前記パターンデータおよび前記学習モデルを少なくとも格納し、前記提供部は、前記パターンデータを少なくとも提供する。

発明の効果

[0024] 本発明によれば、多項目同時測定データから、既知の変異と、未知の変異と、を高精度に検出し、検出した結果を解析して可視化した上で判断根拠を視覚的に示すことが可能な高精度診断システムを提供することが可能となる。

図面の簡単な説明

[0025] [図1]図1は、高精度診断システム1の概要図である。

[図2]図2は、高精度診断システム1の構成図である。

[図3]図3は、高精度診断システム1のコンピュータ2が実行する高精度診断処理の手順を示すフローチャート図である。

[図4]図4は、ユーザ端末3が表示するコンピュータ2が作成した多項目変異部位パターンデータの表示画面の一例である。

[図5]図5は、ユーザ端末3が表示するコンピュータ2が作成した診断データ

の表示画面の一例である。

発明を実施するための形態

[0026] 以下、本発明を実施するための最良の形態について図を参照しながら説明する。なお、これらは一例であって、本発明の技術的範囲は、これに限られるものではない。

[0027] [高精度診断システム 1 の概要]

図 1 は、高精度診断システム 1 の概要を説明するための図である。

[0028] 図 1 に示すように、高精度診断システム 1 は、例えば、遺伝子変異の診断に利用するためのコンピュータシステムであり、コンピュータ 2 と、ユーザ端末 3 と、を少なくとも含むシステムである。コンピュータ 2 とユーザ端末 3 とは、ネットワーク 4 を介して通信可能である。

[0029] 高精度診断システム 1 のコンピュータ 2 は、高精度診断システム 1 を制御するためのコンピュータであって、例えば、物理的に 1 台または複数のコンピュータで実現されてもよいし、クラウドコンピュータのように仮想的な装置で実現されてもよい。

[0030] 高精度診断システム 1 のユーザ端末 3 は、コンピュータ 2 に多項目同時測定データ 101 と、塩基配列データ 102 と、多項目変異部位パターンデータ 103 と、診断データ 104 と、を送受信するための端末であって、パソコンやノートパソコン、スマートフォンやタブレット端末等の携帯端末、スマートグラス等のヘッドマウントディスプレイやスマートウォッチといったウェアラブル端末等であってもよい。

[0031] まず、コンピュータ 2 の取得部 201 は、ユーザ端末 3 から、多項目同時測定データ 101 と、塩基配列データ 102 と、を少なくとも取得する（ステップ S01）。本発明において、多項目同時測定データ 101 とは、例えば、マイクロアレイなどの多項目同時遺伝子関連検査により得られたデータであるものとする。ここで、多項目同時測定データ 101 は、遺伝子の変異箇所ごとに閾値が設定され、変異の有り無しという 0 か 1 かというデータに置換されたものであってよいし、変異箇所の変異型によって定量的に置換され

たものであってもよい。また、塩基配列データ102とは、例えば、WGSを用いてゲノム解析された、遺伝子に変異した変異体に対して本来的にその生物あるいは微生物が備える遺伝子型を有する野生株（Wild Type Strain: WT）の塩基配列データと、既知の変異遺伝子を有する変異株（Mutant: MUT）の塩基配列データと、を少なくとも含む塩基配列データであるものとする。ここで、塩基配列データ102は、多項目同時測定データ101に準じて遺伝子の変異箇所ごとに定量性の有るデータに置換されたものであってよい。なお、該データの取得タイミングについては限定されず、当該データのデータ形式も限定されない。

[0032] 次に、コンピュータ2の学習モデル作成部202は、取得した少なくとも、多項目同時測定データ101と、塩基配列データ102と、からkNNを用いて分類し、LOFを用いて既知の遺伝子変異と傾向が異なる未知の遺伝子変異を検出し、多項目変異部位パターンデータ103を生成する学習モデル10を作成する（ステップS02）。ここで、多項目変異部位パターンデータ103とは、例えば、変異箇所をパネルとして変異箇所の変異型によって定量的に置換されたバーコードパターンであってよい。また、当該データが既知の遺伝子変異であった場合、塩基配列データ102を参照して既知の変異型の名称を紐づけて多項目変異部位パターンデータ103が作成されてもよい。また、当該データが未知の遺伝子変異であった場合、例えば、Atypicalなどの予め定義づけられた名称を紐づけて作成されてもよい。

[0033] 次に、コンピュータ2の診断部203は、学習モデル10に基づいて生成された多項目変異部位パターンデータ103と、学習モデル10の学習データと、からMDSを用いて作成した2次元散布図を診断データ104として生成する（ステップS03）。ここで、診断データ104とは、例えば、学習モデル10が生成した多項目変異部位パターンデータ103が未知の変異を示した場合、学習モデル10の判断根拠を既知の変異との相違を距離として2次元散布図で可視化したデータである。

[0034] 次に、コンピュータ2の格納部204は、取得した多項目同時測定データ

１０１と、塩基配列データ１０２と、生成した多項目変異部位パターンデータ１０３と、診断データ１０４と、を少なくともコンピュータ２に格納する（ステップ０４）。

[0035] 最後に、コンピュータ２の提供部２０５は、格納部２０４が格納した当該データを少なくともユーザ端末３を介して提供する（ステップ０５）。

[0036] 以上が高精度診断システム１の概要である。

[0037] [高精度診断システム１の構成]

図２は、高精度診断システム１の構成を説明するための図である。

[0038] 図２に示すように、高精度診断システム１のコンピュータ２は、制御部（非図示）として、CPU（Central Processing Unit）、GPU（Graphics Processing Unit）、RAM（Random Access Memory）、ROM（Read Only Memory）等を備え、通信部（非図示）として、他の端末や装置等と通信可能にするためのデバイス、例えば、IEEE802.11に準拠したWi-Fi（Wireless Fidelity）対応デバイス等を備える。

また、コンピュータ２は、記憶部（非図示）として、ハードディスクや半導体メモリ、記録媒体、メモリカード等によるデータのストレージ部を備える。ストレージ部はネットワーク通信可能な外部に存在してもよい。

[0039] コンピュータ２において、制御部が所定のプログラムを読み込むことにより、通信部、記憶部と協働して、取得部２０１、学習モデル作成部２０２、診断部２０３、格納部２０４、提供部２０５を実現する。

[0040] ユーザ端末３は、コンピュータ、スマートフォン、タブレット端末等の電子デバイスであり、上述したコンピュータ２と、ネットワーク４を介してデータ通信可能に接続されている。ユーザ端末３は１台である必要はなく、複数台あってもよい。

[0041] ユーザ端末３は、端末制御部（非図示）として、上述したコンピュータ２における制御部と同様に、CPU、GPU、RAM、ROM等を備える。ユ

ーザ端末3は、通信部（非図示）として、他の端末や装置等と通信可能にするためのデバイスを備える。また、ユーザ端末3は、入出力部（非図示）として、データを入出力する入出力デバイス等を備える。また、ユーザ端末3は、記憶部（非図示）として、ハードディスクや半導体メモリ、記録媒体、メモリカード等によるデータのストレージ部を備える。

[0042] 以上が高精度診断システム1の構成である。

[0043] 〔遺伝子変異高精度診断処理〕

図3は、高精度診断システム1のコンピュータ2が実行する遺伝子変異高精度診断処理の手順を示すフローチャート図である。図3に基づいて、コンピュータ2が実行する遺伝子変異高精度診断処理について説明する。

[0044] まず、コンピュータ2の取得部201は、多項目同時測定データ101と、塩基配列データ102と、を少なくとも取得する（ステップS301）。該データはコンピュータ2にすでに保存されているものでもよいし、ユーザ端末3と、公衆回線網等のネットワーク4を介して、データ通信可能に接続し、取得したものでもよい。

[0045] 次に、コンピュータ2の学習モデル作成部202は、取得した少なくとも、多項目同時測定データ101と、塩基配列データ102と、からkNNを用いて既知の遺伝子変異を分類し、LOFを用いて既知の遺伝子変異と傾向が異なる未知の遺伝子変異を検出し、多項目変異部位パターンデータ103を生成する学習モデル10を作成する（ステップS302）。

[0046] 次に、コンピュータ2の学習モデル10は、多項目同時測定データ101から既知の遺伝子変異を分類し、未知の遺伝子変異を検出して多項目変異部位パターンデータ103を生成する（ステップS303）。

[0047] ここで、kNNを用いた既知の遺伝子変異の分類方法を、新型コロナ変異ウイルスを例として更に詳しく説明する。

多項目同時測定データ101は、新型コロナ変異ウイルスの多項目同時遺伝子関連検査により得られたデータであり、遺伝子の変異箇所ごとに閾値が設定され、変異箇所の変異型によって定量性のあるデータ、例えば、変異箇所

をパネルとして示し、パネルの色の濃淡によって出力値がことなり、多項目で検出されたパネルをバーコードパターンとして置換されたものとする。

また、塩基配列データ 102 は、新型コロナウイルスの変異ウィルスの WT から得られた塩基配列データ、あるいは既知の変異遺伝子を有する変異型の塩基配列データ、を少なくとも含む塩基配列データであり、多項目同時測定データ 101 に準じて遺伝子の変異箇所ごとに定量性の有るデータに置換されたものとする。

[0048] 既知の遺伝子変異の分類は、まず、 n 番目の変異部位の出力値（連続値）を α_n で表す。この際、ある変異型の出力はベクトル $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ として n 次元空間のベクトルで表現できる。ウィルス量でベクトルの要素の大きさが変わるため、ベクトル α をベクトルの大きさで割る正規化を行う。これによって、 n 次元の単位円周上に各症例が配置され、原点からどちらの方向を向いているかで各症例が分類される。もし、同じパネルのバーコードパターンで色の濃さ（出力値の大きさ）が異なっていたとしても、パネルのバーコードパターンが同じであれば、同じ方向を向くために同じ変異型として分類される。

学習モデル作成部 202 が行う学習では、上述した塩基配列データ 102 を学習のための教師データとして当該処理を行い、 n 次元空間上に学習データを配置し、既知の変異型の名称を紐づけ、学習モデル 10 を作成する。

新規に取得した多項目同時測定データ 101 を、テストデータ（ベクトル $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ ）として分類する場合は、 i 番目の学習データ（ベクトル α_i ）とテストデータ（ベクトル β ）のユークリッド距離を計算し、最も近い距離の学習データを求めて、学習データの変異型の名称を紐づけて、多項目変異部位パターンデータ 103 を生成する（ステップ S303a）。

[0049] 次に、LOF を用いた未知の遺伝子変異の検出方法を、新型コロナウイルスを例として更に詳しく説明する。

まず、 x から x' への到達可能距離は、下記の数式 1 で定義される。ここで、 $x^{(k)}$ は、学習データ $\{x_i\}_{i=1}^n$ のうち、 x に k 番目に近い症例を表す。

このとき、 x の局所到達可能密度は下記の数式2と定義される。この局所到達可能密度を用いて、 x のLOFは、下記の数式3で定義される。

$LOF_k(x)$ は、 $x^{(i)}$ の局所到達可能密度の平均と x の局所到達可能密度の比である。もし、学習データ $x^{(i)}$ の周辺は密度が高いが、テストデータ x の周辺の密度が低いとき、LOFの値は大きな値となる。この場合、学習データにない未知のパターンであると判断し、未知の遺伝子変異として検出し、多項目変異部位パターンデータ103を生成する（ステップS303b）。

[0050] [数1]

$$RD_k(x, x') = \max\{\|x - x^{(k)}\|, \|x - x'\|\}$$

[0051] [数2]

$$LRD_k(x) = \left(\frac{1}{k} \sum_{i=1}^k RD_k(x^{(i)}, x) \right)^{-1}$$

[0052] [数3]

$$LOF_k(x) = \frac{\frac{1}{k} \sum_{i=1}^k LRD_k(x^{(i)})}{LRD_k(x)}$$

[0053] 図4は、新型コロナ変異ウイルスを対象とした場合の、ユーザ端末3が表示するコンピュータ2が生成した多項目変異部位パターンデータ103の表示画面の一例である。図4に示すようにコンピュータ2が生成した多項目変異部位パターンデータ103は、Delta型などの既知の新型コロナ変異ウイルスの分類と、未知の新型コロナ変異ウイルスの検出と、を行い、変異部位パターン、例えば、2次元データの個々の値を色や濃淡として表現した可視化グラフの一種であるheatmap、として表示される。

[0054] このように、既知の遺伝子変異を分類すること、未知の遺伝子変異を検出すること、が可能となるが、学習モデル10がなぜそのように判断したかの

根拠の提示が難しく、学習モデル 10 の提示した結果を信頼して活用することが難しい。

[0055] そこで、コンピュータ 2 の診断部 203 は、多項目変異部位パターンデータ 103 と、学習モデル 10 の学習データと、から MDS を用いてデータ分布を可視化した 2 次元散布図を診断データ 104 として生成する（ステップ S304）。

[0056] ここで、MDS を用いたデータ分布の可視化方法を、新型コロナ変異ウイルスを例として更に詳しく説明する。

多項目同時測定データ 101 の出力値は n 次元データになるため、各症例は n 次元空間上のひとつの点と分布する。このとき、n 次元データを 2 次元データ（または 3 次元データ）に次元縮約することができれば、散布図で表示することができるため、テストデータがどの既知の変異に近いのか、未知の変異であるのかを直感的に判断できようになり、学習モデル 10 の提示した結果の判断根拠とすることが可能となる。そこで、MDS を用いて次元集約を行う。MDS は次の手順で新しい軸を構成する。

まず、入力 i と入力 j のユークリッド距離からなる距離行列 d_{ij} を求め、原点が n 個の入力データの重心になるように移動する変換行列 Z_{ij} を求める（数式 4）。

[0057] [数4]

$$z_{ij} = -\frac{1}{2} \left(d_{ij}^2 - \sum_{i=1}^n \frac{d_{ij}^2}{n} - \sum_{j=1}^n \frac{d_{ij}^2}{n} + \sum_{i=1}^n \sum_{j=1}^n \frac{d_{ij}^2}{n^2} \right)$$

[0058] 次に、行列 Z_{ij} の固有ベクトルで与えられる軸上の座標値として新しい座標点を決定する。MDS は、データ間のユークリッド距離を保持した形の線形変換であるから、n 次元空間上での各症例の相対的な位置関係を 2 次元空間上で再現したものと解釈できる。

学習データを用いて n 次元データを 2 次元データに次元縮約する変換式を求め、それをテストデータに適用することで、テストデータの 2 次元空間上の位置を求める。多項目同時測定データ 101 の出力値は、例えば（1， 0）

や (0, 1) として、バーコードのパターンとして表現されるため、各症例は n 次元空間上で線形独立な関係にある軸に沿って分布する。

MDS は、各症例のユークリッド距離を保持した形で、正規直交関係にある固有ベクトルによって新しい 2 軸を求める手法であるため、多項目同時測定データ 101 のバーコードパターンの数学的性質を考慮した次元縮約法であることが発明のポイントである。

また、次元縮約の手法として頻繁に用いられる t 分布型確率的近傍埋め込み法 (T-distributed Stochastic Neighbor Embedding, : t-SNE) では、新しいテストデータが入力される度に軸が新たに形成されるため、学習とテストを別々に行うことができない。したがって、固定した既知の変異の 2 次元分布上にテストデータの分布を加えることができる MDSの方が、t-SNE よりも優位であると考えられる。

[0059] 図 5 は、新型コロナ変異ウイルスを対象とした場合の、ユーザ端末 3 が表示するコンピュータ 2 が生成した診断データ 104 の表示画面の一例である。
図 5 に示すようにコンピュータ 2 が生成した診断データ 104 は、Delta 型などの既知の新型コロナ変異ウイルスの分類と、未知の新型コロナ変異ウイルスの検出と、を行った多項目変異部位パターンデータ 103 から、2 次元散布図を診断データとして生成し、学習モデルに基づいて生成された多項目変異部位パターンデータ 103 の妥当性を視覚的に示すことが可能である。

[0060] 以上が、高精度診断処理である。

[0061] このように高精度診断処理によれば、多項目同時遺伝子関連検査に係る多項目同時測定データを活用し、既知の遺伝子変異と、未知の遺伝子変異と、を高精度に分類し、分類した結果を解析して可視化した上で判断根拠を視覚的に示す高精度診断システムを提供することができる。

[0062] 高精度診断システム 1 は、上述した新型コロナ変異ウイルスだけでなく、例えば、がん細胞の遺伝子変異の検出や、特定タンパク質の検出による他の疾

患の診断検査としても有用性がある。

[0063] [がん細胞の遺伝子変異の検出による診断]

癌に関する至適治療法を提案するAIシステムとして利用可能である。近年、がんの増殖に関係する遺伝子異常が次々に明らかになっており、現在は、がん細胞の遺伝子変異の検査を経て、条件が合えば、大きな効果が期待できる分子標的薬を用いる治療が行われている。例えば、肺腺がんに関する遺伝子異常として、EGFR遺伝子変異、ALK融合遺伝子、HER2遺伝子変異、MET遺伝子変異、ROS1融合遺伝子、BRAF遺伝子変異、RET融合遺伝子、RAS遺伝子変異などが知られている。もし、EGFR遺伝子の変異があれば、分子標的薬ゲフィチニブが劇的に効く可能性が高い。など、がん細胞の遺伝子変異と分子標的薬がセットになった診断と治療を施すゲノム医療が行われている。

がん細胞の遺伝子変異の有無を調べる検査の結果は、本発明におけるパネルパターンの出力と同様に考えることができる。そのためがん細胞の遺伝子変異の有無の情報を本発明と同じアルゴリズムで分析することによって、有効な分子標的薬の選択が可能となり、至適治療法を提案するAIシステムを構築して利用することが可能になる、また、既知のパターンと異なると判別できれば分子標的薬が奏効しない症例であると予測することも可能である。

[0064] [特定タンパク質の検出]

それぞれ異なるタンパク質に特異的に反応するものを結合した100種類のカラーコードビーズを使用した「Bio-Plex™サスペンションアレイシステム」などのデータは、新型コロナ多項目同時測定変異データと同様に最大100種類の異なった生体成分（タンパクや核酸）多項目情報が取得される。炎症性サイトカインのパネルを使用して得られる、炎症性サイトカインに関する多項目データを活用して、炎症病態を診断することに、本発明手法の有用性が期待される。

[0065] 以上、本発明の実施形態について説明したが、本発明は上述したこれらの実施形態に限るものではない。また、本発明の実施形態に記載された効果は

、本発明から生じる最も好適な効果を列挙したに過ぎず、本発明による効果は、本発明の実施形態に記載されたものに限定されるものではない

符号の説明

- [0066] 1 高精度診断システム
2 コンピュータ
3 ユーザ端末
4 ネットワーク

請求の範囲

- [請求項1] 多項目同時測定データを活用し、既知の試料と、未知の試料と、を検出し可視化する高精度判定システムであって、
- 未知の試料の多項目同時測定データと、既知の試料の多項目同時測定データと、を少なくとも取得する取得部と、
- 取得した前記未知の試料の多項目同時測定データと、前記既知の試料の多項目同時測定データと、の近似性を解析して定量化し、既知の試料と傾向が異なる未知の試料を分類し、分類データを生成する分類部と、
- 生成された前記分類データを次元圧縮して2次元散布図を判定データとして生成する判定部と、
- 前記未知の多項目同時測定データと、前記既知の多項目同時測定データと、前記分類データと、前記判定データと、を少なくとも格納する格納部と、
- 格納した当該各種データを少なくとも提供する提供部と、を備えることを特徴とする高精度判定システム。
- [請求項2] 前記分類部は、k近傍法（*k-Nearest Neighbor*，*kNN*）を用いて取得した前記未知の試料の多項目同時測定データと、前記既知の試料の多項目同時測定データと、の近似性を解析して定量化し、局所外れ値因子法（*Local Outlier Factor*，*LOF*）を用いて前記既知の試料と前記傾向が異なる未知の試料を分類し、前記判定部は、多次元尺度構成法（*MDS: Multi-Dimensional Scaling*）を用いて、前記分類データから2次元散布図を判定データとして生成する
- 請求項1に記載の高精度判定システム。
- [請求項3] 前記多項目同時測定データは、前記未知の試料および前記既知の試料から得た核酸配列に係るデータである請求項1または請求項2に記載の高精度判定システム。

- [請求項4] 既知の新型コロナ変異ウイルスと、未知の新型コロナ変異ウイルスと、を判定する請求項1または請求項2に記載の高精度判定システム。
- [請求項5] 前記未知の試料および前記既知の試料は、がん細胞である請求項3に記載の高精度判定システム。
- [請求項6] 前記多項目同時測定データは、特定タンパク質の存在量を測定したデータであって、病態を判定する請求項1または請求項2に記載の高精度判定システム。
- [請求項7] 定量化した一つ乃至複数の項目を、一つ乃至複数のパターンとして認識して分類精度を向上する学習モデルを作成する学習モデル作成部と、
前記学習モデルに基づいて、前記分類データからパターンデータを生成するパターンデータ生成部と、を更に備え、
前記格納部は、前記パターンデータおよび前記学習モデルを少なくとも格納し、前記提供部は、前記パターンデータを少なくとも提供する、請求項1に記載の高精度判定システム。
- [請求項8] コンピュータが実行し、多項目同時測定データを活用し、既知の試料と、未知の試料と、を検出し可視化する高精度判定方法であって、
未知の試料の多項目同時測定データと、既知の試料の多項目同時測定データと、を少なくとも取得するステップと、
取得した前記未知の試料の多項目同時測定データと、前記既知の試料の多項目同時測定データと、の近似性を解析して定量化し、既知の試料と傾向が異なる未知の試料を分類し、分類データを生成するステップと、
生成された前記分類データを次元圧縮して2次元散布図を判定データとして生成するステップと、
前記未知の多項目同時測定データと、前記既知の多項目同時測定データと、前記分類データと、前記判定データと、を少なくとも格納する

ステップと、

格納した当該各種データを少なくとも提供するステップと、を備えることを特徴とする高精度判定方法。

[請求項9]

多項目同時測定データを活用し、既知の試料と、未知の試料と、を検出し可視化するコンピュータに、

未知の試料の多項目同時測定データと、既知の試料の多項目同時測定データと、を少なくとも取得するステップ、

取得した前記未知の試料の多項目同時測定データと、前記既知の試料の多項目同時測定データと、の近似性を解析して定量化し、既知の試料と傾向が異なる未知の試料を分類し、分類データを生成するステップ、

生成された前記分類データを次元圧縮して2次元散布図を判定データとして生成するステップ、

前記未知の多項目同時測定データと、前記既知の多項目同時測定データと、前記分類データと、前記判定データと、を少なくとも格納するステップ、

格納した当該各種データを少なくとも提供するステップ、を実行させるためのプログラム。