

(19)日本国特許庁(JP)

## (12)特許公報(B2)

(11)特許番号

特許第7195514号

(P7195514)

(45)発行日 令和4年12月26日(2022.12.26)

(24)登録日 令和4年12月16日(2022.12.16)

(51)国際特許分類

F I

G 1 6 H 50/00 (2018.01)

G 1 6 H 50/00

G 1 6 B 40/30 (2019.01)

G 1 6 B 40/30

請求項の数 14 (全25頁)

(21)出願番号 特願2020-535049(P2020-535049)  
 (86)(22)出願日 平成30年12月21日(2018.12.21)  
 (65)公表番号 特表2021-509502(P2021-509502  
 A)  
 (43)公表日 令和3年3月25日(2021.3.25)  
 (86)国際出願番号 PCT/IB2018/060485  
 (87)国際公開番号 WO2019/135143  
 (87)国際公開日 令和1年7月11日(2019.7.11)  
 審査請求日 令和3年5月25日(2021.5.25)  
 (31)優先権主張番号 15/864,278  
 (32)優先日 平成30年1月8日(2018.1.8)  
 (33)優先権主張国・地域又は機関  
 米国(US)

(73)特許権者 390009531  
 インターナショナル・ビジネス・マシー  
 ンズ・コーポレーション  
 INTERNATIONAL BUSI  
 NESS MACHINES CORPO  
 RATION  
 アメリカ合衆国10504 ニューヨー  
 ク州 アーモンク ニュー オーチャード  
 ロード  
 New Orchard Road, A  
 rmonk, New York 105  
 04, United States of  
 America  
 (74)代理人 100112690  
 弁理士 太佐 種一

最終頁に続く

(54)【発明の名称】 がん確率のライブラリ・スクリーニングのためのコンピュータ実装方法、コンピュータ・プログラム製品、およびシステム(がん確率のライブラリ・スクリーニング)

## (57)【特許請求の範囲】

## 【請求項1】

コンピュータ実装方法であって、

1つまたは複数のプロセッサにより、ある状態と診断もしくは予備診断されるか、または、前記状態と診断されないものとして指定される患者のペプチド・ライブラリを含む生データ・セットを取得することと、

前記1つまたは複数のプロセッサにより、前記生データ・セットを所定数のグループに分割することであり、不適合データ・グループをその他のグループから取り出すことを含む、前記分割することと、

前記1つまたは複数のプロセッサにより、前記その他のグループに対して主成分分析を実行して、前記その他のグループにおける特徴の頻度に基づいて、前記1つまたは複数のプロセッサにより、前記その他のグループのデータにおける共通特徴を識別し、前記その他のグループにおける発生頻度に基づいて、係数により表される主成分を含む前記共通特徴を重み付けすることと、

前記1つまたは複数のプロセッサにより、所定レベルの検証精度を生じる最小数の前記主成分を決定することと、

前記1つまたは複数のプロセッサにより、ロジスティック回帰モデルにおける最良適合のパラメータとして前記最小数の主成分を利用することによって予測モデルを生成することであり、前記予測モデルが、所定の閾値内に前記状態が存在する尤度および前記所定の閾値内に前記状態が存在しない尤度から成る群から選択される2値結果を提供する、前記

10

20

生成することと、

を含み、

前記主成分分析の実行に先立って、前記1つまたは複数のプロセッサにより、すべてのペプチドにわたりペプチドのペプチド値を合算するとともに各発現値を結果の合計値で割って比を演算することによって、前記生データ・セットを含むペプチドを正規化することを含む、

むコンピュータ実装方法。

【請求項2】

コンピュータ実装方法であって、

1つまたは複数のプロセッサにより、ある状態と診断もしくは予備診断されるか、または、前記状態と診断されないものとして指定される患者のペプチド・ライブラリを含む生データ・セットを取得することと、ここで、前記ペプチド・ライブラリが、12-merペプチド・ライブラリを含む、

10

前記1つまたは複数のプロセッサにより、前記生データ・セットを所定数のグループに分割することであり、不適合データ・グループをその他のグループから取り出すことを含む、前記分割することと、

前記1つまたは複数のプロセッサにより、前記その他のグループに対して主成分分析を実行して、前記その他のグループにおける特徴の頻度に基づいて、前記1つまたは複数のプロセッサにより、前記その他のグループのデータにおける共通特徴を識別し、前記その他のグループにおける発生頻度に基づいて、係数により表される主成分を含む前記共通特徴を重み付けすることと、

20

前記1つまたは複数のプロセッサにより、所定レベルの検証精度を生じる最小数の前記主成分を決定することと、

前記1つまたは複数のプロセッサにより、ロジスティック回帰モデルにおける最良適合のパラメータとして前記最小数の主成分を利用することによって予測モデルを生成することであり、前記予測モデルが、所定の閾値内に前記状態が存在する尤度および前記所定の閾値内に前記状態が存在しない尤度から成る群から選択される2値結果を提供する、前記生成することと、

を含むコンピュータ実装方法。

【請求項3】

30

コンピュータ実装方法であって、

1つまたは複数のプロセッサにより、ある状態と診断もしくは予備診断されるか、または、前記状態と診断されないものとして指定される患者のペプチド・ライブラリを含む生データ・セットを取得することと、ここで、前記状態ががんを含む、

前記1つまたは複数のプロセッサにより、前記生データ・セットを所定数のグループに分割することであり、不適合データ・グループをその他のグループから取り出すことを含む、前記分割することと、

前記1つまたは複数のプロセッサにより、前記その他のグループに対して主成分分析を実行して、前記その他のグループにおける特徴の頻度に基づいて、前記1つまたは複数のプロセッサにより、前記その他のグループのデータにおける共通特徴を識別し、前記その他のグループにおける発生頻度に基づいて、係数により表される主成分を含む前記共通特徴を重み付けすることと、

40

前記1つまたは複数のプロセッサにより、所定レベルの検証精度を生じる最小数の前記主成分を決定することと、

前記1つまたは複数のプロセッサにより、ロジスティック回帰モデルにおける最良適合のパラメータとして前記最小数の主成分を利用することによって予測モデルを生成することであり、前記予測モデルが、所定の閾値内に前記状態が存在する尤度および前記所定の閾値内に前記状態が存在しない尤度から成る群から選択される2値結果を提供する、前記生成することと、

を含むコンピュータ実装方法。

50

## 【請求項 4】

前記状態が、乳がんを含む、請求項 3 に記載の方法。

## 【請求項 5】

前記状態が、ステージ 1 の乳がんを含む、請求項 3 又 4 に記載の方法。

## 【請求項 6】

前記 1 つまたは複数のプロセッサにより、前記係数を適用して前記不適合データ・グループに使用する特徴を計算することによって前記予測モデルをテストすることと、

前記 1 つまたは複数のプロセッサにより、前記不適合データ・グループの前記特徴を前記主成分と比較することによって前記予測モデルを調節することと、

をさらに含む、請求項 1 ～ 5 いずれか 1 項に記載のコンピュータ実装方法。

10

## 【請求項 7】

前記分割が、前記グループに割り当てる前記生データ・セットのレコードを選択する乱数を生成することを含む、請求項 1 ～ 6 いずれか 1 項に記載のコンピュータ実装方法。

## 【請求項 8】

前記その他のグループに対して主成分分析を実行することが、

前記 1 つまたは複数のプロセッサにより、トレーニング・データ集合を生成することであり、前記トレーニング・データ集合の各トレーニング・データ集合が、前記その他のグループのうちの 1 つを除くすべてのグループのデータを含む、前記生成することと、

前記 1 つまたは複数のプロセッサにより、各トレーニング・データ集合に対して個別に主成分分析を実行することによって、各トレーニング・データ集合の共通特徴を識別することと、

20

前記 1 つまたは複数のプロセッサにより、前記その他のグループを単一のトレーニング・セットとして利用することにより主成分分析を実行することによって、前記単一のトレーニング・セットの共通特徴を識別することと、

各トレーニング・データ集合の前記共通特徴および前記単一のトレーニング・セットの前記共通特徴について、前記その他のグループの数に等しい回数の相互検証を実行することによって、前記主成分を識別するとともに前記係数を決定することと、

を含む、請求項 1 ～ 7 いずれか 1 項に記載のコンピュータ実装方法。

## 【請求項 9】

前記所定レベルの検証精度が、前記回数に基づく最高の検証精度である、請求項 1 ～ 7 いずれか 1 項に記載のコンピュータ実装方法。

30

## 【請求項 10】

前記 1 つまたは複数のプロセッサにより、所与の個人の 12 - m e r ペプチド・ライブラリを含む新たな生データを取得することと、

前記 1 つまたは複数のプロセッサにより、前記予測モデルを前記新たな生データに適用して、前記所与の個人の 2 値結果を決定することと、

をさらに含む、請求項 1 又は 2 に記載のコンピュータ実装方法。

## 【請求項 11】

前記予測モデルを適用することが、前記係数を利用して前記新たな生データに対する主成分分析を実行することと、ロジスティック回帰を実行して前記 2 値結果を生成することと、を含む、請求項 10 に記載のコンピュータ実装方法。

40

## 【請求項 12】

コンピュータ・プログラムであって、請求項 1 ～ 11 のいずれか 1 項に記載の方法の各ステップをコンピュータに実行させるための、コンピュータ・プログラム。

## 【請求項 13】

請求項 12 に記載のコンピュータ・プログラムを記録した、コンピュータ可読記憶媒体。

## 【請求項 14】

システムであって、

メモリと、

前記メモリと通信するプロセッサと、

50

方法を実行するように前記メモリを介して前記プロセッサにより実行可能なプログラム命令と、  
を備え、前記方法は、請求項 1 ~ 11 のいずれか 1 項に記載の方法の各ステップを含む、  
前記システム。

【発明の詳細な説明】

【技術分野】

【0001】

本出願は、一般に、がん予測モデルのコンピュータ処理方法および装置に関し、より詳細には、患者の生データ・セット（ペプチド・ライブラリ）に基づき、がん予測モデルの生成のための方法および装置に関する。

10

【背景技術】

【0002】

米国においては、約 8 人に 1 人の女性（すなわち、12%）が生涯のうちに、浸潤性乳がん（invasive breast cancer）を患う。2017 年より前、公衆衛生機関は、米国の女性において、新たに 252,710 症例が浸潤性乳がんとして診断され、新たに 63,410 症例が非浸潤性（non-invasive）（上皮内（in situ））乳がんとして診断されると推定していた。上皮内がんの症例では、悪性細胞（malignant cell）が腫瘍として存在する一方、腫瘍が見つかった場所の基底膜（basement membrane）を越えては、転移も浸潤も起こっていない。治療法の改善にも関わらず、2017 年の乳がんによる米国女性の死亡者数は、およそ 40,610 人と推定されるが、これは減少を表しており、1989 年以降は毎年、特に 50 歳以下の女性では、死亡者数が減少している。ただし、乳がんは依然として、米国の女性においては、各種がんのうちで死の主要な原因となっている。女性の新たに診断されるがんのうち、およそ 30% が乳がんとして推定される。乳がんのリスク増大に関する一定の遺伝的指標が存在するものの、およそ 85% の乳がんは、乳がんの家系ではない女性に発症する。現在、乳がんの最も重要なリスク要因は、性別（女性）および年齢（加齢）である。

20

【発明の概要】

【発明が解決しようとする課題】

【0003】

患者の生データ・セット（ペプチド・ライブラリ）に基づき、がん予測モデルを生成する。

30

【課題を解決するための手段】

【0004】

従来技術の欠点は、状態の尤度を予測するコンピュータ・プログラム製品の提供によって克服され、別の利点がもたらされる。このコンピュータ・プログラム製品は、処理回路により可読な記憶媒体であって、方法を実行するように処理回路により実行するための命令を格納した記憶媒体を備える。この方法は、たとえば、1 つまたは複数のプロセッサにより、ある状態と診断もしくは予備診断されるか、または、当該状態と診断されないものとして指定される患者のペプチド・ライブラリ（peptide library）を含む生データ・セットを取得することと、1 つまたは複数のプロセッサにより、生データ・セットを所定数のグループに分割することであり、不適合（holdout）データ・グループをその他のグループから取り出すことを含む、分割することと、1 つまたは複数のプロセッサにより、その他のグループに対して主成分分析を実行して、その他のグループにおける特徴の頻度に基づいて、1 つまたは複数のプロセッサにより、その他のグループのデータにおける共通特徴を識別し、その他のグループにおける発生頻度に基づいて、係数により表される主成分を含む共通特徴を重み付けすることと、1 つまたは複数のプロセッサにより、所定レベルの検証精度（validation accuracy）を生じる最小数の主成分を決定することと、1 つまたは複数のプロセッサにより、ロジスティック回帰モデルにおける最良適合（best fit）のパラメータとして最小数の主成分を利用することによって予測モデルを生成することであり、予測モデルが、所定の閾値内に状態が存在する尤度および所定の閾値内に状態が

40

50

存在しない尤度から成る群から選択される 2 値結果を提供する、生成することと、を含む。

【 0 0 0 5 】

従来技術の欠点は、状態の尤度を予測する方法の提供によって克服され、別の利点もたらされる。この方法は、たとえば、1つまたは複数のプロセッサにより、ある状態と診断もしくは予備診断されるか、または、当該状態と診断されないものとして指定される患者のペプチド・ライブラリを含む生データ・セットを取得することと、1つまたは複数のプロセッサにより、生データ・セットを所定数のグループに分割することであり、不適合データ・グループをその他のグループから取り出すことを含む、分割することと、1つまたは複数のプロセッサにより、その他のグループに対して主成分分析を実行して、その他のグループにおける特徴の頻度に基づいて、1つまたは複数のプロセッサにより、その他のグループのデータにおける共通特徴を識別し、その他のグループにおける発生頻度に基づいて、係数により表される主成分を含む共通特徴を重み付けすることと、1つまたは複数のプロセッサにより、所定レベルの検証精度を生じる最小数の主成分を決定することと、1つまたは複数のプロセッサにより、ロジスティック回帰モデルにおける最良適合のパラメータとして最小数の主成分を利用することによって予測モデルを生成することであり、予測モデルが、所定の閾値内に状態が存在する尤度および所定の閾値内に状態が存在しない尤度から成る群から選択される 2 値結果を提供する、生成することと、を含む。

10

【 0 0 0 6 】

従来技術の欠点は、状態の尤度を予測するシステムの提供によって克服され、別の利点もたらされる。このシステムは、メモリと、メモリと通信する 1つまたは複数のプロセッサと、方法を実行するようにメモリを介して 1つまたは複数のプロセッサにより実行可能なプログラム命令と、を備える。この方法は、たとえば、1つまたは複数のプロセッサにより、ある状態と診断もしくは予備診断されるか、または、当該状態と診断されないものとして指定される患者のペプチド・ライブラリを含む生データ・セットを取得することと、1つまたは複数のプロセッサにより、生データ・セットを所定数のグループに分割することであり、不適合データ・グループをその他のグループから取り出すことを含む、分割することと、1つまたは複数のプロセッサにより、その他のグループに対して主成分分析を実行して、その他のグループにおける特徴の頻度に基づいて、1つまたは複数のプロセッサにより、その他のグループのデータにおける共通特徴を識別し、その他のグループにおける発生頻度に基づいて、係数により表される主成分を含む共通特徴を重み付けすることと、1つまたは複数のプロセッサにより、所定レベルの検証精度を生じる最小数の主成分を決定することと、1つまたは複数のプロセッサにより、ロジスティック回帰モデルにおける最良適合のパラメータとして最小数の主成分を利用することによって予測モデルを生成することであり、予測モデルが、所定の閾値内に状態が存在する尤度および所定の閾値内に状態が存在しない尤度から成る群から選択される 2 値結果を提供する、生成することと、を含む。

20

30

【 0 0 0 7 】

本明細書においては、1つまたは複数の態様に関する方法およびシステムも記載・請求する。さらに、本明細書においては、1つまたは複数の態様に関するサービスを記載するとともに、請求する場合もある。

40

【 0 0 0 8 】

本明細書に記載の技術によって、付加的な特徴および利点の実現される。本発明のいくつかの実施形態において、この方法は、1つまたは複数のプロセッサにより、係数を適用して不適合グループに使用する特徴を計算することによって予測モデルをテストすることと、1つまたは複数のプロセッサにより、不適合グループの特徴を主成分と比較することによって予測モデルを調節することと、を含む。本発明のいくつかの実施形態においては、ペプチド・ライブラリが 12 - m e r ペプチド・ライブラリを含むか、または、状態がステージ 1 の乳がんを含むか、あるいはその両方である。本発明のいくつかの実施形態において、分割は、グループに割り当てる生データ・セットのレコードを選択する乱数を生成することを含む。

50

## 【 0 0 0 9 】

本発明のいくつかの実施形態において、この方法は、主成分分析の実行に先立って、1つまたは複数のプロセッサにより、すべてのペプチドにわたりペプチドのペプチド値を合算するとともに各発現値を結果の合計値で割って比を演算することによって、生データ・セットを含むペプチドを正規化することを含む。

## 【 0 0 1 0 】

本発明のいくつかの実施形態において、その他のグループに対して主成分分析を実行することは、1つまたは複数のプロセッサにより、トレーニング・データ集合を生成することであり、トレーニング・データ集合の各トレーニング・データ集合が、その他のグループのうちの1つを除くすべてのグループのデータを含む、生成することと、1つまたは複数のプロセッサにより、各トレーニング・データ集合に対して個別に主成分分析を実行することによって、各トレーニング・データ集合の共通特徴を識別することと、1つまたは複数のプロセッサにより、その他のグループを単一のトレーニング・セットとして利用することにより主成分分析を実行することによって、単一のトレーニング・セットの共通特徴を識別することと、各トレーニング・データ集合の共通特徴および単一のトレーニング・セットの共通特徴について、その他のグループの数に等しい回数の相互検証を実行することによって、主成分を識別するとともに係数を決定することと、を含む。本発明のいくつかの実施形態において、所定レベルの検証精度は、回数に基づく最高の検証精度である。

## 【 0 0 1 1 】

本発明のいくつかの実施形態において、この方法は、1つまたは複数のプロセッサにより、所与の個人の12-merペプチド・ライブラリを含む新たな生データを取得することと、1つまたは複数のプロセッサにより、予測モデルを新たな生データに適用して、所与の個人の2値結果を決定することと、を含む。本発明のいくつかの実施形態において、予測モデルを適用することは、係数を利用して新たな生データに対する主成分分析を実行することと、ロジスティック回帰を実行して2値結果を生成することと、を含む。

## 【 0 0 1 2 】

本明細書の最後の特許請求の範囲において、1つまたは複数の態様を詳細に示すとともに、一例として明確に請求する。上記および他の目的、特徴、および利点については、添付の図面と併せた以下の詳細な説明によって明らかとなる。

## 【図面の簡単な説明】

## 【 0 0 1 3 】

【図1】本発明のいくつかの実施形態の特定の態様を示したワークフローである。

【図2】本発明のいくつかの実施形態の特定の態様を示したワークフローである。

【図3】本発明の一実施形態の特定の態様を示した図である。

【図4】本発明のいくつかの実施形態の特定の態様を示したワークフローである。

【図5】クラウド・コンピューティング環境において利用可能なコンピューティング・ノードの一実施形態を示した図である。

【図6】本発明の実施形態に係るクラウド・コンピューティング環境を示した図である。

【図7】本発明の実施形態に係る抽象化モデル・レイヤを示した図である。

## 【発明を実施するための形態】

## 【 0 0 1 4 】

添付の図面は、それぞれの全体を通して同じ参照番号が同一または機能的に同様の要素を表し、本明細書に組み込まれるとともにその一部を構成するものであるが、本発明をさらに示すとともに、本発明の詳細な説明と併せて、本発明の原理を説明するのに役立つ。当業者が理解する通り、添付の図面は、理解の容易化のために提供され、本発明の特定の実施形態の態様を示す。本発明は、図面に示す実施形態に限定されない。

## 【 0 0 1 5 】

当業者が理解する通り、本願の全体を通して言及するプログラム・コードは、ソフトウェアおよびハードウェアの両者を含む。たとえば、本発明の特定の実施形態におけるプログラム・コードは、固定機能のハードウェアを含む一方、他の実施形態においては、記載

10

20

30

40

50

の機能のソフトウェアベースの実装を利用する。特定の実施形態は、両種のプログラム・コードを組み合わせる。1つまたは複数のプログラムとも称するプログラム・コードの一例を図5に示すが、一組（少なくとも一組）のプログラム・モジュール42を有するプログラム/ユーティリティ40がメモリ28に格納されていてもよい。

【0016】

本発明の実施形態は、乳がんが最終的に存在する確率の増大を認識するコンピュータ実装方法、コンピュータ・プログラム製品、およびコンピュータ・システムを含み、1つまたは複数のプログラムが12アミノ酸(12-mer)ペプチド・ライブラリ、免疫グロブリンG(IgG)抗体、ならびにさまざまなコンピュータ・データ・モデリングおよびコンピュータ管理を適用することによって、この決定を許容範囲内の精度レベルとする。ペプチド・ライブラリは、さまざまなペプチドを符号化したランダムなデオキシリボ核酸(DNA)配列で構成され、バクテリオファージに取り付けられたターゲットを認識可能である。一方、IgGは、最も豊富な種類の抗体であり、すべての体液中に存在して、細菌およびウイルス感染に対する保護を担う。

【0017】

本発明の実施形態の態様は、既存のコンピューティング技術の改良を表し、コンピューティングと密接に結びついている。具体的に、本発明の実施形態は、大量のデータを扱うとともに、データからロジスティック・モデルを構築する改良された方法を表す。たとえば、本発明の実施形態は、最終結果における観測データ・レートを低減する。プログラム・コードが低効率の2値ビニング手順を使用するのではなく、パターンの構築に利用されるデータを前処理するためである。

【0018】

本発明の実施形態の態様がコンピューティングと密接に結びついているのは、少なくとも、本発明の実施形態によって生成される、トレーニング・データから生成される自動生成自己学習予測モデルを含む電子モデルがコンピューティングの外側では生成不可能であり、また、コンピューティングの外側には存在しないためである。本発明の実施形態において最初に利用されるレコードは、1つまたは複数のデータベースに含まれる1つまたは複数のデータ・セット中の機械可読電子レコードである。また、結果としてのモデルも電子的であり、コンピューティング・リソースを利用して別の電子データ・セットに適用される。プログラム・コードによる別のデータ・セットへの適用によって、将来のある時点(たとえば、所定の時点)において乳がんが存在する可能性の予測のためにモデルが生成された事象または状態の確率でレコードを識別可能な機械可読データ・モデルが得られる本発明の実施形態の特定の態様を実現することは、データの量および性質の両者から個人には不可能である。

【0019】

本発明の実施形態は、結果を提供可能な速度のために個人および既存のシステムでは不可能なユーティリティを提供する。有用化のため、本発明の実施形態のプログラム・コードは、限られた期間内で、モデルの生成および更新を行うとともに、結果(モデルに適合するレコードの識別)を提供する。たとえば、個人が医療機関を訪れるシナリオにおいて、個人および医療機関は、データ(ペプチド)ライブラリ中の個々の値で電子的に表されるように、予測モデルが求めるデータに一致するレコード中の項目を個人が有するかに関する情報の取得の恩恵を受けることになる。この情報が訪問中に提供され得ない場合は、ほぼ間違いなく、個人または医療機関にとって有用とはならない。このため、本発明の実施形態において、プログラム・コードは、個人の所与のサンプルを分析するとともに、予測モデルを実時間または近実時間で適用する。これにより、本発明の実施形態は、ライブラリに記録された大量のペプチドが、本発明の実施形態のプログラム・コードにより生成された予測モデルと適合するかに基づいて、電子データ・ライブラリの実時間分析を可能にする。

【0020】

本発明の実施形態は、演算効率および有効性を増大させる一定の利点をもたらすことが

ら、コンピュータ技術と密接に結びついた利点および改良を提供する。たとえば、以下により詳しく説明する通り、本発明の実施形態は、予想質問結果に基づいて分散処理を利用することにより、重要な分析成果物のための時間を短縮する。この分散処理によれば、プログラム・コードは、複数の分析プロセスを同時に実行可能である。本発明の特定の実施形態の一部は、クラウド・アーキテクチャに移行させることで、サービスとしてのソフトウェア（SaaS）の提供物としてユーザが利用可能となる。クラウド・アーキテクチャにおけるリソースの無制限の演算能力は、データが豊富な環境におけるシステムの効率要求を満たすためのプログラム・コードによる同時質問およびプロセスの分散の補助に適している。

#### 【0021】

また、本発明の実施形態は、機械学習を利用することから、コンピュータ技術と密接に結びついた利点および改良を提供する。データ高密度環境における事象（たとえば、状態）識別の既存の手法に対する本発明のいくつかの実施形態の有利な一態様では、利用可能なツールにおける選択肢を制限する手法である機械学習問題の代わりに、統計的問題として、その他何らかの方法により事象識別および認識の問題に取り組む。機械学習の利用により、本発明の実施形態では、事象を含むレコードを識別可能であるが、当該事象を直接識別する情報は存在しない。たとえば、機械学習の使用により、プログラム・コードは、診断未確定患者の12-merペプチド・ライブラリを用いて、乳がんを患う確率で個々の患者を識別可能である。すなわち、患者がすでにこの病気を患っていることをデータが示しているわけではない。場合により、プログラム・コードは、機械学習を利用することにより、個人と関連するデータにおいて反対のことが示される場合に、当該個人が乳がんを患う閾値確率を有する旨を示すことができる。このため、プログラム・コードは、1つまたは複数のメモリ・デバイスに格納された既存の確立データを単に識別して読み出しているわけではない。むしろ、プログラム・コードは、パターンを確立し、機械学習アルゴリズムを継続的にトレーニングしてパターンを適用し、結果の予測モデルを利用して、データによりまだ明確に示されていない事象のインスタンスを識別する。

#### 【0022】

本発明の一実施形態においては、統計的手順である主成分分析（PCA）を1つまたは複数のプログラムが利用して、トレーニング・データ集合中の多くの個人に一致する1つまたは複数の特徴と関連する概念または成分集合を決定する。PCAは、場合により相関する変数の観測集合を主成分と称する線形無相関の変数の値集合に変換する直交変換を含む。主成分の数は、元の変数の数以下である。このため、本発明の一実施形態において、プログラム・コードは、親概念を利用して、複数の副成分を生成する。この変換は、最初の主成分が考え得る最大の分散を有し（すなわち、可能な限り多くのデータ変動性を考慮し）、後続の各成分が先行成分に直交する制約下で考え得る最大の分散を有するように規定される。結果としてのベクトルは、無相関直交基底集合である。PCAは、元の変数の相対スケーリングの影響を受けやすい。

#### 【0023】

本発明の種々実施形態の態様は、予測モデルの生成および適用に基づいて個人の乳がんの確率を予測するコンピュータ実装方法、コンピュータ・システム、またはコンピュータ・プログラム製品、あるいはその組み合わせを利用する既存のシステムに対する利点を提供する。既存の方法では、遺伝的多型（polymorphism）に対する乳がんマーカー、DNAの遺伝暗号（genetic code）の変化、組織の採取、過剰発現のバイオマーカーを検出する尿サンプルまたは母乳サンプル、あるいはその両方、質量分析の利用による術前サンプルの分析、遺伝子発現プロファイルの分析、リボ核酸（RNA）からの遺伝子発現データの分析、ならびに反応を誘発する免疫療法薬（Immunotherapeutic agent）もしくはワクチンの導入に依拠していた。これら既存の方法に対して、本発明の実施形態においては、12-merペプチドの免疫徴候サンプル（immunosignature sample）の分析ならびに乳がん患者の計測可能リスクを示すパターンの生成および認識の両者を行うプログラム・コードのトレーニングに部分的に基づいて、1つまたは複数のプログラムが予測モデルを

10

20

30

40

50



生成して適用する。

【 0 0 2 4 】

図 1 は、本発明の特定の実施形態のいくつかの態様の全体ワークフロー 1 0 0 を提供する。図 2 は、図 1 のワークフロー 1 0 0 に含まれる態様を別途詳細化したワークフロー 2 0 0 である。

【 0 0 2 5 】

最初に図 1 を参照して、本発明のいくつかの実施形態においては、1 つまたは複数のプログラムが生電子データ・セットを取得する ( 1 1 0 )。本発明の一実施形態において、生電子データは、抗ヒト I g G 抗体を用いた 1 2 - m e r ペプチド・ライブラリ・スクリーニングによる生のデータを含む。データ・セットには、最大 1 2 万 ( 1 2 0 K ) 個の値を含んでいてもよく、各値は、ある方法 ( グリーン色素 ( green dye ) の使用を含む蛍光標識 ( fluorescent labeling ) が挙げられるが、これに限定されない) により分離された 1 2 アミノ酸長のペプチドを表す。データ・セットには、プレステージ 1 もしくはステージ 1 の乳がんと過去に診断された個人ならびにこのように診断されていない個人に関するデータを含む。データ・セット中のレコードは、この特性化を示すようにラベル付与される。

10

【 0 0 2 6 】

1 つまたは複数のプログラムは、データ・セットを利用して、プレステージ 1 またはステージ 1 の乳がんを有する個人と関連するデータを、この診断のないデータにより表される個人から区別する予測モデルを生成する ( 1 2 0 )。本発明の一実施形態において 1 つまたは複数のプログラムにより生成された予測モデルは、分類器または分類アルゴリズムとして理解可能である。予測モデルを生成するプロセスについては、図 2 においてより詳しく論じる。ただし、予測モデルの生成において、1 つまたは複数のプログラムは、データ・セットをトレーニング・データとして利用するとともに、このデータ・セットにより提供されるトレーニングに基づいて、予測モデルを生成する。

20

【 0 0 2 7 】

1 つまたは複数のプログラムは、プレステージ 1 またはステージ 1 の乳がんと過去に診断されたかが不明であり、現時点でこの診断がないかも不明な個人と関連するデータを取得する ( 1 3 0 )。本発明の一実施形態において、データは、この個人の 1 2 - m e r ライブラリにおける各ペプチドの値を含む。このデータは、1 2 0 K 個の値を含んでいてもよい。

30

【 0 0 2 8 】

本発明の一実施形態において、1 つまたは複数のプログラムは、予測モデルを個人と関連するデータに適用して、個人と関連するデータがある状態のある確率を示しているかを決定する ( 1 4 0 )。この状態は、個人 (たとえば、腫瘍のある個人) が乳がんと診断される ( 所与の範囲内の ) 尤度または予測モデルの適用から所与の期間内に個人が乳がんと診断される尤度、あるいはその両方のうちの 1 つまたは複数であってもよい。たとえば、本発明のいくつかの実施形態は、将来のある時点で識別される状態の確率を返すモデルを生成するものであり、1 つまたは複数のプログラムがモデルを適用した 2 年後が挙げられるが、これに限定されない。

40

【 0 0 2 9 】

図 2 は、図 1 と比較してより詳細なワークフロー 2 0 0 であり、本発明のいくつかの実施形態の種々態様を示している。図 1 に示すように、本発明のいくつかの実施形態においては、1 つまたは複数のプログラムが生電子データを取得し ( 1 1 0 )、このデータを利用して予測モデルを生成する ( 1 2 0 )。図 2 は、このモデルの生成に関する別途詳細を提供する。

【 0 0 3 0 】

図 2 を参照して、本発明の一実施形態において、1 つまたは複数のプログラムは、1 ) 過去にステージ 1 の乳がんと診断されるか、もしくは、乳がんの予備診断を受けた個人との関連または 2 ) 乳がんと明確に診断されていない ( すなわち、従来の診断手段を利用し

50

て、乳がんであるか、もしくは、がんの初期ステージであることが示されていない)個人との関連、という2つのカテゴリの一方に属することに応じて各個人のレコードにラベルが付与されたデータ・セットを取得する(210)。後述の通り、データには、12-mer ペプチドの識別を含み、ペプチド・ライブラリとして理解可能である。本発明の実施形態において、1つまたは複数のプログラムは、データを処理して、モデル生成の効率を向上させる。本発明のいくつかの実施形態において、1つまたは複数のプログラムは、アミノ酸配列部(たとえば、dim: 1100 + by 120k+)および特徴部(たとえば、dim: 1100 + by 5 (サンプルID、ステータス、ソース、Predx、およびウェハを含む))にデータを分離する。

#### 【0031】

1つまたは複数のプログラムは、レコードをグループに対してランダムに割り当てることにより、レコードを所定数のグループに分離する(220)。本発明のいくつかの実施形態において、1つまたは複数のプログラムは、複数の乱数シードを生成し、シードごとに、レコードを所定数のグループに分離する。たとえば、乱数が25の場合、1つまたは複数のプログラムは、各25番目のレコードをレコードのうちの1つに割り当てることになる。1つまたは複数のプログラムは、さまざまなシードを連続的に生成し、これらのシードを利用して、レコードのグループに対するランダムな割り当てを行うことになる。グループは、各グループがその他の各グループと等しい(または、同じような)量のデータを含むように、1つまたは複数のプログラムにより生成される。本発明の一実施形態においては、ランダム化に基づいて、1つまたは複数のプログラムがレコードを6つのグループに分離する。

#### 【0032】

本発明の実施形態において、1つまたは複数のプログラムがデータを6グループのデータ・セット(たとえば、3つのチャンネルによる2K-mer)に割り当てた場合は、名称がファイル名に従うグループ数(たとえば、6)に等しい数のサブリストを伴って、アミノ酸配列部および特徴部の両方がリスト・オブジェクトとしてRに格納され得る。サンプルID-ラベル対は、すべてのサブデータ・セットにより共有されて、テーブル構成していてもよい。サンプルIDおよびステータス(または、ラベル)は、アミノ酸配列データをインデックス化するキーとして組み合わせられる。クラスまたはサンプルID等のフィルタリングのため、1つまたは複数のプログラムは、この小さなテーブルにおいてフィルタリングを実行した後、列インデックスを用いて、対応するアミノ酸配列データを抽出することができる。1つまたは複数のプログラムは、アミノ酸配列データをフィルタリングした場合(これにより、すべてのデータ・セットによるサンプルの共有を保つ場合)、テーブルと同じ順序になるように列の順序を変更する。本発明の実施形態において、不均衡なクラスを扱うため、1つまたは複数のプログラムは、ダウン・サンプリングおよびアップ・サンプリングを利用して、トレーニング用のクラスの均衡分散を生成する。

#### 【0033】

1つまたは複数のプログラムは、不適合として生成したグループのうちの1つを選択し、(不適合未満の)グループと同数の相互検証の実行に利用されるその他のグループを指定する(各検証では、検証の反復ごとに、その他の各グループをテスト/検証セットとして使用する一方、(不適合およびテスト・セット未満の)その他のグループをトレーニング・セットとして利用する)(230)。本明細書において論じるPCA分析(たとえば、250)の一部として、1つまたは複数のプログラムは、(分離グループ未満の)グループ数と同じ数(すなわち、回数)だけ、この検証を実行する。このため、1つまたは複数のプログラムは、6つのグループを生成した場合、5回の相互検証を実行する。この検証の実行においては、相互検証に含まれるグループ(すなわち、最初に差し引かれたグループ未満のグループ総数)のうち、別の1つのグループがテスト・グループとして取り出される。1つまたは複数のプログラムは、その他のグループ(すなわち、最初に差し引かれたグループおよびテスト・グループ未満のすべてのグループ)をトレーニングに利用するが、テスト・グループは、検証の実行に利用される。1つまたは複数のプログラムは、

10

20

30

40

50

これらのグループをそれぞれ検証セットとして利用すべく循環的に、（最初に差し引かれたグループ未満の）グループ数に等しい回数だけ、この動作を繰り返す。このため、相互検証の回数は、トレーニング・データ集合の数と同等である。

【 0 0 3 4 】

6つのグループの例に戻る。1つまたは複数のプログラムは、6つのグループによって、第1のグループを取り出す。1つまたは複数のプログラムは、その他5つのグループを利用して、5回の相互検証を実行する。この検証の実行において、1つまたは複数のプログラムは、5つのグループのうちの1つをテスト・グループとして選択し、このグループで検証を実行する一方、その他のグループをトレーニング・データとして利用する。1つまたは複数のプログラムは、5つのグループのうち、テスト・グループとして指定されるグループを変更して、この検証ステップを繰り返す。1つまたは複数のプログラムは、5つのグループすべてがテスト・グループとして機能するまで循環的に、この動作を繰り返す。

10

【 0 0 3 5 】

図2を再び参照して、本発明のいくつかの実施形態において、1つまたは複数のプログラムは、サンプル（または、電子レコードにより表される個人）ごとに、約10万（120K）個のペプチド全体の値を合算した後、各発現値を合計で割って比を演算することにより、グリーン色素により識別されたライブラリ中の120K個のペプチドを正規化する（240）。値の正規化によって、1つまたは複数のプログラムは、個人および120Kペプチド・ライブラリの全体にわたって、発現値を比較することができる。本発明の一実施形態において、正規化のため、1つまたは複数のプログラムは、個人のサンプルごとに120K個のペプチド全体で発現値を加算した後、各発現値を合計で割る。

20

【 0 0 3 6 】

図2に示す実施形態では、120Kという数字を一例として使用するが、当業者が理解する通り、さまざまなサンプルにおいて識別するとともに本発明の実施形態において利用するペプチドの数は、これより多くすることも可能であるし、少なくすることも可能である。

【 0 0 3 7 】

本発明のいくつかの実施形態において、1つまたは複数のプログラムは、各トレーニング・セット（上記指定の通り、グループによりテスト・グループが除外された場合は、最初の不適合グループ未満）について、PCAを利用することにより、トレーニング・セット中の（電子レコードにより表される）多くの個人に共通する多くの特徴を決定する（250）。PCA分析は、使用される乱数シードごとに、相互検証ステップ全体で安定した値を提供する。当業者が理解する通り、PCAは、場合により関連する変数の観測集合を主成分と称する線形無相関の変数の値集合に変換する直交変換を含む。主成分の数は、元の変数の数以下である。このため、本発明の一実施形態において、プログラム・コードは、テスト・データに基づき、（正規化により）識別された発現値を利用して、複数の共通特徴を生成する。この変換は、最初の主成分が考え得る最大の分散を有し（すなわち、可能な限り多くのデータ変動性を考慮し）、後続の各成分が先行成分に直交する制約下で考え得る最大の分散を有するように規定される。結果としてのベクトルは、無相関直交基底集合である。PCAは、元の変数の相対スケーリングの影響を受けやすい。

30

40

【 0 0 3 8 】

図2を再び参照するに、本発明の一実施形態において、1つまたは複数のプログラムは、PCAを実行することにより、テスト・データにおいて表された個人と関連する特徴を決定する（250）。PCAにおいて、1つまたは複数のプログラムは、テスト・データにおいて最も代表的な特徴を決定する。PCAの部分的な結果として、1つまたは複数のプログラムにより生成された最終的なモデルには、個人に共通する特徴の累積的な理解を含んでいてもよい。1つまたは複数のプログラムは、このデータを利用して、分散が最も大きな特徴を推定するとともに、当該項の順序を決定するようにしてもよい。

【 0 0 3 9 】

50

1つまたは複数のプログラムは、相互検証の中でPCA（たとえば、250）を動作させる。相互検証のすべてのステップについて、1つまたは複数のプログラムは、トレーニング・セットに対してPCAを動作させ、このPCA結果による加重をモデルに送って予測を行う前に、加重を用いて検証を変換することにより、検証精度を演算する。本発明の一実施形態において、1つまたは複数のプログラムは、相互検証ステップのほか、最終的な評価ステップ（たとえば、後述の260）ごとに、各トレーニング・セットに対してPCAを実行するとともに、サブデータ・セットごとに結果のリストをファイルに書き込む。1つまたは複数のプログラムは、このPCA結果ファイルの内容を相互検証およびモデリングに利用する場合に、当該ファイルをメモリに読み込む。

#### 【0040】

個々のテスト・データ・グループのPCAの完了後、1つまたは複数のプログラムは、グループ（たとえば、5回の例における不適合未満の全5つのグループ）を連結トレーニング・データ・セットとして組み合わせるとともに、PCAを実行する（260）。図3は、トレーニング・セットに対して1つまたは複数のプログラムにより取得されたPCA結果の一例であって、回数1～5は、個々のトレーニング・セットに関する結果であり、回数6は、相互検証全体に関する結果である。

#### 【0041】

1つまたは複数のプログラムは、ロジスティック回帰（LR）モデルにおける最良適合のパラメータとして、PCAにより識別された（たとえば、優性に基づいて選択された）多くの特徴を利用して予測モデルを生成することにより、（予測）2値結果（たとえば、元のラベル付与データと同様に、ある程度の誤差またはコントロール・データ内のステージ1の乳がん）を取得する（270）。この態様の結果がPCA係数の生成である（たとえば、270）。図2に示すように、本発明のいくつかの実施形態において、1つまたは複数のプログラムは、（2値）クラス・ラベル（この場合は、ステージ1またはコントロール）の予測を目的として、PCAをデータ・セットに適用するとともに、結果としての主成分を利用して、ロジスティック回帰モデルを構築する。図2および図3に示すように、1つまたは複数のプログラムは、相互検証を利用して、モデルに含まれる最良の主成分数を調節する（たとえば、10は最初の10成分を意味し、100は最初の100成分を意味する）。1つまたは複数のプログラムは、回数に対して最高の検証を生じる最小数（最良数）を選択する。たとえば、合計6つのグループおよび5回の検証を利用する本発明の実施形態において、1つまたは複数のプログラムは、最高の5回検証精度を生じる最小数（最良数）を選択する。1つまたは複数のプログラムは、このパラメータを用いたモデルに基づいて、テストの精度を評価する。これら本発明の6グループの実施形態において、1つまたは複数のプログラムは、相互検証を実行して、12-mer（グリーン）による複数の主成分を有するLRモデルが最大95%の平均検証精度を生じ、テスト・セットに対する評価においても高い精度が持続するものと決定する。

#### 【0042】

本発明のいくつかの実施形態において、1つまたは複数のプログラムは、単一のPCA導出特徴を利用して、モデルを生成する（たとえば、データ中の分散を説明する最も優勢な特徴を利用する）。たとえば、一部のモデルでは、平均検証精度が95.08%（0.005%）、ROC曲線下側の平均検証面積（AUC）が96.02%（0.006%）、テスト精度が96.72%、テストAUCが95.55%となり得るため、12-merデータに由来する最初の3つの主成分しか含まない可能性がある。本発明のいくつかの実施形態において、1つまたは複数のプログラムは、最高の検証精度および検証AUCを生じるパラメータを選択する。

#### 【0043】

1つまたは複数のプログラムは、PCA係数を利用して不適合データ（元から不適合であったグループ）に使用する特徴を計算することにより、予測モデル（最良適合モデル）をテストおよび調節する（280）。この調節は、上述の繰り返しPCA手順とともに、機械学習として理解され得る。1つまたは複数のプログラムは、情報が既知のデータ・セ

10

20

30

40

50

ットを利用してさまざまなステップを繰り返すことにより、モデルが状態有無の決定に利用する認識特徴（パターン）の精度を調節して向上させ得る。

#### 【 0 0 4 4 】

1つまたは複数のプログラムは、所与の個人の12-merペプチド・ライブラリを含む新たなデータのサンプルを取得する（290）。1つまたは複数のプログラムは、調節済み予測モデルを適用して、個人をスコアリングすることにより、潜在的なステージ1の乳がんの有無の結果を確率で予測する（295）。1つまたは複数のプログラムは、生成PCA係数を用いてPCA演算を実行するとともにLRを実行することにより、調節済み予測モデルを適用して（2値）予測（たとえば、モデルに基づくステージ1の確率またはステージ1の取得不可能な確率）を生成する。このため、1つまたは複数のプログラムは、予測モデルの調節パラメータとして主成分を利用する。

10

#### 【 0 0 4 5 】

図4は、本発明のいくつかの実施形態の特定の態様を示したワークフロー400である。具体的に、1つまたは複数のプログラムは、得られたデータを6つのグループに分割する一実施形態を示す。本発明のいくつかの実施形態において、1つまたは複数のプログラムは、個人の12-merペプチド・ライブラリ発現を表し、1)基準がんの存在または可能性を示すステージ1または予測ステージ1および2)第1のカテゴリの指標がないことを示すコントロール、という1つまたは2つのカテゴリに分割されたデータを取得する（410）。1つまたは複数のプログラムは、ランダムな異なる乱数シードを使用して（個人を表す）レコードを異なるグループに割り当てることにより、データを6つの等サイズ・グループに分割する（420）。

20

#### 【 0 0 4 6 】

1つまたは複数のプログラムは、6つのグループのうちの1つを不適合として指定するとともに、その他のグループを相互検証に使用する（430）。上述の通り、すべてのグループが検証セットとして使用されるまで、1つまたは複数のプログラムが検証を循環的に5回繰り返すため、検証は5回の相互検証である。1つまたは複数のプログラムは、約120K個のペプチド全体の値を合算した後、各発現値を合計で割って比を演算することにより、ライブラリ中の120K個のペプチドの発現値を正規化する（440）。本発明の一実施形態において、正規化のため、1つまたは複数のプログラムは、個人のサンプルごとに120K個のペプチド全体で発現値を加算した後、各発現値を合計で割る。1つまたは複数のプログラムは、PCAを実行することにより、データの所与のトレーニング・セットにおいて表される多くの個人に共通する多くの特徴を決定して、（トレーニング・セットにおいて表される多くの個人以下の）多くの特徴を生成する（450）。データには4つのトレーニング・セットが存在し、各トレーニング・セットは、不適合グループではない5つのグループのうちの4つから成る。1つまたは複数のプログラムは、各トレーニング・セットに対してPCAを実行する。

30

#### 【 0 0 4 7 】

1つまたは複数のプログラムは、5回分をトレーニング・データ組み合わせ集合として組み合わせるとともに、PCAを組み合わせ集合に適用して、多くの特徴を決定する（460）。1つまたは複数のプログラムは、相互検証においてPCAを動作させる場合、全部で、（6つのサブデータ・セット）\*（テスト用の全体として5回分+1cvセット）=36回のPCAを実行する。1つまたは複数のプログラムは、2値結果（ステージ1またはコントロール）を予測するロジスティック回帰（LR）モデルの最良適合のパラメータとして多くの特徴を通すことにより、最高の5回検証精度を生じる最小数の特徴を選択して、モデルに適合させる（470）。1つまたは複数のプログラムは、トレーニング・データから計算されたPCA係数を利用して不適合データに使用する特徴を演算することにより、結果としての最良適合モデルを不適合データに対してテスト/調節する（480）。

40

#### 【 0 0 4 8 】

モデルが生成されたら、当該モデルを使用して、個人の12-merペプチド・ライブ

50

ラリ発現の新たな任意サンプルをスコアリングすることにより、潜在的なステージ 1 の乳がんの有無の結果を確率で予測することができる。この新たなサンプルに対してスコアリングを演算するため、1 つまたは複数のプログラムは、トレーニング・データによる係数を（再度）使用し、L R モデルに適用して予測することにより、P C A 演算を実行する。

【 0 0 4 9 】

本発明の実施形態は、ある状態と診断もしくは予備診断されるか、または、当該状態と診断されないものとして指定される患者のペプチド・ライブラリを含む生データ・セットを取得する 1 つまたは複数のプロセッサにより実行される 1 つまたは複数のプログラムを含むコンピュータ実装方法、コンピュータ・プログラム製品、およびコンピュータ・システムを含む。1 つまたは複数のプログラムは、生データ・セットを所定数のグループに分割するが、この分割は、不適合データ・グループをその他のグループから取り出すことを含む。1 つまたは複数のプログラムは、その他のグループに対して主成分分析を実行して、その他のグループにおける特徴の頻度に基づいて、1 つまたは複数のプロセッサにより、その他のグループのデータにおける共通特徴を識別し、その他のグループにおける発生頻度に基づいて、係数により表される主成分を含む共通特徴を重み付けする。1 つまたは複数のプログラムは、所定レベルの検証精度を生じる最小数の主成分を決定する。1 つまたは複数のプログラムは、ロジスティック回帰モデルにおける最良適合のパラメータとして最小数の主成分を利用することによって予測モデルを生成するが、この予測モデルは、所定の閾値内に状態が存在する尤度および所定の閾値内に状態が存在しない尤度から成る群から選択される 2 値結果を提供する。

【 0 0 5 0 】

また、本発明のいくつかの実施形態において、1 つまたは複数のプログラムは、係数を適用して不適合グループに使用する特徴を計算することによって予測モデルをテストするとともに、不適合グループの特徴を主成分と比較することによって予測モデルを調節する。ペプチド・ライブラリは、1 2 - m e r ペプチド・ライブラリを含んでいてもよい。上記状態は、ステージ 1 の乳がんを含んでいてもよい。

【 0 0 5 1 】

本発明のいくつかの実施形態において、1 つまたは複数のプログラムは、分割する場合、グループに割り当てる生データ・セットのレコードを選択する乱数を生成する。

【 0 0 5 2 】

本発明のいくつかの実施形態において、1 つまたは複数のプログラムは、主成分分析の実行に先立って、すべてのペプチドにわたりペプチドのペプチド値を合算するとともに各発現値を結果の合計値で割って比を演算することにより、生データ・セットを含むペプチドを正規化する。

【 0 0 5 3 】

本発明のいくつかの実施形態において、1 つまたは複数のプログラムは、その他のグループに対して主成分分析を実行する場合、トレーニング・データ集合を生成するが、このトレーニング・データ集合の各トレーニング・データ集合は、その他のグループのうちの 1 つを除くすべてのグループのデータを含む。1 つまたは複数のプログラムは、各トレーニング・データ集合に対して個別に主成分分析を実行することによって、各トレーニング・データ集合の共通特徴を識別する。1 つまたは複数のプログラムは、その他のグループを単一のトレーニング・セットとして利用することにより主成分分析を実行することによって、単一のトレーニング・セットの共通特徴を識別する。また、1 つまたは複数のプログラムは、各トレーニング・データ集合の共通特徴および単一のトレーニング・セットの共通特徴について、その他のグループの数に等しい回数の相互検証を実行することによって、主成分を識別するとともに係数を決定する。

【 0 0 5 4 】

本発明のいくつかの実施形態において、所定レベルの検証精度は、回数に基づく最高の検証精度である。

【 0 0 5 5 】

本発明のいくつかの実施形態において、１つまたは複数のプログラムは、所与の個人の１２-mer ペプチド・ライブラリを含む新たな生データを取得する。１つまたは複数のプログラムは、予測モデルを新たな生データに適用して、所与の個人の２値結果を決定する。本発明のいくつかの実施形態において、１つまたは複数のプログラムは、予測モデルを適用する場合、係数を利用して新たな生データに対する主成分分析を実行するとともに、ロジスティック回帰を実行して２値結果を生成する。

【００５６】

ここで図５を参照して、この図は、コンピューティング・ノードの一例を模式的に示しており、クラウド・コンピューティング・ノード１０が可能である。クラウド・コンピューティング・ノード１０は、好適なクラウド・コンピューティング・ノードの一例に過ぎず、本明細書に記載の本発明の実施形態の使用または機能の範囲に関する制限の示唆を何ら意図したものではない。いずれにしろ、クラウド・コンピューティング・ノード１０は、実装または上述の機能のいずれかの実行、あるいはその両方が可能である。本発明の一実施形態において、PCAを実行する処理デバイスを含むコンピューティング・リソースは、たとえば１つまたは複数のクラウド・コンピューティング・ノード１０の一部として理解可能であり（図５）、クラウド・コンピューティング・ノード１０の一部の例でない場合は、クラウド・コンピューティング・ノード１０の態様を含む１つまたは複数の一般的なコンピューティング・ノードの一部として理解可能である。

【００５７】

クラウド・コンピューティング・ノード１０には、他の多くの汎用または専用コンピューティング・システム環境または構成と作用するコンピュータ・システム／サーバ１２が存在する。コンピュータ・システム／サーバ１２との併用に適し得る周知のコンピューティング・システム、環境、または構成、あるいはその組み合わせの例としては、パーソナル・コンピュータ・システム、サーバ・コンピュータ・システム、シン・クライアント、シック・クライアント、手持ち式もしくはラップトップ・デバイス、マルチプロセッサ・システム、マイクロプロセッサベースのシステム、セット・トップ・ボックス、プログラム可能な家電製品、ネットワークPC、ミニコンピュータ・システム、メインフレーム・コンピュータ・システム、ならびに上記システムもしくはデバイスのいずれかを含む分散クラウド・コンピューティング環境が挙げられるが、これらに限定されない。

【００５８】

コンピュータ・システム／サーバ１２は、プログラム・モジュール等、コンピュータ・システムにより実行されるコンピュータ・システム実行可能命令の一般的背景にて記述されていてもよい。一般的に、プログラム・モジュールとしては、特定のタスクの実行または特定の抽象データ型の実装を行うルーチン、プログラム、オブジェクト、コンポーネント、ロジック、データ構造等が挙げられる。コンピュータ・システム／サーバ１２は、通信ネットワークを通じてつながったりモート処理デバイスによりタスクが実行される分散クラウド・コンピューティング環境において実現されるようになっていてもよい。分散クラウド・コンピューティング環境においては、メモリ記憶デバイスを含むローカルおよびリモートの両コンピュータ・システム記憶媒体にプログラム・モジュールが配置されていてもよい。

【００５９】

図５に示すように、クラウド・コンピューティング・ノード１０として利用可能なコンピュータ・システム／サーバ１２は、汎用コンピューティング・デバイスの形態で示している。コンピュータ・システム／サーバ１２の構成要素としては、１つまたは複数のプロセッサまたは処理ユニット１６、システム・メモリ２８、およびシステム・メモリ２８を含むさまざまなシステム構成要素をプロセッサ１６に結合するバス１８が挙げられるが、これらに限定されない。

【００６０】

バス１８は、任意の複数種類のバス構造のうちの１つまたは複数を表し、多様なバス・アーキテクチャのいずれかを用いたメモリ・バスもしくはメモリ・コントローラ、周辺機

10

20

30

40

50

器用バス、アクセラレイテッド・グラフィックス・ポート、ならびにプロセッサもしくはローカル・バスが挙げられる。非限定的な一例として、このようなアーキテクチャとしては、業界標準アーキテクチャ（ＩＳＡ）バス、マイクロ・チャネル・アーキテクチャ（ＭＣＡ）バス、拡張ＩＳＡ（ＥＩＳＡ）バス、ビデオ・エレクトロニクス規格協会（ＶＥＳＡ）ローカル・バス、および周辺機器相互接続（ＰＣＩ）バスが挙げられる。

【００６１】

コンピュータ・システム／サーバ１２は通常、多様なコンピュータ・システム可読媒体を含む。このような媒体は、コンピュータ・システム／サーバ１２によりアクセス可能な任意の利用可能な媒体であってもよく、揮発性および不揮発性媒体、リムーバブルおよび非リムーバブル媒体の両者を含む。

【００６２】

システム・メモリ２８は、ランダム・アクセス・メモリ（ＲＡＭ）３０またはキャッシュ・メモリ３２、あるいはその両方等、揮発性メモリの形態のコンピュータ・システム可読媒体を含み得る。コンピュータ・システム／サーバ１２は、他のリムーバブル／非リムーバブル、揮発性／不揮発性コンピュータ・システム可読媒体をさらに含んでもよい。非リムーバブル不揮発性磁気媒体（図示しないが、通常は「ハード・ドライブ」と称する）に対する読み書き用として、ストレージ・システム３４を設けることも可能であるが、一例に過ぎない。図示はしていないものの、リムーバブル不揮発性磁気ディスク（たとえば、「フロッピー・ディスク」）に対する読み書き用の磁気ディスク・ドライブならびにＣＤ－ＲＯＭ、ＤＶＤ－ＲＯＭ、もしくは他の光学媒体等のリムーバブル不揮発性光ディスクに対する読み書き用の光ディスク・ドライブを設けることも可能である。このような場合は、１つまたは複数のデータ媒体インターフェースによって、それぞれをバス１８に接続可能である。以下に別途図示および説明する通り、メモリ２８は、本発明の実施形態の機能を実行するように構成された一組（たとえば、少なくとも一組）のプログラム・モジュールを有する少なくとも１つのプログラム製品を含んでもよい。

【００６３】

非限定的な一例として、一組（少なくとも一組）のプログラム・モジュール４２を有するプログラム／ユーティリティ４０のほか、オペレーティング・システム、１つまたは複数のアプリケーション・プログラム、他のプログラム・モジュール、およびプログラム・データがメモリ２８に格納されていてもよい。オペレーティング・システム、１つもしくは複数のアプリケーション・プログラム、他のプログラム・モジュール、およびプログラム・データ、またはこれらの何らかの組み合わせはそれぞれ、ネットワーキング環境の一実施態様を含んでもよい。プログラム・モジュール４２は一般的に、本明細書に記載の本発明の実施形態の機能または方法、あるいはその両方を実行する。

【００６４】

また、コンピュータ・システム／サーバ１２は、キーボード、ポインティング・デバイス、ディスプレイ２４等の１つもしくは複数の外部デバイス１４、ユーザによる当該コンピュータ・システム／サーバ１２との相互作用を可能にする１つもしくは複数のデバイス、または当該コンピュータ・システム／サーバ１２による１つもしくは複数の他のコンピュータ・システム／サーバ１２との通信を可能にする任意のデバイス（たとえば、ネットワーク・カード、モデム等）、あるいはその組み合わせと通信するようになっていてもよい。このような通信は、入出力（Ｉ／Ｏ）インターフェース２２を介して発生し得る。さらに、コンピュータ・システム／サーバ１２は、ネットワーク・アダプタ２０を介して、ローカル・エリア・ネットワーク（ＬＡＮ）、一般的なワイド・エリア・ネットワーク（ＷＡＮ）、またはパブリック・ネットワーク（たとえば、インターネット）、あるいはその組み合わせ等の１つまたは複数のネットワークと通信可能である。図示のように、ネットワーク・アダプタ２０は、バス１８を介して、コンピュータ・システム／サーバ１２の他の構成要素と通信する。図示はしていないものの、コンピュータ・システム／サーバ１２と併せて、他のハードウェアまたはソフトウェア構成要素、あるいはその両方を使用することも可能であることが了解されるものとする。その例としては、マイクロコード、デバイス

10

20

30

40

50



・ドライバ、冗長処理ユニット、外部ディスク・ドライブ・アレイ、RAIDシステム、テープ・ドライブ、およびデータ記録ストレージ・システム等が挙げられるが、これらに限定されない。

【0065】

本開示はクラウド・コンピューティングに関する詳細な説明を含むものの、本明細書に列挙の教示内容の実装は、クラウド・コンピューティング環境に限定されないことが了解されるものとする。むしろ、本発明の実施形態は、その他任意の種類の既知のコンピューティング環境または今後開発されるコンピューティング環境と併せて実現可能である。

【0066】

クラウド・コンピューティングは、最小限の管理労力またはサービス提供者との相互作用で迅速に設定および公開し得る設定可能なコンピューティング・リソース（たとえば、ネットワーク、ネットワーク帯域幅、サーバ、処理、メモリ、ストレージ、アプリケーション、仮想マシン、およびサービス）の共用プールへの簡便なオンデマンド・ネットワーク・アクセスを可能にするサービス提供のモデルである。このクラウド・モデルには、少なくとも5つの特性、少なくとも3つのサービス・モデル、および少なくとも4つの配置モデルを含み得る。

【0067】

特性は、以下の通りである。

【0068】

オンデマンド・セルフサービス：クラウド利用者は、サービス提供者との人間の相互作用の必要なく、必要に応じて自動的に、サーバ時間およびネットワーク・ストレージ等のコンピューティング機能を一方的に設定可能である。

【0069】

広範なネットワーク・アクセス：各機能は、ネットワーク上で利用可能であり、シン/シックを問わず異種クライアント・プラットフォーム（たとえば、携帯電話、ラップトップ、およびPDA）による使用を促進する標準的な機構を通じてアクセスされる。リソース・プーリング：マルチテナント・モデルを用いることにより、複数の利用者に対してサービスするために、提供者のコンピューティング・リソースがプールされる。その際、要求に応じて、さまざまな物理的リソースおよび仮想的リソースが動的に割り当ておよび再割り当てされる。利用者は一般的に、提供されるリソースの正確な場所を制御も把握もできないが、より高い抽象化レベル（たとえば、国、州、またはデータセンタ）では場所を特定し得る点において、ある意味で場所の独立性が存在する。迅速な柔軟性：各機能は、場合によっては自動的に、高速スケール・アウトするように迅速かつ柔軟に設定され、高速スケール・インするように迅速に公開され得る。利用者にとって、設定に利用可能な機能は、見かけ上は制限なく、いつでも如何なる量でも購入可能である。

【0070】

サービス測定：クラウド・システムは、サービスの種類（たとえば、ストレージ、処理、帯域幅、およびアクティブ・ユーザ・アカウント）に適した何らかの抽象化レベルでの測定機能を利用することによって、リソースの使用を自動的に制御および最適化する。リソースの使用量は、モニタリング、制御、および報告によって、利用サービスの提供者および利用者の両者に透明性をもたらし得る。

【0071】

サービス・モデルは、以下の通りである。

【0072】

サービスとしてのソフトウェア（SaaS）：利用者に提供される機能は、クラウド・インフラ上で動作する提供者のアプリケーションを使用する。これらのアプリケーションは、ウェブ・ブラウザ等のシン・クライアント・インターフェースを通じて、さまざまなクライアント・デバイスからアクセス可能である（たとえば、ウェブベースの電子メール）。利用者は、ネットワーク、サーバ、オペレーティング・システム、ストレージ、あるいは個々のアプリケーション機能等、下層のクラウド・インフラを管理も制御もしない。

10

20

30

40

50

ただし、ユーザ固有の限定されたアプリケーション構成の設定については、この限りではない。

【 0 0 7 3 】

サービスとしてのプラットフォーム（ P a a S ）：利用者に提供される機能は、提供者がサポートするプログラミング言語およびツールを用いて作成された利用者作成または取得アプリケーションをクラウド・インフラ上に配置する。利用者は、ネットワーク、サーバ、オペレーティング・システム、またはストレージ等、下層のクラウド・インフラを管理も制御もしない。ただし、配置されたアプリケーションと、場合によっては、アプリケーション・ホスティング環境の構成とを制御する。

【 0 0 7 4 】

サービスとしてのインフラ（ I a a S ）：利用者に提供される機能は、処理、ストレージ、ネットワーク、および他の基本的なコンピューティング・リソースを設定するが、利用者は、オペレーティング・システムおよびアプリケーション等、任意のソフトウェアを配置および実行可能である。利用者は、下層のクラウド・インフラを管理も制御もしない。ただし、オペレーティング・システム、ストレージ、配置されたアプリケーションを制御するとともに、場合によっては、選択されたネットワーク・コンポーネント（たとえば、ホストのファイアウォール）を限定的に制御する。

【 0 0 7 5 】

配置モデルは、以下の通りである。

【 0 0 7 6 】

プライベート・クラウド：このクラウド・インフラは、ある組織のために単独で運用される。また、当該組織またはサード・パーティにより管理され、オンプレミスまたはオフプレミスとして存在し得る。

【 0 0 7 7 】

コミュニティ・クラウド：このクラウド・インフラは、複数の組織により共有され、懸案事項（たとえば、ミッション、セキュリティ要件、ポリシー、およびコンプライアンス事項）を共有する特定のコミュニティをサポートする。また、当該組織またはサード・パーティにより管理され、オンプレミスまたはオフプレミスとして存在し得る。

【 0 0 7 8 】

パブリック・クラウド：このクラウド・インフラは、一般の人々または大規模な業界団体が利用可能で、クラウド・サービスを販売する組織により所有される。

【 0 0 7 9 】

ハイブリッド・クラウド：このクラウド・インフラは、2つ以上のクラウド（プライベート、コミュニティ、またはパブリック）の組み合わせであり、それぞれに固有のエンティティは維持する一方、データおよびアプリケーションの移植性（たとえば、クラウド間の負荷分散のためのクラウド・バースティング）を実現する標準または個別の技術により結合される。

【 0 0 8 0 】

クラウド・コンピューティング環境は、無国籍、低結合、モジュール方式、およびセマンティック相互運用性に焦点を当てたサービス指向型である。クラウド・コンピューティングの中心には、相互接続ノードのネットワークを含むインフラがある。

【 0 0 8 1 】

ここで図6を参照して、この図は、例示的なクラウド・コンピューティング環境50を示している。図示のように、クラウド・コンピューティング環境50は、たとえば個人用デジタル補助装置（PDA）もしくは携帯電話54A、デスクトップ・コンピュータ54B、ラップトップ・コンピュータ54C、または自動車コンピュータ・システム54N、あるいはその組み合わせ等、クラウド利用者が使用するローカルのコンピューティング・デバイスの通信を可能にする1つまたは複数のクラウド・コンピューティング・ノード10を具備する。ノード10は、互いに通信するようにしてもよい。これらは、上述のプライベート、コミュニティ、パブリック、もしくはハイブリッド・クラウド、またはその組

10

20

30

40

50

み合わせ等、1つまたは複数のネットワークにおいて物理的または仮想的にグループ化されていてもよい（図示せず）。これにより、クラウド・コンピューティング環境50は、クラウド利用者がローカルのコンピューティング・デバイス上でリソースを維持する必要のないサービスとしてのインフラ、プラットフォーム、またはソフトウェア、あるいはその組み合わせを提供することができる。図6に示すコンピューティング・デバイス54A～54Nの種類は、例示を意図したものに過ぎず、コンピューティング・ノード10およびクラウド・コンピューティング環境50は、（たとえば、ウェブ・ブラウザを用いることにより）任意の種類のネットワークまたはネットワーク・アドレス指定可能な接続、あるいはその両方において、任意の種類のコンピュータ・デバイスと通信可能であることが了解される。

10

#### 【0082】

ここで図7を参照して、この図は、クラウド・コンピューティング環境50（図6）が提供する一組の機能的抽象化レイヤを示している。図7に示すコンポーネント、レイヤ、および機能は、例示を意図したものに過ぎず、本発明の実施形態はこれらに限定されないことが予め了解されるものとする。図示のように、以下のレイヤおよび対応する機能が提供される。

#### 【0083】

ハードウェアおよびソフトウェア・レイヤ60は、ハードウェアおよびソフトウェア・コンポーネントを含む。ハードウェア・コンポーネントの例としては、メインフレーム61、RISC（縮小命令セット・コンピュータ）アーキテクチャ・ベースのサーバ62、サーバ63、ブレード・サーバ64、記憶デバイス65、およびネットワーク/ネットワークング・コンポーネント66が挙げられる。いくつかの実施形態において、ソフトウェア・コンポーネントとしては、ネットワーク・アプリケーション・サーバ・ソフトウェア67およびデータベース・ソフトウェア68が挙げられる。

20

#### 【0084】

仮想化レイヤ70は、抽象化レイヤを提供し、考え得る仮想エンティティの例としては、仮想サーバ71、仮想ストレージ72、仮想ネットワーク73（仮想プライベート・ネットワークを含む）、仮想アプリケーション/オペレーティング・システム74、および仮想クライアント75が挙げられる。

#### 【0085】

一例において、管理レイヤ80は、後述の機能を提供するようにしてもよい。リソース設定81は、クラウド・コンピューティング環境におけるタスクの実行に利用されるコンピューティング・リソース等のリソースの動的な調達を提供する。測定・価格設定82は、クラウド・コンピューティング環境においてリソースが利用される場合のコスト追跡ならびにこれらリソースの消費に対する請求もしくはインボイスを提供する。一例において、これらのリソースには、アプリケーション・ソフトウェア・ライセンスを含み得る。セキュリティは、クラウド利用者およびタスクの識別情報確認のほか、データおよび他のリソースの保護を提供する。ユーザ・ポータル83は、利用者およびシステム管理者に対してクラウド・コンピューティング環境へのアクセスを提供する。サービス・レベル管理84は、所要サービス・レベルを満足するように、クラウド・コンピューティング・リソースの割り当ておよび管理を提供する。サービス・レベル合意（SLA）計画・実行85は、SLAに従って将来的な要件が見込まれるクラウド・コンピューティング・リソースの事前の取り決めおよび調達を提供する。

30

40

#### 【0086】

ワークロード・レイヤ90は、クラウド・コンピューティング環境を利用可能な機能の例を提供する。このレイヤから提供可能なワークロードおよび機能の例としては、マッピングおよびナビゲーション91、ソフトウェア開発・ライフサイクル管理92、仮想教室/教育提供93、データ解析処理94、トランザクション処理95、および予測モデル生成96が挙げられる。

#### 【0087】

50

本発明は、考え得る任意の技術的詳細統合レベルにおけるシステム、方法、またはコンピュータ・プログラム製品、あるいはその組み合わせであってもよい。コンピュータ・プログラム製品は、本発明の態様をプロセッサに実行させるコンピュータ可読プログラム命令が格納された（１つまたは複数の）コンピュータ可読記憶媒体を具備していてもよい。

【 0 0 8 8 】

コンピュータ可読記憶媒体としては、命令実行デバイスが使用する命令を保持および格納し得る有形デバイスが可能である。コンピュータ可読記憶媒体は、たとえば電子記憶デバイス、磁気記憶デバイス、光学記憶デバイス、電磁記憶デバイス、半導体記憶デバイス、またはこれらの任意好適な組み合わせであってもよいが、これらに限定されない。コンピュータ可読記憶媒体のより具体的な例の非網羅的な一覧には、携帯型コンピュータ・ディスクレット、ハード・ディスク、ランダム・アクセス・メモリ（RAM）、リードオンリー・メモリ（ROM）、消去・プログラム可能リードオンリー・メモリ（EPROMもしくはフラッシュ・メモリ）、スタティック・ランダム・アクセス・メモリ（SRAM）、携帯型コンパクト・ディスク・リードオンリー・メモリ（CD-ROM）、デジタル多用途ディスク（DVD）、メモリ・スティック、フロッピー・ディスク、パンチカードもしくは命令が記録された溝中の隆起構造等の機械的符号化デバイス、ならびにこれらの任意好適な組み合わせを含む。本明細書において、コンピュータ可読記憶媒体は、電波等の自由伝搬電磁波、導波路等の送信媒体を伝搬する電磁波（たとえば、光ファイバ・ケーブルを通過する光パルス）、またはワイヤを通じて送信される電気信号等、本質的に一過性の信号としては解釈されないものとする。

【 0 0 8 9 】

本明細書に記載のコンピュータ可読プログラム命令は、コンピュータ可読記憶媒体から各コンピューティング／処理デバイスにダウンロードすることも可能であるし、ネットワーク、たとえばインターネット、ローカル・エリア・ネットワーク、ワイド・エリア・ネットワーク、または無線ネットワーク、あるいはその組み合わせを介して外部コンピュータまたは外部記憶デバイスにダウンロードすることも可能である。ネットワークは、送信銅ケーブル、送信光ファイバ、無線送信、ルータ、ファイアウォール、スイッチ、ゲートウェイ・コンピュータ、またはエッジ・サーバ、あるいはその組み合わせを含んでいてもよい。各コンピューティング／処理デバイスのネットワーク・アダプタ・カードまたはネットワーク・インターフェースは、ネットワークからコンピュータ可読プログラム命令を受信し、当該コンピュータ可読プログラム命令を転送して、各コンピューティング／処理デバイス内のコンピュータ可読記憶媒体に格納する。

【 0 0 9 0 】

本発明の動作を実行するコンピュータ可読プログラム命令は、アセンブラ命令、命令セット・アーキテクチャ（ISA）命令、マシン命令、マシン依存命令、マイクロコード、ファームウェア命令、状態設定データ、集積回路の設定データ、あるいはSmalltalk（R）、C++等のオブジェクト指向プログラミング言語ならびにCプログラミング言語もしくは類似のプログラミング言語等の手続き型プログラミング言語を含む１つまたは複数のプログラミング言語の任意の組み合わせで記述されたソース・コードまたはオブジェクト・コードであってもよい。コンピュータ可読プログラム命令は、独立型ソフトウェア・パッケージとして全部または一部をユーザのコンピュータ上で実行してもよいし、一部をユーザのコンピュータ上、一部をリモート・コンピュータ上で実行してもよいし、全部をリモート・コンピュータまたはサーバ上で実行してもよい。後者のシナリオでは、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）等、任意の種類のネットワークを通じてリモート・コンピュータがユーザのコンピュータに接続されていてもよいし、（たとえば、インターネット・サービス・プロバイダを用いることによりインターネットを通じて）外部コンピュータに接続されていてもよい。いくつかの実施形態においては、本発明の態様を実行するため、コンピュータ可読プログラム命令の状態情報を利用して電子回路をカスタマイズすることにより、たとえばプログラマブル論理回路、フィールドプログラマブル・ゲート・アレイ（FPGA）、またはプ

ログラマブル・ロジック・アレイ ( P L A ) を含む電子回路がコンピュータ可読プログラム命令を実行するようにしてもよい。

【 0 0 9 1 】

本明細書においては、本発明の実施形態に係る方法、装置 ( システム ) 、およびコンピュータ・プログラム製品のフローチャート図またはブロック図あるいはその両方を参照して、本発明の態様を説明している。フローチャート図またはブロック図あるいはその両方の各ブロックならびにフローチャート図またはブロック図あるいはその両方のブロックの組み合わせは、コンピュータ可読プログラム命令により実装可能であることが了解される。

【 0 0 9 2 】

これらのコンピュータ可読プログラム命令は、コンピュータまたは他のプログラム可能データ処理装置のプロセッサを介して実行される命令が、フローチャートまたはブロック図あるいはその両方の1つまたは複数のブロックに規定の機能 / 動作を実装する手段を生成するように、汎用コンピュータ、専用コンピュータ、または他のプログラム可能データ処理装置のプロセッサに提供されてマシンを作り出すものであってよい。また、これらのコンピュータ可読プログラム命令は、命令が格納されたコンピュータ可読記憶媒体が、フローチャートまたはブロック図あるいはその両方の1つまたは複数のブロックに規定の機能 / 動作の態様を実装する命令を含む製造品を含むように、コンピュータ可読記憶媒体に格納され、コンピュータ、プログラム可能データ処理装置、または他のデバイス、あるいはその組み合わせに対して特定の状態で機能するように指示し得るものであってよい。

【 0 0 9 3 】

また、コンピュータ可読プログラム命令は、コンピュータ、他のプログラム可能装置、または他のデバイス上での実行される命令が、フローチャートまたはブロック図あるいはその両方の1つまたは複数のブロックに規定の機能 / 動作を実装するように、コンピュータ実装プロセスを作り出すべく、コンピュータ、他のプログラム可能データ処理装置、または他のデバイスにロードされ、コンピュータ、他のプログラム可能装置、または他のデバイス上で一連の動作ステップを実行させるものであってよい。

【 0 0 9 4 】

図中のフローチャートおよびブロック図は、本発明の種々実施形態に係るシステム、方法、およびコンピュータ・プログラム製品の考え得る実施態様のアーキテクチャ、機能、および動作を示している。この点、フローチャートまたはブロック図の各ブロックは、特定の論理機能を実装する1つまたは複数の実行可能命令を含む命令のモジュール、セグメント、または一部を表し得る。いくつかの代替実施態様において、ブロックに記載の機能は、図面に記載の順序から外れて発生するようになっていてもよい。たとえば、連続して示す2つのブロックが実際のところは、関与する機能に応じて、実質的に同時に実行されるようになっていてもよいし、場合により逆の順序で実行されるようになっていてもよい。また、ブロック図またはフローチャート図あるいはその両方の各ブロックならびにブロック図またはフローチャート図あるいはその両方のブロックの組み合わせは、特定の機能または動作を実行する専用ハードウェアベースのシステムにより実装することも可能であるし、専用ハードウェアおよびコンピュータ命令の組み合わせを実行することも可能であることに留意されたい。

【 0 0 9 5 】

本明細書において使用する専門用語は、特定の実施形態を説明することを目的としているに過ぎず、何ら限定する意図はない。本明細書において、単数形「ある ( a ) 」、「一 ( a n ) 」、および「その ( t h e ) 」は、文脈上の別段の明確な指定のない限り、複数形も同様に含むことが意図される。本明細書において使用する場合の用語「備える ( c o m p r i s e s ) 」または「備えている ( c o m p r i s i n g ) 」、あるいはその両方は、記載の特徴、整数、ステップ、動作、要素、またはコンポーネント、あるいはその組み合わせの存在を指定するものの、1つまたは複数の他の特徴、整数、ステップ、動作、要素、コンポーネント、またはその群、あるいはその組み合わせの存在または追加を除外するものではないことがさらに了解される。

10

20

30

40

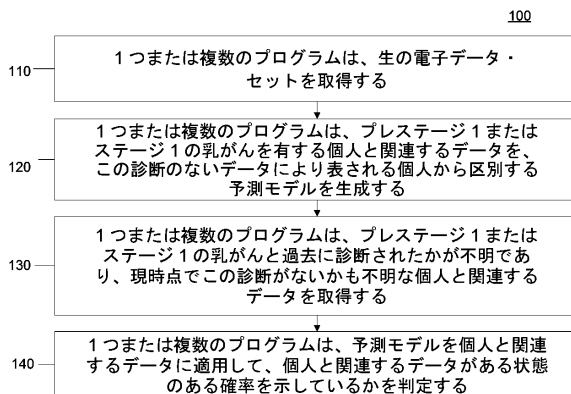
50

## 【 0 0 9 6 】

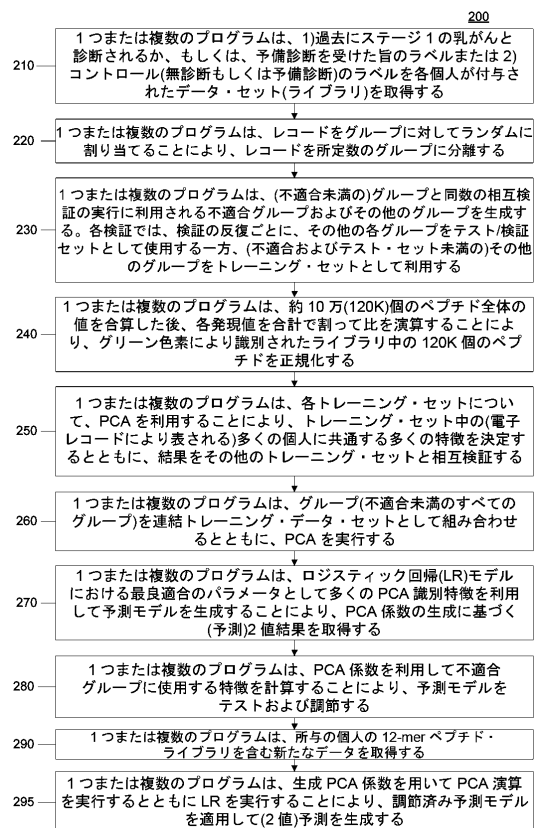
以下の特許請求の範囲において、すべてのミーンズ・プラス・ファンクション要素またはステップ・プラス・ファンクション要素の対応する構造、材料、動作、および同等物が存在する場合は、具体的に請求される他の特許請求の範囲に係る要素との組み合わせで機能を実行する任意の構造、材料、または動作を含むことが意図される。1つまたは複数の実施形態の記載は、例示および説明を目的として提示したものであるが、網羅性または開示の形態への限定を意図したものではない。当業者には、多くの改良および変形が明らかとなるであろう。上記実施形態は、種々態様および実際の用途を最もよく説明するとともに、当業他者が企図される特定の使用に適した種々改良を伴う種々実施形態を理解できるように選定および記載したものである。

## 【 図面 】

## 【 図 1 】



## 【 図 2 】



10

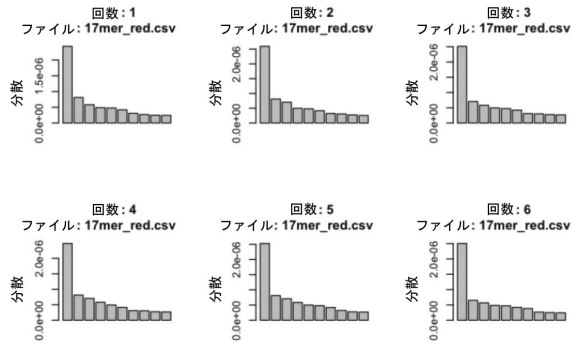
20

30

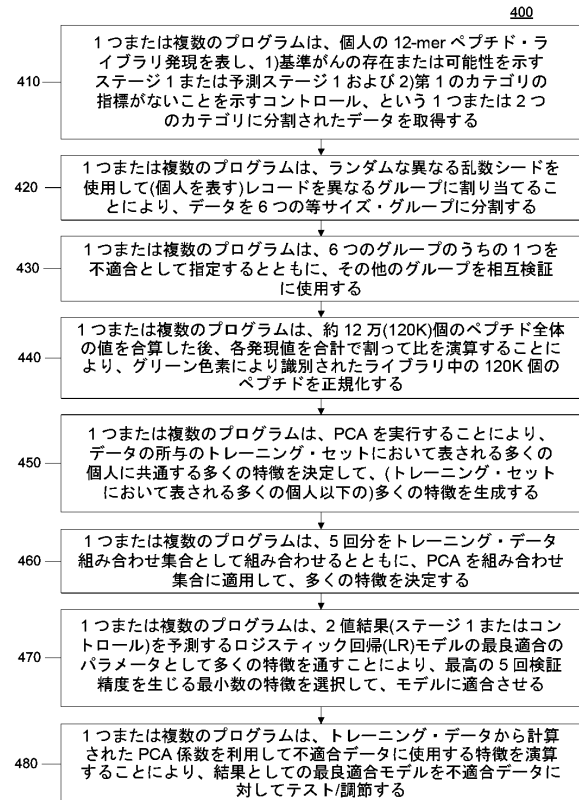
40

50

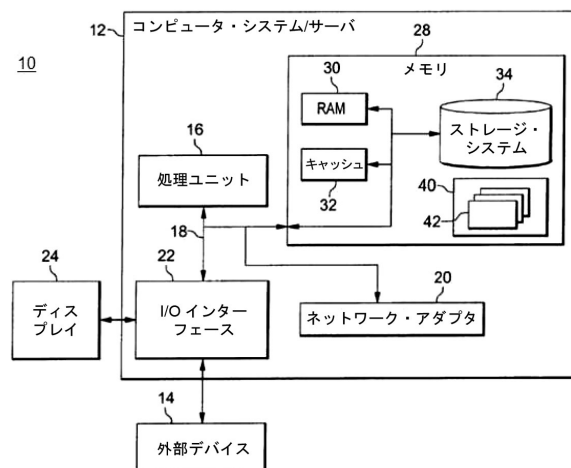
【図 3】



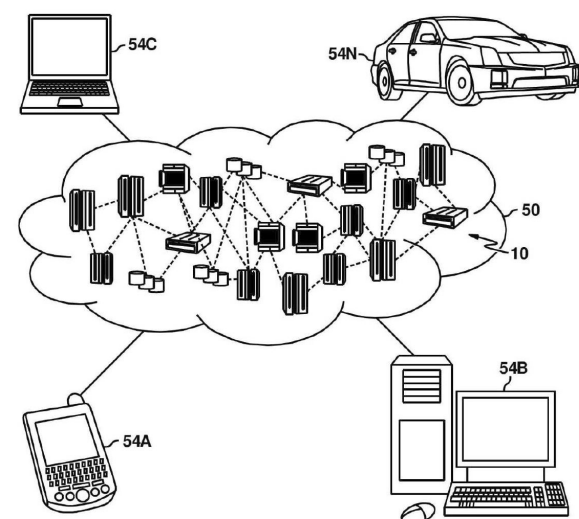
【図 4】



【図 5】



【図 6】



10

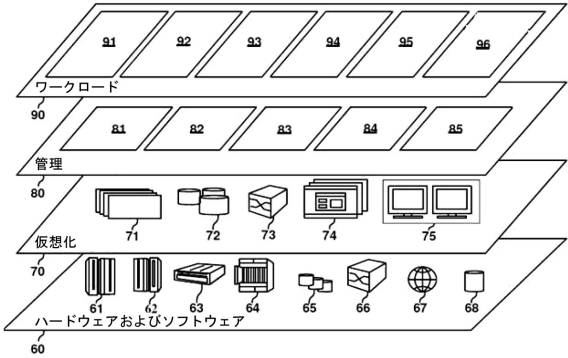
20

30

40

50

【図 7】



10

20

30

40

50



---

フロントページの続き

(72)発明者 チャッタジー、アビジット

アメリカ合衆国 1 0 5 0 4 - 1 7 8 4 ニューヨーク州 アーモンク ノース・キャッスル・ドライブ 1

(72)発明者 ワン、ウェンディ

アメリカ合衆国 1 0 5 0 4 - 1 7 8 4 ニューヨーク州 アーモンク ノース・キャッスル・ドライブ 1

審査官 森田 充功

(56)参考文献 特表 2 0 1 6 - 5 3 7 6 9 9 ( J P , A )

(58)調査した分野 (Int.Cl. , D B 名)

G 1 6 H 1 0 / 0 0 - 8 0 / 0 0

G 1 6 B 4 0 / 3 0