(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)公表番号
特表2022-518286
(P2022-518286A)
(43)公表日　令和4年3月14日(2022.3.14)

(51)国際特許分類　FI
G06N 20/00 (2019.01)　G06N 20/00　130
G06N 3/08 (2006.01)　G06N 3/08

審査請求　未請求　予備審査請求　未請求　（全35頁）

(21)出願番号　特願2021-543210(P2021-543210)
(86)(22)出願日　令和2年1月23日(2020.1.23)
(85)翻訳文提出日　令和3年7月26日(2021.7.26)
(86)国際出願番号　PCT/EP2020/051647
(87)国際公開番号　WO2020/156924
(87)国際公開日　令和2年8月6日(2020.8.6)
(31)優先権主張番号　19153977.4
(32)優先日　平成31年1月28日(2019.1.28)
(33)優先権主張国・地域又は機関
欧州特許庁(EP)
(81)指定国・地域　AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,
最終頁に続く

(71)出願人　590000248
コーニンクレッカ フィリップス エヌヴェ
Ｋｏｎｉｎｋｌｉｊｋｅ Ｐｈｉｌｉｐｓ Ｎ．Ｖ．
オランダ国 ５６５６ アーヘー アインドーフェン ハイテック キャンパス ５２
Ｈｉｇｈ Ｔｅｃｈ Ｃａｍｐｕｓ ５２，５６５６ ＡＧ Ｅｉｎｄｈｏｖｅｎ，Ｎｅｔｈｅｒｌａｎｄｓ
(74)代理人　110001690
特許業務法人Ｍ＆Ｓパートナーズ
(72)発明者　ウェーゼ ロルフ ユルゲン
オランダ国 ５６５６ アーエー アインドーフェン ハイ テック キャンパス ５
最終頁に続く

(54)【発明の名称】　被訓練モデルへの母集団記述子の関連付け

(57)【要約】
モデルは、例えば医用画像処理又は医用信号処理のタスクのために、訓練データセットで訓練される。母集団記述子を被訓練モデルに関連付け、それら母集団記述子を使用して、モデルが適用されるレコードが母集団記述子と合致するかどうかを判定するためのシステム及びコンピュータ実施方法が提供される。母集団記述子は、訓練データセットにわたる１つ又は複数の特性特徴の分布を特徴付け、特性特徴は、訓練レコード及び／又は被訓練モデルが訓練レコードに適用されたときに提供されるモデル出力を特徴付ける。例えば、モデルは、母集団記述子と合致するレコードのみに適用されるか、又は、合致しないレコードにモデルを適用したモデル出力が、信頼できない可能性があるとしてフラグを立てられる。

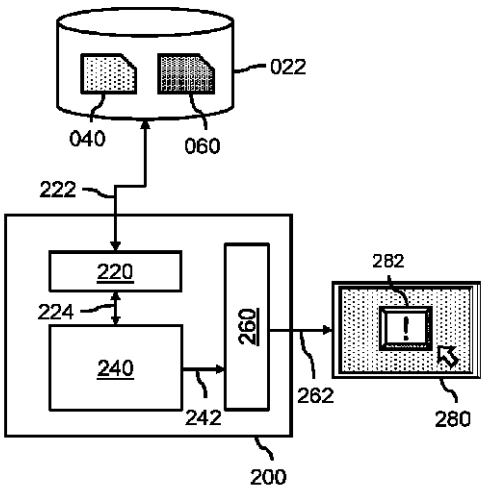


Fig. 2

【特許請求の範囲】
【請求項１】
訓練データセットの訓練レコードを処理するためのシステムであって、前記システムは、被訓練モデルのための訓練データセットであって、前記訓練データセットは複数の訓練レコードを備え、前記被訓練モデルは、レコードに適用されたときにモデル出力を提供するように前記訓練データセットで訓練される、訓練データセットにアクセスするためのデータインターフェースと、
プロセッササブシステムとを備え、
前記プロセッササブシステムは、
前記訓練データセットの訓練レコードについての１つ又は複数の特性特徴を判定することであって、前記１つ又は複数の特性特徴は、前記訓練レコード、及び／又は前記被訓練モデルが前記訓練レコードに適用されたときに提供されるモデル出力を特徴付け、前記特性特徴は、前記訓練レコード、並びに／又は前記訓練レコード及び／若しくは前記モデル出力に関連付けられたメタデータに基づく、判定すること、
前記訓練データセットにわたる前記１つ又は複数の特性特徴の分布を特徴付ける母集団記述子を判定することであって、前記判定することは、前記被訓練モデルの中間出力にアクセスしない、判定すること、
前記被訓練モデルを少なくとも１つのレコードに適用するエンティティが、前記少なくとも１つのレコードが前記母集団記述子と合致するかどうかを、前記少なくとも１つのレコードについて前記１つ又は複数の特性特徴を判定することにより、及び前記１つ又は複数の特性特徴の、前記訓練データセットにわたる前記特性特徴の前記分布からの逸脱を判定することにより判定することを可能にするために、前記母集団記述子を前記被訓練モデルに関連付けること
を行う、システム。
【請求項２】
前記母集団記述子が、前記訓練データセットにわたる前記１つ又は複数の特性特徴の確率分布を示す確率データを備える、請求項１に記載のシステム。
【請求項３】
前記母集団記述子は、前記被訓練モデルを適用する前記エンティティが前記少なくとも１つのレコードを合致するものと分類することを可能にするために、前記訓練レコードの前記１つ又は複数の特性特徴で訓練された１クラス分類器を備える、請求項１に記載のシステム。
【請求項４】
訓練レコードが画像を備え、前記プロセッササブシステムが、前記画像から特性特徴を抽出することにより、及び／又は前記画像に関連付けられたメタデータから特性特徴を判定することにより、前記訓練レコードの前記１つ又は複数の特性特徴を判定する、請求項１から３のいずれか一項に記載のシステム。
【請求項５】
訓練レコードが画像を備え、前記被訓練モデルが、前記画像のセグメント化を提供するように訓練され、前記プロセッササブシステムが、前記画像の前記セグメント化の形状記述子を抽出することによって前記訓練レコードの特性特徴を判定する、請求項１から４のいずれか一項に記載のシステム。
【請求項６】
前記訓練レコードについての前記特性特徴は、前記被訓練モデルを訓練する入力として使用された訓練レコードの特徴、前記被訓練モデルを訓練する入力として使用されなかったメタデータ、及び／又は前記被訓練モデルを適用したモデル出力である、請求項１から５のいずれか一項に記載のシステム。
【請求項７】
被訓練モデルが適用される少なくとも１つのレコードを処理するためのシステムであって、前記システムは、

前記被訓練モデルに関連付けられた母集団記述子であって、前記被訓練モデルは訓練データセットで訓練され、前記訓練データセットは複数の訓練レコードを備え、前記母集団記述子は、前記訓練データセットにわたる前記訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの前記１つ又は複数の特性特徴は、前記訓練レコード、及び／又は前記被訓練モデルを前記訓練レコードに適用したときに提供されるモデル出力を特徴付け、前記１つ又は複数の特性特徴は、前記少なくとも１つのレコード、並びに／又は前記少なくとも１つのレコード及び／若しくは前記少なくとも１つのレコードに対する前記被訓練モデルのモデル出力に関連付けられたメタデータに基づいて判定され、前記母集団記述子は、前記被訓練モデルの中間出力にアクセスすることなく前記特性特徴に基づいて判定されたものである、母集団記述子、及び、
前記被訓練モデルが適用される前記少なくとも１つのレコード、
にアクセスするためのデータインターフェースと、
プロセッササブシステムとを備え、
前記プロセッササブシステムは、
前記少なくとも１つのレコードの前記１つ又は複数の特性特徴を判定すること、
前記少なくとも１つのレコードの前記１つ又は複数の特性特徴の、前記訓練データセットにわたる前記特性特徴の前記分布からの逸脱を判定して、前記少なくとも１つのレコードが前記母集団記述子と合致するかどうかを判定すること、
前記少なくとも１つのレコードが前記被訓練モデルの前記母集団記述子と合致しない場合、前記非合致を示す出力信号を生成すること
を行う、システム。

【請求項８】

前記出力信号を感知可能な方式でユーザに対してレンダリングするレンダリングデバイスに前記出力信号を出力するための出力インターフェースをさらに備える、請求項７に記載のシステム。

【請求項９】

前記出力インターフェースがさらに、前記母集団記述子への前記非合致を生じさせた、１つ又は複数の判定された特性特徴の逸脱を出力する、請求項８に記載のシステム。

【請求項１０】

前記データインターフェースが、複数の被訓練モデルに関連付けられた母集団記述子にアクセスし、
前記プロセッササブシステムが、
前記少なくとも１つのレコードが、それぞれの被訓練モデルに関連付けられた母集団記述子と合致するかどうかを判定することと、
合致する場合、前記少なくとも１つのレコードに適用するために前記複数の被訓練モデルのうちの前記それぞれの被訓練モデルを選択することと
を行う、請求項７から９のいずれか一項に記載のシステム。

【請求項１１】

前記プロセッササブシステムがさらに、前記出力信号に応答して前記被訓練モデルのモデルパラメータを適合する、請求項７から１０のいずれか一項に記載のシステム。

【請求項１２】

訓練データセットの訓練レコードを処理するコンピュータ実施方法であって、前記コンピュータ実施方法は、
被訓練モデルの訓練データセットにアクセスするステップであって、前記訓練データセットは複数の訓練レコードを備え、前記被訓練モデルは、レコードに適用されたときにモデル出力を提供するように前記訓練データセットで訓練される、ステップと、
前記訓練データセットの訓練レコードについての１つ又は複数の特性特徴を判定するステップであって、前記１つ又は複数の特性特徴は、前記訓練レコード、及び／又は前記被訓練モデルが前記訓練レコードに適用されたときに提供されるモデル出力を特徴付け、前記特性特徴は、前記訓練レコード、並びに／又は前記訓練レコード及び／若しくは前記モデ

ル出力に関連付けられたメタデータに基づいて判定される、ステップと、
前記訓練データセットにわたる前記１つ又は複数の特性特徴の分布を特徴付ける母集団記述子を判定するステップであって、前記判定することは、前記被訓練モデルの中間出力にアクセスしない、ステップと、
前記被訓練モデルを少なくとも１つのレコードに適用するエンティティが、前記少なくとも１つのレコードが前記母集団記述子と合致するかどうかを、前記少なくとも１つのレコードについて前記１つ又は複数の特性特徴を判定することにより、及び前記１つ又は複数の特性特徴の、前記訓練データセットにわたる前記特性特徴の前記分布からの逸脱を判定することにより判定することを可能にするために、前記母集団記述子を前記被訓練モデルに関連付けるステップと、
を有する、コンピュータ実施方法。
【請求項１３】
被訓練モデルが適用される少なくとも１つのレコードを処理するコンピュータ実施方法であって、前記コンピュータ実施方法は、
前記被訓練モデルに関連付けられた母集団記述子であって、前記被訓練モデルは訓練データセットで訓練されており、前記訓練データセットは複数の訓練レコードを備え、前記母集団記述子は、前記訓練データセットにわたる前記訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの前記１つ又は複数の特性特徴は、前記訓練レコード、及び／又は前記被訓練モデルを前記訓練レコードに適用したときに提供されるモデル出力を特徴付け、前記１つ又は複数の特性特徴は、前記少なくとも１つのレコード、並びに／又は前記少なくとも１つのレコード及び／若しくは前記少なくとも１つのレコードに対する前記被訓練モデルのモデル出力に関連付けられたメタデータに基づいて判定され、前記母集団記述子は、前記被訓練モデルの中間出力にアクセスすることなく前記特性特徴に基づいて判定されたものである、母集団記述子、及び、
前記被訓練モデルが適用される前記少なくとも１つのレコード
にアクセスするステップと、
前記少なくとも１つのレコードの前記１つ又は複数の特性特徴を判定するステップと、
前記少なくとも１つのレコードの前記１つ又は複数の特性特徴が前記訓練データセットにわたるそれらの分布に従って発生する尤度を判定することにより、前記少なくとも１つのレコードが前記母集団記述子と合致するかどうかを判定するステップと、
前記少なくとも１つのレコードが前記被訓練モデルの前記母集団記述子と合致しない場合、前記非合致を示す出力信号を生成するステップと、
を有する、コンピュータ実施方法。
【請求項１４】
請求項１２又は１３に記載のコンピュータ実施方法をプロセッサシステムに行わせる命令を表す一時的又は非一時的なデータを備える、コンピュータ可読媒体。
【請求項１５】
被訓練モデルと、前記被訓練モデルに関連付けられた母集団記述子とを表す一時的又は非一時的なデータを備えるコンピュータ可読媒体であって、前記被訓練モデルは訓練データセットで訓練されており、前記訓練データセットは複数の訓練レコードを備え、前記母集団記述子は、前記訓練データセットにわたる前記訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの前記１つ又は複数の特性特徴は、前記訓練レコード、及び／又は前記被訓練モデルを前記訓練レコードに適用したときに提供されるモデル出力を特徴付け、前記１つ又は複数の特性特徴は、前記少なくとも１つのレコード、並びに／又は前記少なくとも１つのレコード及び／若しくは前記少なくとも１つのレコードに対する前記被訓練モデルのモデル出力に関連付けられたメタデータに基づいて判定され、前記母集団記述子は、前記被訓練モデルの中間出力にアクセスすることなく前記特性特徴に基づいて判定されたものである、コンピュータ可読媒体。
【発明の詳細な説明】
【技術分野】

【０００１】
本発明は、訓練データセットの訓練レコード、例えば医用レコード又は画像を処理するためのシステムに関する。本発明はさらに、被訓練モデルが適用される少なくとも１つのレコードを処理するためのシステムに関する。本発明はさらに、訓練レコードを処理する方法、及び少なくとも１つのレコードを処理する方法に関する。本発明はさらに、いずれかのコンピュータ実施方法を行うための命令を備えた、又は被訓練モデルを備えた、コンピュータ可読媒体に関する。
【背景技術】
【０００２】
（医用）画像処理や（医用）信号処理における多くのタスクについて、知識のハードコード表現及び／又は静的に構成された表現が、機械学習（ＭＬ）及び／又は深層学習（ＤＬ）方法などの人工知能（ＡＩ）を使用して作られた解決法に置き換えられつつある。例えば、機械学習を応用するのに特に適した分野の１つは、自動的な（医用）画像のセグメント化である。実際には、完全に自動化された方式で画像セグメント化を行うことは困難であることが判明しており、特に、放射線療法の計画、術前計画等の様々な臨床用途に必要とされる精度をもたらすことは難しい。
【０００３】
例えば、（大きな）訓練データセットがグラウンドトゥルースでアノテーションされる教師あり学習方式などの様々なＭＬ／ＤＬツールを使用して、訓練データセットの下にあるパターン及び相関関係を自動的に判定することができる。作成されたモデルは、通例、普通は訓練データから切り離された実行環境で、それまでに見られたことのないデータに対して実行時に実行される。例えば、実行環境を訓練データから切り離す理由の１つは、訓練データが実行環境に不必要に記憶されるのを防ぐためであり、実際、記憶及び処理の要件は、学習フェーズでは実行フェーズよりも数桁大きくなり得る。別の理由は、訓練データセットが、データの所有権、機密性、プライバシー、同意、データ保管場所／地理、又はデータの開示や配布を禁止若しくは制限する何らかのデータ使用法の合意に関する、法的規定及び／又は契約上の合意から生じる要件を有し得ることである。
【０００４】
見られたことのないレコードにモデルを適用した出力の精度は、モデルが訓練された訓練データセットに大きく依存し得る。例えば、被訓練モデルが適用されるレコードが、訓練データセット中の優勢な事例に十分に近いか又は類似していれば、モデル出力が正確であることが予想され、一方、レコードが訓練データセットとそれほど近くなければ、結果の精度が下がるか、又は少なくともその精度に関してより多くの不確実性が生じ得る。例えば、モデルが適用されるレコードは、訓練データセット中の一般的でない事例やまれな条件に相当することがある。モデル出力の潜在的な不正確性は、例えば画像分析パイプラインなど、教師なしで完全に自動的に機能することが期待されるパイプラインで特に関連する。しかし、対話型の用途についても、不正確なモデル出力の可能性に対処することは重要である。
【発明の概要】
【発明が解決しようとする課題】
【０００５】
例えば、モデルが訓練された訓練データセットのレコードから大幅に異なるレコード、例えば比較的遠いレコード、にモデルが適用される場合に、被訓練モデルが不正確であり得るモデル出力を提供する可能性に対処することが望ましい。例えば、低下した精度が予想され得るレコード、例えば、モデル入力及び／又は出力の面で訓練データセットの母集団に合致しないレコードへのモデルの適用を検出すること、及び／又はそれにフラグを立てることが有益であり得る。例えば、そのような事例を検出する、及び／又はフラグを立てることができれば、モデル出力を自動化された方式で使用する助けとなるだけでなく、対話型の用途においてモデル出力が不正確であり得ることをユーザに警告して、例えば警告無しに不正確な結果を表示することによってシステムへの信頼が低下することを回避する

助けともなる。
【課題を解決するための手段】
【０００６】
以下の本発明の態様は、被訓練モデルに関連付けられた母集団記述子を判定することを伴う。母集団記述子は、被訓練モデルを訓練するのに使用された訓練データセットの訓練レコードの１つ又は複数の特性特徴の分布を特徴付けられる。訓練レコードの特性特徴は、当該訓練レコード、被訓練モデルを当該レコードに適用したモデル出力、又はその両方を特徴付ける。母集団記述子は、例えば、レコードが母集団記述子に合致するかどうかを判定するために、訓練データセットにわたる１つ又は複数の特性特徴の逸脱を判定するために使用される。例えば、合致するレコードにモデルを適用したモデル出力は、合致しないレコードにモデルを適用したモデル出力よりも信頼できる。非合致の場合は、例えばユーザに警告するために、出力信号が生成される。
【０００７】
発明の第１の態様によれば、訓練データセットの訓練レコードを処理するためのシステムが提供される。このシステムは、
－　データインターフェースであって、
－　被訓練モデルの訓練データセットであって、訓練データセットは複数の訓練レコードを備え、被訓練モデルは、レコードに適用されたときにモデル出力を提供するように訓練データセットで訓練される、訓練データセット
にアクセスするためのデータインターフェースと、
－　プロセッササブシステムであって、
－　訓練データセットの訓練レコードについての１つ又は複数の特性特徴を判定することであって、前記１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルが訓練レコードに適用されたときに提供されるモデル出力を特徴付け、前記特性特徴は、訓練レコード、並びに／又は訓練レコード及び／若しくはモデル出力に関連付けられたメタデータに基づく、判定すること、
－　訓練データセットにわたる１つ又は複数の特性特徴の分布を特徴付ける母集団記述子を判定することであって、前記判定することは、被訓練モデルの中間出力にアクセスしない、判定すること、
－　被訓練モデルを少なくとも１つのレコードに適用するエンティティが、当該少なくとも１つのレコードが母集団記述子と合致するかどうかを、少なくとも１つのレコードについて１つ又は複数の特性特徴を判定することにより、及び前記１つ又は複数の特性特徴の、訓練データセットにわたる前記特性特徴の分布からの逸脱を判定することにより判定することを可能にするために、母集団記述子を被訓練モデルに関連付けること
を行うように構成されたプロセッササブシステムとを備える。
【０００８】
本発明のさらなる態様によれば、被訓練モデルが適用される少なくとも１つのレコードを処理するためのシステムが提供される。このシステムは、
－　データインターフェースであって、
－　被訓練モデルに関連付けられた母集団記述子であって、被訓練モデルは訓練データセットで訓練されており、訓練データセットは複数の訓練レコードを備え、母集団記述子は、訓練データセットにわたる訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルを訓練レコードに適用したときに提供されるモデル出力を特徴付け、１つ又は複数の特性特徴は、少なくとも１つのレコード、並びに／又は少なくとも１つのレコード及び／若しくは少なくとも１つのレコードに対する被訓練モデルのモデル出力に関連付けられたメタデータに基づいて判定され、母集団記述子は、被訓練モデルの中間出力にアクセスすることなく特性特徴に基づいて判定されたものである、母集団記述子、及び
－　被訓練モデルが適用される少なくとも１つのレコード、
にアクセスするためのデータインターフェースと、

－　プロセッササブシステムであって、
－　少なくとも１つのレコードの１つ又は複数の特性特徴を判定すること、
－　少なくとも１つのレコードの１つ又は複数の特性特徴の、訓練データセットにわたる前記特性特徴の分布からの逸脱を判定して、当該少なくとも１つのレコードが母集団記述子と合致するかどうかを判定すること、
－　少なくとも１つのレコードが被訓練モデルの母集団記述子と合致しない場合、前記非合致を示す出力信号を生成すること
を行うように構成されたプロセッササブシステムとを備える。
【０００９】
本発明のさらなる態様によれば、訓練データセットの訓練レコードを処理するコンピュータ実施方法が提供される。この方法は、
－　被訓練モデルの訓練データセットにアクセスするステップであって、訓練データセットは複数の訓練レコードを備え、被訓練モデルは、レコードに適用されたときにモデル出力を提供するように訓練データセットで訓練される、ステップと
－　訓練データセットの訓練レコードについての１つ又は複数の特性特徴を判定するステップであって、前記１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルが訓練レコードに適用されたときに提供されるモデル出力を特徴付け、前記特性特徴は、訓練レコード、並びに／又は訓練レコード及び／若しくはモデル出力に関連付けられたメタデータに基づいて判定される、ステップと、
－　訓練データセットにわたる１つ又は複数の特性特徴の分布を特徴付ける母集団記述子を判定するステップであって、前記判定することは、被訓練モデルの中間出力にアクセスしない、ステップと、
－　被訓練モデルを少なくとも１つのレコードに適用するエンティティが、当該少なくとも１つのレコードが母集団記述子と合致するかどうかを、少なくとも１つのレコードについて１つ又は複数の特性特徴を判定することにより、及び前記１つ又は複数の特性特徴の、訓練データセットにわたる前記特性特徴の分布からの逸脱を判定することにより判定することを可能にするために、母集団記述子を被訓練モデルに関連付けるステップとを有する。
【００１０】
本発明のさらなる態様によれば、被訓練モデルが適用される少なくとも１つのレコードを処理するコンピュータ実施方法が提供される。この方法は、
－　被訓練モデルに関連付けられた母集団記述子であって、被訓練モデルは訓練データセットで訓練されており、訓練データセットは複数の訓練レコードを備え、母集団記述子は、訓練データセットにわたる訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルを訓練レコードに適用したときに提供されるモデル出力を特徴付け、１つ又は複数の特性特徴は、少なくとも１つのレコード、並びに／又は少なくとも１つのレコード及び／若しくは少なくとも１つのレコードに対する被訓練モデルのモデル出力に関連付けられたメタデータに基づいて判定され、母集団記述子は、被訓練モデルの中間出力にアクセスすることなく特性特徴に基づいて判定されたものである、母集団記述子、及び
－　被訓練モデルが適用される少なくとも１つのレコード
にアクセスするステップと、
－　少なくとも１つのレコードの１つ又は複数の特性特徴を判定するステップと、
－　少なくとも１つのレコードの１つ又は複数の特性特徴が訓練データセットにわたるそれらの分布に従って発生する尤度を判定することにより、当該少なくとも１つのレコードが母集団記述子と合致するかどうかを判定するステップと、
－　少なくとも１つのレコードが被訓練モデルの母集団記述子と合致しない場合、前記非合致を示す出力信号を生成するステップとを有する。
【００１１】
本発明のさらなる態様によれば、いずれか又は両方のコンピュータ実施方法をプロセッサ

システムに行わせるように構成された命令を表す一時的又は非一時的なデータを備えるコンピュータ可読媒体が提供される。

【００１２】

本発明のさらなる態様によれば、被訓練モデルと、被訓練モデルに関連付けられた母集団記述子とを表す一時的又は非一時的なデータを備えるコンピュータ可読媒体が提供され、被訓練モデルは訓練データセットで訓練され、訓練データセットは複数の訓練レコードを備え、母集団記述子は、訓練データセットにわたる訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルを訓練レコードに適用したときに提供されるモデル出力を特徴付け、１つ又は複数の特性特徴は、少なくとも１つのレコード、並びに／又は少なくとも１つのレコード及び／若しくは少なくとも１つのレコードに対する被訓練モデルのモデル出力に関連付けられたメタデータに基づいて判定され、母集団記述子は、被訓練モデルの中間出力にアクセスすることなく特性特徴に基づいて判定されたものである。

【００１３】

上記の手段は、訓練データセットで訓練される被訓練モデルを伴う。訓練データと呼ばれることもある訓練データセットは、訓練事例と呼ばれることもある複数の訓練レコードを備える。記載される技術は、例えば、医用画像、患者レコード、又は最適化すべき機械の性能の測定値や、金融情報などの非医用データ等を備える、幅広い訓練データセットに適用することができる。

【００１４】

被訓練モデルは、レコードに適用されたときにモデル出力を提供するように訓練される。例えば分類、回帰、クラスタリングなど、機械学習から知られている任意の適切なタイプのモデル出力が、記載される技術と組み合わせられる。多くのそのような適切な機械学習モデルが当該技術分野で知られている。モデルを訓練するための様々な技術が、当該技術分野でそれ自体知られており、使用され得る。訓練は、教師あり又は半教師ありであり、例えば、訓練レコードの一部又はすべてが、入力物体と望まれる出力値とを備える。訓練は教師なしであってもよい。訓練レコードのモデル出力を教師あり学習に使用する場合、望まれる出力値、及び／又は被訓練モデルを訓練レコードに適用した出力が使用される。

【００１５】

上記特徴は、訓練レコードの１つ又は複数の特性特徴が判定されることも伴う。特性特徴は、訓練レコード、被訓練モデルを訓練レコードに適用したモデル出力、又はその両方を特徴付ける。例えば、特性特徴は、訓練レコードの特徴、例えば、本質的な特徴、導出された特徴、若しくはメタデータ、及び／又は被訓練モデルをレコードに適用した結果の特徴に対応するか又はそれに基づく。特性特徴は、すべての訓練レコード、又は、少なくとも訓練データセットの訓練レコードの実質的な部分に対して判定されたときに訓練データセットの母集団を記述することが分かっている又は予想される、訓練レコード又はモデル出力の主要な特徴である。例えば、特性特徴は、モデルの訓練への入力として使用された訓練レコードの特徴であるが、モデルを訓練するために必ずしも使用されなかったメタデータ、又はモデルを適用したモデル出力である。一般に、特性特徴は、例えばユーザによって事前に選択される。ただし、それに代えて又はそれに加えて、例えば潜在的な特徴の主成分分析に基づいて、特性特徴の自動化された選択を行うことが可能である。レコードは１つの特性特徴を有し得るが、他の実施形態では、レコードは複数の特性特徴、例えば、少なくとも又はせいぜい３、５、１０個の特徴を有する。

【００１６】

上記の特徴は母集団記述子にも関する。母集団記述子は、訓練データセットにわたる１つ又は複数の特性特徴の分布を特徴付ける。通例、母集団記述子は、訓練レコードに関連付けられた特性特徴に基づいて判定される。それにより、レコード、例えば非訓練レコードを与えられたとき、母集団記述子は、レコードの１つ又は複数の特性特徴の、訓練データセットにわたるそれら特性特徴の分布からの逸脱を判定して、例えばレコードが母集団記述子に合致するかどうかを判定することを可能にする。例えば、特性特徴に基づいて、レ

コードが外れ値であるか否か、例えば、訓練母集団中又は訓練データセットが母集団から引き出された当該母集団中で、レコードが発生する可能性が低いかどうか、又は発生したかどうかを判定することが可能であり得る。

【００１７】

母集団記述子にレコードが合致することは、そのレコードが何らかの意味で正常なレコード、例えば訓練セットの範囲以内にあるレコードであることを意味する。それが該当する場合、被訓練モデルを適用したことから得られたモデル出力が信頼できることが判定され得る。例えば、レコードが、訓練セット中で十分に表された母集団からのものであるか、モデル出力が、訓練セット中で好適に表されたタイプのものであるか、又はその両方であり得る。そのような場合、モデル出力は正確であることが予想される。レコードが母集団記述子と合致しないと判定された場合、これは、レコードが被訓練モデルにとって範囲から外れていることを示し得、例えば、被訓練モデルを適用したモデル出力は、それほど信頼できないと予想される。例えば、レコードが、訓練セット中で実際よりも過小に表された母集団からのものである、及び／又は、モデル出力が、訓練セット中で実際よりも過小に表されたタイプのものであり得る。

【００１８】

一例として、磁気共鳴（ＭＲ）画像中の正常な心臓をセグメント化するように訓練されたニューラルネットワークが、大動脈の転位を伴う重篤な先天性心臓疾患事例に適用されたときに十分に機能しないことがあり得、例えばこのまれな条件が訓練データセット中で表されていないことがあり得る。実際、母集団記述子との合致又は非合致は、レコードが被訓練モデルにとって仕様範囲内であるか仕様範囲外であるかどうか、例えば、概念的に、レコードが訓練セットによって定められる凸包の内側にあるか外側にあるかどうかを示す。

【００１９】

本発明者らは、上記で説明したような母集団記述子がレコードの合致を判定するのに特に適し得ることを見いだした。母集団記述子は、訓練レコード及び／又はモデル出力を特徴付ける特性特徴に基づくことから、幅広い機械学習モデルに適用することができ、大概の場合、モデルを訓練するために使用される特定の機械学習モデルに依存しない。さらに、母集団記述子は一般に訓練データセット及び／又は被訓練モデルと比べて大きさが制限され、そのため、母集団記述子を訓練データセットに関連付けることは、特に合致を判定するために訓練データセットを記憶する場合と比べて、記憶及び処理の面でかかるオーバーヘッドが比較的少ない。さらに、母集団記述子は通例、集約されたデータを備えるため、母集団記述子を合致に使用することで、合致を判定するために訓練データセットを記憶することに伴う様々な所有権、機密性、プライバシー、同意、及び／又は保管場所／地理の制約を回避することができる。

【００２０】

母集団記述子が被訓練モデルに関連付けられることにより、被訓練モデルを少なくとも１つのレコードに適用するエンティティが、その少なくとも１つのレコードが母集団記述子に合致するかどうかを判定することを可能にし得る。例えば、母集団記述子は、任意の適切な方式で、例えば、ファイルヘッダ、ＸＭＬ要素等として母集団記述子を被訓練モデル自体に含めることにより、又は母集団記述子を別個のファイルとして提供することにより、又は任意の他の方式で、被訓練モデルに関連付けられたメタデータとして符号化される。

【００２１】

被訓練モデルが適用される少なくとも１つのレコードの処理中に、母集団記述子を使用して、当該少なくとも１つのレコードの１つ又は複数の特性特徴の、訓練データセットにわたる前記特性特徴の分布からの逸脱を判定する。このために、最初に少なくとも１つのレコードの１つ又は複数の特性特徴が判定され、例えば、被訓練モデルが適用されるレコードごとに１つ又は複数の特性特徴が判定される。モデル出力を特徴付ける特性特徴について、特性特徴を判定することは、被訓練モデルを少なくとも１つのレコードに適用して、

少なくとも１つのモデル出力、例えば被訓練モデルが適用される各レコードについてのモデル出力を得ることを含む。逸脱を判定することにより、少なくとも１つのレコードが、被訓練モデルに関連付けられた母集団記述子と合致するかどうかが判定される。逸脱は、偏差値、例えば、閾値と比較可能な連続値、及び／又は逸脱の判定、例えば逸脱若しくは非逸脱の２値判定を含む。

【００２２】

合致を判定することは、１つ又は複数の合致検査を行うことを含み、各検査は、特性特徴のそれぞれのサブセット、例えば１つ、複数個、又はすべての特性特徴に対して行われる。例えば、合致は、母集団記述子を判定するために使用された各特性特徴に対して別々に、又は、母集団記述子を判定するために使用されたすべての特性特徴に対してまとめて判定される。ただし、これは必要ではなく、例えば合致は、母集団記述子を判定するために使用された特性特徴のサブセットに対して判定されてもよい。

【００２３】

判定された特性特徴に基づいて、少なくとも１つのレコードの１つ又は複数の特性特徴の、訓練データセットにわたるそれら特性特徴の分布からの逸脱が判定される。したがって、当該少なくとも１つのレコードが母集団記述子と合致するかどうかが判定され得る。レコードが母集団記述子と合致しない場合、非合致を示す出力信号が生成される。

【００２４】

様々な実施形態において、合致は単一のレコードに対して判定され得る。これは、やはり上記で論じたように、その特定のレコードに被訓練モデルを適用したモデル出力が、信頼できる、正確である、仕様範囲内である等と予想されるかどうかを示すことができる。出力信号は、例えば、モデルによって提供されたモデル出力を使用するのを控えるようにユーザに警告する等のために使用される。母集団記述子によって合致を判定することは、モデルの適用を不必要に制約することを回避することができる。例えば、モデルがそれを元に訓練された特定の年齢階層、民族性、地理等の明示的／手動的なコーディングが回避され得る。その結果、被訓練モデルは、母集団記述子に従って訓練コホート及び地理を超えて使用することができ、なおモデルの範囲内にある追加的な対象者への適用を可能にする。

【００２５】

様々な実施形態において、合致は、複数のレコードのグループに対して判定されてよく、グループとしての複数のレコードの、母集団記述子との合致の査定を提供する。これは、レコードのグループ、例えば患者のコホート、が訓練セットを表すかどうかを示す。例えば、これによりモデル／概念のドリフトを検出することが可能になる。興味深い点として、このようにすると、モデルが適用される際の条件、例えば、それに対応する患者が訓練セットに含まれていた地理、の明示的なコーディングを必要とせずに、モデル／概念のドリフトが検出可能となる。さらに、被訓練モデルの適用が、訓練セットの特定の設定、例えばその地理、に不必要に制限されることが防止され得る。実際、モデルは訓練セットよりも広く適用可能であり、母集団記述子により、モデルが適用されるレコードがもはや訓練セットに十分に合致しないことを検出することが可能になる。この場合も出力信号を使用してユーザに警告することができる。

【００２６】

任意で、出力信号を使用して、被訓練モデルのモデルパラメータを適合することができる。例えば、モデルパラメータは、信号によって示される、訓練データセットと比較したモデルが適用されるレコードのタイプの変化にモデルを適合させるように適合される。例えば、１つ又は複数のモデルパラメータが増加又は減少される、或いは１つ又は複数のモデルフラグが設定又はリセットされる、例えば、モデルパラメータが、特異性を犠牲にしてモデルの感度を増大させるように、又はその逆に適合される。具体例として、個々の又はグループとしてのいくつかのレコードの非合致、例えばモデル／概念のドリフトは、通常よりも高い特定の疾患の診断の発生率を示し、したがってその疾患の大流行を示唆し、その時点で、それによって特異性が低くなる結果となっても、大流行を事前に抑制するのを

支援するためにモデルの感度を一時的に上げることが有用であり得る。一実施形態では、モデルパラメータは、被訓練モデルが適用される複数のレコードにわたって蓄積された出力信号、例えば見られたことのないレコードのセット又はコホートに合わせて適合される。例えば、モデルパラメータは、例えば毎週、４週間ごとなど、周期的に調節される。

【００２７】

任意で、被訓練モデルは、ニューラルネットワーク、例えば深層ニューラルネットワーク又は畳み込みニューラルネットワーク（ＣＮＮ）である。そのようなニューラルネットワークは、それ自体知られている訓練方法を使用して、例えば後退代入及び／又は確率的勾配降下を使用して、訓練される。しかし、例えばＩＤ３若しくはＣ４．５を使用して訓練された決定木、サポートベクトルマシン（ＳＶＭ）、ランダムフォレスト、ｋ平均クラスタリング、又はオートエンコーダなどの、他の知られている種類のモデル及び訓練方法が等しく適用可能である。

【００２８】

任意で、母集団記述子は、例えば被訓練モデルに関連付けるために符号化され、この母集団記述子の符号化は、１つ又は複数の特性特徴の特徴識別子を備える。被訓練モデルを使用するエンティティに対してレコードの特性特徴をどのように算出するかを示すのに適した任意の特徴識別子が使用されてよく、それらは例えば、レコード若しくはモデル出力の特徴のインデックス若しくは名前、特性特徴を算出するための機能の識別、例えば、「ｎｕｍ＿ｐｉｘｅｌｓ＿ｄａｒｋｅｒ＿ｔｈａｎ（１００）」若しくは「＜特徴値１＞＋＜特徴値２＞＋＜モデル出力１＞」、又は、レコード、関連付けられたメタデータ、及び／若しくは被訓練モデルをレコードに適用したモデル出力から、特性特徴を算出するための実行可能コードである。特徴識別子を含めることは、低減された設定及び／又は向上した柔軟性を有するレコードの処理を可能にするという利点を有し得る。

【００２９】

任意で、母集団記述子は、訓練データセットにわたる１つ又は複数の特性特徴の確率分布を示す確率データを備える。確率データは、判定された特性特徴が確率データに従って発生する確率を判定するために使用される。例えば、確率データは、訓練データセットがそこからサンプリングされた訓練母集団内で特性特徴のセットが発生する尤度の推定値を判定することを可能にする。

【００３０】

任意で、確率データは、１つ又は複数の特性特徴の確率分布のパラメータを備える。パラメータは、訓練データセットに最もよく一致する所与の確率分布、例えば正規分布又はガウス分布のパラメータである。例えば、確率データは、少なくとも１つの特性特徴の一変量確率分布、例えば正規分布の中央値及び標準偏差、指数分布の率等のパラメータを備える。

【００３１】

任意で、確率データは、特性特徴の１つ又は複数の分位数、例えば中央値、又は１つ若しくは複数の三分位数、四分位数、百分位数等を備える。例えば、確率データは、特性特徴、例えば連続変数として表された特性特徴の各百分位数を備える。例えば、ガウス分布に従わない連続変数の場合、分位数を使用することは、空間及び／又は計算要件の点で、及び／又は例えば外れ値、歪み等に対する不感性などにおけるそのロバスト性の点で、特に有利であり得る。

【００３２】

任意で、確率データは、１つ又は複数の特性特徴のヒストグラムを備える。例えば、ヒストグラムは、少なくとも１つの特性特徴についての訓練データセットにわたる値又は値範囲の発生回数を備える。

【００３３】

任意で、確率データは、複数の特性特徴の同時確率分布を示すデータを備える。これにより、レコードの複数の特性特徴がこの確率データに従って発生する確率、例えば同時確率を判定することが可能になる。例えば、確率データは、複数の特性特徴の多変量確率分布

のパラメータ、複数の特性特徴の組み合わせの発生のヒストグラム等を備える。

【００３４】

被訓練モデルが適用される少なくとも１つのレコードを処理する際、あるレコードに対して判定された確率データ及び特性特徴を使用して、判定された特性特徴がその確率データに従って発生する確率を判定する。確率は、例えば、判定された特性特徴が確率データに従って発生する絶対確率である。確率は、相対確率、例えば、特定の特性特徴又は特性特徴のセットに対応する現在値がその特徴又は特徴のセットの最も一般的な値と比較して発生する確率である。確率は、ある特性特徴又は特性特徴のセットについて、現在値よりも高い、又は低い値、又はより一般的でない値が訓練セット中で発生する確率である。一般に、確率の使用は、母集団記述子との合致についてのロバストで詳細な査定を提供することを可能にする。さらに、確率の使用は、説明可能なフィードバックを提供することを可能にして、何に基づいて合致の判定が行われたかをユーザが理解できるようにする。

【００３５】

任意で、母集団記述子は、２つの特性特徴間の相関を示す相関値を備える。相関値は、例えば、２つの特性特徴が同時に比較的高い値を有する可能性が高いこと、又は一方の特性特徴が低い値を有するときに別の特性特徴が通常高い値を有することを示す。例えば、相関値は、ピアソン若しくはスピアマン相関又は類似のものである。相関値は、特性特徴間の共分散尺度である。例えば、母集団記述子は、複数のそれぞれの第１及び第２の特性特徴間の共分散を提供する共分散行列を備える。複数の特性特徴の確率を使用することは、特性特徴の互いとの共発生を考慮に入れられるようにすることにより合致判定の精度を向上させ、例えば、訓練データセット中で、ある特性特徴の低い値が、別の特性特徴の高い値と組み合わせられることが滅多にない場合、そのような特徴値の組み合わせは非合致を示し得る。ただし、複数の特性特徴についての確率推定値は、例えば特性特徴が統計的に独立している／相関していないと仮定することにより、個々の特性特徴の確率分布を示すデータに基づいて判定される。

【００３６】

任意で、確率データは、訓練データセットにわたる１つ又は複数の特性特徴の時間的変動性を示すデータを備える。例えば、訓練データセットは、時系列データを備える。時間的変動性データは、例えばモデル出力の陽性率についての、例えば中央値及び時間的分散基準値を備える。例えば、訓練データセットからのデータは、１日、１週間、又は１カ月の枠で時間枠が設定される。例えば、使用中に、時間的変動性データへの複数のレコードの合致が、例えば実際の陽性率の著しい逸脱を判定することに基づいて判定される。そのような場合は、出力信号、例えばモデルドリフト警告が生成され、例えば臨床ユーザ又はデータ科学の専門家などのユーザに出力される可能性がある。

【００３７】

確率データは、複数の特性特徴についての、又は場合によっては重複している複数の特性特徴のセットについての別々の確率データを備える。したがって、確率データは、複数の特徴又は特徴のセットに関する合致検査を可能にする。合致を検査するには、判定された確率を予め定義された閾値と比較して合致を検査し、例えば、確率が閾値を下回る場合、非合致が判定される。複数の確率が判定されてもよく、例えばそれら各々が閾値と比較され、予め定義された数の値が閾値を満たす場合に非合致が判定され、閾値が何倍かに増大され、結果が閾値と比較される等する。

【００３８】

任意で、母集団記述子は、１クラス分類器を備える。当技術分野でそれ自体知られているように、１値分類器又はクラスモデラーとも称される１クラス分類器は、物体のより大きなセットの中で特定のクラスの物体を識別するように構築される。そのような１クラス分類器は通例、主としてそのクラスの物体で訓練される。１クラス分類器は、少なくとも１つのレコードを合致していると分類するために使用され、訓練レコードの特性特徴のセットで訓練される。これにより、比較的複雑な判定境界を使用して合致判定を行うことを可能にし、及び／又は合致検査の実行の向上した自動化を可能にし、例えば分類器が周期的

に再訓練される。文献から知られている様々な１クラス分類器が使用される。任意で、１クラス分類器は、１クラスサポートベクトルマシン（ＳＶＭ）を備える。

【００３９】

任意で、訓練レコードは画像を備える。例えば、画像は、例えばＣＴ又はＭＲＩスキャナから得られた医用画像である。プロセッササブシステムは、画像から、例えば実際の画像データから、例えば画像のピクセルから、特性特徴を抽出することによって訓練レコードの１つ又は複数の特性特徴を判定するように構成される。例えば、特性特徴は、平均グレースケール値など、画像全体の特徴、例えば静的な特徴、又は画像の特定部分若しくは画像中で認識される特定物体の特徴である。プロセッササブシステムは、画像に関連付けられたメタデータから特性特徴を判定するように構成される。例えば、メタデータは、画像データ以外の、例えば画像のピクセル以外の、任意のデータである。例えば、メタデータは、撮像条件、例えば画像サイズ、焦点長、日付、時刻等、及び／又は画像上のエンティティ、例えば撮像された患者の患者データを記述する。そのようなメタデータの具体例は、ＤＩＣＯＭ（Ｄｉｇｉｔａｌ Ｉｍａｇｉｎｇ ａｎｄ Ｃｏｍｍｕｎｉｃａｔｉｏｎｓ ｉｎ Ｍｅｄｉｃｉｎｅ）メタデータである。

【００４０】

任意で、被訓練モデルは、画像のセグメント化を提供するように訓練される。例えば、そのようなセグメント化は、画像を複数のピクセルセグメントに区分する。画像セグメントは、例えば、撮像された物体又は撮像された物体の輪郭、例えば２Ｄ輪郭、又は表面、例えば３Ｄ表面を表すことができる。プロセッササブシステムは、画像のセグメント化の形状記述子を抽出することによって訓練レコードの特性特徴を判定するように構成されてよい。例えば、形状記述子は、画像中で認識された形状の簡略表現、例えば合同不変形状記述子、本質的形状記述子等である。任意で、特性特徴は、例えばそれぞれの形状記述子、例えば面積、体積、向き、場所等から判定された、認識された形状の性質である。

【００４１】

任意で、訓練データセットの訓練中に、訓練データセットのレコードに重みが割り当てられる。そのような場合、母集団記述子は、それら割り当てられた重みに従って判定され得る。例えば、訓練データセットの訓練中に、例えば様々なクラスの訓練レコードの数間の不均衡を低減するために、訓練データセットが再平衡される。例えば、そのような再平衡は、訓練中に使用するための訓練データセットのレコードを複製することを含む。それらの重みは母集団記述子中で考慮されてよく、例えばヒストグラムを構築する時に、訓練中に複製されたレコードが、それが複製された回数に対応する重みと共にヒストグラムに追加される。例えば、母集団記述子は、再平衡された訓練データセットに基づいて判定される。

【００４２】

任意で、集団記述子を判定するシステムは、機械学習モデルを訓練データセットで訓練し、それにより被訓練モデルを得る。訓練と、母集団記述子の判定とは通例両方とも訓練データセットにアクセスするので、これは、訓練データセットの複製／共有がより少なくて済むという利点を有する。ただし、興味深い点として、母集団記述子の判定では、通例、機械学習モデルの訓練の内部、例えば機械学習モデルの中間出力にアクセスする必要はなく、したがって、母集団記述子を判定するためのモジュールはモデルの訓練とは比較的独立して配置され得る。

【００４３】

任意で、少なくとも１つのレコードを処理するシステムは、被訓練モデルを少なくとも１つのレコードに適用して少なくとも１つのモデル出力を得るようにさらに構成される。いくつかの実施形態では、システムは、非合致信号が生成されない場合のみ、被訓練モデルを少なくとも１つのレコードに適用する。他の実施形態では、システムは、非合致信号の場合でも被訓練モデルを適用し、例えば、システムは、合致又は非合致の通知と共に少なくとも１つのモデル出力を提供する。モデル出力は、１つ又は複数の特性特徴、例えばモデル出力を特徴付ける特性特徴を判定するために使用される。

【００４４】
任意で、少なくとも１つのレコードを処理するシステムは、少なくとも１つのレコードの非合致を示す出力信号を出力するための出力インターフェースをさらに備える。出力信号は、感覚的に感知可能な方式で出力信号をレンダリングするレンダリングデバイスに出力される。信号を出力するために、レンダリングデバイスは、出力信号が、例えばポップアップボックス、レコードの一覧中での当該少なくとも１つのレコードの強調表示、それぞれのモデル出力の色分け等の形態で、表示される画面を備える。例えば警告音を再生するスピーカなど、他の種類のレンダリングデバイスがそれに代えて又はそれに加えて使用される。
【００４５】
出力信号のレンダリングを通じて、ユーザは、少なくとも１つのレコードが母集団記述子と一致しないというフィードバックを得ることができる。例えば、出力信号は、被訓練モデルを少なくとも１つのレコードに適用した少なくとも１つのモデル出力のレンダリングと並べてレンダリングされる。いくつかの実施形態では、例えばユーザがモデルをレコードに適用することを要求した時に、モデル出力が、判定された合致又は非合致と共にレンダリングされる。他の実施形態では、非合致が判定された場合、モデル出力は抑制されるか、又は全く判定されない。さらに他の実施形態では、モデル出力は合致しないレコードのみについて示され、例えば、その合致が判定されたレコードは自動的に処理され、合致しないレコードはユーザによる点検のために提示される。そのため、一部のモデル出力は信頼性が低い可能性があることを様々な方式でユーザに注意喚起することができる。
【００４６】
任意で、母集団記述子への非合致を生じさせた、１つ又は複数の判定された特性特徴の逸脱が、感覚的に感知可能な方式で出力される。これにより、ユーザが、何故非合致が判定されたかについての洞察を得ることが可能になる。例えば、特性特徴の特定のセットの組み合わせに基づいて非合致が判定されたことが報告される。
【００４７】
任意で、逸脱の度合いが報告され、例えば、逸脱は、１つ又は複数の特性特徴が母集団記述子の確率データに従って発生する確率を含む。例えば、あるレコードが、現在のレコードよりも特性特徴に対して極端な値を有する確率が報告され、例えば、「訓練セット中では患者の１％のみが女性であり、現在の患者よりも背が高かった」又は「訓練セットはそれほど幼い幼児に対してそれほど大きい肺を含んでいない」という警告が与えられる。多くの変形例が明らかとなろう。
【００４８】
任意で、複数の被訓練モデルに関連付けられた母集団記述子が、被訓練モデルを使用するシステムのデータインターフェースによってアクセスされる。複数の被訓練モデルは、例えば、それらが訓練された訓練データセット、及び／又は例えばＳＶＭ若しくはニューラルネットワークなどのモデルのタイプ、及び／又はモデルを訓練するのに使用されたパラメータにおいて、異なり得る。システムのプロセッササブシステムは、少なくとも１つのレコードが、それぞれの被訓練モデルに関連付けられた母集団記述子と合致するかどうかを判定し、合致する場合、複数の被訓練モデルの当該それぞれの被訓練モデルを選択するように構成されてよい。例えば、プロセッササブシステムは、被訓練モデルごとに、少なくとも１つのレコードがそのレコードのモデル母集団記述子と合致するかどうかを判定し、それにより合致する被訓練モデルのセットを得る。それぞれの被訓練モデルは、合致する被訓練モデルのセットから選択される。選択された被訓練モデルは、例えば、選択された被訓練モデルをあるレコードに適用したモデル出力として、そのレコードに対する総合モデル出力を判定するために使用される。複数の被訓練モデルが選択される。例えば、総合モデル出力は、複数の選択された被訓練モデルをレコードに適用したモデル出力から、例えばモデル出力同士の多数決や平均化等により、判定される。このようにすると、当該レコードに対して最も適当な１つ又は複数のモデルが使用されるため、全体的に精度が増大したモデル出力が得られる。また、複数の被訓練モデルのモデル出力を、例えば選択さ

れた１つ又は複数の被訓練モデルが強調表示された状態で、提示して、ユーザが使用するモデル出力をより多くの情報に基づいて決定できるようにすることも可能である。

【００４９】

被訓練モデルの選択は、合致／非合致の判定に基づく。また、被訓練モデルの選択は、合致／非合致の判定につながる中間結果、例えば、レコードの１つ又は複数の特性特徴がそれぞれの確率データに従って発生する判定確率や、それぞれの１クラス分類器がレコードを合致しているとして分類することの信頼値等に基づく。例えば、最も高い確率、信頼性等を持つモデルが選択される。モデル出力は、それぞれの確率、信頼性等と共にユーザに提示される。

【００５０】

一般に、被訓練モデルは、訓練レコードを特徴付ける特性特徴の合致に基づいて、又は被訓練モデルを訓練レコードに適用したモデル出力を特徴付ける特性特徴の合致に基づいて、又は混合方式でそれら両方に基づいて、選択される。例えば、第１の可能性は、例えば同じ種類のモデル及び／又はモデルパラメータを使用して訓練された、複数の訓練データセットの中で最もよく一致する入力データ表現に基づいて、事前に被訓練モデルを選択することを可能にする。第２の可能性は、それぞれのモデルのモデル出力の中でモデル出力の最も高い尤度に基づいて、事後に被訓練モデルを選択することを可能にする。この第２の可能性は、例えば、複数の異なるモデルが同じデータセットで訓練された場合に使用される。

【００５１】

任意で、合致を判定することはさらに、１つ又は複数の特性特徴が１つ又は複数の予め定義されたドメイン制約を満たすかどうかを判定することに基づく。これは、様々な種類のドメイン知識が判定に組み込まれることを可能にし、その全体的な精度を上げる。例えば、ドメイン制約は、例えば年齢は０～１３０の間である、体重は正である等の、特性パラメータの取り得る値の範囲を備える。

【００５２】

当業者には、上述の本発明の実施形態、実施態様、及び／又は任意選択の態様の２つ以上が、有用と考えられる任意の方式で組み合わされてよいことを認識されよう。

【００５３】

訓練データセットの訓練レコードを処理するためのシステム又は被訓練モデルが適用される少なくとも１つのレコードを処理するためのシステムの変更及び変形が、本記載に基づいて当業者によって実施され得る。対応するシステムの記載される変更及び変形に対応する、任意のコンピュータ実施方法及び／又は任意のコンピュータプログラム製品の変更及び変形が、本記載に基づいて当業者によって実施され得る。

【００５４】

本発明のこれら及び他の態様は、以下の説明で添付図面を参照しながら例として説明される実施形態から明らかになり、またそれを参照してさらに解説される。

【図面の簡単な説明】

【００５５】

【図１】訓練データセットにわたる１つ又は複数の特性特徴の分布を特徴付ける母集団記述子を判定するために、訓練データセットの訓練レコードを処理するためのシステムの図である。

【図２】被訓練モデルが適用される少なくとも１つのレコードを処理し、当該少なくとも１つのレコードが母集団記述子と合致するかどうかを判定するためのシステムの図である。

【図３ａ】母集団記述子、この場合は確率データ、がどのように訓練データセットに対して判定され、レコードの合致を判定するために使用されるかの詳細な例を示す図である。

【図３ｂ】母集団記述子がどのように判定され、この場合はヒストグラム比較を行うことによって、複数のレコードの合致を判定するために使用されるかの詳細な例を示す図である。

【図４】母集団記述子、この場合は訓練画像のセグメント化のＳｈａｐｅ　ＤＮＡ、がどのように判定され、レコードに対するモデル出力、この場合は画像のセグメント化、の合致を判定するために使用されるかの詳細な例を示す図である。
【図５】訓練データセットにわたる１つ又は複数の特性特徴の分布を特徴付ける母集団記述子を判定するために訓練データセットの訓練レコードを処理するコンピュータ実施方法の図である。
【図６】被訓練モデルが適用される少なくとも１つのレコードを処理し、当該少なくとも１つのレコードが母集団記述子と合致するかどうかを判定する、コンピュータ実施方法の図である。
【図７】データを備えるコンピュータ可読媒体の図である。
【発明を実施するための形態】
【００５６】
図は純粋に模式的なものであり実際の縮尺で描かれていないことが留意されるべきである。図面では、すでに説明された要素に相当する要素は、同じ参照符号を有する場合がある。
【００５７】
参照符号の一覧
以下の参照符号の一覧は、図面の解釈を助けるために提供され、特許請求の範囲を制限するものと解釈されるべきではない。
０２０、０２２　データストレージ
０３０　訓練データセット
０４０　１つ又は複数のレコード
０６０　母集団記述子
１００　訓練レコードを処理するためのシステム
１２０　データインターフェース
１２２、１２４　データ通信
１４０　プロセッササブシステム
２００　被訓練モデルを使用するためのシステム
２２０　データインターフェース
２２２、２２４　データ通信
２４０　プロセッササブシステム
２４２　出力信号
２６０　ディスプレイ出力インターフェース
２６２　表示データ
２８０　ディスプレイ
２８２　警告ダイアログボックス
３００　訓練データ
３１０　再平衡された訓練データ
３２０　母集団記述子（ヒストグラム）
３３０　レコード
３４０　被訓練モデル（ニューラルネットワーク）
３５０　合致検査
３６０　モデル出力（結果）
３７０　合致結果
３３０’　複数のレコード
３５０’　合致検査（ヒストグラム比較）
３６０’　モデル出力（結果）
３７０’　コホート合致結果
４００　訓練データ（画像）
４１０　拡張された訓練データ

４２０　特性特徴（Ｓｈａｐｅ　ＤＮＡ）
４３０　母集団記述子（１クラスＳＶＭ）
４４０　レコード（画像）
４５０　被訓練モデル（畳み込みニューラルネットワーク）
４６０　特性特徴（Ｓｈａｐｅ　ＤＮＡ）
４７０　合致検査（１クラス分類）
４８０　モデル出力（セグメント化）
４９０　合致データ（分類）
５００　訓練レコードを処理するための方法
５１０　訓練データセットにアクセスする
５２０　特性特徴を判定する
５３０　母集団記述子を判定する
５４０　母集団記述子を被訓練モデルに関連付ける
６００　少なくとも１つのレコードを処理する方法
６１０　母集団記述子、レコードにアクセスする
６２０　特性特徴を判定する
６３０　合致を判定する
６４０　非合致を示す出力信号を生成する
７００　コンピュータ可読媒体
７１０　非一時的なデータ

【００５８】
図１は、被訓練モデルに関連付けられた母集団記述子を判定するために被訓練モデルを処理するシステム１００を示す。被訓練モデルは、訓練データセットで訓練される。被訓練モデルは、レコードに適用されたときにモデル出力を提供するように訓練される。モデル母集団は、被訓練モデルが訓練される訓練レコードに関連付けられた特性特徴に基づいて判定される。

【００５９】
システム１００は、データ通信１２４を介して内部通信する、データインターフェース１２０及びプロセッササブシステム１４０を備える。プロセッササブシステム１４０は、システム１００の動作中に、データインターフェース１２０を使用して、被訓練モデルが訓練される訓練データセット０３０にアクセスするように構成される。任意で、プロセッササブシステム１４０は、システム１００の動作中に、データインターフェース１２０を使用して、被訓練モデルにアクセスするように構成される。訓練データセット０３０は、複数の訓練レコードを備え得る。例えば、図１に示すように、データインターフェース１２０は、前記データ０３０を備える外部データストレージ０２０へのアクセス１２２を提供する。代替として、データ０３０は、システム１００の一部である内部データストレージからアクセスされる。代替として、データ０３０は、ネットワークを介して別のエンティティから受け取られる。一般に、データインターフェース１２０は、ローカル又はワイドエリアネットワーク、例えばインターネットへのネットワークインターフェース、内部又は外部データストレージへのストレージインターフェース等の様々な形態を取り得る。データストレージ０２０は、任意の知られている適切な形態を取り得る。

【００６０】
プロセッササブシステム１４０は、システム１００の動作中に、訓練データセット０３０の訓練レコードの１つ又は複数の特性特徴を判定するようにさらに構成される。１つ又は複数の特性特徴は、訓練レコード、及び／又は訓練レコードに被訓練モデルを適用したモデル出力を特徴付ける。プロセッササブシステムは、訓練レコードに関連付けられた特性特徴に基づいて母集団記述子０６０を判定する。母集団記述子０６０は、訓練データセット０３０にわたる１つ又は複数の特性特徴の分布を特徴付ける。プロセッササブシステム１４０はさらに、被訓練モデルを少なくとも１つのレコードに適用するエンティティが、当該少なくとも１つのレコードが母集団記述子に合致するかどうかを、少なくとも１つの

レコードの１つ又は複数の特性特徴を判定することに基づいて判定することを可能にするために、母集団記述子０６０を被訓練モデルに関連付ける。そのようなエンティティの一例は、図２のシステム２００である。

【００６１】

母集団記述子０６０は、システム１００によってデータストレージ０２０又は他の場所に記憶される、ネットワークを介して送られる、などされる。一般に、母集団記述子０６０は、被訓練モデルと同じデータコンテナ、例えば同じファイルに記憶されるが、被訓練モデルに関連付けられた別個の母集団記述子０６０として提供される。例えば、いくつかの実施形態では、被訓練モデルは、例えば母集団記述子０６０にアクセス可能なＵＲＬを含むことにより、母集団記述子０６０にリンクし、又は母集団記述子０６０が被訓練モデルにリンクする。様々な他の関連付け手段が等しく考えられ、当業者の理解の範囲内にある。

【００６２】

システム１００の動作の様々な詳細及び態様は、その任意選択の態様を含めて、図３ａ及びその他を参照してさらに解説される。

【００６３】

一般に、システム１００は、例えばラップトップ若しくはデスクトップベースのワークステーション、又はサーバなどの、単一のデバイス若しくは装置として具現化されるか、又はその中にある。デバイス又は装置は、適当なソフトウェアを実行する１つ又は複数のマイクロプロセッサを備える。例えば、プロセッササブシステムは、単一の中央演算処理装置（ＣＰＵ）によって具現化されるが、そのようなＣＰＵ及び／又はグラフィック処理ユニット（ＧＰＵ）などの他の種類の処理ユニットの組み合わせ又はシステムによって具現化される。ソフトウェアは、ダウンロードされて、及び／又は対応するメモリ、例えばＲＡＭなどの揮発性メモリ若しくはフラッシュなどの不揮発性メモリに記憶される。代替として、システムの機能単位、例えばデータインターフェース及びプロセッササブシステムは、プログラム可能論理の形態のデバイス又は装置として、例えばフィールドプログラム可能ゲートアレイ（ＦＰＧＡ）として、実施される。一般に、システムの各機能単位は、回路の形態で実施される。システム１００は、例えば分散サーバなどの種々のデバイス又は装置を伴う、例えばクラウドコンピューティングの形態などの分散型で実施されてもよいことが留意される。

【００６４】

図２は、被訓練モデルが適用される少なくとも１つのレコードを処理すると共に、少なくとも１つのレコードが被訓練モデルに関連付けられた母集団記述子と合致するかどうかを判定するためのシステム２００を示す。システム２００は、データ通信２２４を介して内部通信する、データインターフェース２２０及びプロセッササブシステム２４０を備える。プロセッササブシステム２４０は、システム２００の動作中に、データインターフェース２２０を使用して、図１を参照して説明したように被訓練モデルに関連付けられた母集団記述子０６０にアクセスすると共に、被訓練モデルが適用される少なくとも１つのレコード０４０にアクセスするように構成される。任意で、プロセッササブシステム２４０は、システム２００の動作中に、データインターフェース２２０を使用して、母集団記述子０６０が関連付けられた被訓練モデルにアクセスするようにも構成される。例えば、同じく図２に示すように、データインターフェース２２０は、前記データ０４０、０６０を備える外部データストレージ０２２へのアクセス２２２を提供する。代替として、データ０４０、０６０は、内部データストレージからアクセスされる。代替として、データ０４０、０６０は、ネットワークを介して受け取られる。一般に、データインターフェース２２０は、ローカル又はワイドエリアネットワーク、例えばインターネット、へのネットワークインターフェース、内部又は外部データストレージへのストレージインターフェース等の様々な形態を取り得る。データストレージ０２２は、任意の知られている適切な形態を取る。

【００６５】

プロセッササブシステム２４０はさらに、システム２００の動作中に、少なくとも１つのレコード０４０１つ又は複数の特性特徴を判定し、少なくとも１つのレコード０４０の１つ又は複数の特性特徴の、訓練データセットにわたる前記特性特徴の分布からの逸脱を判定して、少なくとも１つのレコード０４０が母集団記述子０６０と合致するかどうかを判定し、少なくとも１つのレコード０４０が被訓練モデルの母集団記述子と合致しない場合、前記非合致を示す出力信号２４２を生成するように構成される。任意選択の構成要素として、システム２００は、ディスプレイ出力インターフェース２６０、又はディスプレイ２８０などのレンダリングデバイスに出力信号２４２を出力するための任意の他の種類の出力インターフェースを備える。例えば、ディスプレイ出力インターフェース２６０は、ディスプレイ２８０の表示データ２６２を生成し、それによってディスプレイ２８０が、出力信号を、例えば画面上の警告ダイアログボックス２８２として、感覚的に感知可能な方式でレンダリングする。

【００６６】

一般に、システム２００は、例えばラップトップ若しくはデスクトップベースのワークステーションなどの単一のデバイス若しくは装置として具現化されるか、若しくはその中にあり、又は（医用）撮像装置若しくはデバイスの一部として、又はモバイルデバイスとして具現化される。デバイス又は装置は、適当なソフトウェアを実行する１つ又は複数のマイクロプロセッサを備える。例えば、プロセッササブシステムは、単一の中央演算処理装置（ＣＰＵ）によって具現化されるが、そのようなＣＰＵ及び／又は他の種類の処理ユニットの組み合わせ又はシステムによって具現化される。ソフトウェアは、ダウンロードされて、及び／又は対応するメモリ、例えばＲＡＭなどの揮発性メモリ若しくはフラッシュなどの不揮発性メモリに記憶される。代替として、システムの機能単位、例えばデータインターフェース及びプロセッササブシステムは、プログラム可能論理の形態のデバイス又は装置として、例えばフィールドプログラム可能ゲートアレイ（ＦＰＧＡ）として、実施される。一般に、システムの各機能単位は、回路の形態で実施される。システム２００は、例えばクライアント・サーバ実施のクライアント及びサーバなどの種々のデバイス又は装置を伴う、分散型で実施されてもよいことが留意される。

【００６７】

図３ａは、母集団記述子が、例えば図１のシステム１００によってどのように判定されるか、及び／又は例えば図２のシステム２００によってレコードの合致を判定するために使用されるかの、詳細であるが非制限的な例を示す。例えば、ここに記載される技術は、画像、例えばコンピュータ断層撮影（ＣＴ）スキャン、磁気共鳴（ＭＲ）スキャン、超音波（ＵＳ）スキャン、又はＸ線のような医用画像、の分類に使用される。ここに記載される技術は、例えば、血圧、心電図（ＥＣＧ）データ、呼吸数等の、集中治療室（ＩＣＵ）内の監視機器で獲得された時系列データにも適用される。例えば、そのような画像又は時系列データは、いくつかの潜在的な疾患又は病状、例えばｂ１，．．．，ｂｍ、に関する分類に使用される。

【００６８】

訓練データセットＴＤ３００は、複数の訓練レコード又はサンプルＩ１，．．．，Ｉｎを備える。例えば、訓練データセットＴＤは、患者レコード、監視信号、画像、画像シーケンス、画像及び信号の組み合わせ等を備える。訓練レコードは、そのレコードに対応する測定パラメータ、例えばＭＲ撮像シーケンスの種類、造影剤の存在等の撮像パラメータを記述する１つ又は複数の特徴ｓｉを備える。訓練レコードは、当該モデルがそれらを近似するように訓練される、いくつかの意図されるモデル出力／アノテーションＢｉを備える。例えば、そのようなアノテーションは、個人の個別の疾患又は病状の存在を示す。訓練レコードは、そのレコードが参照する対象者、例えば撮像された患者などの患者の１つ又は複数の特徴ｔｉも備える。例えば、そのような特徴は、性別、年齢、体重、身長、喫煙者／非喫煙者等を含む。訓練データセットＴＤのレコードは、それに応じてｓｉ，ｔｉ，Ｉｉ，Ｂｉと表される。

【００６９】

任意で、訓練データセットの特定部分が、訓練中により頻繁に使用されるか、又は訓練中に異なる重みを得てもよく、その結果、再平衡された訓練データセットＲＴＤ３１０を得る。例えば、訓練データセットＴＤが特定の分類に関して不均衡を有する場合には、発生がより少ない分類を有するレコードが、再平衡された訓練データセットＲＴＤ内で複製される。例えば、分類のあるクラスは、訓練データセット中で別のクラスよりも発生する頻度がはるかに少なく、例えばまれな疾患を表すクラスであり得る。このまれなクラスの認識を向上させるために、例えば、このまれなクラスのレコードを重みに対応する回数だけコピーすることにより、より高い重みがそれに割り当てられる。ただし、これは任意選択であり、再平衡された訓練データセットＲＴＤは、訓練データセットＴＤと同じであってもよい。上記と同様に、再平衡された訓練データセットＲＴＤのレコードは、ｓ$_i$，ｔ$_i$，Ｉ$_i$，Ｂ$_i$と表される。

【００７０】

被訓練モデルＮＮ３４０は、再平衡された訓練データセットＲＴＤで、又は再平衡が行われない場合は元のデータセットＴＤで訓練される。例えば、母集団記述子を判定するシステムは、データセット自体で機械学習モデルを訓練するか、又は別のエンティティから被訓練モデルＮＮを取得する。この例では、被訓練モデルＮＮは、レコードＲ３３０のアノテーションｂの予測をモデル出力として提供するように訓練されたニューラルネットワークであってよく、レコードＲ３３０は、例えば、撮像パラメータｓ及び患者パラメータｔを伴う画像Ｉを含み、その結果、モデル出力ＲＥＳ３６０、例えばｂ＝Ｆ$_{NN}$（Ｉ，ｓ，ｔ）を得る。例えば、被訓練モデルＮＮは、知られているニューラルネットワーク訓練方法を使用して、例えば後退代入及び／又は確率的勾配降下に基づいて、訓練される。値ｓ、ｔは、画像Ｉに関連付けられたメタデータとみなされてよく、この例では、このメタデータも被訓練モデルに入力される。ただし、教師なし訓練で学習されたモデルＮＮを含む、他の種類の被訓練モデルをここで適用することができる。

【００７１】

訓練データＲＴＤ／ＴＤに基づき、母集団記述子が判定され、この場合、母集団記述子は、訓練データセット３００の１つ又は複数の特性特徴の確率データＰＤ３２０を備える。訓練レコードの特性特徴は、訓練レコード自体に基づく。例えば、特性特徴は、訓練レコードの１つ又は複数の特徴、例えば撮像パラメータｓ及び／又は患者パラメータｔを備える。特性特徴は、訓練レコードの複数の特徴から導出されてもよく、例えば、２つの特徴の最大値や平均等である。いくつかの実施形態では、特性特徴は、訓練レコードＲＴＤ／ＴＤの特徴の厳密なサブセット、例えば特徴のうちせいぜい半分、又はせいぜい１、５、若しくは１０個の特徴に基づく。ただし、すべての特徴を使用することも可能である。特性特徴は、訓練レコードから算出される。一実施形態では、訓練レコードＲＴＤ／ＴＤの一部又はすべてが画像を備え、特性特徴が、例えば平均グレースケール値、色のスペクトル等として、画像から算出、例えば抽出される。

【００７２】

いくつかの実施形態では、特性特徴は、メタデータ、例えば、訓練レコードＲＴＤ／ＴＤに関連付けられているがモデルＮＮの訓練時に使用されないメタデータに基づく。例えば、訓練レコードは、モデルＮＮが訓練される画像を備え、例えば、モデルは、物体認識又は画像のセグメント化を提供する。そのようなモデルＮＮは、例えば、撮像条件及び／又は撮像対象の物体に関する情報などの、画像に関連付けられたメタデータを使用しない場合もあるが、そのようなメタデータはそれでも訓練データセットを特徴付けるのに有用であり得る。

【００７３】

様々な実施形態において、特性特徴は、被訓練モデルＮＮを訓練レコードに適用したモデル出力に基づく。例えば、特性特徴は、モデル出力に含まれる分類又は回帰を備える。モデル出力から特性特徴が算出される。一例として、データセットＴＤ／ＲＴＤの訓練レコードは、被訓練モデルがそのセグメント化を提供するように訓練される画像を備える。そのようなセグメント化から特性特徴が算出されてよく、例えば、特性特徴は、撮像された

物体の寸法、例えば心臓などの撮像された物体の体積を示す。教師あり学習用の訓練レコードの特性特徴は、訓練データセットＴＤ／ＲＴＤからの意図されるモデル出力に、及び／又は被訓練モデルＮＮを適用したモデル出力に基づく。モデル出力を特性特徴として使用することは必要でない。いくつかの実施形態では、レコードの特性特徴は、被訓練モデルＮＮの適用を伴わずに、又は適用とは別に判定される。

【００７４】

例えば、特性特徴として使用されるまれなモデル出力は、特にそのまれなモデル出力につながったレコードの特性特徴と組み合わせられたときに、訓練データセットにレコードが合致していないことを示し得る。教師なし学習の場合、訓練レコードのモデル出力は通例、被訓練モデルＮＮを訓練レコードに適用することによって取得される。教師あり学習の場合、訓練レコードのモデル出力は、訓練データセットＴＤ／ＲＴＤに含まれる望まれる出力、被訓練モデルＮＮの実際のモデル出力、又はそれら２つの組み合わせである。

【００７５】

確率データＰＤは、被訓練モデルを適用するエンティティが、あるレコードの１つ又は複数の特性特徴が発生する確率を判定することを可能にし、例えば、一部若しくはすべてのモデル出力ｂ、撮像パラメータｓ及び／又は患者パラメータｔに関係する、１つ又は複数の確率分布ｐ（ｂ，ｓ，ｔ）、ｐｓ（ｓ）及び／又はｐｔ（ｔ）又はその周辺分布を推定することを可能にする。

【００７６】

確率データＰＤは、ヒストグラムの構築などの知られている技術を使用して判定される。確率データＰＤは、例えば、２値データ又は離散データ（例えば１００個未満の取り得る値を伴う）の各クラスにおける百分率を備える。確率データＰＤは、正規分布に従う連続データについての平均、中央値、分散、又は同様の統計的記述子を備える。確率データＰＤは、より正確な確率分布、例えば、例えば年齢、体重、身長などの連続パラメータについての、０．．．１００のような百分位数の値を備えることもできる。確率データＰＤは、２つの特性特徴の間の相関を示す相関値、例えばパラメータの組み合わせに対応する共分散行列を含むこともできる。このような種類の各々の確率データは、モデル入力、例えば撮像パラメータｓ、患者パラメータｔ、モデル出力、又は組み合わせ、を指し得る。

【００７７】

論じられたように、訓練データセットの特定部分が、訓練中により頻繁に使用されるか、又は訓練中に異なる重みを得てもよく、その結果、再平衡された訓練データセットＲＴＤ３１０を得る。例えば、再平衡は、データの性質を平衡させるために訓練中に訓練レコードの特定のセットのいくつかのコピーを作成すること、例えばまれな分類を持つレコードを複製することを含む。確率データＰＤは、それら割り当てられた重みに従って判定される。例えば、確率データＰＤは、再平衡された訓練データセットＲＴＤに基づくか、又は、確率データＰＤは、訓練データセットＴＤであるが同じ重みが適用されたものに基づき、例えば、確率データＰＤのヒストグラム中で各自の重みに対応する回数だけ項目が計数される。例えば、これは、モデルがそれらまれな事例に適するかどうかの判断を向上させ、例えばそのような再平衡を行わないと、訓練セット中にあるまれな事例が比較的少ないために、母集団記述子との非合致が判定されることがある。

【００７８】

確率データＰＤに基づいて、合致検査ＣＣＨ３５０を行って、レコードＲ、例えば訓練データセットＴＤ中にないレコード、が母集団記述子と合致するかどうかを判定する。結果Ｃ？３７０は、被訓練モデルＮＮをレコードＲに適用したことから得られたモデル出力ＲＥＳの信頼性を示し、例えば、モデル出力と並べてユーザに提示される。合致検査ＣＣＨを行うために、確率データＰＤを判定するためにも使用されたレコードＲの１つ又は複数の特性特徴が判定、例えば選択される。例えば、別個のパラメータ又はパラメータの組み合わせに基づき、例えば、確率データＰＤのヒストグラムから確率を取得することにより、又は確率データＰＤの統計的記述子によって与えられる確率密度関数を評価することにより、そのパラメータ又は組み合わせが確率データＰＤに従って発生する確率が判定され

る。任意で、そのように判定された確率を、最大の確率、例えばヒストグラムの中の最も高い度数、最大値の確率密度等、で割って、０～１の間の信頼値を得る。

【００７９】

合致検査ＣＣＨの結果Ｃ？は、出力インターフェースを介してレンダリングデバイスに提供されてよく、例えば結果は画面上に提示される。例えば、１つ又は複数のパラメータ又はパラメータの組み合わせに対応する判定された合致又は非合致がレンダリングされる。それに代えて又はそれに加えて、合致を判定するために算出された値、例えば判定された確率が表示されてよい。図に示されるのは、特性特徴の３つのセットに関する３つの合致検査の結果であり、すなわち、例えばレコードＲの年齢特徴に基づく、「１」と表記された第１の合致検査、例えばレコードＲの体型指数（ＢＭＩ）特徴に基づく、「２」と表記された第２の合致検査、及び、例えばレコードＲの特徴並びにモデル出力ＲＥＳに基づく全体的な合致検査である、「３」と表記された第３の合致検査である。合致は、ここではチェックボックスを用いて示されるが、色コードなどの他の表示が、それに代えて又はそれに加えて使用され得る。非合致が判定された場合、例えば特定の値が閾値を下回る場合、非合致を示す出力信号が生成され、例えば、被訓練モデルＮＮが適用可能でない旨の警告が提示される。

【００８０】

具体例では、図３ａに関して説明される技術が、医用信号処理、例えば集中治療室（ＩＣＵ）に関する予測的分析で使用される。そのような用途では、訓練データセットＴＤは、アノテーションが付加された時系列データを備えてよく、訓練データの訓練レコードは、様々な入力メタパラメータ、例えばいくつかの実施形態ではモデルＮＮの訓練に使用されないデータを備える。例えば、入力メタパラメータは、患者人口統計学的パラメータ、例えば、年齢、性別、体重、人種の１つ又は複数；バイタルパラメータ、例えば血圧、心拍数、呼吸数の１つ又は複数；検査値、及び／又は上述したものの時間的変動特性を備える。例えば、入力データセットは、ＩＣＵ遭遇事例（encounters）のＭＩＭＩＣデータベースのレコードを備える。合致検査ＣＣＨは、例えば、成人のみで訓練されたモデルに誤って供給される小児患者データや、訓練データセット中で見られたことが全くないか滅多に見られない入力パラメータの組み合わせを、査定する、検出する、及び／又はフラグを立てることを可能にする。例えば、そのような合致検査は、患者が、異なる事例の混合、異なる症状の有病率及び／又は異なる重篤性を呈することが予想される、例えば救急部門などの他の状況から得たレコードＲにモデルＮＮを適用する際に行われる。

【００８１】

別の具体例では、図３ａに関して説明した技術が、事象を予測するのを可能にするモデルＮＮを訓練するために使用される。例えば、訓練データセットＴＤは、バイタルパラメータ、検査値、患者の人口統計的情報、介入、及び／又は他のパラメータを備え、例えばデータセットは、ＭＩＭＩＣ及び／又はｅＩＣＵなどのキュレートされた訓練データセットからのレコードを備える。例えば、モデルＮＮは、血行動態不安定性表示モデルである。このコンテキストにおいて、本発明者らは、成人データで訓練されたモデルＮＮに小児の入力データＲを供給すると、性能の低下を招き得ることを見いだした。本明細書に記載される技術と組み合わせて年齢を訓練データセットＴＤの特性パラメータとして使用すると、訓練データセットＴＤと合致しない患者、例えば、年齢が例えば訓練データセットＴＤ中で閾値百分位数を下回る患者、にフラグを立てることが可能となる。興味深い点として、そのために、モデルＮＮを訓練する時、及び／又は確率データＰＤを得るために被訓練モデルを処理する時、未成年の対象者データが後にアルゴリズムに供給され得ることを知ることは必ずしも必要でない。例えば、本明細書に記載される技術は、被訓練モデルＮＮを使用しながら、そのモデルが適用されるレコードＲが特定の特性特徴に関して合致しない、例えば範囲外にあり、よって信頼できない可能性があることの自動化された判定を可能にする。

【００８２】

図３ｂは、母集団記述子が、例えば図１のシステム１００によってどのように判定される

か、及び／又は例えば図２のシステム２００によって使用されるかの、詳細であるが非制限的な例を示す。この例では、システム２００は、複数のレコードの合致を判定する。

【００８３】

図３ｂの例は少なくとも部分的に図３ａに基づいてよく、例えば、訓練データＴＤ３００；再平衡された訓練データＲＴＤ３１０；母集団記述子ＰＤ３２０、例えばヒストグラムなどの確率データ；及び／又は被訓練モデルＮＮ３４０、例えばニューラルネットワークは、図３ａと同様である。

【００８４】

この例では、被訓練モデルを複数のレコードＲ３３０’に適用して、結果ＲＥ３６０’を得る。また、複数のレコードＲに合致検査ＨＣ３５０’が行われ、例えば、例えばグループとしてのレコードＲが、母集団記述子に含まれる確率データＰＤと合致するかどうかが判定される。言い換えると、個人のコホートに対する、被訓練モデル、この場合はニューラルネットワークＮＮの性能が検討される。例えば、非合致の判定を使用して、例えば訓練データセットＴＤに対するレコードＲのモデル及び／又は概念ドリフトを検出する。

【００８５】

複数のレコードＲが確率データＰＤと合致するかどうかを判定することは、ユニットＨＣ３５０’で、複数のレコードＲの各々に判定された１つ又は複数の特性特徴が確率データに従って発生する確率を判定することを含む。例えば、合致を判定することは、訓練レコードＲについての確率データ、例えば確率関数の推定値

【数１】

$$\tilde{p}$$

（ｂ，ｓ，ｔ）

【数２】

$$\tilde{p}_s$$

（ｓ）

及び／又は

【数３】

$$\tilde{p}_t$$

（ｔ）、

を判定し、この確率データを、訓練データセットＴＤに判定された情報ＰＤと比較することを含む。例えば、確率関数は、モデル出力ｂ、撮像パラメータｓ、及び／又は患者パラメータｔに基づく。合致を判定するために、例えば、確率データＰＤ中の訓練データセットＴＤと、レコードＲとの両方に対してヒストグラムが判定され、その場合、ユニットＨＣは、例えばＦｒａｎｋ Ｃ．Ｐｏｒｔｅｒ、「Ｔｅｓｔｉｎｇ Ｃｏｎｓｉｓｔｅｎｃｙ ｏｆ Ｔｗｏ Ｈｉｓｔｏｇｒａｍｓ」、２００８、ａｒＸｉｖ：０８０４．０３８０［ｐｈｙｓｉｃｓ．ｄａｔａ－ａｎ］（参照により本明細書に組み込まれる）に記載されるように、ヒストグラム整合性試験を行って、両方のヒストグラムが同じ母集団を表すかどうかを試験する。合致を判定することは、例えば、訓練データセットＴＤの確率データＰＤを、使用中に遭遇されたレコードＲと比較する統計試験、例えばＷｉｌｋｉｎｓｏｎ又はＳｔｕｄｅｎｔ試験を行うことを含む。

【００８６】

図３ａのように、母集団記述子ＰＤへの複数のレコードＲの合致の判定の結果ＣＣ？３７０’は、例えば、非合致の場合のみ、合致の場合のみ、又は両方の場合に、画面などのレンダリングデバイスに出力される。例えば、複数のレコードＲ、例えば患者のコホートの合致又は非合致は、例えば図３ａに関して説明されたように判定されたＲの個々のレコードの合致又は非合致と並べて、及び／又はレコードＲの判定されたモデル出力と並べて、レンダリングされる。これにより、ユーザが、任意のレコードに対する被訓練モデルＮＮのモデル出力を信頼できるかどうかを、そのコホートが概して被訓練モデルＮＮの訓練データセットＴＤと合致するかどうか、したがって、被訓練モデルが概して当該任意のレコードの設定において信頼できるモデル出力を提供することが予想されるかどうかと、任意のレコード自体が訓練データセットＴＤと合致するかどうか、したがって、モデルが任意のレコードに対して信頼できるモデル出力を提供することが予想されるかどうかとの両方を考慮することにより、よりよく判定することが可能になる。

【００８７】

被訓練モデルＮＮを使用するシステムのプロセッササブシステムはさらに、非合致を示す出力信号に応答して被訓練モデルＮＮのモデルパラメータを適合するように構成される。例えば、分類閾値が、例えば適合率と再現率との間を平衡させるように適合される。一実施では、非合致は、ある疾患の有病率の増大を示し、例えば非合致は、被訓練モデルＮＮによって出力されたある疾患の診断が、訓練データセットＴＤ中よりもレコードＲ中でより頻繁に発生すると判定することに基づく。そのような場合、被訓練モデルＮＮのモデルパラメータは、例えば、感度、例えば再現率を増大するように調節される。

【００８８】

特例の例として、被訓練モデルＮＮは、入力画像に基づいて疾患を診断するように訓練され、例えば、被訓練モデルＮＮは、画像に基づいて結核の存在／不在を検出するように訓練される。本発明者らは、１つの特定の国に入って来る労働年齢のスクリーニング母集団から得た訓練データセットＴＤに基づいて、このタスクのための正確な被訓練モデルＮＮを構築することができた。しかし、この被訓練モデルＮＮが、例えば別の地理にある病弱な高齢者などの異なるコンテキストで適用されると、モデルの精度が低下することがある。そのような場合、図３ａの合致検査ＣＣＨで、例えば年齢、性別、人種、ＢＭＩ、又はそれらの組み合わせに関して、例えば外れ値を自動的に査定して、外れ値にフラグを立てることを可能にする。そのために、被訓練モデルＮＮが特定の地理にのみ適用する明示的な／手動のコーディングは必要ない場合があり、さらに、被訓練モデルは、そのような手動でコーディングされた訓練コホート及び地理に必ずしも制限されない。その結果、被訓練モデルＮＮは、より広い範囲の対象者に使用され得る。さらに、訓練時よりもモデル出力の有病率が高い場合、例えば結核の診断では、モデルパラメータはこの新しい状況を加味するように適合される。例えば、より高い有病率は結核の地域的な大流行を示し得る。可能な応答として、特異性を犠牲にしてモデルの感度が一時的に上げられる。

【００８９】

より一般的には、合致検査ＨＣは、被訓練モデルＮＮの実行設定からの入力及び／又は出力の分布を、訓練データセットＴＤ中の分布と比較することを可能にする。例えば、訓練データセットと比較されたレコードＲは、過去１か月にわたるデータの移動平均を備える。訓練データセットＴＤにわたる確率データ、例えば陽性率の中央値及び時間的分散基準値に基づいて、合致検査ＨＣで、実行フェーズで実際の陽性率が大幅に逸脱したときには必ずフラグを立てることができる。例えば、そのような場合、出力信号、例えばモデルドリフト警告が生成され、例えば臨床ユーザ又はデータ科学の専門家などのユーザに出力される可能性がある。

【００９０】

図４は、母集団記述子が、例えば図１のシステム１００によってどのように判定されるか、及び／又は例えば図２のシステム２００によってレコードの合致を判定するためにどのように使用されるかの、詳細であるが非制限的な別の例を示す。

【００９１】

図４に示す例では、訓練データセットＴＩ４００が、例えば対応するアノテーションＢｉと共に複数の画像Ｉｉを備える。例えば、画像のアノテーションは、画像のセグメント化、例えば画像内の臓器のセグメント化、を含む。任意で、訓練データセットＴＩでモデルを訓練する前に、まず拡張された訓練データセットＡＴＩ４１０が訓練データセットＴＩから導出される。例えば、知られているデータ拡張技術を使用して、画像Ｉｉ及び／又はアノテーションＢｉを変換し、その結果、例えば対応するアノテーションと共に、増加した数の画像を備える拡張された訓練データセットＡＴＩを得る。例えば非線形変換が使用され、例えば、訓練データセットＴＤの２つの画像の厳格でない位置合わせと、訓練データセットＴＤ中の残りの画像への変換の適用とによって生成された非線形変換が使用される。例えば、非線形変換は、Ｔｏｍ Ｖｅｒｃａｕｔｅｒｅｎら、「Ｎｏｎ－ｐａｒａｍｅｔｒｉｃ Ｄｉｆｆｅｏｍｏｒｐｈｉｃ Ｉｍａｇｅ Ｒｅｇｉｓｔｒａｔｉｏｎ ｗｉｔｈ ｔｈｅ Ｄｅｍｏｎｓ Ａｌｇｏｒｉｔｈｍ」、Ｐｒｏｃｅｅｄｉｎｇｓ ｏｆ ＭＩＣＣＡＩ ２００７；Ｓｐｒｉｎｇｅｒ ＬＮＣＳ ｖｏｌ ４７９２（参照により本明細書に組み込まれる）に開示されるように使用される。

【００９２】

訓練データセットＴＩ、又は拡張が使用される場合は拡張された訓練データセットＡＴＩが、モデルＣＮＮ４５０、この例では畳み込みニューラルネットワーク（ＣＮＮ）を訓練するために使用される。例えば、各自の対応するアノテーションを伴う訓練画像が訓練で使用されて、例えば、Ｔｏｍ Ｂｒｏｓｃｈ、Ａｘｅｌ Ｓａａｌｂａｃｈ、「Ｆｏｖｅａｌ ｆｕｌｌｙ ｏｎｖｏｌｕｔｉｏｎａｌ ｎｅｔｓ ｆｏｒ ｍｕｌｔｉ－ｏｒｇａｎ ｓｅｇｍｅｎｔａｔｉｏｎ」、Ｐｒｏｃｅｅｄｉｎｇｓ Ｖｏｌｕｍｅ １０５７４、Ｍｅｄｉｃａｌ Ｉｍａｇｉｎｇ ２０１８：Ｉｍａｇｅ Ｐｒｏｃｅｓｓｉｎｇ；１０５７４０Ｕ；２０１８（参照により本明細書に組み込まれる）に開示されるようにＦ－ｎｅｔを訓練する。被訓練モデルＣＮＮは、ボクセル分類器である。例えば、被訓練モデルＣＮＮは、画像を備えるレコードが入力されると、その画像のセグメント化を提供するように訓練される。ニューラルネットワークの他に、様々な他の種類の被訓練モデルが使用されてよく、例えば、被訓練モデルは、例えばＯｌｉｖｉｅｒ Ｅｃａｂｅｒｔら、「Ｓｅｇｍｅｎｔａｔｉｏｎ ｏｆ ｔｈｅ ｈｅａｒｔ ａｎｄ ｇｒｅａｔ ｖｅｓｓｅｌｓ ｉｎ ＣＴ Ｉｍａｇｅｓ ｕｓｉｎｇ ａ ｍｏｄｅｌ－ｂａｓｅｄ ａｄａｐｔａｔｉｏｎ ｆｒａｍｅｗｏｒｋ」、Ｍｅｄｉｃａｌ Ｉｍａｇｅ Ａｎａｌｙｓｉｓ；１５（６）；２０１１（参照により本明細書に組み込まれる）に開示されるようなモデルに基づくセグメント化モデルである。

【００９３】

訓練データセットＴＩ、又は拡張が使用される場合には拡張された訓練データセットＡＴＩに基づいて、データセットの訓練レコードの特性特徴が判定される。本例では、訓練画像Ｉｉの特性特徴は、画像Ｉｉ及び／又はセグメント化Ｂｉ内で表される形状の、１つ又は複数の形状記述子ＳＤ４２０を備える。例えば、形状記述子は、１つ又は複数のＳｈａｐｅＤＮＡ特徴を備える。ＳｈａｐｅＤＮＡ特徴は、スケールの不変性及び／又は非剛体形状の形状検索の性能が重要である状況で特に有用であり得る。特性特徴を判定することは、セグメント化Ｂｉの幾何学的性質を分析することを含む。例えば、スペクトル形状分析を使用して特性特徴を判定する。スペクトル形状分析技術は、例えば、Ｍａｒｃ Ｎｉｅｔｈａｍｍｅｒら、「Ｇｌｏｂａｌ Ｍｅｄｉｃａｌ Ｓｈａｐｅ Ａｎａｌｙｓｉｓ Ｕｓｉｎｇ ｔｈｅ Ｌａｐｌａｃｅ－Ｂｅｌｔｒａｍｉ Ｓｐｅｃｔｒｕｍ」、Ｐｒｏｃｅｅｄｉｎｇｓ ＭＩＣＣＡＩ ２００７、Ｓｐｒｉｎｇｅｒ ＬＮＣＳ ｖｏｌ ４７９１（参照により本明細書に組み込まれる）から知られている。例えば、そのような分析は、スペクトル、例えばＬａｐｌａｃｅ－Ｂｅｌｔｒａｍｉ演算子の固有値及び／又は固有関数、を使用して、様々な形状記述子、例えばＳｈａｐｅＤＮＡ、を導出することができる。

【００９４】

スペクトル形状分析によって導出される特徴に代えて又はそれに加えて、形状記述子ＳＤ

は、変換パラメータ及び／又は形状固有モードを、例えばモデルに基づくセグメント化と組み合わせて備える。形状記述子ＳＤは、セグメント化Ｂ¡から他の形で導出されてもよく、例えば形状記述子は、画像中で識別された物体、例えば臓器、の寸法、体積等を備える。例えば臓器の最小／最大サイズなど、手動で定義された制限がそのような形状記述子に適用される。

【００９５】

形状記述子ＳＤを判定することは、例えば、慣性テンソルのモーメントの固有ベクトルに対応するオイラー角を使用して、又はコサインの行列を使用して、画像の座標系に対する形状の向きを査定することを含む。これは、例えば長球又は長円の腫瘍などの、非球形のセグメント化結果に特に有用であり得る。形状記述子ＳＤは、任意で、例えば画像座標系に対する、又は画像中の他の物体若しくは構造物に対する、物体の場所、例えば重心を含む。

【００９６】

形状記述子ＳＤは、追加的な特性特徴、例えば画像ＴＩに関連付けられたメタデータに基づく特性と組み合わせられる。例えば、図３ａに関して説明したように、撮像プロトコルの詳細ｓ及び／又は撮像対象者のパラメータｔなどのメタデータが含まれる。例えば、特性特徴は焦点長を含み、それにより、例えば標準レンズ／望遠レンズの写真で訓練されたモデルが、魚眼レンズで撮影された画像に適用されるときに、非合致を検出することを可能にする。医用画像処理の用途には、同様に線源検出器距離（ＳＤＤ）が特性パラメータとして使用される。興味深い点として、いくつかの実施形態では、そのようなメタデータは、被訓練モデルＣＮＮに入力されないが、それでも合致を判定するために使用される。

【００９７】

さらに、例えば画像の平均及び／又は中央グレー値、ノイズ尺度等、一部の特性特徴は、画像自体から導出される。例えば、そのような導出された特徴は、医用画像処理、例えば自動化された臓器セグメント化の状況で、訓練データセット中の画像よりも信号対雑音比が悪い入力画像や、訓練時に見られなかった予期しないＭＲＩコントラストにフラグを立てるために有用であり得る。さらに、様々なメタデータ又は導出された特徴の使用は、操作／改ざん、又は不正行為、例えば被訓練モデルの特定の出力を強制するように操作されている画像、を検出するために使用され得る。

【００９８】

形状記述子ＳＤを含む特性特徴は、母集団記述子、この例では１クラスサポートベクトルマシン（ＳＶＭ）などの１クラス分類器ＳＶＭ４３０を決定するために使用される。１クラス分類器ＳＶＭは、モデルを使用するときにレコードが合致すると分類することができる。例えば、分類器は、セグメント化の結果が、例えば形状の性質に関して、予想の範囲内にあるかどうかを検証することができる。例えば、Ａｘｅｌ Ｓａａｌｂａｃｈら、「Ｆａｉｌｕｒｅ Ａｎａｌｙｓｉｓ ｆｏｒ Ｍｏｄｅｌ－Ｂａｓｅｄ Ｏｒｇａｎ Ｓｅｇｍｅｎｔａｔｉｏｎ Ｕｓｉｎｇ Ｏｕｔｌｉｅｒ Ｄｅｔｅｃｔｉｏｎ」、Ｐｒｏｃｅｅｄｉｎｇｓ Ｖｏｌｕｍｅ ９０３４、Ｍｅｄｉｃａｌ Ｉｍａｇｉｎｇ ２０１４：Ｉｍａｇｅ Ｐｒｏｃｅｓｓｉｎｇ；９０３４０８；２０１４（参照により本明細書に組み込まれる）に開示される外れ値検出手法が使用される。

【００９９】

被訓練モデルＣＮＮを使用するためのシステムでは、モデルＣＮＮは、新しい画像Ｉ４４０に適用され、その結果、モード出力、例えばセグメント化ＳＥＧ４８０を得る。上記と同様に、アノテーションの特性特徴ＳＤ’４６０、例えばセグメント化ＳＥＧの形状記述子が例えばスペクトル形状分析を使用して判定される。合致検査ＣＣＬ４７０では、１クラス分類器ＳＶＭを使用してラベルＬＢＬ４９０を判定し、アノテーションＳＥＧを正常、例えば訓練データセットＴＩに合致している、又は外れ値として分類する。

【０１００】

様々な実施形態において、画像ＴＩ、Ｉのセグメント化は、複数の連結成分を備えることがある。そのような場合、形状記述子ＳＤを判定することは、画像ＴＩ、Ｉのセグメント

化に対して、それ自体知られているように連結成分分析を実行することを含む。いくつかの実施形態では、連結成分の１つ、例えば最も大きい物体が選択され、その選択された連結成分の形状記述子ＳＤが判定される。例えば、これは特に、主要な物体から切り離された検出結果を提供し得るボクセル分類器が使用される場合に有用である。他の実施形態では、例えばセグメント化の結果、脊椎などのいくつかの物体が得られることが予想される場合、形状記述子ＳＤが判定されてよく、及び／又は、いくつかの、例えばすべての識別された連結成分に合致判定が行われる。

【０１０１】

図５は、訓練データセットの訓練レコードを処理するコンピュータ実施方法５００のブロック図を示す。方法５００は、図１のシステム１００の動作に対応し得る。ただし、方法５００は別のシステム、装置、又はデバイスを使用して行われてもよいという点で、これは制限ではない。

【０１０２】

方法５００は、「訓練データセットにアクセスする」という動作において、被訓練モデルの訓練データセットにアクセスするステップ５１０を有し、訓練データセットは複数の訓練レコードを備え、被訓練モデルは、レコードに適用されたときにモデル出力を提供するように訓練データセットで訓練される。方法５００は、「特性特徴を判定する」という動作において、訓練データセットの訓練レコードについての１つ又は複数の特性特徴を判定するステップ５２０をさらに有し、前記１つ又は複数の特性特徴は、訓練レコード及び／又は被訓練モデルが訓練レコードに適用されたときに提供されるモデル出力を特徴付ける。方法５００は、「母集団記述子を判定する」という動作において、訓練データセットにわたる１つ又は複数の特性特徴の分布を特徴付ける母集団記述子を判定するステップ５３０をさらに有する。方法５００は、「母集団記述子を被訓練モデルに関連付ける」という動作において、被訓練モデルを少なくとも１つのレコードに適用するエンティティが、当該少なくとも１つのレコードが母集団記述子と合致するかどうかを、少なくとも１つのレコードについて１つ又は複数の特性特徴を判定することにより、及び前記１つ又は複数の特性特徴の、訓練データセットにわたる前記特性特徴の分布からの逸脱を判定することにより判定することを可能にするために、母集団記述子を被訓練モデルに関連付けるステップをさらに有する。

【０１０３】

図６は、被訓練モデルを使用するためのコンピュータ実施方法６００のブロック図を示す。方法６００は、図２のシステム２００の動作に対応し得る。ただし、方法６００は別のシステム、装置、又はデバイスを使用して行われてもよいという点で、これは制限ではない。

【０１０４】

方法６００は、「母集団記述子、レコードにアクセスする」という動作において、被訓練モデルに関連付けられた母集団記述子であって、被訓練モデルは訓練データセットで訓練されており、訓練データセットは複数の訓練レコードを備え、母集団記述子は、訓練データセットにわたる訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルを訓練レコードに適用したときに提供されるモデル出力を特徴付ける、母集団記述子と、被訓練モデルが適用される少なくとも１つのレコードとにアクセスするステップ６１０を有する。方法６００は、「特性特徴を判定する」という動作において、少なくとも１つのレコードの１つ又は複数の特性特徴を判定するステップ６２０をさらに有する。方法６００は、「合致を判定する」という動作において、当該少なくとも１つのレコードが母集団記述子と合致するかどうかを、少なくとも１つのレコードの１つ又は複数の特性特徴が訓練データセットにわたるそれらの分布に従って発生する尤度を判定することによって判定するステップ６３０をさらに有する。例えば、そのような判定は、判定された特性特徴が、母集団記述子に含まれる確率データに従って発生する確率ｐ／ｐｍａｘを求め、求めた確率を閾値Ｔと比較することを伴う。方法６００は、「非合致を示す出力信号を生成する」とい

う動作において、少なくとも１つのレコードが被訓練モデルの母集団記述子と合致しない場合、前記非合致を示す出力信号を生成するステップ６４０をさらに有する。

【０１０５】

概して、図５の方法５００及び／又は図６の方法６００の動作は、該当する場合、例えば入力／出力関係などによって特定の順序が必要とされる場合に、任意の適切な順序で、例えば連続して、同時に、又はそれらの組み合わせで行われてよいことが理解されよう。

【０１０６】

方法は、コンピュータ実施方法として、専用ハードウェアとして、又は両者の組み合わせとしてコンピュータ上で実施されてよい。図７にも示されるように、コンピュータに対する命令、例えば実行可能コードが、一連の機械可読物理的マーク７１０の形態で、及び／又は異なる電気的、例えば磁気的、若しくは光学的性質若しくは値を有する一連の要素として、コンピュータ可読媒体７００に記憶されてよい。実行可能コードは、一時的又は非一時的な方式で記憶される。媒体７００は、その代わりに又はそれに加えて、被訓練モデルと、被訓練モデルに関連付けられた母集団記述子とを表すデータを記憶してもよく、被訓練モデルは訓練データセットで訓練されており、訓練データセットは複数の訓練レコードを備え、母集団記述子は、訓練データセットにわたる訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルを訓練レコードに適用したときに提供されるモデル出力を特徴付ける。コンピュータ可読媒体の例は、メモリデバイス、光学記憶装置、集積回路、サーバ、オンラインソフトウェア等を含む。図７は、光学ディスク７００を示す。

【０１０７】

以下の番号付きの項は、企図される非制限的な実施形態を含む。

１．訓練データセットの訓練レコードを処理するためのシステムであって、このシステムは、

－　データインターフェースであって、

－　被訓練モデルの訓練データセットであって、訓練データセットは複数の訓練レコードを備え、被訓練モデルは、レコードに適用されたときにモデル出力を提供するように訓練データセットで訓練される、訓練データセット

にアクセスするためのデータインターフェースと、

－　プロセッササブシステムであって、

－　訓練データセットの訓練レコードについての１つ又は複数の特性特徴を判定することであって、前記１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルが訓練レコードに適用されたときに提供されるモデル出力を特徴付ける、判定すること、

－　訓練データセットにわたる１つ又は複数の特性特徴の分布を特徴付ける母集団記述子を判定すること、

－　被訓練モデルを少なくとも１つのレコードに適用するエンティティが、当該少なくとも１つのレコードが母集団記述子と合致するかどうかを、少なくとも１つのレコードについて１つ又は複数の特性特徴を判定することにより、及び前記１つ又は複数の特性特徴の、訓練データセットにわたる前記特性特徴の分布からの逸脱を判定することにより判定することを可能にするために、母集団記述子を被訓練モデルに関連付けること

を行うプロセッササブシステムと、を備える、システム。

２．プロセッササブシステムが、訓練データセットで機械学習モデルを訓練し、それにより被訓練モデルを得るように構成される、先行する項に記載のシステム。

３．被訓練モデルが適用される少なくとも１つのレコードを処理するためのシステムであって、このシステムは、

－　データインターフェースであって、

－　被訓練モデルに関連付けられた母集団記述子であって、被訓練モデルは訓練データセットで訓練されており、訓練データセットは複数の訓練レコードを備え、母集団記述子は、訓練データセットにわたる訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モ

デルを訓練レコードに適用したときに提供されるモデル出力を特徴付ける、母集団記述子、
－　被訓練モデルが適用される少なくとも１つのレコード、
にアクセスするためのデータインターフェースと、
－　プロセッササブシステムであって、
－　少なくとも１つのレコードの１つ又は複数の特性特徴を判定すること、
－　当該少なくとも１つのレコードの１つ又は複数の特性特徴の、訓練データセットにわたる前記特性特徴の分布からの逸脱を判定して、少なくとも１つのレコードが母集団記述子と合致するかどうかを判定すること、
－　当該少なくとも１つのレコードが被訓練モデルの母集団記述子と合致しない場合、前記非合致を示す出力信号を生成すること
を行うプロセッササブシステムと、を備える、システム。
４．訓練データセットの訓練レコードを処理するコンピュータ実施方法であって、この方法は、
－　被訓練モデルの訓練データセットにアクセスするステップであって、訓練データセットは複数の訓練レコードを備え、被訓練モデルは、レコードに適用されたときにモデル出力を提供するように訓練データセットで訓練される、ステップと、
－　訓練データセットの訓練レコードについての１つ又は複数の特性特徴を判定するステップであって、前記１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルが訓練レコードに適用されたときに提供されるモデル出力を特徴付ける、ステップと、
－　訓練データセットにわたる１つ又は複数の特性特徴の分布を特徴付ける母集団記述子を判定するステップと、
－　被訓練モデルを少なくとも１つのレコードに適用するエンティティが、当該少なくとも１つのレコードが母集団記述子と合致するかどうかを、少なくとも１つのレコードについて１つ又は複数の特性特徴を判定することにより、及び前記１つ又は複数の特性特徴の、訓練データセットにわたる前記特性特徴の分布からの逸脱を判定することにより判定することを可能にするために、母集団記述子を被訓練モデルに関連付けるステップと、を有する、方法。
５．被訓練モデルが適用される少なくとも１つのレコードを処理するコンピュータ実施方法であって、この方法は、
－　被訓練モデルに関連付けられた母集団記述子であって、被訓練モデルは訓練データセットで訓練されており、訓練データセットは複数の訓練レコードを備え、母集団記述子は、訓練データセットにわたる訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルを訓練レコードに適用したときに提供されるモデル出力を特徴付ける、母集団記述子、
－　被訓練モデルが適用される少なくとも１つのレコード
にアクセスするステップと、
－　少なくとも１つのレコードの１つ又は複数の特性特徴を判定するステップと、
－　少なくとも１つのレコードの１つ又は複数の特性特徴が訓練データセットにわたるそれらの分布に従って発生する尤度を判定することにより、少なくとも１つのレコードが母集団記述子と合致するかどうかを判定するステップと、
－　少なくとも１つのレコードが被訓練モデルの母集団記述子と合致しない場合、前記非合致を示す出力信号を生成するステップとを有する、方法。
６．被訓練モデルと、被訓練モデルに関連付けられた母集団記述子とを表す一時的又は非一時的なデータを備えるコンピュータ可読媒体であって、被訓練モデルは訓練データセットで訓練されており、訓練データセットは複数の訓練レコードを備え、母集団記述子は、訓練データセットにわたる訓練レコードについての１つ又は複数の特性特徴の分布を特徴付け、訓練レコードの１つ又は複数の特性特徴は、訓練レコード、及び／又は被訓練モデルを訓練レコードに適用したときに提供されるモデル出力を特徴付ける、コンピュータ可

読媒体。

【０１０８】

例、実施形態、又は任意選択の特徴は、非制限的と示されてもそうでなくても、特許請求の範囲に記載される本発明を制限するものと解釈すべきでない。

【０１０９】

上述の実施形態は、本発明を制限するのではなく例示するものであり、当業者は、添付の特許請求の範囲から逸脱することなく多くの代替実施形態を考案することができることが留意されるべきである。特許請求の範囲において、括弧内に置かれた参照符号は、請求項を制限するものと解釈すべきでない。動詞「有する、含む」及びその活用形の使用は、請求項内で述べられるもの以外の要素又は段階の存在を排除しない。単数形は、複数のそのような要素の存在を排除しない。要素の列挙又は群の前にあるとき、「～の少なくとも１つ」などの表現は、その列挙又はグループからの要素のすべて又は任意のサブセットの選択を表す。例えば、「Ａ、Ｂ、及びＣの少なくとも１つ」という表現は、Ａのみ、Ｂのみ、Ｃのみ、ＡとＢの両方、ＡとＣの両方、ＢとＣの両方、又はＡ、Ｂ、及びＣのすべてを含むものと理解すべきである。本発明は、いくかの別個の要素を備えるハードウェアを用いて、及び適宜プログラムされたコンピュータを用いて実施されてよい。いくつかの手段を列挙する装置の請求項では、それら手段のいくつかが、１つの同じハードウェア品によって具現化されてよい。特定の手段が相互に異なる従属請求項に記載されているという単なる事実は、それら手段の組み合わせを有利に使用できないことを意味するものではない。

【図面】

【図１】

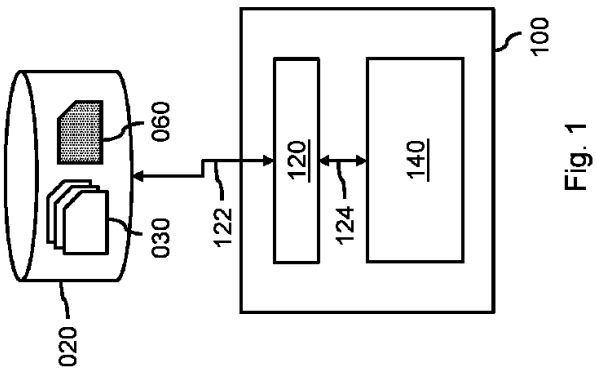


Fig. 1

【図２】

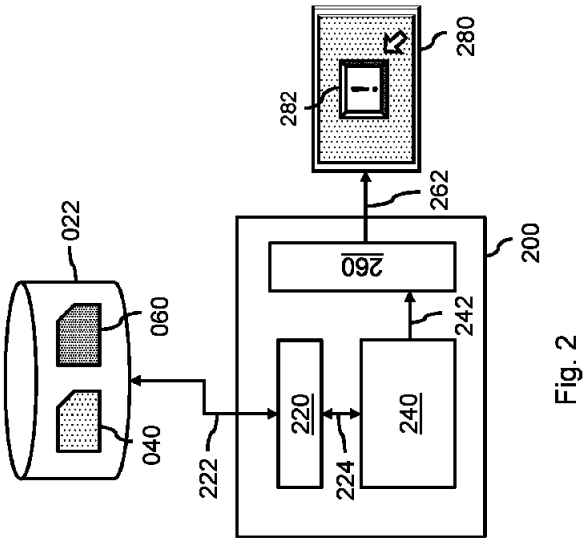


Fig. 2

【図３ａ】

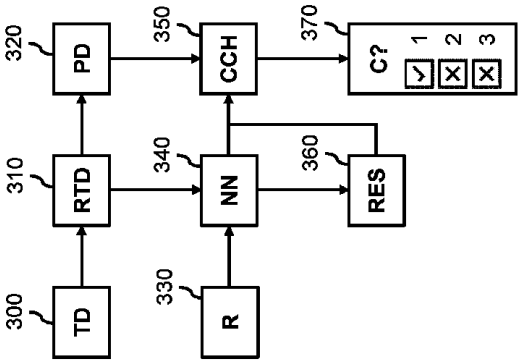


Fig. 3a

【図３ｂ】

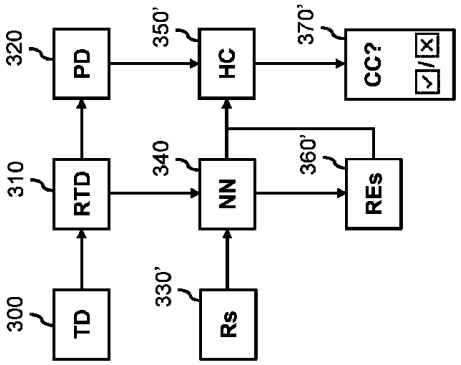


Fig. 3b

【図４】

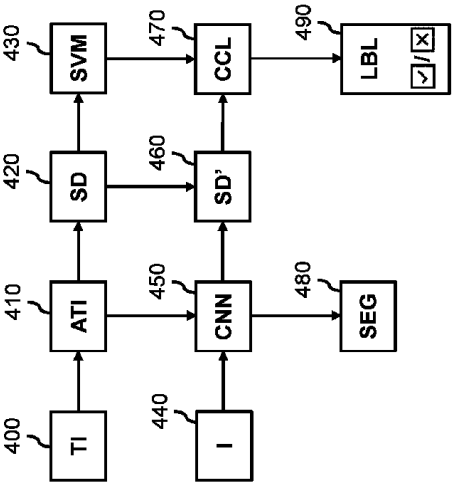


Fig. 4

【図５】

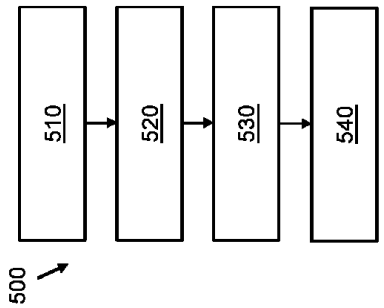


Fig. 5

【図６】

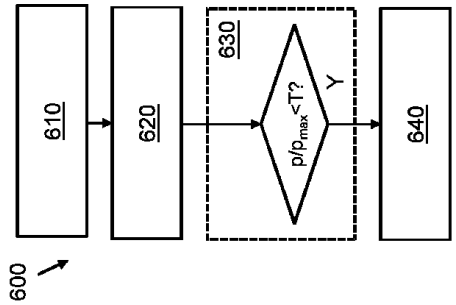


Fig. 6

【図７】

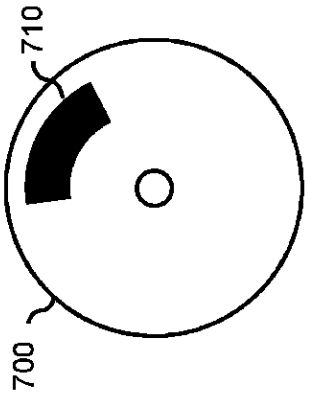


Fig. 7

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/051647

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06N3/04 G06N3/08 G06N20/10
ADD. G06N5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Igor M. Quintanilha ET AL: "Detecting Out-of-Distribution Samples using Low-order Deep Feature Statistics", , 21 December 2018 (2018-12-21), XP055603220, Retrieved from the Internet: URL:https://openreview.net/pdf?id=rkgpCoRctm [retrieved on 2019-07-08] the whole document ----- -/--	1-15

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 April 2020	24/04/2020

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Moro Pérez, Gonzalo

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/051647

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SAALBACH AXEL ET AL: "Failure analysis for model-based organ segmentation using outlier detection", PROGRESS IN BIOMEDICAL OPTICS AND IMAGING, SPIE - INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING, BELLINGHAM, WA, US, vol. 9034, 21 March 2014 (2014-03-21), pages 903408-903408, XP060031681, ISSN: 1605-7422, DOI: 10.1117/12.2041922 ISBN: 978-1-5106-0027-0 cited in the application the whole document	1-15
A	RAZA HAIDER ET AL: "Learning with covariate shift-detection and adaptation in non-stationary environments: Application to brain-computer interface", 2015 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), IEEE, 12 July 2015 (2015-07-12), pages 1-8, XP033222223, DOI: 10.1109/IJCNN.2015.7280742 the whole document	1-15
A	DAN HENDRYCKS ET AL: "A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 7 October 2016 (2016-10-07), XP081059792, the whole document	1-15
A	VICTORIA HODGE ET AL: "A Survey of Outlier Detection Methodologies", ARTIFICIAL INTELLIGENCE REVIEW, vol. 22, no. 2, 1 October 2004 (2004-10-01), pages 85-126, XP055176642, ISSN: 0269-2821, DOI: 10.1023/B:AIRE.0000045502.10941.a9 the whole document	1-15

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

フロントページの続き

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

(72)発明者　ウィッシュマン　ハンス‐アロイス
オランダ国　５６５６　アーエー　アインドーフェン　ハイ　テック　キャンパス　５