

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2024-540041

(P2024-540041A)

(43)公表日 令和6年10月31日(2024.10.31)

|                                   |                      |             |
|-----------------------------------|----------------------|-------------|
| (51)国際特許分類                        | F I                  | テーマコード (参考) |
| G 0 6 T 7/00 (2017.01)            | G 0 6 T 7/00 3 5 0 B | 5 L 0 9 6   |
| G 0 6 V 10/764 (2022.01)          | G 0 6 T 7/00 6 3 0   |             |
| G 0 6 V 20/69 (2022.01)           | G 0 6 V 10/764       |             |
| G 0 6 N 20/00 (2019.01)           | G 0 6 V 20/69        |             |
| G 0 6 N 3/08 (2023.01)            | G 0 6 N 20/00        |             |
| 審査請求 未請求 予備審査請求 未請求 (全47頁) 最終頁に続く |                      |             |

|                   |                                                                                                                                                                                             |         |                                                                                  |
|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|----------------------------------------------------------------------------------|
| (21)出願番号          | 特願2024-525144(P2024-525144)                                                                                                                                                                 | (71)出願人 | 523109150                                                                        |
| (86)(22)出願日       | 令和4年10月19日(2022.10.19)                                                                                                                                                                      |         | パナケイア・テクノロジーズ・リミテッ<br>ド                                                          |
| (85)翻訳文提出日        | 令和6年6月12日(2024.6.12)                                                                                                                                                                        |         | イギリス国、シービー１・２エルエー<br>ケンブリッジ、ステーション・ロード、<br>ソールズベリー・ハウス                           |
| (86)国際出願番号        | PCT/GB2022/052663                                                                                                                                                                           | (74)代理人 | 110003708                                                                        |
| (87)国際公開番号        | WO2023/073346                                                                                                                                                                               |         | 弁理士法人鈴榮特許総合事務所                                                                   |
| (87)国際公開日         | 令和5年5月4日(2023.5.4)                                                                                                                                                                          | (72)発明者 | ラハルジャ、ムハンマド・パンドゥ                                                                 |
| (31)優先権主張番号       | 21205556.0                                                                                                                                                                                  |         | イギリス国、シービー１・２エルエー<br>ケンブリッジ、ステーション・ロード、<br>ソールズベリー・ハウス、パナケイア・<br>テクノロジーズ・リミテッド気付 |
| (32)優先日           | 令和3年10月29日(2021.10.29)                                                                                                                                                                      | (72)発明者 | リー、シュースー                                                                         |
| (33)優先権主張国・地域又は機関 | 欧州特許庁(EP)                                                                                                                                                                                   |         | イギリス国、シービー１・２エルエー<br>最終頁に続く                                                      |
| (81)指定国・地域        | AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA<br>,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(<br>AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,A<br>T,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR<br>,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,<br>最終頁に続く |         |                                                                                  |

(54)【発明の名称】 分類の方法

## (57)【要約】

第１のクラスに対応する第１の入力データを備える第１のデータセットを取得することと、  
第１のクラスに対応する第２の入力データを備える第２のデータセットを取得することと、  
第１のデータセットを使用して少なくとも１つの分類器をトレーニングすることと、  
第２のデータセットからの第２の入力データを少なくとも１つの分類器に入力すること、および第１の分類と第２の分類とを備える分類モデルを提供することと、ここにおいて、第１の分類が、第２の分類よりも、第１のクラス中にあることになる第１のクラスに対応する第２の入力データのより大きい比率を予測する、を備える、コンピュータ実装方法。

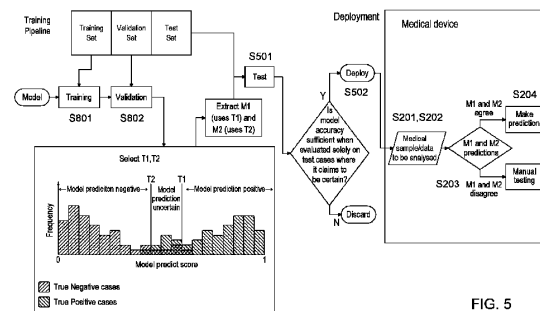


FIG. 5

## 【特許請求の範囲】

## 【請求項 1】

第 1 のクラスに対応する第 1 の入力データを備える第 1 のデータセットを取得することと、

前記第 1 のクラスに対応する第 2 の入力データを備える第 2 のデータセットを取得することと、

前記第 1 のデータセットを使用して少なくとも 1 つの分類器をトレーニングすることと、

前記第 2 のデータセットからの前記第 2 の入力データを前記少なくとも 1 つの分類器に入力すること、および第 1 の分類と第 2 の分類とを備える分類モデルを提供することと、  
ここにおいて、前記第 1 の分類が、前記第 2 の分類よりも、前記第 1 のクラス中にあることになる前記第 1 のクラスに対応する前記第 2 の入力データのより大きい比率を予測する

を備える、コンピュータ実装方法。

## 【請求項 2】

前記第 1 の分類が、第 1 の分類器を適用することと、第 1 の予測しきい値を使用することとを備え、前記第 2 の分類が、前記第 1 の分類器を適用することと、第 2 の予測しきい値を使用することとを備え、前記第 1 の予測しきい値を決定することが、前記第 1 のクラスに対応する前記第 2 のデータセットからの入力について前記第 1 の分類器の最も高い出力値を決定することと、前記第 1 の予測しきい値を、前記最も高い出力値よりも小さくまたはそれに等しく設定することとを備え、前記第 2 の予測しきい値を決定することが、第 2 のクラスに対応する前記第 2 のデータセットからの入力について最も低い出力値を決定することと、前記第 2 の予測しきい値を、前記最も低い出力値よりも大きくまたはそれに等しく設定することとを備える、請求項 1 に記載の方法。

## 【請求項 3】

前記少なくとも 1 つの分類器が少なくとも 2 つの分類器を備え、前記分類モデルを提供することは、前記少なくとも 2 つの分類器のうちの第 1 の分類器を決定することと、ここにおいて、前記第 1 の分類器が、前記少なくとも 2 つの分類器の、前記第 1 のクラス中にあることになる前記第 1 のクラスに対応する前記第 2 の入力データの最も大きい比率を予測する、前記少なくとも 2 つの分類器のうちの第 2 の分類器を決定することと、ここにおいて、前記第 2 の分類器が、前記少なくとも 2 つの分類器の、第 2 のクラス中にあることになる前記第 2 のクラスに対応する前記第 2 の入力データの最も大きい比率を予測する、を備える、請求項 1 に記載の方法。

## 【請求項 4】

前記分類モデルを提供することが、第 1 の予測しきい値と第 2 の予測しきい値とを決定することとを備え、前記第 1 の予測しきい値と前記第 2 の予測しきい値とを決定することが、前記第 1 の予測しきい値についての値と前記第 2 の予測しきい値についての値との複数の可能な組合せを取得することと、前記可能な組合せの各々を使用して分類モデルの精度を決定することと、所定の精度値よりも高い精度を有する組合せを選択することとを備える、請求項 2 または 3 に記載の方法。

## 【請求項 5】

第 1 のクラスに対応する第 1 の入力データを備える第 1 のデータセットを取得することと、

前記第 1 のクラスに対応する第 2 の入力データを備える第 2 のデータセットを取得することと、

前記第 1 のデータセットを使用して分類器をトレーニングすることと、

前記第 2 のデータセットからの前記第 2 の入力データを少なくとも 1 つの分類器に入力すること、および第 1 の分類を備える分類モデルを提供することと、ここにおいて、前記第 1 の分類を備える前記分類モデルを提供することが、前記第 1 のクラスに対応する予測しきい値を決定することとを備える、

を備える、コンピュータ実装方法。

【請求項 6】

前記第 1 のデータセットが、第 1 の集団に対応する第 1 の入力データを備え、前記第 2 のデータセットが、第 2 の集団に対応する第 2 の入力データを備える、請求項 1 から 5 のいずれかに記載の方法。

【請求項 7】

入力データを取得することと、

前記入力データを分類モデルに入力することと、前記分類モデルが第 1 の分類と第 2 の分類とを適用し、ここにおいて、前記第 1 の分類が、前記第 2 の分類よりも、第 1 のクラスに対応する入力データのより大きい比率を正しく予測するように構成される、

10

前記第 1 の分類および前記第 2 の分類の出力に基づいて前記入力データについての分類予測を決定することと

を備える、コンピュータ実装方法。

【請求項 8】

前記第 1 の分類が、第 1 の分類器を適用することと、第 1 の予測しきい値を使用することとを備え、前記第 2 の分類が、前記第 1 の分類器を適用することと、第 2 の予測しきい値を使用することとを備え、ここにおいて、前記第 1 の予測しきい値が前記第 2 の予測しきい値よりも高い、請求項 1 または 7 に記載の方法。

【請求項 9】

前記第 1 の分類が第 1 の分類器を適用することを備え、前記第 2 の分類が第 2 の分類器を適用することを備える、請求項 7 に記載の方法。

20

【請求項 10】

前記第 1 の分類が第 1 の予測しきい値を使用し、前記第 2 の分類が第 2 の予測しきい値を使用し、ここにおいて、前記第 1 の予測しきい値が前記第 2 の予測しきい値よりも高い、請求項 9 に記載の方法。

【請求項 11】

前記入力データが組織の画像を備え、ここにおいて、前記分類を決定することが、医療診断に係る情報を決定することを備える、請求項 7 から 9 のいずれかに記載の方法。

【請求項 12】

入力データを取得することと、

30

前記入力データを分類モデルに入力することと、前記分類モデルが第 1 の分類を適用し、ここにおいて、前記第 1 の分類が、第 1 の分類器を適用することと、第 1 のクラスに対応する予測しきい値を使用することとを備える、

前記第 1 の分類の出力に基づいて前記第 1 のクラスについての分類予測を決定することと

を備える、分類のコンピュータ実装方法。

【請求項 13】

請求項 7 から 12 のいずれかに記載の方法を実施するように構成された、1 つまたは複数のプロセッサ

を備える、分類システム。

40

【請求項 14】

請求項 1 から 7 のいずれかに記載の方法に従ってトレーニングされる分類モデルを備える分類システム。

【請求項 15】

コンピュータに請求項 1 から 12 のいずれかに記載の方法を実施させるように構成されたコンピュータ可読コードを備えるキャリア媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、分類の方法、分類のためのシステム、およびトレーニングの方法に関する。

50

## 【背景技術】

## 【0002】

分類器モデルは、広範囲の分野における適用例を有する。たとえば、分類器モデルは、様々な医療適用例において使用され得る。分類器モデルは、たとえば生理的測定値を入力としてとることによって、医療診断を提供するために使用され得る。別の例では、分類器モデルは、たとえば、入力組織画像が癌性細胞を含んでいるかどうかを予測するために、または特定のバイオマーカーが組織画像に関連付けられるかどうかを予測するために、医療画像データを分析するために使用され得る。

## 【0003】

そのような分類器モデルは、トレーニングデータセットを使用してトレーニングされ得る。たとえば、トレーニングデータセットは、ラベリングされた分類予測とともに、多数の例示的な入力を備え得る。トレーニングデータセットに関してモデル化された分類器モデルを使用する分類方法は、しかしながら、新しい入力データ上に展開されたとき、あるレベルの不正確さを有することになる。分類方法の精度は、その方法によって行われた総予測の数の比率としての正しい予測の数によって測定され得る。これは、たとえばテストデータセットに関して査定され得、ここで、正しい予測は、データセット中の各入力について知られる。

## 【0004】

いくつかの使用事例、たとえば医療使用事例では、高い精度が必要とされる。分類方法の精度は、分類器モデル性能を改善することによって改善され得る。そのような改善は、たとえば新しいトレーニングデータセットに関する分類器モデルのさらなるトレーニングによって達成され得る。しかしながら、いくつかの場合には、たとえばさらなるトレーニングデータが容易に取得され得ない場合、分類器モデル性能に対するさらなる改善は可能でない。

## 【発明の概要】

## 【0005】

次に、非限定的な実施形態によるシステムおよび方法が、添付の図を参照しながら説明される。

## 【図面の簡単な説明】

## 【0006】

【図1】一実施形態による、分類の方法の概略図。

【図2(a)] データセットについての分類器モデル出力の概略図。

【図2(b)] 第1の例による方法の概略図。

【図3] 第1の例による方法の概略図。

【図4] 畳み込みニューラルネットワークの一例を示す図。

【図5] 第1の例による、トレーニングパイプラインと、トレーニングされた分類システムの展開とを示す概略図。

【図6] 第1の予測しきい値(predictive threshold)を示す概略図。

【図7] 第2の予測しきい値を示す概略図。

【図8(a)] 第1の例による、分類のためのシステムをトレーニングする方法の概略図。

【図8(b)] 第1の例による、分類のためのシステムをトレーニングする方法の概略図。

【図9] 完全な感度から完全な特異度への遷移の概略図。

【図10(a)] 第1の例の修正による、分類のためのシステムをトレーニングする方法の概略図。

【図10(b)] 例示的な最適化プロセスを示す図。

【図10(c)] しきい値組合せについて計算された精度値を示す図。

【図11(a)] 第2の例による、分類の方法を示す図。

【図11(b)] 入力データを含む方法の概略図。

10

20

30

40

50

【図 1 2 ( a )】第 2 の例による、分類システムをトレーニングする方法の概略図。

【図 1 2 ( b )】第 2 の例による、トレーニングパイプラインと、トレーニングされた分類システムの展開とを示す概略図。

【図 1 2 ( c )】第 2 の例の修正による、分類器をトレーニングする方法の概略図。

【図 1 3】第 3 の例による、分類の方法の概略図。

【図 1 4】分類の例示的な方法の概略図。

【図 1 5】例示的な分類システムの概略図。

【図 1 6】第 4 の例による、分類の方法を示す図。

【発明を実施するための形態】

【 0 0 0 7 】

第 1 の態様によれば、

第 1 のクラスに対応する第 1 の入力データを備える第 1 のデータセットを取得することと、

第 1 のクラスに対応する第 2 の入力データを備える第 2 のデータセットを取得することと、

第 1 のデータセットを使用して少なくとも 1 つの分類器をトレーニングすることと、

第 2 のデータセットからの第 2 の入力データを少なくとも 1 つの分類器に入力すること、および第 1 の分類と第 2 の分類とを備える分類モデルを提供することと、ここにおいて、第 1 の分類が、第 2 の分類よりも、第 1 のクラス中にあることになる第 1 のクラスに対応する第 2 の入力データのより大きい比率を予測する、  
を備える、コンピュータ実装方法が提供される。

【 0 0 0 8 】

一例では、第 2 の分類は、第 1 の分類よりも、第 2 のクラス中にあることになる第 2 のクラスに対応する第 2 の入力データのより大きい比率を予測する。

【 0 0 0 9 】

一例では、第 1 のクラスは陰性クラスであり、第 2 のクラスは陽性クラスである。

【 0 0 1 0 】

一例では、第 1 の分類は、第 1 の分類器を適用することと、第 1 の予測しきい値を使用することとを備え、第 2 の分類は、第 1 の分類器を適用することと、第 2 の予測しきい値を使用することとを備える。第 1 の予測しきい値は第 2 の予測しきい値よりも高い。

【 0 0 1 1 】

一例では、第 1 の分類器はバイナリ分類器である。第 1 の予測しきい値を決定することは、第 1 のクラスに対応する第 2 のデータセットからの入力についてバイナリ分類器の最も高い出力値を決定することと、第 1 の予測しきい値を、最も高い出力値よりも小さくまたはそれに等しく設定することとを備え得、第 2 の予測しきい値を決定することは、第 2 のクラスに対応する第 2 のデータセットからの入力について最も低い出力値を決定することと、第 2 の予測しきい値を、最も低い出力値よりも大きくまたはそれに等しく設定することとを備える。

【 0 0 1 2 】

一例では、少なくとも 1 つの分類器は少なくとも 2 つの分類器を備え、分類モデルを提供することは、少なくとも 2 つの分類器のうちの第 1 の分類器を決定することと、ここにおいて、第 1 の分類器が、少なくとも 2 つの分類器の、第 1 のクラス中にあることになる第 1 のクラスに対応する第 2 の入力データの最も大きい比率を予測する、少なくとも 2 つの分類器のうちの第 2 の分類器を決定することと、ここにおいて、第 2 の分類器が、少なくとも 2 つの分類器の、第 2 のクラス中にあることになる第 2 のクラスに対応する第 2 の入力データの最も大きい比率を予測する、を備える。

【 0 0 1 3 】

一例では、分類モデルを提供することは、第 1 の予測しきい値と第 2 の予測しきい値とを決定することとを備え、第 1 の予測しきい値と第 2 の予測しきい値とを決定することは、第 1 の予測しきい値についての値と第 2 の予測しきい値についての値との複数の可能な組

10

20

30

40

50

合せを取得することと、可能な組合せの各々を使用して分類モデルの精度を決定することと、所定の精度値よりも高い精度を有する組合せを選択することとを備える。

【0014】

一例では、第1のデータセットは、第1の入力データのための分類ラベルを備え、第2のデータセットは、第2の入力データのための分類ラベルを備える。

【0015】

一例では、第1の分類器と第2の分類器とは、異なるモデルアーキテクチャに対応し、ならびに／あるいは異なる入力データに基づいておおよび／または異なるハイパーパラメータを使用して更新されたパラメータに対応する。

【0016】

一例では、第1のデータセットは、第1の集団に対応する第1の入力データを備え、第2のデータセットは、第2の集団に対応する第2の入力データを備える。

【0017】

別の態様によれば、

第1のクラスに対応する第1の入力データを備える第1のデータセットを取得することと、

第1のクラスに対応する第2の入力データを備える第2のデータセットを取得することと、

第1のデータセットを使用して分類器をトレーニングすることと、

第2のデータセットからの第2の入力データを分類器に入力すること、おおよび第1の分類を備える分類モデルを提供することと、ここにおいて、第1の分類を備える分類モデルを提供することが、第1のクラスに対応する予測しきい値を決定することとを備える、を備える、コンピュータ実装方法が提供される。

【0018】

一例では、予測しきい値を決定することは、第1のクラスに対応する第2のデータセットからの入力について分類器の最も高い出力値を決定することと、予測しきい値を、最も高い出力値よりも小さくまたはそれに等しく設定することとを備える。

【0019】

一例では、予測しきい値を決定することは、第1のクラスに対応する第2のデータセットからの入力について最も低い出力値を決定することと、予測しきい値を、最も低い出力値よりも大きくまたはそれに等しく設定することとを備える。

【0020】

一例では、予測しきい値を決定することは、予測しきい値についての複数の可能な値を取得することと、可能な値の各々を使用して分類モデルについてのスコアを決定することと、所定のスコアよりも高いスコアを有する値を選択することとを備える。スコアは、第1のクラス中にあるものとして分類される、第1のクラスに対応しないデータの比率の測度であり得る。複数の可能な値は、第1のクラスに対応する第2のデータセットからの入力についての分類器の最も高い出力値と第1のクラスに対応する第2のデータセットからの入力についての分類器の最も低い出力値との間の値のセットを備え得る。

【0021】

別の態様によれば、

入力データを取得することと、

入力データを分類モデルに入力することと、分類モデルが第1の分類と第2の分類とを適用し、ここにおいて、第1の分類が、第2の分類よりも、第1のクラスに対応する入力データのより大きい比率を正しく予測するように構成される、

第1の分類および第2の分類の出力に基づいて入力データについての分類予測を決定することと

を備える、コンピュータ実装方法が提供される。

【0022】

別の態様によれば、

10

20

30

40

50

入力データを取得することと、

入力データを分類モデルに入力することと、分類モデルが第 1 の分類と第 2 の分類とを適用し、ここにおいて、第 1 の分類が第 1 の予測しきい値を使用し、第 2 の分類が第 2 の予測しきい値を使用し、ここにおいて、第 1 の予測しきい値が第 2 の予測しきい値よりも高い、

第 1 の分類および第 2 の分類の出力に基づいて入力データについての分類予測を決定することと

を備える、コンピュータ実装方法が提供される。

【0023】

一例では、第 1 の分類は、第 1 の分類器を適用することと、第 1 の予測しきい値を使用することとを備え、第 2 の分類は、第 1 の分類器を適用することと、第 2 の予測しきい値を使用することとを備え、ここにおいて、第 1 の予測しきい値は第 2 の予測しきい値よりも高い。

【0024】

一例では、第 1 の分類は第 1 の分類器を適用することとを備え、第 2 の分類は第 2 の分類器を適用することとを備える。

【0025】

一例では、入力データは組織の画像を備え、ここにおいて、分類を決定することは、医療診断に係る情報を決定することとを備える。

【0026】

一例では、第 1 の分類および第 2 の分類の出力に基づいて入力データについての分類予測を決定することは、

第 1 の分類の出力と第 2 の分類の出力とが同じである場合、第 1 の分類または第 2 の分類の出力である分類予測を出力することと、

分類予測がないと決定することとを備える。

【0027】

別の態様によれば、

入力データを取得することと、

入力データを分類モデルに入力することと、分類モデルが第 1 の分類を適用し、ここにおいて、第 1 の分類が、第 1 の分類器を適用することと、第 1 のクラスに対応する予測しきい値を使用することとを備える、

第 1 の分類の出力に基づいて第 1 のクラスについての分類予測を決定することとを備える、分類のコンピュータ実装方法が提供される。

【0028】

一例では、分類予測を決定することは、

第 1 の分類の出力が第 1 のクラスである場合、第 1 のクラスである分類予測を出力することと、

第 1 の分類の出力が第 1 のクラスでない場合、分類予測がないと決定することとを備える。

【0029】

別の態様によれば、上記の方法を実施するように構成された、1 つまたは複数のプロセッサを備える分類システムが提供される。

【0030】

別の態様によれば、上記の方法に従ってトレーニングされる分類モデルを備える分類システムが提供される。

【0031】

別の態様によれば、コンピュータに上記の方法を実施させるように構成されたコンピュータ可読コードを備えるキャリア媒体が提供される。それらの方法は、コンピュータ実装方法である。実施形態によるいくつかの方法がソフトウェアによって実装され得るので、

10

20

30

40

50

いくつかの実施形態は、任意の好適なキャリア媒体上の汎用コンピュータに提供されたコンピュータコードを包含する。キャリア媒体は、フロッピー（登録商標）ディスク、CDROM、磁気デバイスまたはプログラマブルメモリデバイスなどの任意の記憶媒体、あるいは、任意の信号、たとえば、電気信号、光信号またはマイクロ波信号などの任意の一時媒体を備えることができる。キャリア媒体は、非一時的コンピュータ可読記憶媒体を備え得る。

#### 【0032】

分類器モデルは、広範囲の分野における適用例を有する。たとえば、分類器モデルは、様々な医療適用例において使用され得る。分類器モデルは、たとえば生理的測定値を入力としてとることによって、医療診断を提供するために使用され得る。別の例では、分類器モデルは、たとえば、入力組織画像が癌性細胞を含んでいるかどうかを予測するために、または特定のバイオマーカーが組織画像に関連付けられるかどうかを予測するために、医療画像データを分析するために使用され得る。

#### 【0033】

そのような分類器モデルは、トレーニングデータセットを使用してトレーニングされ得る。たとえば、トレーニングデータセットは、ラベリングされた分類出力とともに、多数の例示的な入力を備え得る。トレーニングデータセットに関してモデル化された分類器モデルを使用する分類方法は、しかしながら、新しい入力データ上に展開されたとき、あるレベルの不正確さを有することになる。

#### 【0034】

たとえば、バイナリ分類器モデルは、各データポイントについて0から1の間のスコアを出力し、このスコアは、予測しきい値に基づいて分類（陽性または陰性）に変換される。スコアがしきい値を上回る場合、出力予測は陽性である。スコアがしきい値を下回る場合、出力予測は陰性である。しきい値は、たとえば、0.5であり得る。バイナリ分類器モデルは、たとえば、全スライド画像が癌性細胞を含んでいるかどうかを予測するために使用され得る。入力全スライド画像は、分類器モデルによって与えられるスコアが0.5よりも大きいかまたはそれに等しい場合のみ、癌性細胞を含んでいると予測される。

#### 【0035】

分類器モデル出力は、分類方法の出力（陽性または陰性）を生成するために予測しきい値と比較される。分類方法の精度は、総分類の数の比率としての正しい分類の数によって測定され得る。これは、たとえばテストデータセットに関して査定され得、ここで、正しい分類（陽性または陰性）は、データセット中の各入力について知られる。

#### 【0036】

いくつかの使用事例、たとえば医療使用事例では、高い精度が必要とされる。分類の精度は、分類器モデル性能を改善することによって改善され得、ここで、分類器モデル性能は、（以下でより詳細に説明される）AUCによって測定され得る。そのような改善は、たとえば新しいトレーニングデータセットに関する分類器モデルのさらなるトレーニングによって達成され得る。しかしながら、いくつかの場合には、分類器モデル性能に対するさらなる改善は可能でなく、たとえば、さらなるトレーニングデータが容易に取得され得ない。

#### 【0037】

図1は、一実施形態による、分類の方法の概略図を示す。

#### 【0038】

S101において、入力データが取得される。入力データは、たとえば、生理的測定値または入力医療画像であり得る。

#### 【0039】

S102において、第1の分類と第2の分類とが適用される。第1の分類と第2の分類とは、各々、陽性クラスまたは陰性クラスへの分類を出力する。陽性クラスは第2のクラスとも呼ばれ、陰性クラスは第1のクラスとも呼ばれる。陽性クラスは、第1の分類および第2の分類について同じである。たとえば、陽性クラスは、医療診断、たとえば病気の

10

20

30

40

50

診断であり得る。代替的に、陽性クラスは、たとえば組織の入力画像についての、癌性細胞の検出、または特定のバイオマーカーの指示であり得る。第1の分類は、第2の分類よりも多くの陰性事例を正しく予測するように構成される。第2の分類は、第1の分類よりも多くの陽性事例を正しく予測するように構成される。第2の分類は、第1の分類よりも高い真陽性率を有するように構成され、ここで、真陽性率は、方法が正しく陽性として予測する陽性事例の比率である。

【0040】

次いで、S103において、分類予測が、第1の分類および第2の分類の出力に基づいて決定される。特に、第1の分類の出力と第2の分類の出力とが同じである場合、分類予測は、第1の分類（または第2の分類）の出力として出力される。第1の分類の出力と第2の分類の出力とが異なる場合、分類予測は出力されない。次いで、何らかのさらなるステップ、たとえば手動分類が実施され得る。分類予測の精度は、予測がより不確実である事例が除外されるので、改善される。

10

【0041】

次に、分類の例示的な方法が説明される。

#### 第1の例

分類方法の性能の様々なインジケータがある。分類方法の感度は、真陽性率（TPR）であり、方法が正しく陽性として予測する陽性事例の比率である。分類方法の特異度は、真陰性率（TNR）であり、モデルが正しく陰性として予測する陰性事例の比率である。方法の特異度はまた、 $1 - FPR$ に等しく、ここで、FPRは偽陽性率である。FPRは、方法が間違っ

20

【0042】

て間違っ陽性として予測する陰性事例の比率である。偽陰性率（FNR）は、モデルが間違っ陰性として予測する陽性事例の比率であり、 $1 - TPR$ に等しい。

上記で説明されたように、バイナリ分類器モデルは、各データポイントについて0から1の間のスコアを出力し、このスコアは、予測しきい値に基づいて分類（陽性または陰性）に変換される。スコアがしきい値を上回る場合、出力分類は陽性である。予測スコアがしきい値を下回る場合、出力分類は陰性である。多くの分類器モデル、たとえば、深層ニューラルネットワーク、サポートベクターマシン、ランダムフォレストなどでは、分類方法の感度および特異度は、単に予測しきい値をシフトすることによって、分類器モデルをトレーニングした後に調整され得る。図2（a）は、データセットについての分類器モデル出力の概略図である。陽性クラスについての分類器モデル出力（0から1までの確率値）は、0.05の個別区間において、x軸に沿ったものである。データセットについての各分類器モデル出力の頻度は、y軸に対応する。真値が陽性（1）である事例に対応する出力は、より暗い陰影で示されているが、真値が陰性（0）である事例についての出力は、より明るい陰影で示されている。

30

【0043】

より高い予測しきい値が、より高い特異度をもつ分類方法を生じ、それは、より多くの陰性事例を正しく予測することができるが、より多くの陽性事例を正しく予測することができない。図2（a）中の線「特異度しきい値」は、データセット中のすべての真陰性事例が正しく予測されるしきい値を示す。

40

【0044】

より低い予測しきい値と組み合わせられた、同じ分類器モデルが、より高い感度をもつ分類方法を生じ、それは、より多くの陽性事例を正しく予測することができるが、より多くの陰性事例を正しく予測することができない。図2（a）中の線「感度しきい値」は、データセット中のすべての陽性事例が正しく予測されるしきい値を示す。「特異度」しきい値は「感度」しきい値よりも高い。

【0045】

分類器モデルとより高いしきい値とを使用する分類方法は、分類器モデルとより低いしきい値とを使用する分類方法よりも多くの陰性事例を正しく予測するように構成される。さらに、分類器モデルの感度しきい値に基づく分類方法は、分類器モデルの特異度しきい

50

値に基づく分類方法よりも多くの陽性事例を正しく予測するように構成される。

【0046】

図2(b)は、第1の例による方法を示す。組織の画像に対応する入力画像データに関して分類が実施される。入力画像データ41を含むプロセスの概略図が、図3に示されている。しかしながら、他の例では、たとえば、(非組織医療画像などの)他のタイプの画像、(年齢、性別、健康状態などの)患者データ、液体生検データ(たとえばDNA配列データ)、または血液テストデータ(たとえば血球の数)のうちの1つまたは複数を含む、他のタイプの入力データが使用され得ることを理解されよう。

【0047】

この例では、入力画像は、それが特定のバイオマーカー、ERバイオマーカーに対応するかどうかに関して分類される。入力画像は、それがERバイオマーカーに対応する場合、陽性クラスに分類される。しかしながら、他の例では、たとえば、画像が異なるバイオマーカーに、またはたとえば癌性細胞に対応するかどうかを分類するために、他のタイプの分類が実施され得ることを理解されよう。

【0048】

第1の例による方法は、第1の分類と第2の分類とを使用する。第1の分類と第2の分類とは同じ分類器を使用する。しかしながら、予測しきい値は、第2の分類についてよりも第1の分類について高い。したがって、第1の分類は、第2の分類よりも多くの陰性事例を正しく予測するように構成される。第2の分類は、第1の分類よりも多くの陽性事例を正しく予測するように構成され、言い換えれば、第2の分類は、第1の分類よりも高い真陽性率を有するように構成される。

【0049】

S201において、入力データが取得される。この例では、入力データは、複数のピクセルを備える入力画像データ41である。入力画像データ41は、この例では、図3に示されているように、組織の全スライド画像からの画像タイルである。以下の説明では、ピクセルデータは、(高さ×幅×3の寸法の)赤緑青であるが、ピクセルデータは、代替的に、たとえば(高さ×幅×1の寸法の)グレースケールであり得る。入力画像データは、第1の数のピクセルを備え、ここで、第1の数は、高さ×幅に等しい。画像データは、(組織学的切片とも呼ばれる)組織の画像をキャプチャする、顕微鏡に取り付けられたデジタルカメラを使用して収集され得る。

【0050】

本明細書で説明される特定の例では、入力画像データ41は、ヘマトキシリンおよびエオシン染色を用いて染色された組織学的切片の画像のタイルを備える。ヘマトキシリンおよびエオシン染色を用いて染色された組織学的切片の全スライド画像(WSI)の一例が図3(WSI)に示されている。グリッドがWSI上に重ねられ、ここで、各グリッド正方形がタイルに対応する。全スライド画像スキャナが、組織スライス全体を走査し、たとえば、約60000ピクセルの高さ×60000ピクセルの幅を備えるヘマトキシリンおよびエオシン染色を用いて染色された組織学的切片の画像を生じる。全スライド画像は、次いで、連続部分、またはタイルにスプリットされる。画像部分(タイル)は、固定の高さおよび幅を有する。それらの部分は、画像内で連続するか、または重複し得る。たとえば、画像タイルサイズは、512×512ピクセルであり得る。全スライド画像はタイルにスプリットされ、各タイルは、S201において、入力として別々にとられる。他のタイルサイズもちろん使用され得る。たとえば、2のべき乗に対応するタイルサイズ、たとえば、128×128、256×256、512×512、または1024×1024ピクセルが使用され得る。各WSI画像は異なるサイズのものであり得、したがって、入力WSIのサイズに応じて、異なる数の部分が入力WSIから抽出され得る。元の画像データは、入力、すなわち元のピクセル値としてとられ得る。代替的に、たとえば色正規化(normalisation)のために、何らかの前処理が元のピクセル値に対して実施され得る。細胞セグメンテーションステップが最初に実施され得、非癌組織のみを含んでいるタイルが破棄される。事前トレーニングされたモデルが、細胞セグメンテーションのために使

10

20

30

40

50

用され得る。

【0051】

S202において、入力画像データ41（画像タイル）は、分類モデルへの入力としてとられる。分類モデルは第1の分類器を備える。この例では、第1の分類器は、第1の畳み込みニューラルネットワーク（CNN）40である。例示的な第1の畳み込みニューラルネットワーク40が図4に示されている。

【0052】

第1の畳み込みニューラルネットワーク40は、少なくとも1つの畳み込み層を備える第1の部分46と、1次元ベクトルを入力としてとる第2の部分47、分類部分とを備える。第2の部分47は、少なくとも1つの全結合層を備える。第1のCNN40は、深度および視野の様々な層において画像に適用される畳み込みフィルタを備える第1の部分46と、その後続く、全結合層および／またはデータ低減のためのプーリング層を備える第2の部分47とを備える。フィルタ重みは、トレーニング段階中に学習されるトレーニング可能なパラメータである。より低いレベルのフィルタが、エッジおよびプロブなどの粗い構造を検出し、より深いレベルが、形状およびテクスチャのようなより複雑なプロパティをキャプチャし、最後に、最上層が、バイオマーカーの識別情報に関する関心オブジェクトに対して一般化するために学習する。

【0053】

第1のCNN40は、バイナリ分類器である。第1のCNN40は、入力画像タイルが特定の分子バイオマーカーに関連付けられるか否かを決定するために使用される。タイルは、第1のCNN40に提出される。ピクセルごとのデータが第1のCNN40に直接入力される。各タイルについて、CNNは、タイルが陽性クラスに割り当てられる、すなわち、タイルがバイオマーカーに関連付けられる、確率を出力する。

【0054】

図4に示されている第1のCNN40は、残差ネットワークアーキテクチャに基づく。残差ニューラルネットワークが、1つまたは複数のスキップ接続（skip connection）を備える。図は、簡単のために、少数の層を示すが、第1のCNN40は、たとえば、100個超の層を備え得る。

【0055】

CNN中の第1の層は、図では「畳み込み層1」とラベリングされた畳み込み層である。第1の層中の各フィルタは、入力データの深度に一致する深度を有する。たとえば、入力データがRGBである場合、第1の層中のフィルタ深度は3である。簡単のために、図4に示されているCNNは、1（すなわち、グレースケール入力データ）の入力データ深度を有する。

【0056】

第1の層の出力ボリュームは、いくつかのファクタによって決定される。

【0057】

第1の層の出力ボリュームの深度は、フィルタの数に対応する。たとえば、第1の層中に32個のフィルタがあり得、したがって、第1の層の出力は32の深度を有する。したがって、後続の層中のフィルタは、32の深度を有することになる。

【0058】

出力ボリュームの高さおよび幅は、入力の高さおよび幅と、フィルタの受容野サイズ（高さおよび幅の両方）と、フィルタストライドとによって決定される。ストライドが1であるとき、フィルタは、一度に1つのピクセルをスライドする。ストライドが2であるとき、フィルタは、一度に2つのピクセルをスライドし、より小さい出力ボリュームを生成する。境界において使用されるゼロパディングも出力サイズに影響を及ぼすことになる。

【0059】

各フィルタは、入力の幅および高さに沿って移動される。各位置において、入力値と重み値との要素ごとの乗算が行われ、その結果は、各位置についての単一の出力値を与えるために加算される。1つのフィルタについての出力値は、2Dアレイを形成する。層中の

10

20

30

40

50

すべてのフィルタからの出力アレイは、深度次元に沿って積み重ねられ、得られたボリュームが次の層に入力される。

【0060】

各畳み込み層の後に、活性化層が続く。活性化層は、要素ごとの活性化関数を適用し、サイズを不変のままにする。活性化層は、簡単のために図に示されていない。たとえば、モデルは、要素ごとの活性化関数を適用する1つまたは複数のReLU（正規化線形ユニット（rectified linear unit））層を備え得る。バッチ正規化層も、各畳み込み層の後に実装され、ここで、活性化層はバッチ正規化層の後に実装される。モデルは、畳み込み層と、バッチ正規化層と、活性化層とを備えるユニット、または、たとえば、第1の畳み込み層と、第1のバッチ正規化層と、第2の畳み込み層と、第2のバッチ正規化層と、活性化層とを備えるユニットを備える。

10

【0061】

第1のCNN40は、出力が入力よりも小さい寸法を有する、たとえば、高さおよび／または幅が層への入力よりも小さい、複数の層を備える。このようにして、深度が増加する間に、高さと幅とがいくつかの層を通して減少する。第1のCNN40は、「エンコーダ／デコーダ」構造を有し、それにより、層は、最初に、（たとえば、ストライドサイズなどのフィルタハイパーパラメータを介して）深度を増加させる間に高さと幅とを減少させ、次いで、（たとえば、プーリング層および／または双線形アップサンプリング層を介して）深度を減少させる間に高さと幅とを増加させる。これは、層の出力サイズを示す図4に示されている。

20

【0062】

第1のCNN40は、空間サイズを変動させるために含まれる1つまたは複数のプーリング層をさらに備える。プーリング層は、幅および／または高さを増加させ、出力の深度を減少させるために使用され得る。プーリング層は、「平均プーリング」層であり得る。平均プーリング層は、空間広がり（extent）およびストライドを有するフィルタを備え、これは、入力にわたって移動され、各位置において平均値をとる。しかしながら、平均以外の関数、たとえば、最大プーリングが使用され得る。アップサンプリング層、たとえば、1つまたは複数の双線形アップサンプリング層が、追加または代替として、高さおよび／または幅を増加させるために含まれ得る。

【0063】

第1のCNN40は、少なくとも1つのスキップ接続をさらに備える。実際には、モデルは複数のスキップ接続を備え得るが、簡単のために、少数の層と単一のスキップ接続とが図4に示されている。第2の層「畳み込み層2」は、出力mと呼ばれる出力を生成する。第4の層「畳み込み層4」は、出力mと同じ寸法を有する出力oを生成する。「畳み込み層5」への入力は、第1の層の出力mならびに第4の層の出力oから生成される。以前の層から後の層に直接出力を入力することが、「スキップ接続」である。この例における出力は、ピクセルごとの加算によって組み合わせられる。代替的に、連結が使用され得、ここで、出力は、たとえば、異なるサイズである。1つまたは複数のスキップ接続を使用して、ダウンストリームからの情報がアップストリームに直接供給される。これは、ネットワーク全体にわたって高レベルのグローバルな視覚特徴と地域の視覚特徴とを維持する。以前の層から後の層に直接特徴を入力し、1つまたは複数の介在層をスキップすることは、コンテキストを提供する。

30

40

【0064】

平坦化層が、最後の畳み込み層の後に含まれる。平坦化層は、最後の畳み込み層からの出力データを、次の層に入力するために1次元ベクトルxに変換する。この例における平坦化層より前の層は、CNNの第1の部分46を形成する。

【0065】

1つまたは複数の全結合層が平坦化層の後に含まれる。最後の全結合層が、陽性クラスに対応する値を出力する。確率値を与えるために、出力において活性化関数が、たとえば、シグモイド関数が適用される。活性化関数は、最後の全結合層からの出力を入力として

50

とり、確率に正規化する。したがって、活性化関数は、陽性クラスに対応する 1 から 0 の間の値を出力する。（1 つまたは複数の）全結合層と活性化関数とは、第 1 の CNN 40 の第 2 の部分 47 を形成する。

#### 【0066】

各タイルについて、CNN 40 は、タイルが陽性クラスに割り当てられる、すなわち、タイルがバイオマーカーに関連付けられる、確率を出力する。バイオマーカーは、特定の病理学的または生理的プロセス、病気、診断、療法または予後が識別され得る、自然発生分子、遺伝子、または特性である。本明細書で説明される特定の例では、バイオマーカーは、癌バイオマーカー、すなわち、特定のタイプの癌、または特に有効な癌治療が識別され得る、自然発生分子、遺伝子、または特性である。さらに、本明細書で説明される例では、バイオマーカーは、分子バイオマーカーである。この例では、特定のバイオマーカーは ER バイオマーカーである。しかしながら、他のバイオマーカーの例が、以下でより詳細に説明される。

10

#### 【0067】

S 203 において、第 1 のしきい値が第 1 の分類器 40 の出力に適用される。言い換えれば、第 1 の CNN 40 から出力された確率値は、第 1 のしきい値と比較される。図 3 に示されている例では、第 1 の分類器から出力された確率値は、0.764 である。第 1 のしきい値は、より高い予測しきい値 T1 である。この例では、第 1 のしきい値 T1 は 0.6 である。第 1 の分類器の出力が、より高い予測しきい値よりも高い場合、第 1 の分類出力は、陽性クラスに対応する 1 である。第 1 の分類器の出力が、より高い予測しきい値よりも低い場合、第 1 の分類出力は 0 である。図 6 は、分類器モデル予測スコアに対する頻度のプロット上に線 T1 によって示される第 1 の予測しきい値を示す概略図である。

20

#### 【0068】

S 203 において、第 2 のしきい値も第 1 の分類器の出力に適用される。言い換えれば、第 1 の CNN 40 から出力された確率値は、別々に、第 2 のしきい値と比較される。第 2 のしきい値は、より低い予測しきい値 T2 である。この例では、第 2 のしきい値 T2 は 0.45 である。第 1 の分類器の出力が、より低い予測しきい値よりも高い場合、第 2 の分類出力は、陽性クラスに対応する 1 である。第 2 の分類器の出力が、より低い予測しきい値よりも低い場合、第 2 の分類出力は 0 である。第 2 の予測しきい値は、図 7 中の線 T2 によって示される。

30

#### 【0069】

第 1 の分類についての予測しきい値 T1 は、第 2 の分類についての予測しきい値 T2 よりも高い。したがって、第 1 の分類は、第 2 の分類よりも多くの陰性事例を正しく予測するように構成される。第 2 の分類は、第 1 の分類よりも多くの陽性事例を正しく予測する。

#### 【0070】

第 1 の分類についての第 1 の予測しきい値 T1 は、第 2 の分類についての第 2 の予測しきい値 T2 よりも高い。しきい値の値は、システムの使用より前に実施されるトレーニング段階中に決定される。トレーニング段階は、以下で図 8 (a) に関して説明される。

#### 【0071】

S 204 において、最後の分類予測が、第 1 の分類出力および第 2 の分類出力に基づいて決定される。このステップにおいて、第 1 の分類が 1 を出力し、第 2 の分類が 1 を出力する場合、分類予測は 1 である。第 1 の分類が 0 を出力し、第 2 の分類が 0 を出力する場合、分類予測は 0 である。第 1 の分類と第 2 の分類とが異なる分類を出力する場合、分類予測は行われない。最後の分類を決定するために、何らかのさらなるステップが実施され得る。たとえば、画像は、専門家によって手動で分類され得る。

40

#### 【0072】

本方法では、2 つの異なる予測しきい値を使用することによって、2 つの異なる分類は、単一のトレーニングプロセスを用いて、単一のトレーニングされた分類器を使用して実施され得る。第 2 の分類 M2 は、より低い予測しきい値 T2 を使用し、したがって、高い

50

感度をもつ分類を提供し、より多くの信頼できる陽性予測を生成する。より低い予測しきい値は、ここでは陰性しきい値とも呼ばれる。第1の分類M1は、より高い予測しきい値T1を使用し、したがって、高い特異度をもつ分類を提供し、より多くの信頼できる陰性予測を生成する。より高い予測しきい値は、ここでは陽性しきい値とも呼ばれる。2つの分類を同時に展開することによって、0.5のしきい値とともに、第1の分類器を使用して単一の分類を適用することによってよりも高い程度の確実性が、分類予測において達成され得る。

【0073】

図8(a)は、第1の例による、分類のためのシステムをトレーニングする方法の概略図である。本方法は、上記で説明された第1の例による、分類方法において使用され得るシステムをトレーニングするために使用される。

10

【0074】

各々が複数の入力画像を備える、トレーニングデータセットと検証データセットとが、トレーニング方法において使用される。それらの画像は、分類方法中の使用のための意図されたタイプの入力画像に対応する。ここで説明される例では、入力画像は、上記で説明されたように、ヘマトキシリンおよびエオジン染色を用いて染色された組織学的切片の画像からの画像タイルである。したがって、トレーニングデータセットと検証データセットは、各々、ヘマトキシリンおよびエオジン染色を用いて染色された組織学的切片の画像からの画像タイルを備える。

【0075】

各画像タイルは、モデルが検出することになる特定のバイオマーカーに各画像タイルが対応するか否かに応じてラベリングされる。上記で説明されたように、特定のバイオマーカーは、この例では、ERバイオマーカーである。各画像タイルは、それがERバイオマーカーに対応する場合、1でラベリングされ、それがERバイオマーカーに対応しない場合、0でラベリングされる。

20

【0076】

ラベルを生成するために、たとえば、IHC染色プロセスからの情報が使用され得る。いくつかのデータセットについて、専門家は、IHC染色された画像を検討し、ターゲット画像のERステータスを、たとえば、それらがメタデータとしてすでに利用可能でない場合、決定し得る。これらは、次いで、トレーニング中に、対応するH&E画像のためのグラウンドトゥールスラベルとして使用される。遺伝的な、トランスクリプトーム的な、および／または免疫学的な分析の手段を通じた患者からのヒト試料の様々なテストが、追加または代替として使用され得る。これらのテストは、液体および／または固体の形態で生検材料(biopsy)と呼ばれるヒト試料に対して行われ、これは、次いで、試料の分子ステータスを通知するための処置を受ける。結果は、次いで、対応するH&E画像についてラベル1または0を生成するために、専門家によって、組織生検のための病理学者、液体生検のための血液学者、細胞学試料のための細胞病理学者、遺伝的な／トランスクリプトーム的な分析のための遺伝学者などによって、分析される。注釈は、トレーニングされた病理学者によって実施され得る。

30

【0077】

データセットを準備するために、入力画像に対して実施されることが意図されるものと同じ画像前処理が実施される。したがって、各全スライド画像について、複数の画像部分(タイル)が取得される。データセットから非癌組織のみを含んでいるタイルを破棄するために細胞セグメンテーションが使用され得る。

40

【0078】

画像注釈は、全スライド画像の各タイルについて別々に実施され得る。代替的に、注釈は、全スライド画像について実施され得、次いで、画像からの各タイルが、全スライド画像に与えられた注釈でラベリングされる。したがって、この場合、各タイルをそれらの対応するスライドのラベルとペアリングするステップが実施される。

【0079】

50

したがって、トレーニングデータセットと検証データセットとは、各々、複数の画像タイルを備え、各タイルは、モデルが検出することになる特定のバイオマーカーに各タイルが対応するか否かに応じてラベリングされる。

【0080】

S801において、トレーニングデータセット中の画像タイルは、第1の分類器、第1のCNN40に提出され、それは、推論（展開）段階に関して上記で説明されたものと同じ様式で、画像が陽性クラスに割り当てられる確率を生成する。

【0081】

第1のCNN40は、関連付けられたパラメータベクトル $\theta_1$ を有する。パラメータは、第1のCNN40の第1の部分中の畳み込み層のすべてのためのフィルタ重みならびに第1のCNN40の第2の部分のための重みを含む。S801におけるトレーニングプロセスの目的は、トレーニングデータセット中の画像の注釈と画像についての第1の分類器の出力との間の差が最小化されるようなパラメータベクトル $\theta_1'$ を見つけることである。

【0082】

最適なパラメータは、最初に、 $\theta_1$ としてランダム値を割り当て、次いで、損失の勾配

【0083】

【数1】

$$\frac{\partial D1}{\partial \theta_1}$$

10

20

【0084】

を算出することと算出された勾配を使用して $\theta_1$ を更新することによって $\theta_1$ を連続的に更新することによって、算出される。D1は、損失関数を表し、これは、「タイルごとの」損失である。この例では、クロスエントロピー損失が使用される。

【0085】

【数2】

$$-y_i \log(p_i) - (1 - y_i) \log(1 - p_i)$$

【0086】

ここで、 $y_i$ は、注釈値（1または0）であり、 $p_i$ は、トレーニングデータセット中の例示的な $i$ についての分類器出力値である。分類器モデルのトレーニング可能なパラメータの各々に関する損失の勾配が逆伝搬を通して決定される。勾配は、次いで、オプティマイザ関数を使用して、更新されたパラメータを決定するために使用される。更新方法のこのファミリーは、概して以下のように反復的に定義される勾配降下法（GD）として知られており、

【0087】

【数3】

$$\theta_1' = \theta_1 - \mu_1 \frac{\partial D1}{\partial \theta_1}$$

30

40

【0088】

ここで、 $\mu_1$ は、パラメータがどのくらい迅速に更新されるかを定義する学習レートである。更新は、バッチ平均に基づいて実施され得る。たとえば、8つのタイルのバッチサイズが使用され得る。8つのタイルにわたる平均値損失が計算され、その平均損失に基づいてモデルパラメータが更新される。したがって、分類器は、パラメータが更新される前に8つのタイルに適用される。

【0089】

Adam最適化アルゴリズムが使用され得る。しかしながら、選択される最適化戦略は、使用事例に対する各戦略の性能に依存し得る。たとえば、以下の最適化方法のうちの1

50

つが選択され得る。

- ・ 確率的勾配降下法 (SGD)
- ・ AdaDelta
- ・ Adam
- ・ AdaMax
- ・ Nesterov Adamオプティマイザ
- ・ RMSProp

トレーニングは、複数のエポックにわたって実施され得る。エポックの数は、分類器がトレーニング中にトレーニングデータセット全体に適用される回数である。各トレーニングエポック中に、トレーニングデータセット中のトレーニング例は、ランダムな順序でモデルに入力され得、したがって、トレーニング例は各エポックについて異なる順序で入力される。

10

【0090】

分類器モデル性能が、検証データセットを使用して各エポックの後に評価され得る。検証データセットからのデータの異なるサブセットが、各エポックの後に評価のために使用され得る。各トレーニングエポックの後に、0.5のしきい値に基づく分類器モデル精度が決定され得る。検証データセットからの各画像タイルは第1の分類器に提出され、第1の分類器は、画像が陽性クラスに割り当てられる確率を生成する。分類器モデル出力は、次いで、画像についての分類を与えるために0.5のしきい値と比較される。次いで、画像についての注釈値（ラベル）と比較することによって、分類が正しいかどうか決定される。精度は、検証データについて実施される分類の数で除算された正しい分類の数として決定される。精度値は、分類器モデル性能の指示である。分類器モデルが所望の性能を達成すると、トレーニングは停止され（すなわち、さらなるエポックは実施されず）、分類器モデルパラメータは固定される。所望の性能は、最小精度（たとえば0.8または0.9）を満たすことによって、または現在のエポックと前のエポックとの間の精度の変化が所定の値よりも小さいことによって決定され得る。精度を監視する代わりに、AUCまたは損失が、監視するために使用され、たとえば同様の様式で実施され得る。

20

【0091】

次いで、S802において、トレーニングされたモデルを仮定すれば、検証データセットを使用して、予測しきい値T1およびT2が決定される。このステップにおいて、検証データセットは大部分がテストセットと同様であることが仮定される。S802において、検証データセット中の画像タイルは、第1の分類器、第1のCNN40に提出され、それは、推論（展開）段階に関して上記で説明されたものと同じ様式で、画像が陽性クラスに割り当てられる確率を生成する。次いで、分類器モデル予測スコア、すなわち、分類器モデルから出力された確率値が、画像のラベルと比較される。この例では、検証データセット結果から、（陰性クラスにおける）ラベル0をもつ画像についての最も高い確率値が、第1の予測しきい値T1としてとられる。（陽性クラスにおける）ラベル1をもつ画像についての最も低い確率値が、第2の予測しきい値T2としてとられる。したがって、検証データセット上で第1の分類器モデルを動作させたことの結果から決定されるように、第1の予測しきい値T1は特異度しきい値に設定され、第2の予測しきい値T2は感度しきい値に設定される。第1の予測しきい値T1を特異度しきい値に設定し、第2の予測しきい値T2を感度しきい値に設定することによって、分類方法は、陽性クラスと陰性クラスの両方についての信頼できる予測を生成することができる。

30

40

【0092】

上記で図2（b）に関して説明されたように、第1の分類が第1のしきい値T1を使用して適用され、第2の分類が第2のしきい値T2を使用して適用される、完全分類方法の性能は、さらなるテストデータセットを使用して査定され得る。たとえば、完全分類方法の精度は、テストデータセットから決定され得る。

【0093】

この例では、第1の予測しきい値T1は特異度しきい値に設定され、第2の予測しきい

50

値  $T_2$  は感度しきい値に設定される。しかしながら、代替例では、第 1 の予測しきい値  $T_1$  は特異度しきい値よりも低く設定され、第 2 の予測しきい値  $T_2$  は感度しきい値よりも高く設定されるが、依然として、第 1 の予測しきい値  $T_1$  が第 2 の予測しきい値  $T_2$  よりも高いことを保証する。これは、分類方法を使用して予測され得る事例の数を増加させる。たとえば、不確実性またはエラーのコストを利益が上回る場合、そのような不確実性またはエラーの許容できるレベルが導入され得ることは、医療診断システムにおいて一般的である。感度／特異度コンテキストでは、これは、(M2 によって使用される) より低い予測しきい値  $T_2$  を、感度しきい値をわずかに上回る値に増加させ、(M1 によって使用される) より上の予測しきい値  $T_1$  を、特異度しきい値をわずかに下回って減少させることによって、達成され得る。これを行うことによって、分類方法が予測することができる事例の数が増加される。

10

#### 【0094】

比較的良好な性能をもつトレーニングされた分類器モデルを仮定すれば、予測しきい値は、上記で説明されたように S802 において決定される。検証セットが大部分がテストセットと同様であると仮定する、上記で説明された例では、両方のクラスについての信頼できる予測を生成するために、検証データセットについて、 $T_1$  は特異度しきい値に設定され、 $T_2$  は感度しきい値に設定される。感度しきい値に設定されたしきい値の場合、分類器モデルは、検証データセット上で 100% PPV を達成する。PPV は、陽性予測値であり、 $100 \times TPR / (TPR + FPR)$  によって与えられる。特異度しきい値に設定されたしきい値の場合、分類器モデルは、検証データセット上で 100% NPV を達成し、ここで、NPV は、陰性予測値 (NPV) であり、 $100 \times TNR / (FNR + TNR)$  によって与えられる。

20

#### 【0095】

図 5 は、第 1 の例による、トレーニングパイプラインと、トレーニングされた分類システムの展開とを示す概略図である。第 1 の分類器は、上記で説明されたように、トレーニングデータセットを使用して S801 においてトレーニングされる。第 1 の分類器は、次いで、検証データセット上で動作され、図 5 に示されているようにモデル予測スコアのセットを生成する。次いで、第 1 のしきい値  $T_1$  と第 2 のしきい値  $T_2$  とが、モデル予測スコアのセットに基づいて、S802 において選択される。第 1 のしきい値  $T_1$  と第 2 のしきい値  $T_2$  とは、自動化された様式で決定される。たとえば、第 1 の予測しきい値  $T_1$  は特異度しきい値に設定され、第 2 の予測しきい値  $T_2$  は感度しきい値に設定される。

30

#### 【0096】

次いで、S501 において、分類方法は、分類方法の精度を評価するために、テストデータセット上で実施される。S502 において、第 1 の分類と第 2 の分類とが合致するテスト事例のみに関して評価される、テストデータセットに関する分類精度が、意図された目的のために十分であるかどうか決定される。たとえば、意図された適用例に応じて、0.90 または 0.95 の精度が十分であり得る。精度が十分である場合、システムは、上記で図 2 (b) に関して説明されたように展開される。十分でない場合、システムは破棄される。

#### 【0097】

展開中に、入力画像が、S201 において受信され、S202 において第 1 の分類器によって分析される。第 1 の分類および第 2 の分類の予測が、S203 において出力される。予測が合致する場合、S204 において最後の分類予測が行われる。合致しない場合、手動テストなど、さらなるステップが実施され得る。

40

#### 【0098】

予測分類器モデルの性能は、以下のうちの一方または両方によって測定され得る。

- ・ 精度：特定の予測しきい値 (prediction threshold value) (たとえば 0.5) を用いて分類器モデルによって行われた、正しい予測の数／予測の数、および

- ・ AUC：すべての予測しきい値にわたる精度の測度として働く、受信者動作特性曲線 (ROC 曲線) 下面積。

50

## 【0099】

R O C 曲線は、変動するしきい値についての偽陽性率（F P R）に対する真陽性率（T P R）のプロットである。真陽性率は、感度としても知られており、分類器モデルがしきい値を用いて正しく陽性として予測する陽性事例の比率である。偽陽性率は、分類器モデルがしきい値を用いて間違っ陽性として予測する陰性事例の比率である。トレーニングされた分類器モデルの（0.5 など、特定のしきい値についての）精度および A U C は、テストデータセットを使用して評価され得る。医療使用事例では、臨床等級の医療システムは、たとえば、0.94～0.98 の精度と A U C とを有する分類器モデルを使用し得る。

## 【0100】

より低い精度と A U C とを有する分類器モデルは、意図された使用がいくらかの誤差マージンを考慮に入れるか、または、真の予測の利益が偽の予測のリスクおよび影響を上回るか、または、予測を確かめる人間がループ中にいる、適用例のために使用され得る。しかしながら、多くの使用事例では、そのような分類器モデルを有することは、たとえば、人間が予測を確かめなければならないので、デバイスが経済的に実現不可能であることを意味する。これは、臨床診療へのデバイスの採用を妨げることがある。分類器モデル性能は、たとえば新しいトレーニングデータセットに関する分類器モデルのさらなるトレーニングによって改善され得る。しかしながら、いくつかの場合には、分類器モデル性能に対するさらなる改善は可能でなく、たとえば、さらなるトレーニングデータが容易に取得され得ない。

## 【0101】

しかしながら、分類方法の感度および特異度は、分類器モデルとともに使用される予測しきい値を選択することによって調整され得る。図9は、R O C 曲線を右から左に横断することによって観測されるような、完全な感度（偽陰性なし）から完全な特異度（偽陽性なし）への遷移の概略図を示す。

## 【0102】

上記で説明された分類方法では、検証データセットに関して査定される、高いまたは完全な感度（少数の偽陰性または偽陰性なし）を有する第1の分類と、高いまたは完全な特異度（少数の偽陽性または偽陽性なし）を有する第2の分類とが、使用される。したがって、分類方法の精度が改善され得る。

## 【0103】

上記で説明された例では、第1の予測しきい値と第2の予測しきい値とは、検証データセットを使用して決定される。しかしながら、トレーニングセットと検証セットとテストセットとを使用する代替例では、S 8 0 1 においてトレーニングセットを使用してトレーニングが実施され、トレーニングの性能が検証セットを使用して検証され、次いで、テストセットを使用してしきい値が決定される。

## 【0104】

図8（a）に関して説明されたトレーニング方法では、トレーニングデータセットと検証データセットとは、同じ集団に対応する。しかしながら、代替実施形態では、トレーニングデータセットと検証データセットとは、異なる集団に対応し得る。たとえば、トレーニングデータセットは、大量のデータが利用可能である一般集団に対応し得る。検証データセットは、より小さい量のデータが利用可能であり、意図された使用事例の集団により密接に対応する、特定の集団に対応し得る。たとえば、トレーニングデータセットは、ある国または地域出身の人々に対応し得、検証データセットは、別の国または地域出身の人々に対応し得る。検証データセットは、特定の病院における患者に対応し得る。検証データセットは、年齢範囲など、特定の特色または特性をもつ患者に対応し得る。このようにして、第1の分類器は、大きいトレーニングデータセットを使用してトレーニングされ、言い換えれば、第1の分類器のパラメータは、大きいトレーニングデータセットを使用して更新される。検証データセットは、次いで、第1のしきい値と第2のしきい値とを決定するために使用される。これは、「較正」段階と見なされ得、第1のしきい値と第2のし

10

20

30

40

50

きい値とは、使用が意図される集団により近い集団に対応する検証データセットに基づいて決定される。

【0105】

図8(b)は、第1の例による、分類のためのシステムをトレーニングする方法の概略図であり、検証データセットは、トレーニングデータセットとは異なる集団に対応する。本方法は、上記で説明された第1の例による、分類方法において使用され得るシステムをトレーニングするために使用される。

【0106】

各々が複数の入力画像を備える、トレーニングデータセットと検証データセットとが、トレーニング方法において使用される。それらの画像は、分類方法中の使用のための意図されたタイプの入力画像に対応する。ここで説明される例では、入力画像は、上記で説明されたように、ヘマトキシリンおよびエオジン染色を用いて染色された組織学的切片の画像からの画像タイルである。したがって、トレーニングデータセットと検証データセットは、各々、ヘマトキシリンおよびエオジン染色を用いて染色された組織学的切片の画像からの画像タイルを備える。トレーニングデータセットは、第1の集団P1に対応する画像を備え、検証データセットは、第2の集団P2に対応する画像を備える。第2の集団P2はターゲット集団であり、言い換えれば、第2の集団P2は、第1の集団P1よりも、意図された使用事例のための集団により密接して対応する。たとえば、トレーニングデータセットは、すべての年齢の患者からの組織の画像を備え得るが、検証データセットは、特定の年齢範囲における患者からの組織の画像を備える。

【0107】

各画像タイルは、上記で説明されたように、モデルが検出することになる特定のバイオマーカーに各画像タイルが対応するか否かに応じてラベリングされる。上記で説明されたように、特定のバイオマーカーは、この例では、ERバイオマーカーである。各画像タイルは、それがERバイオマーカーに対応する場合、1でラベリングされ、それがERバイオマーカーに対応しない場合、0でラベリングされる。したがって、トレーニングデータセットと検証データセットとは、各々、複数の画像タイルを備え、各タイルは、検出されることになる特定のバイオマーカーに各タイルが対応するか否かに応じてラベリングされる。

【0108】

S801において、トレーニングデータセット中の画像タイルは、第1の分類器、第1のCNN40に提出され、それは、画像が陽性クラスに割り当てられる確率を生成する。上記で説明されたように、トレーニングデータセット中の画像の注釈と画像についての第1の分類器の出力との間の差が最小化されるように、更新されたパラメータベクトル $\theta_1$ が決定される。トレーニングされていない第1の分類器は、第1の集団P1に対応するトレーニングデータセットを使用してトレーニングされる。S801の出力は、第1の集団P1に関してトレーニングされる、トレーニングされた分類器である。

【0109】

上記で説明されたように、各トレーニングエポックの後に、0.5のしきい値に基づく分類器精度が決定され得る。検証データセットからの各画像タイルは第1の分類器に提出され、第1の分類器は、画像が陽性クラスに割り当てられる確率を生成する。分類器出力は、次いで、画像についての分類を与えるために0.5のしきい値と比較される。結果は、モデル予測スコアを示す、図では「結果1」とラベリングされた第1のグラフに概略的に示されている。0.5を上回るスコアのすべてが陽性として予測され、0.5を下回るスコアのすべてが陰性として予測される。真の分類も第1のグラフに示されており、より暗い陰影が、真値が陽性であることを示す。

【0110】

画像についての注釈値（ラベル）と比較することによって、分類が正しいかどうか決定される。精度は、検証データについて実施される分類の数で除算された正しい分類の数として決定される。精度値は、分類器モデル性能の指示である。分類器モデルが所望の性

能を達成すると、トレーニングは停止され（すなわち、さらなるエポックは実施されず）、分類器パラメータは固定される。

【0111】

次いで、S802において、トレーニングされたモデルを仮定すれば、検証データセットを使用して、予測しきい値T1およびT2が決定される。検証データセットは、第2の集団P2、ターゲット集団に対応する。したがって、S802は較正プロセスと見なされ得る。トレーニングされたモデルは入力としてとられ、予測しきい値が決定される。

【0112】

前に説明されたように、S802において、検証データセット中の画像タイルは、第1の分類器に提出され、それは、推論（展開）段階に関して上記で説明されたものと同じ様式で、画像が陽性クラスに割り当てられる確率を生成する。次いで、分類器モデル予測スコア、すなわち、分類器モデルから出力された確率値が、画像のラベルと比較される。分類器予測スコアは、グラフ「結果2」に示されている。この例では、検証データセット結果から、（陰性クラスにおける）ラベル0をもつ画像についての最も高い確率値が、第1の予測しきい値T1としてとられる。（陽性クラスにおける）ラベル1をもつ画像についての最も低い確率値が、第2の予測しきい値T2としてとられる。したがって、検証データセット上で第1の分類器モデルを動作させたことの結果から決定されるように、第1の予測しきい値T1は特異度しきい値に設定され、第2の予測しきい値T2は感度しきい値に設定される。第1の予測しきい値T1を特異度しきい値に設定し、第2の予測しきい値T2を感度しきい値に設定することによって、分類方法は、陽性クラスと陰性クラスの両方についての信頼できる予測を生成することができる。

【0113】

上記で図2（b）に関して説明されたように、第1の分類が第1のしきい値T1を使用して適用され、第2の分類が第2のしきい値T2を使用して適用される、完全分類方法の性能は、さらなるテストデータセットを使用して査定され得る。たとえば、完全分類方法の精度は、テストデータセットから決定され得る。この例におけるテストデータセットも、第2の集団P2に対応する。

【0114】

この例では、第1の予測しきい値T1は特異度しきい値に設定され、第2の予測しきい値T2は感度しきい値に設定される。しかしながら、代替例では、第1の予測しきい値T1は特異度しきい値よりも低く設定され、第2の予測しきい値T2は感度しきい値よりも高く設定されるが、依然として、第1の予測しきい値T1が第2の予測しきい値T2よりも高いことを保証する。

【0115】

S802における較正のステップは、モデルが、たとえばターゲット病院集団に基づいて調整されること、言い換えれば集団シフトについて調整されることを可能にする。これは、集団変化から生じる予測スコアの分布変化について調整すべき予測しきい値T1およびT2を決定することによって行われる。集団変化について調整すべきモデルの重みを調整する中間ステップも含まれ得る。たとえば、第1のトレーニングデータセットを使用して実施されるトレーニングプロセスの後にさらなるトレーニングプロセスを実施するために、第2の集団P2に対応する第2のトレーニングデータセットが使用され得る。このさらなるトレーニングプロセスでは、第1の分類器のパラメータは、ターゲット集団に対応するデータに基づいてさらに調整される。次いで、第1のしきい値と第2のしきい値とが、上記で説明されたようにS802において決定される。

【0116】

上記で説明された例では、第1の分類器がトレーニングされ、次いで、ターゲット集団に従って、第1の予測しきい値と第2の予測しきい値とが決定される。較正段階S802中に、入力画像は、第1の分類器への入力としてとられる。予測の前に最後の層から出力された値が、予測スコアとして出力される。したがって、ターゲット集団予測スコア分布が生成される。次いで、第1のしきい値と第2のしきい値とが決定される。

## 【 0 1 1 7 】

表 1 は、第 1 の例による、分類の第 1 の方法を使用して生成された結果を示し、それは、テストデータセットからの画像を E R バイオマーカーに対応するものとして分類する。方法は、第 1 および第 2 のしきい値とともに、上記で説明された分類器モデルを使用する。表 1 は、比較例による、分類の第 1 の方法を使用して生成された結果をも示し、それは、テストデータセットからの画像を E R バイオマーカーに対応するものとして分類する。方法は、0.5 の単一のしきい値とともに、上記で説明されたものと同じ分類器モデルを使用する。

## 【 0 1 1 8 】

表 1 は、第 1 の例による、分類の第 2 の方法を使用して生成された結果をも示し、それは、テストデータセットからの画像を P R バイオマーカーに対応するものとして分類する。方法は、第 1 および第 2 のしきい値とともに、上記で説明された分類器モデルを使用する。表 1 は、比較例による、分類の第 2 の方法を使用して生成された結果をも示し、それは、テストデータセットからの画像を P R バイオマーカーに対応するものとして分類する。方法は、0.5 の単一のしきい値とともに、上記で説明されたものと同じ分類器モデルを使用する。

## 【 0 1 1 9 】

## 【 表 1 】

|     | 比較例           |               | 第1の例          |               |
|-----|---------------|---------------|---------------|---------------|
|     | ERバイオ<br>マーカー | PRバイオ<br>マーカー | ERバイオ<br>マーカー | PRバイオ<br>マーカー |
| 精度  | 0.82234957    | 0.768361582   | 0.870090634   | 0.826086957   |
| F1  | 0.89084507    | 0.836653386   | 0.926243568   | 0.889655172   |
| AUC | 0.825141878   | 0.756788921   | 0.823986358   | 0.721239607   |
| FPR | 0.484848485   | 0.47008547    | 0.689655172   | 0.530612245   |
| FNR | 0.106007067   | 0.113924051   | 0.010989011   | 0.044444444   |
| PPV | 0.887719298   | 0.79245283    | 0.870967742   | 0.832258065   |
| NPV | 0.53125       | 0.696629213   | 0.857142857   | 0.793103448   |
| 感度  | 0.893992933   | 0.886075949   | 0.989010989   | 0.955555556   |
| 特異度 | 0.515151515   | 0.52991453    | 0.310344828   | 0.469387755   |
| PLR | 1.843860424   | 1.884925201   | 1.434065934   | 1.800854701   |
| NLR | 0.205778424   | 0.214985708   | 0.035409035   | 0.09468599    |

表 1

## 【 0 1 2 0 】

上記の表 1 では、P L R は陽性尤度比であり、N L R は陰性尤度比である。

## 第 1 の例の修正

図 10 (a) は、第 1 の例の修正による、分類のためのシステムをトレーニングする方法の概略図であり、より低い予測しきい値  $T_2$  は、感度しきい値を上回る値に設定され得、より上の予測しきい値  $T_1$  は、特異度しきい値を下回る値に設定され得る。これを行うことによって、モデルが予測することができる事例の数が、増加され得る。

【0121】

S1001において、第1の分類器40は、上記でS801に関して説明されたものと同じ様式で、トレーニングデータセットを使用してトレーニングされる。

【0122】

S1002において、感度しきい値と特異度しきい値とは、上記でS802に関して説明されたものと同じ様式で、検証データセットを使用して決定される。

【0123】

S1003において、第1のしきい値と第2のしきい値とは、最適化プロセスを使用して決定される。第1の例の修正による方法では、第1のしきい値および第2のしきい値をそれぞれ特異度および感度しきい値として設定する代わりに、第1のしきい値と第2のしきい値とを決定するために最適化プロセスが実施される。

【0124】

次に、S1003において実施され得る例示的な最適化プロセスが、図10(b)に関して説明される。

【0125】

S901において、可能な値の第1のセットが第1のしきい値について取得され、可能な値の第2のセットが第2のしきい値について取得される。第1のセットおよび第2のセット中の値のすべてが、感度しきい値よりも大きいかまたはそれに等しく、特異度しきい値よりも小さいかまたはそれに等しい。可能な値の第1のセットは、特異度しきい値を含むが、感度しきい値を含まない、感度しきい値と特異度しきい値との間の固定数の均等に離間した値をとることによって取得される。可能な値の第2のセットは、感度しきい値を含むが、特異度しきい値を含まない、感度しきい値と特異度しきい値との間の固定数の均等に離間した値をとることによって取得される。

【0126】

この例では、特異度しきい値は0.65であり、感度しきい値は0.4である。0.025の増分における均等に離間した値のセットが、値の第1のセットおよび値の第2のセットとしてとられる。値の第1のセットは、値{0.425、0.450、0.475、0.500、0.525、0.550、0.575、0.600、0.625、0.650}を備える。値の第2のセットは、値{0.400、0.425、0.450、0.475、0.500、0.525、0.550、0.575、0.600、0.625}を備える。

【0127】

S902において、可能な第1のしきい値と可能な第2のしきい値との複数の組合せの各々について、精度値が計算される。精度は、テストデータセットから決定される。このステップにおいて、第1のセットからの各値が、第2のセットからの各より低い値と組み合わせられる。テストデータセットからの入力、第1の分類器を通して動作され、第1の分類器からの出力確率値は、分類予測を与えるために、組合せにおける第1のしきい値および第2のしきい値と比較される。次いで、組合せについての精度が、分類予測の総数で除算された正しい分類予測の数として決定される。これは、複数の組合せの各々について繰り返される。

【0128】

このステップの結果は、本例では、図10(c)に示されている。第1のセットからの各値{0.425、0.450、0.475、0.500、0.525、0.550、0.575、0.600、0.625、0.650}が、グリッドの上部に沿って示されている。第2のセットからの各値{0.400、0.425、0.450、0.475、0

10

20

30

40

50

． 5 0 0、0． 5 2 5、0． 5 5 0、0． 5 7 5、0． 6 0 0、0． 6 2 5} が、グリッドの左側に沿って示されている。グリッド中の残りのセルは、（可能な組合せに対応するセルの場合）精度値を示すか、または、（第 2 のセットからの値が第 1 のセットからの値よりも低い組合せに対応するセルの場合）空である。第 1 のセットからの各値は、第 2 のセットからの各より低い値と組み合わせられ、対応する精度値を有する組合せ、すなわち、（0． 4 2 5、0． 4 0 0）、（0． 4 5 0、0． 4 0 0）、（0． 4 5 0、0． 4 2 5）、．．．（0． 6 5 0、0． 6 0 0）、（0． 6 5 2、0． 6 0 0）を生じる。各組合せについて、第 1 のしきい値は第 2 のしきい値よりも高い。図 1 0（c）は、それらの組合せの各々について計算された精度値を示す。

【0 1 2 9】

10

S 9 0 3 において、意図された適用例についての所望の精度よりも大きいかまたはそれに等しい精度スコアを生じる、第 1 のしきい値と第 2 のしきい値との組合せのすべてが取り出される。所望の精度は、所定の精度値である。この例では、所望の精度は 0． 9 である。この例では、組合せ（0． 5 7 5、0． 4 0 0）、（0． 6 0 0、0． 4 0 0）、（0． 6 0 0、0． 4 2 5）、（0． 6 2 5、0． 4 0 0）、（0． 6 2 5、0． 4 2 5）、（0． 6 2 5、0． 4 5 0）、（0． 6 2 5、0． 4 7 5）、（0． 6 5 0、0． 4 0 0）、（0． 6 5 0、0． 4 2 5）、（0． 6 5 0、0． 4 5 0）、（0． 6 5 0、0． 4 7 5）は、0． 9 よりも大きい精度を生じる。したがって、これらの組合せは、S 9 0 3 において取り出される。

【0 1 3 0】

20

S 9 0 4 において、取り出された組合せからの 1 つの組合せが選択される。組合せは、カバレッジに基づいて選択される。たとえば、テストデータセットからの最も大きい数の結果についての分類を予測した、取り出された組合せのうちの組合せが、選択される。代替的に、第 1 のしきい値と第 2 のしきい値との間の最も大きい差を有する組合せが、選択される。この例では、組合せ（0． 6 2 5、0． 4 7 5）が、最も大きいカバレッジを有するものとして選択される。したがって、S 1 0 0 3 において、第 1 のしきい値は 0． 6 2 5 に設定され、第 2 のしきい値は 0． 4 7 5 に設定される。

第 2 の例

図 1 1（a）は、第 2 の例による、分類の方法を示す。図 1 1（a）に関して説明される方法では、分類方法は、第 1 の例に関して説明されたものと同じ入力画像データに対して実施される。図 1 1（b）は、入力データを含む方法の概略図を示す。

30

【0 1 3 1】

第 2 の例による方法も、第 1 の分類と第 2 の分類とを使用する。しかしながら、第 1 の分類と第 2 の分類とは、異なる分類器を使用する。第 1 の分類は、第 2 の分類のために使用される第 2 の分類器よりも多くの陰性事例を正しく予測するように構成された第 1 の分類器を使用する。第 2 の分類器は、第 1 の分類器よりも高い真陽性率を有するように構成される。第 2 の実施形態では、2 つの別個の分類器がトレーニングされ、そのうちの一方が高い特異度と低い感度とを有し、他方が高い感度と低い特異度とを有する。

【0 1 3 2】

S 1 1 0 1 において、入力データが、上記で S 2 0 1 に関して説明されたものと同じ様式で取得される。

40

【0 1 3 3】

S 1 1 0 2 において、入力画像データ 4 1（画像タイル）が、分類モデルへの入力としてとられる。分類モデルは、第 1 の分類器と第 2 の分類器とを備える。したがって、入力画像データ 4 1（画像タイル）は、第 1 の分類器への入力としてとられ、別々に、第 2 の分類器への入力としてもとられる。この例では、第 1 の分類器は、上記で図 4 に関して説明されたように、畳み込みニューラルネットワーク（CNN）4 0 である。第 2 の分類器も、上記で図 4 に関して説明されたように、畳み込みニューラルネットワーク（CNN）4 0 である。この例では、第 1 の分類器と第 2 の分類器とは、同じモデルアーキテクチャを有するが、異なるトレーニングデータセットを使用してトレーニングされ、したがって

50

、異なるパラメータ値を有し得る。上記で図 4 に関して説明された、分類器の構造は、同じである。しかしながら、代替実施形態では、第 1 の分類器と第 2 の分類器とは、異なるモデルアーキテクチャを有する。たとえば、第 1 の分類器と第 2 の分類器とは、異なる数の層、層中の異なる数のフィルタ、または異なるスキップ接続を有し得る。

【0134】

S1103において、第 1 のしきい値が第 1 の分類器 40 a の出力に適用される。言い換えれば、第 1 の CNN 40 a から出力された確率値は、第 1 のしきい値と比較される。図 11 (b) に示されている例では、第 1 の分類器 40 a から出力された確率値は、0.764 である。この例では、第 1 の分類器 40 a とともに使用される第 1 の予測しきい値は、検証データセットから決定された、第 1 の分類器 40 a についての特異度しきい値である。第 1 の分類器の出力が、第 1 の予測しきい値よりも高い場合、第 1 の分類出力は 1 である。第 1 の分類器の出力が、第 1 の予測しきい値よりも低い場合、第 1 の分類出力は 0 である。

10

【0135】

S1103において、第 2 のしきい値が第 2 の分類器 40 b の出力に適用される。言い換えれば、第 2 の CNN 40 b から出力された確率値は、別々に、第 2 のしきい値と比較される。この例では、第 2 の分類器 40 b とともに使用される第 2 の予測しきい値は、検証データセットから決定された、第 2 の分類器 40 b についての特異度しきい値である。第 2 の分類器 40 b の出力が、第 2 の予測しきい値よりも高い場合、第 2 の分類出力は 1 である。第 2 の分類器の出力が、第 2 の予測しきい値よりも低い場合、第 2 の分類出力は 0 である。

20

【0136】

この例では、第 1 の分類器 40 a とともに使用される第 1 の予測しきい値は、検証データセットから決定された、第 1 の分類器 40 a についての特異度しきい値であり、第 2 の分類器 40 b とともに使用される第 2 の予測しきい値は、検証データセットから決定された、第 2 の分類器 40 b についての特異度しきい値である。これらのしきい値がどのように決定されるかは、以下で図 12 に関して説明される。しかしながら、代替例では、第 1 のしきい値と第 2 のしきい値とは、同じ所定の値に設定される。たとえば、第 1 のしきい値と第 2 のしきい値とは、0.5 に設定され得る。

【0137】

30

第 1 の分類は、第 2 の分類よりも多くの陰性事例を正しく予測するように構成される。第 2 の分類は、第 1 の分類よりも多くの陽性事例を正しく予測するように構成される。より多くの陰性事例を正しく予測するように構成された、第 1 の分類のために使用される第 1 の分類器 40 a は、展開より前に実施されるトレーニング段階中に取得される。より多くの陽性事例を正しく予測するように構成された、第 2 の分類のために使用される第 2 の分類器 40 b も、トレーニング段階中に取得される。第 2 の分類器 40 b は、第 1 の分類器よりも高い真陽性率を有するように選択される。

【0138】

S1104において、分類予測が、第 1 の分類出力および第 2 の分類出力に基づいて決定される。このステップにおいて、第 1 の分類が 1 を出力し、第 2 の分類が 1 を出力する場合、分類予測は 1 である。第 1 の分類が 0 を出力し、第 2 の分類が 0 を出力する場合、分類予測は 0 である。第 1 の分類と第 2 の分類とが異なる分類を出力する場合、分類予測は行われない。次いで、最後の分類を決定するために、何らかのさらなるステップが実施され得る。たとえば、画像は、専門家によって手動で分類され得る。

40

【0139】

図 12 (a) は、第 2 の例による、分類システムをトレーニングする方法の概略図である。本方法は、上記で説明された第 2 の例による、分類方法において使用され得る分類システムをトレーニングするために使用される。

【0140】

各々が複数の画像を備える、トレーニングデータセットと検証データセットとが、使用

50

される。それらの画像は、意図されたタイプの入力画像に対応する。前に説明されたように、トレーニングデータセットと検証データセットとが同じ集団に対応し得るか、または、トレーニングデータセットが第1の集団に対応し得、検証データセットが第2の、ターゲット集団に対応し得る。

【0141】

ここで説明される例では、入力画像は、ヘマトキシリンおよびエオジン染色を用いて染色された組織学的切片の画像からの画像タイルである。したがって、トレーニングデータセットと検証データセットは、各々、ヘマトキシリンおよびエオジン染色を用いて染色された組織学的切片の画像からの画像タイルを備える。第1の例に関して説明されたように、各画像タイルはまた、モデルが検出することになる特定のバイオマーカーに各画像タイルが対応するか否かに応じてラベリングされる。

10

【0142】

ここで説明される例では、トレーニングデータセットは、5つの別個のトレーニングデータサブセットにスプリットされる。トレーニングデータセット中の各画像は、5つの別個のトレーニングデータサブセットのうちの1つにランダムに割り振られる。したがって、各トレーニングデータサブセットは、異なる画像を備える。各サブセットは、同じ、または実質的に同じ数の画像を有する。

【0143】

S1201において、5つの別個の分類器A、B、C、DおよびEがトレーニングされる。各分類器は、同じモデルアーキテクチャを有するが、異なるトレーニングデータサブセットを使用する。

20

【0144】

第1のトレーニングデータサブセットからの画像タイルは分類器Aに提出され、分類器Aは、陽性クラスに割り当てられる確率を生成する。分類器Aは、関連付けられたパラメータベクトル $\theta_A$ を有する。最適なパラメータは、 $\theta_A$ としてランダム値を割り当て、次いで、損失の勾配

【0145】

【数4】

$$\frac{\partial DA}{\partial \theta_A}$$

30

【0146】

を算出することと算出された勾配を使用して $\theta_A$ を更新することによって $\theta_A$ を連続的に更新することによって、算出される。DAは、損失関数を表し、これは、「タイルごとの」損失である。前に説明されたように、クロスエントロピー損失が使用される。第1の分類器のトレーニング可能なパラメータの各々に関する損失の勾配が逆伝搬を通して決定される。勾配は、次いで、前に説明されたように、オプティマイザ関数を使用して、更新されたパラメータを決定するために使用される。

【0147】

第2のトレーニングデータサブセットからの画像タイルは分類器Bに提出され、分類器Bは、陽性クラスに割り当てられる確率を生成する。分類器Bは、関連付けられたパラメータベクトル $\theta_B$ を有する。最適なパラメータは、 $\theta_B$ としてランダム値を割り当て、次いで、損失の勾配

40

【0148】

【数5】

$$\frac{\partial DB}{\partial \theta_B}$$

【0149】

を算出することと算出された勾配を使用して $\theta_B$ を更新することによって $\theta_B$ を連続的

50

に更新することによって、算出される。この例では、各分類器をトレーニングするために、同じ形式の損失関数が使用される。モデルのトレーニング可能なパラメータの各々に関する損失の勾配が逆伝播を通して決定される。勾配は、次いで、前に説明されたように、オプティマイザ関数を使用して、更新されたパラメータを決定するために使用される。この例では、各分類器について同じ学習レート  $\mu$  が使用される。

【0150】

第3、第4および第5のトレーニングデータサブセットは、同様の様式で、それぞれ分類器C、DおよびEをトレーニングするために使用される。したがって、この例では、5つの分類器間の唯一の差は、S1201におけるトレーニングプロセスのために使用されるトレーニングデータである。

【0151】

各分類器について、モデル性能は、前に説明されたように、検証データセットからのデータを使用して評価される。トレーニングが所望の性能を達成すると、モデルパラメータは、各分類器について固定される。

【0152】

S1202において、検証データセットからの同じデータが、5つの分類器の各々に入力される。次いで、予測スコア、すなわち、分類器から出力された確率値が、画像ラベルと比較される。次いで、各分類器について、特異度しきい値と感度しきい値とが計算される。各分類器について、検証データセット結果は、（陰性クラスにおける）ラベル0をもつ画像についての最も高い出力確率値を決定するために使用され、これは特異度しきい値である。各分類器について、（陽性クラスにおける）ラベル1をもつ画像についての最も低い出力確率値が、感度しきい値として決定される。最も高い特異度しきい値に対応する分類器が第1の分類器として選択され、第1の分類器は、他の分類器よりも多くの陰性事例を正しく予測することができる。最も低い感度しきい値をもつ分類器が第2の分類器として選択され、第2の分類器は、他の分類器よりも多くの陽性事例を正しく予測することができる。第1の分類器とともに使用されることになる第1の予測しきい値は、選択された分類器の特異度しきい値に設定される。第2の分類器とともに使用されることになる第2の予測しきい値は、選択された分類器の感度しきい値に設定される。この第2の例では、2つのモデル、陰性事例を予測するための第1のモデルと陽性事例を予測するための第2のモデルとが選択される。次いで、特異度しきい値が第1のモデルについて選択され、感度しきい値が第2のモデルについて選択される。この手法は、いくつかのモデルは陰性のみを予測し、陽性を予測しないことが極めて得意であるが、他のモデルは陽性を予測し、陰性を予測しないことが極めて得意である、適用例のために使用され得る。

【0153】

次いで、上記で図11(a)に関して説明されたように、第1の分類が第1の分類器を使用して適用され、第2の分類が第2の分類器を使用して適用される、完全分類の性能が、テストデータセットを使用して査定され得る。

【0154】

上記の例では、5つの別個の分類器がトレーニングされ、最も高い特異度しきい値に対応する分類器が第1の分類器として選択され、最も低い感度しきい値をもつ分類器が第2の分類器として選択される。しかしながら、代替例では、2つの分類器のみがトレーニングされ、より高い特異度しきい値に対応する分類器が第1の分類器として選択され、他方の分類器が第2の分類器として選択される。代替例では、3つ以上の分類器がトレーニングされ、最も高い特異度しきい値に対応する分類器が第1の分類器として選択され、最も低い感度しきい値をもつ分類器が第2の分類器として選択される。

【0155】

上記の例では、最も高い特異度しきい値に対応する分類器が第1の分類器として選択され、最も低い感度しきい値をもつ分類器が第2の分類器として選択される。しかしながら、代替的に、0.5のしきい値を使用する、最も高い特異度に対応する分類器が、第1の分類器として選択され、0.5のしきい値を使用する、最も高い感度をもつ分類器が、第

10

20

30

40

50

2 の分類器として選択される。次いで、第 1 の予測しきい値と第 2 の予測しきい値とが、両方とも 0.5 に設定される。

【0156】

上記の例では、複数の分類器の各々は、トレーニングデータセットの別個のサブセットを用いてトレーニングされ、ここで、各画像は、サブセットのうちの 1 つにランダムに割り振られる。しかしながら、代替例では、画像は、画像の 1 つまたは複数のプロパティに基づいて割り振られ得る。たとえば、サブセットのうちの 1 つまたは複数は、他のものよりも多くの陽性例を有し得る。たとえば、画像は、サブセットのうちの 2 つが 80 % 陽性例を有し、サブセットのうちの 1 つが 50 % 陽性例を有し、サブセットのうちの 2 つが 20 % 陽性例を有するように割り振られ得る。上記の例では、サブセットは実質的に同じ数の画像を有するが、代替的に、サブセットのうちのいくつかは、他のものよりもはるかに多い画像を有し得る。

10

【0157】

上記の例では、複数の分類器の各々は、同じモデルアーキテクチャを有する。しかしながら、代替例では、分類器の一部または全部が、異なるモデルアーキテクチャを有することができ、たとえば、分類器は、異なる数の層、層中の異なる数のフィルタ、または異なるスキップ接続を有し得る。

【0158】

上記の例では、複数の分類器の各々は、同じハイパーパラメータを使用してトレーニングされる。しかしながら、代替例では、分類器の一部または全部が、異なるハイパーパラメータを使用してトレーニングされ得、たとえば、学習レートは異なり得る。

20

【0159】

図 12 (b) は、第 2 の例による、トレーニングパイプラインと、トレーニングされた分類システムの展開とを示す概略図である。2 つの分類器は、上記で説明されたように、トレーニングデータセットを使用して S1201 においてトレーニングされる。分類器は、次いで、検証データセット上で動作され、図 12 (b) に示されているように各分類器についてモデル予測スコアのセットを生成する。次いで、第 1 の分類器と第 2 の分類器とが、モデル予測スコアのセットに基づいて選択される。第 1 の分類器と第 2 の分類器とは、自動化された様式で決定される。たとえば、検証データセット結果に基づいて決定されるように、第 1 の分類器は、最も高い特異度しきい値に対応する分類器として選択され、第 2 の分類器は、最も低い感度しきい値をもつ分類器として選択される。

30

【0160】

次いで、S501 において、完全分類方法は、方法の精度を評価するために、テストデータセット上で実施される。S502 において、それが確実であると主張する（すなわち、第 1 の分類と第 2 の分類とが合致するかどうかにかかわらず）テスト事例のみに関して評価される、テストデータセットに関する精度が、意図された目的のために十分であるかどうか決定される。精度が十分である場合、システムは、上記で図 11 (a) に関して説明されたように展開される。十分でない場合、システムは破棄される。

【0161】

展開中に、入力画像が、S1101 において受信され、S1102 において第 1 の分類器および第 2 の分類器によって分析される。第 1 の分類および第 2 の分類の予測が、S1103 において出力される。予測が合致する場合、S1104 において最後の分類予測が行われる。合致しない場合、手動テストなど、さらなるステップが実施され得る。

40

第 2 の例の修正

図 12 (c) は、第 2 の例の修正による、分類器をトレーニングする方法の概略図である。

【0162】

S1401 において、複数の分類器が、上記で S1201 に関して説明されたものと同じ様式でトレーニングされる。

【0163】

50

S 1 4 0 2において、第1の分類器と第2の分類器とが、上記でS 1 2 0 2に関して説明されたものと同じ様式で選択される。S 1 2 0 2に関して説明されたように、このステップの一部として、第2の分類器の感度しきいと第1の分類器の特異度しきい値とが、検証データセットを使用して決定される。

【0164】

上記で説明された第2の例では、第1の分類器40aとともに使用される第1の予測しきい値は、検証データセットから決定された、第1の分類器40aについての特異度しきい値として設定され、第2の分類器40bとともに使用される第2の予測しきい値は、検証データセットから決定された、第2の分類器40bについての感度しきい値として設定される。しかしながら、第2の例の修正による方法では、第1のしきい値と第2のしきい値とを決定するために最適化プロセスが実施される。

10

【0165】

S 1 4 0 3において、第1のしきい値と第2のしきい値とは、最適化プロセスを使用して決定される。上記で図10(b)に関して説明された例示的な最適化プロセスは、このステップにおいて実施され得る。

【0166】

可能な値の第1のセットが第1のしきい値について取得され、可能な値の第2のセットが第2のしきい値について取得される。この例では、第1のセット中の値のすべては、第1の分類器の感度しきい値よりも大きいかまたはそれに等しく、第1の分類器の特異度しきい値よりも小さいかまたはそれに等しいが、第2のセット中の値のすべては、第2の分類器の感度しきい値よりも大きいかまたはそれに等しく、第2の分類器の特異度しきい値よりも小さいかまたはそれに等しい。たとえば、可能な値の第1のセットは、特異度しきい値と感度しきい値とを含む、第1の分類器についての感度しきい値から第1の分類器についての特異度しきい値までの固定数の均等に離間した値をとることによって取得され得る。可能な値の第2のセットは、感度しきい値と特異度しきい値とを含む、第2の分類器の感度しきい値から第2の分類器の特異度しきい値までの固定数の均等に離間した値をとることによって取得され得る。

20

【0167】

次いで、可能な第1のしきい値と可能な第2のしきい値との複数の組合せの各々について、精度が決定される。このステップにおいて、第1のセットからの各値が、第2のセットからの各値と組み合わせられる。テストデータセットからの入力、第1の分類器および第2の分類器を通して動作される。分類予測を与えるために、第1の分類器モデル出力は、組合せにおける第1のしきい値と比較され、第2の分類器モデル出力は、組合せにおける第2のしきい値と比較される。次いで、精度は、組合せの各々について決定される。

30

【0168】

適用例についての所望の精度よりも大きいかまたはそれに等しい精度スコアを生じる、第1のしきい値と第2のしきい値との組合せのすべてが取り出され、取り出された組合せからの組合せが、カバレッジに基づいて選択される。たとえば、取り出された組合せの中から、テストデータセットからの最も多くの結果を分類した組合せが、選択される。

40

第3の例

第1の例では、分類方法はタイルごとに適用される。画像タイルが、全スライド画像からとられ、第1の分類器への入力としてとられる。第1のしきい値を使用する第1の分類と第2のしきい値を使用する第2の分類とが適用される。次いで、画像タイルがERバイオマーカーに対応するかどうかに関する分類予測が決定される。

【0169】

図13は、第3の例による、分類の方法の概略図であり、分類方法は画像ごとに適用される。言い換えれば、全スライド画像がERバイオマーカーに対応するかどうかに関する分類予測が取得される。

【0170】

S 1 5 0 1において、入力データが取得される。入力画像データは、この例では、前に

50

説明されたように、組織の全スライド画像からの画像タイルである。この例では、図 13 に関して示されているように、全スライド画像から複数の画像タイルが取得される。非癌組織のみを含んでいるタイルを破棄するために細胞セグメンテーションが使用され得る。

【0171】

S1502において、各入力画像タイルが、第1の分類器を備える分類モデルへの入力として別々にとられ、第1の分類器は、この例では、第1の例に関して説明されたように、第1のCNN40である。第1の分類器は、各画像タイルについて、その画像タイルがERバイオマーカーに関連付けられる、確率を出力する。

【0172】

次いで、アグリゲーションステップが実施される。この例では、アグリゲーションステップは、プーリング演算子を適用することを備える。最も高い確率値に対応する画像タイルが識別される。この画像タイルは、「トップランクの」画像タイルと呼ばれる。この「トップランクの」画像タイルは、次いで、分類方法の後続のステップにおいて画像を分類するために使用される。

【0173】

S1503において、第1のしきい値が、トップランクの画像タイルについて第1の分類器40の出力に適用される。言い換えれば、トップランクの画像タイルについて第1のCNN40から出力された確率値は、第1の分類の出力を与えるために、第1のしきい値と比較される。第2のしきい値が、トップランクの画像タイルについて第1の分類器40の出力に適用される。言い換えれば、トップランクの画像タイルについて第1のCNN40から出力された確率値は、第2の分類を与えるために、別々に、第2のしきい値と比較される。第1のしきい値と第2のしきい値とは、第1の例に関して前に説明されたように、トレーニング段階中に決定される。

【0174】

S1504において、全スライド画像についての最後の分類予測が、第1の分類出力および第2の分類出力に基づいて決定される。このステップにおいて、第1の分類が1を出力し、第2の分類が1を出力する場合、分類予測は1である。第1の分類が0を出力し、第2の分類が0を出力する場合、分類予測は0である。第1の分類と第2の分類とが異なる分類を出力する場合、分類予測は行われない。最後の分類を決定するために、何らかのさらなるステップが実施され得る。たとえば、画像は、専門家によって手動で分類され得る。

【0175】

第3の例では、展開中に、第1のCNN分類器40は、タイルごとに確率を返し、これらの個々のスコアは、最大値演算子を通してアグリゲートされる。平均化など、他のトレーニング不可能なアグリゲーション関数が、代替的に使用され得る。第1の分類器40は、第1の例の場合と同じ様式でトレーニングされる。全スライド画像についての分類予測は、単一の「代表的な」タイルについての分類予測としてとられる。

【0176】

代替例では、トレーニング可能なアグリゲーション関数が使用される。入力全スライド画像がバイオマーカーに関連付けられるかどうかの指示が、複数のタイルからのデータをアグリゲーションモデルに入力することによって決定される。アグリゲーションモデルは、長短期記憶(LSTM)ネットワークなど、リカレントニューラルネットワークを備え得る。第1の分類器40の出力は、画像タイルの低減されたセット、たとえば、第1の分類器40から出力された確率に基づく上位k個のタイルを選択するために使用され得、ここで、kは2よりも大きい整数である。画像タイルの低減されたセットは、分類モデルへの入力としてとられる。最初に、第1のトレーニングされたモデルを使用して、画像タイルの各々に対応する特徴ベクトルが生成される。これらの特徴ベクトルは、次いで、たとえばLSTMを備え得る第2のトレーニングされたモデルへの、シーケンス中の入力としてとられる。シーケンス中の最後の入力に対応するLSTMの出力は、全結合ニューラルネットワーク層に入力され、全結合ニューラルネットワーク層は、入力画像が(陽性クラ

10

20

30

40

50

スに対応する) E R バイオマーカーに関連付けられる確率を出力する。この出力値は、次いで、第 1 の分類の出力を与えるために第 1 のしきい値と比較され、第 2 の分類の出力を与えるために第 2 のしきい値と比較される。次いで、全スライド画像についての分類予測が、第 1 の分類および第 2 の分類の出力に基づいて決定される。トレーニング中に、トレーニングデータセットからの画像タイルは、推論中と同じ方法で、分類モデルへの入力としてとられる。出力は、たとえば前に説明されたようにクロスエントロピー損失関数とオプティマイザ関数とを使用して、第 1 のモデルと第 2 のモデルとをトレーニングするために、全スライド画像のためのラベルと比較される。代替例では、分類モデルは、第 1 のしきい値を用いた、第 1 のトレーニングされたモデルと第 2 のトレーニングされたモデルとを備える第 1 の分類と、第 2 のしきい値を用いた、異なる第 1 のトレーニングされたモデルと異なる第 2 のトレーニングされたモデルとを備える第 2 の分類とを適用することを備え得る。

10

#### 第 4 の例

上記で説明された例では、すべてのクラスについて、予測カバレッジ、すなわち、分類モデルが予測することができる事例の数の、低減を伴う、改善された精度が取得される。しかしながら、いくつかの適用例では、分類の方法は、クラスのうちのいくつかのみ、たとえば陰性クラスのみまたは陽性クラスのみを予測するために使用される。たとえば、分類の方法は、画像が H E R 2 バイオマーカーに対応するかどうかを分類するために使用され得、ここで、モデルは、陰性クラスを予測するためにのみ使用される。

20

【0177】

そのような例では、分類の方法は、単一の分類器と単一の予測しきい値とを使用し得る。いくつかのクラスについて、改善された精度が取得される。たとえば、分類器が第 1 のクラスを予測するためにのみ使用され、ここで、第 1 のクラスについて、改善された精度が取得される。第 1 のクラスに対応するしきい値が、本方法において使用される。分類が第 1 のクラスを予測する場合、第 1 のクラスの分類予測が行われる。分類予測が第 1 のクラスを予測しない場合、分類予測は行われない。

【0178】

前に説明されたように、より高い予測しきい値が、より高い特異度をもつ分類方法を生じ、それは、より多くの陰性事例を正しく予測することができるが、より多くの陽性事例を正しく予測することができない。たとえば、図 2 (a) 中の線「特異度しきい値」は、データセット中のすべての真陰性事例が正しく予測されるしきい値を示す。より低い予測しきい値と組み合わせられた、同じ分類器が、より高い感度をもつ分類方法を生じ、それは、より多くの陽性事例を正しく予測することができるが、より多くの陰性事例を正しく予測することができない。図 2 (a) 中の線「感度しきい値」は、データセット中のすべての陽性事例が正しく予測されるしきい値を示す。「特異度」しきい値は「感度」しきい値よりも高い。

30

【0179】

図 16 は、第 4 の例による方法を示す。第 1 の例に関して説明されたように、組織の画像に対応する入力画像データに関して分類が実施される。入力画像は、それが H E R 2 バイオマーカーに対応しない場合、陰性クラスに分類される。

40

【0180】

第 4 の例による方法は、第 1 の分類を使用する。

【0181】

S 1601 において、入力データが、第 1 の例に関して S 201 において説明されたものと同じ様式で取得される。

【0182】

S 1602 において、入力データが、分類モデルへの入力としてとられる。分類モデルは第 1 の分類器を適用する。第 1 の分類器は、第 1 の例において S 202 に関して説明されたように、C N N であり得る。

【0183】

50

S 1 6 0 3において、しきい値が第1の分類器の出力に適用される。言い換えれば、第1の分類器から出力された確率値は、しきい値と比較される。しきい値は、この例では、陰性クラスについてのしきい値に対応する。たとえば、しきい値は、データセットを使用して決定された感度しきい値に対応し得る。

【0184】

S 1 6 0 4において、陰性クラスについての最後の分類予測が、S 1 6 0 3から出力された第1の分類に基づいて決定される。このステップにおいて、第1の分類が0を出力する場合、分類予測は0である。第1の分類が1を出力する場合、分類予測は行われない。最後の分類を決定するために、何らかのさらなるステップが実施され得る。たとえば、画像は、専門家によって手動で分類され得る。

10

【0185】

この例では、第1の分類は、陰性クラスを予測するために使用される。代替例では、第1の分類は、陽性クラスを予測するために使用される。この場合、より高い予測しきい値が陽性クラスについてのしきい値に対応するように、より高い予測しきい値が使用される。たとえば、しきい値は、データセットを使用して決定された特異度しきい値に対応し得る。

【0186】

図17は、第4の例による、分類のためのシステムをトレーニングする方法の概略図である。本方法は、上記で説明された第1の例による、分類方法において使用され得るシステムをトレーニングするために使用される。

20

【0187】

各々が複数の入力データを備える、トレーニングデータセットと検証データセットとが、トレーニング方法において使用される。前に説明されたように、入力データは、分類方法中の使用のための意図されたタイプの入力データに対応する。前に説明されたように、トレーニングデータセットは第1の集団に対応し得、検証データセットはターゲット集団に対応し得る。この例では、第1の例に関して説明されたように、入力データは、モデルが検出することになる特定のバイオマーカーに組織の画像が対応するか否かに応じてラベリングされる、組織の画像を備える。上記で説明されたように、特定のバイオマーカーは、この例では、HER2バイオマーカーである。各画像タイルは、それがバイオマーカーに対応する場合、1でラベリングされ、それがバイオマーカーに対応しない場合、0でラベリングされる。

30

【0188】

S 1 7 0 1において、トレーニングデータセット中の画像タイルは、第1の分類器、第1のCNN40に提出され、それは、画像が陽性クラスに割り当てられる確率を生成する。上記で説明されたように、トレーニングデータセット中の画像の注釈と画像についての第1の分類器の出力との間の差が最小化されるように、更新されたパラメータベクトル $\theta_{1'}$ が決定される。

【0189】

次いで、S 1 7 0 2において、トレーニングされたモデルを仮定すれば、検証データセットを使用して、予測しきい値が決定される。検証データセット中の画像タイルは、第1の分類器、第1のCNN40に提出され、それは、推論（展開）段階に関して上記で説明されたものと同じ様式で、画像が陽性クラスに割り当てられる確率を生成する。次いで、分類器モデル予測スコア、すなわち、分類器モデルから出力された確率値が、画像のラベルと比較される。この例では、検証データセット結果から、（陽性クラスにおける）ラベル1をもつ画像についての最も低い確率値が、予測しきい値としてとられる。したがって、検証データセット上で第1の分類器モデルを動作させたことの結果から決定されるように、予測しきい値は感度しきい値に設定される。予測しきい値を感度しきい値に設定することによって、分類方法は、陰性クラスについての正確な予測、言い換えれば、偽陰性率を生成することができ、すなわち、モデルが間違っ

40

50

## 【0190】

上記で説明された例では、予測しきい値は、検証データセットから決定された感度しきい値に設定される。しかしながら、いくつかの例では、予測しきい値は、感度しきい値よりも高い値に設定される。たとえば、予測しきい値は、最適化プロセスを使用して決定され得る。しきい値について、可能な値のセットが取得される。この例では、セット中の値のすべては、第1の例の修正に関して前に説明されたように、第1の分類器の感度しきい値よりも大きいかまたはそれに等しく、第1の分類器の特異度しきい値よりも小さいかまたはそれに等しい。

## 【0191】

次いで、各可能なしきい値についてスコアが決定される。スコアは、たとえばFNRまたはNPVであり得る。テストデータセットからの入力、第1の分類器を通して動作される。第1の分類器モデル出力は、分類予測を与えるためにしきい値と比較される。次いで、たとえば予測のセットについてのFNRを決定することによって、しきい値についてスコアが決定される。これは、各可能なしきい値について繰り返される。

10

## 【0192】

適用例についての所望のスコアよりも大きいかまたはそれに等しいスコアを生じるしきい値のすべてが取り出され、取り出されたしきい値からのしきい値が、カバレッジに基づいて選択される。たとえば、取り出されたしきい値の中から、テストデータセットからの最も多くの結果を分類したしきい値が選択されるか、または、最も高いしきい値が選択される。

20

## 【0193】

陽性クラスを予測するために第1の分類が使用される例では、予測しきい値は、偽陽性率を低減するために、検証データセットから決定された特異度しきい値に設定され得る。代替的に、しきい値は、上記で説明されたものと同様の様式で、最適化プロセスによって設定され得る。

## 【0194】

第4の例では、分類の方法は、単一の分類器を使用し、ターゲット集団に従って、たとえばターゲット集団に対応する検証データセットを使用して、しきい値を調整し得る。その例は、1つのしきい値をもつ1つの分類器を使用し、ここで、分類の方法は、たとえば陰性クラスのみまたは陽性クラスのみを予測することが意図される。いくつかの例では、分類器はトレーニングセットに関してトレーニングされ、トレーニングの性能は検証セットに関して検証され、次いで、しきい値はテストセットに関して設定される。

30

## 【0195】

表2は、第4の例による、分類の第1の方法を使用して生成された結果を示し、それは、テストデータセットからの画像をHER2バイオマーカーに対応しないものとして分類する。方法は、単一のしきい値が陰性クラスに対応する、上記で説明された分類器モデルを使用する。表2は、比較例による、分類の第1の方法を使用して生成された結果をも示し、それは、テストデータセットからの画像をHER2バイオマーカーに対応するものとして分類する。方法は、0.5の単一のしきい値とともに、上記で説明されたものと同じ分類器モデルを使用する。

40

## 【0196】

【表 2】

|     | 比較例         | 第4の例        |
|-----|-------------|-------------|
| 精度  | 0.647482014 | 0.867521368 |
| F1  | 0.319444444 | 0.114285714 |
| AUC | 0.612370114 | 0.558557127 |
| FPR | 0.331914894 | 0.009852217 |
| FNR | 0.465116279 | 0.935483871 |
| PPV | 0.227722772 | 0.5         |
| NPV | 0.88700565  | 0.873913043 |
| 感度  | 0.534883721 | 0.064516129 |
| 特異度 | 0.668085106 | 0.990147783 |
| PLR | 1.611508646 | 6.548387097 |
| NLR | 0.696193157 | 0.944792168 |

表2

【0197】

概して癌および／または患者の特定の分子プロファイルを理解することによって、中でも、ホルモン療法、免疫療法または標的薬治療を含む癌に対して行われる様々な処置が通知され得る。中でも、突然変異ステータス、受容体ステータス、またはコピー数多型を含む診断、治療または予後マーカーのいずれかを含む様々な医療的に関連するバイオマーカーは、患者がいくつかの療法にどのくらいよく応答することになるかを決定するために識別され得る。突然変異ステータス、受容体ステータス、またはコピー数多型は、分子バイオマーカーの例である。たとえば、いくつかの場合には、分子バイオマーカーは、タンパク質発現レベルであり得る。たとえば、特定のバイオマーカーは、エストロゲン受容体（ER）、プロゲステロン受容体（PR）またはヒト上皮成長因子受容体（HER2）であり得る。これらのピラーバイオマーカーは、乳癌に特異的である。それらは、乳癌の予後のための最も重要なバイオマーカーであり、標的療法に基づく。ERおよびHER2は、最も一般的に、それぞれ癌治療のタモキシフェンおよびハーセプチンに関連付けられる。患者は、これらの治療のための適合性を決定するためにこれらの2つのバイオマーカーについてテストされ得る。

【0198】

上記で説明された例では、入力画像がERバイオマーカーに関連付けられるかどうかに関して、入力画像が分類される。代替的に、方法は、入力画像タイルが、HER2バイオマーカーに関連付けられるのか、PRバイオマーカーに関連付けられるのか、（肺腺癌に関連付けられる）EGFRバイオマーカーに関連付けられるのか、（結腸腺癌に関連付け

られる) M S Iに関連付けられるのかを決定するために使用され得る。様々な分子バイオマーカーは、いくつかの癌を、胸部または結腸直腸などのカテゴリーに分類するために使用され得る。たとえば乳癌は、各々が E R、P R および H E R 2 のステータスに基づいて決定される、5つの異なる分子「サブタイプ」を有する。たとえば、E R、P R および H E R 2 がすべて陰性である場合、分子サブタイプは、「基底細胞様 (basal-like)」である。したがって、複数の分子バイオマーカーの存在または不在を決定することによって、分子サブタイプが予測され得る。「分子サブタイプ」は、存在または不在に基づいて、あるいは、いくつかの場合には、1つのバイオマーカーまたはバイオマーカーのセットのレベルに基づいて、特定のタイプの癌をカテゴリー分類する方法である。

#### 【0199】

分類方法は、様々な他のバイオマーカーを検出するために使用され得る。たとえば、抗原 K i - 6 7 も、癌の攻撃性を示す細胞増殖のためのマーカーとしてますますテストされている。したがって、特定のバイオマーカーは、代替的に、K i - 6 7 であり得る。たとえば、H E R 2 - および E R + 乳癌の腫瘍増殖を査定するときの等級方式における核分裂像数の代替として、K i 6 7 核抗原の I H C 染色に基づくラベリングインデックスが、他の I H C マーカーとともに使用され得る。それは、アジュバント化学療法のための任意の要件など、治療的な決定のための追加情報を提供し得る。様々な研究では、それは、生存の強力な予測子であることが示された。たとえば、P R E D I C T は、初期浸潤乳癌のための異なる治療が手術後の生存率をどのように改善し得るかを示すオンラインツールである。P R E D I C T モデル性能は、予後マーカーとしての K i 6 7 の関与とともに改善された。I H C 染色された K i 6 7 スライドを解釈するための手動スコアリング方法は、腫瘍の外周においてなど、ランダムに選択された関心領域中の浸潤細胞をカウントすることと、すべての浸潤腫瘍細胞に関する K i 6 7 染色の割合を決定することを含む。上記で説明された従来の分子プロファイリング技法と同様に、このプロセスは、労働集約型であり、人的エラーを受けやすく、観測者間／観測者内にオープンである。たとえば、H & E 画像から K i 6 7 インデックスを予測することによって、そのようなプロセスは、より短縮され、精度が潜在的に改善され得る。

#### 【0200】

分類方法は、癌の診断、治療および／または予後について関連する特定のバイオマーカーの自動プロファイリングを提供することができる。特定のバイオマーカーは、他の例の中でも、突然変異ステータス、受容体ステータスまたはコピー数多型であり得る。プロファイリングは、説明された例では全スライド H & E 画像から実施されるが、他の画像または入力が使用され得る。

#### 【0201】

バイオマーカーは、分子であるか、または、たとえば、特定の分子の量など、1つまたは複数の分子のうちの1つに関連付けられた特性であり得る。いくつかの場合には、バイオマーカーは、特定の癌治療に関連付けられた分子である。バイオマーカーは、臨床的に実行可能な遺伝子変異であり得る。この例では、特定のバイオマーカーは E R バイオマーカーである。他のバイオマーカーの例が、以下でより詳細に説明される。

#### 【0202】

代替的に、分類予測は、他のプロパティに基づき得る。いくつかの例では、分類方法は、たとえば画像が癌性細胞を含んでいるかどうかを分類する。

#### 【0203】

上記で説明された例では、ヘマトキシリンおよびエオシン染色を用いて染色された組織学的切片の画像が使用されるが、様々な方法を使用して取得された様々なタイプの組織画像が、説明された方法を使用して処理され得る。たとえば、代替的に、免疫組織化学 (I H C) 染色を受けた組織学的切片の画像が、入力としてとられ得る。I H C 染色は、組織切片の細胞中の抗原を選択的に識別することに関与する。抗体は、生物組織中の抗原に特異的に結合する。染色は、抗体と抗原の相互作用の視覚化を可能にする。たとえば、色素生産性免疫組織化学 (C I H) を使用して、抗体は、発色反応を触媒することができる酵

10

20

30

40

50

素に活用される。

【0204】

上記で説明された例では、バイナリ分類方法が実施される。しかしながら、代替例では、分類の方法は、多方向分類のために使用され得る。多方向分類では、分類モデルは、複数のクラスの各クラスについて第1の分類および第2の分類を適用し、ここで、第2の分類は、そのクラスについて、第1の分類よりも高い真陽性率を有するように構成される。検証データセットからの結果を仮定すれば、クラスについての真陽性率は、クラス中にあるものとして方法が正しく予測するクラス中の事例の比率として決定される。

【0205】

次に、3つのクラスへの分類が実施される一例が説明される。3方向分類器モデルが入力に適用される。第1のクラスについての第1の分類は、第1のクラスについての出力を第1の予測しきい値と比較することを備える。第1のクラスについての第2の分類は、第1のクラスについての出力を第2の予測しきい値と比較することを備える。第1の予測しきい値は第2の予測しきい値よりも高い。第2のクラスについての第1の分類は、第2のクラスについての出力を第3の予測しきい値と比較することを備える。第2のクラスについての第2の分類は、第2のクラスについての出力を第4の予測しきい値と比較することを備える。第3の予測しきい値は第4の予測しきい値よりも高い。第3のクラスについての第1の分類は、第3のクラスについての出力を第5の予測しきい値と比較することを備える。第3のクラスについての第2の分類は、第3のクラスについての出力を第6の予測しきい値と比較することを備える。第5の予測しきい値は第6の予測しきい値よりも高い。第1の予測しきい値は、第1のクラス中にある（すなわち、第2または第3のクラス中にある）入力について、検証データセット結果からの第1のクラスについての最も高い出力確率値として設定される。第2の予測しきい値は、第1のクラス中の入力について、検証データセット結果からの第1のクラスについての最も低い確率値として設定される。したがって、検証データセット上で分類器を動作させたことの結果から決定されるように、第1の予測しきい値T1は、第1のクラスについての特異度しきい値に設定され、第2の予測しきい値T2は、第1のクラスについての感度しきい値に設定される。第3の予測しきい値は、第2のクラス中にある入力について、検証データセット結果からの第2のクラスについての最も高い出力確率値として設定される。第4の予測しきい値は、第2のクラス中の入力について、検証データセット結果からの第2のクラスについての最も低い確率値として設定される。第5の予測しきい値は、第3のクラス中にある入力について、検証データセット結果からの第3のクラスについての最も高い出力確率値として設定される。第6の予測しきい値は、第3のクラス中の入力について、検証データセット結果からの第3のクラスについての最も低い確率値として設定される。

【0206】

次に、3つのクラスへの分類が実施される代替例が説明される。第1のクラスについての第1の分類は、第1のバイナリ分類器を適用することと、第1の予測しきい値を使用することとを備え、第1のクラスについての第2の分類は、第1の分類器を適用することと、第2の予測しきい値を使用することとを備え、ここにおいて、第1の予測しきい値は第2の予測しきい値よりも高い。第1のバイナリ分類器は、入力が第1のクラス中にある（陽性クラス）のか第1のクラス中にある（陰性クラス）のかを決定するようにトレーニングされる。第2のクラスについての第1の分類は、第2のバイナリ分類器を適用することと、第3の予測しきい値を使用することとを備え、第2のクラスについての第2の分類は、第2の分類器を適用することと、第4の予測しきい値を使用することとを備え、ここにおいて、第3の予測しきい値は第4の予測しきい値よりも高い。第2のバイナリ分類器は、入力が第2のクラス中にある（陽性クラス）のか第2のクラス中にある（陰性クラス）のかを決定する。第3のクラスについての第1の分類は、第3のバイナリ分類器を適用することと、第5の予測しきい値を使用することとを備え、第3のクラスについての第2の分類は、第3の分類器を適用することと、第6の予測しきい値を使用することとを備え、ここにおいて、第5の予測しきい値は第6の予測しきい値よりも高い。第2のバイナリ分

10

20

30

40

50

類器は、入力が第3のクラス中にある（陽性クラス）のか第3のクラス中にある（陰性クラス）のかを決定する。第1の例に関して説明されたように、各分類器がトレーニングデータセットに関して別々にトレーニングされ、しきい値が選択される。

【0207】

次に、3つのクラスへの分類が実施される代替例が説明される。第1のクラスについての第1の分類は、第1のバイナリ分類器を適用することと、第1の予測しきい値を使用することとを備え、第1のクラスについての第2の分類は、第2のバイナリ分類器を適用することと、第2の予測しきい値を使用することとを備える。第1のバイナリ分類器と第2のバイナリ分類器とは、入力が第1のクラス中にある（陽性クラス）のか第1のクラス中にある（陰性クラス）のかを決定するようにトレーニングされる。第2のクラスについての第1の分類は、第3のバイナリ分類器を適用することと、第3の予測しきい値を使用することとを備え、第2のクラスについての第2の分類は、第4のバイナリ分類器を適用することと、第4の予測しきい値を使用することとを備える。第3のバイナリ分類器と第4のバイナリ分類器とは、入力が第2のクラス中にある（陽性クラス）のか第2のクラス中にある（陰性クラス）のかを決定するようにトレーニングされる。第3のクラスについての第1の分類は、第5のバイナリ分類器を適用することと、第5の予測しきい値を使用することとを備え、第3のクラスについての第2の分類は、第6のバイナリ分類器を適用することと、第6の予測しきい値を使用することとを備える。第5のバイナリ分類器と第6のバイナリ分類器とは、入力が第3のクラス中にある（陽性クラス）のか第3のクラス中にある（陰性クラス）のかを決定するようにトレーニングされる。上記で第2の例に関して説明されたように、各クラスに対応する3つまたはそれ以上の分類器が、トレーニングデータセットに関して別々にトレーニングされ、第1の分類および第2の分類のために使用される分類器が、各クラスについて選択される。各クラスについて、陽性クラスは、2つの分類器について同じである。

【0208】

図14は、分類の方法の概略図である。方法は、バイナリ予測値だけでなく、テストデータセットに対する、予測の前の実施に基づいてその予測が正しいことが確実であるかどうかのインジケータをも出力する、予測医療デバイス（PMD：predictive medical device）を展開することを備える。PMDが確実である事例に関しては、その事例について予測が与えられる。PMDが確実でない事例に関しては、予測が与えられず、その事例は標準治療（standard of care）へと続くことになる。

【0209】

S1301において、分析されるべき入力医療サンプルまたはデータが取得される。S1302において、入力がPMDに入力される。S1303において、予測値（1または0）が、確実性ととともに、出力される。第1の分類と第2の分類とが合致する場合、確実性は1であり、方法はS1305に移動する。標準治療分析を避けて、分析（予測値）が提供される。第1の分類と第2の分類とが合致しない場合、確実性は0であり、方法はS1304に移動する。次いで、手動検査を含み得る標準治療分析が実施される。この分析は、時間がかかるかまたは費用がかかることがある。S1305またはS1304のいずれかにおいて、分析が提供されると、方法は、S1306に移動し、後続のプロセスを実施する。

【0210】

図15は、一実施形態による、分類システム1の概略図を示す。システム1は、入力11と、プロセッサ3と、作業メモリ9と、出力13と、ストレージ7とを備える。システム1は、入力画像データを取り、出力を生成する。出力は、診断情報を備え得る。特に、出力は、入力画像がバイオマーカーに関連付けられるかどうかの指示であり得る。

【0211】

システム1は、コンピューティングシステム、たとえば、汎用コンピュータなどのエンドユーザシステム、またはサーバである。

【0212】

10

20

30

40

50

プロセッサ 3 は、ストレージ 7 に結合され、RAM 9 を含む作業メモリにアクセスする。プロセッサ 3 は、作業メモリに記憶されたコード中の命令に応答し、それを処理する、論理回路を備え得る。

【0213】

コンピュータプログラム 5 は、不揮発性メモリ 7 に記憶される。不揮発性メモリ 7 は、プロセッサ 3 によってアクセスされ、記憶されたコード 5 は、プロセッサ 3 によって取り出され、実行される。特に、実行されたとき、本明細書で説明される方法を実施するコンピュータプログラムコード 5 は、作業メモリ 9 に記憶されたソフトウェア製品として表される。プロセッサ 3 によるコード 5 の実行は、本明細書で説明される実施形態を実装させることになる。

10

【0214】

プロセッサ 3 はまた、入力モジュール 11 と出力モジュール 13 とにアクセスする。入力および出力モジュールまたはインターフェース 11、13 は、単一の構成要素であり得るか、または別個の入力インターフェース 11 と別個の出力インターフェース 13 とに分割され得る。

【0215】

入力モジュール 11 は、データを受信するための入力 15 に接続される。入力 15 は、外部記憶媒体からまたは通信ネットワークを通してデータを受信するための受信機であり得る。代替的に、入力 15 は、画像キャプチャ装置などのハードウェアを備え得る。代替的に、入力 15 は、記憶された画像ファイルからデータを読み取り得、これは、システム上に、あるいはフロッピーディスク、CD ROM、磁気デバイスまたはプログラマブルメモリデバイスなど、別個の記憶媒体上に記憶され得る。

20

【0216】

出力モジュール 13 に接続されるのは、出力 17 である。出力 17 は、視覚ディスプレイなどのハードウェアを備え得る。代替的に、出力は、外部記憶媒体にまたは通信ネットワークを通してデータを送信するための送信機であり得る。代替的に、出力 17 は、記憶された画像ファイルにデータを書き込み得、これは、システム上に、あるいはフロッピーディスク、CD ROM、磁気デバイスまたはプログラマブルメモリデバイスなど、別個の記憶媒体上に記憶され得る。

【0217】

ストレージ 7 は、プロセッサ 3 に通信可能に結合される。ストレージ 7 は、プロセッサ 3 によって実行されたときにコード 5 によって使用されるデータを含んでいることがある。示されているように、ストレージ 7 は、デバイス中に含まれているローカルメモリである。しかしながら、代替的に、ストレージ 7 は、たとえば、（インターネットなどの）通信ネットワークを介してリモートでアクセスされ得るクラウドベースメモリを使用して完全にまたは部分的にリモートに位置し得る。コード 5 はまた、ストレージ 7 に記憶される。コード 5 は、実行されたとき、作業メモリ 9 中に配置される。

30

【0218】

システム 1 は、データを入力および出力するためのハードウェアをもつ共通システム中に位置し得る。代替的に、システム 1 は、別個のユニット（たとえば、画像キャプチャデバイス）から送信された画像データを受信し、別の別個のユニット（たとえば、画面を備えるユーザコンピュータ）に出力データを送信するリモートシステム 1 であり得る。たとえば、システムは、データを受信および送信するクラウドコンピューティングシステム上に実装され得る。

40

【0219】

説明されるシステムでは、デバイス中に位置する単一のプロセッサ 3 が使用されるが、システムは、同じシステム中に位置することリモートに位置することもある 2 つまたはそれ以上のプロセッサを備え得、これらは、処理の異なる部分を実施し、それらの間でデータを送信するように構成される。たとえば、システムは、グラフィカル処理ユニット（GPU）と一般的な中央処理ユニット（CPU）とを備え得る。様々な動作は GPU によ

50

って実装されるが、他の動作はCPUによって実装される。たとえば、行列演算がGPUによって実施され得る。

【0220】

メモリへのソフトウェアのローディングとストレージユニット7へのデータの記憶のための通常のプロシージャが適用される。コード5は、元の機器中に埋め込まれ得るか、あるいは製造後に全体としてまたは部分的に提供され得る。たとえば、コードは、全体として、ダウンロードの形態であり得るコンピュータプログラム製品として導入され得るか、または光ディスクなどのコンピュータプログラム記憶媒体を介して導入され得る。代替的に、既存のダイアログマネージャソフトウェアへの修正が、説明される実施形態の特徴を提供するために更新またはプラグインによって行われ得る。

10

【0221】

分類モデルをトレーニングする方法と、トレーニングされた分類モデルを使用する分類の方法とが、同じシステム上で実施され得る。たとえば、分類モデルをトレーニングする方法と、トレーニングされた分類モデルを使用する分類の方法とは、図15に関して説明されたようにシステム上で実施され得、ここで、トレーニングされた分類モデルは、トレーニングされるとメモリ7に記憶される。代替的に、分類モデルをトレーニングする方法は第1のシステム上で実施され得、トレーニングされた分類モデルは第2のシステムに記憶され、したがって、分類の方法は第2のシステム上で実施される。

【0222】

説明される実施形態が任意のコンピューティングシステムに適用可能であるが、図15に示されている例示的なコンピューティングシステムは、本明細書で説明される実施形態を実施することが可能な手段を提供することが諒解されよう。使用中に、システム1は、データ入力11を通してデータを受信する。プロセッサ3上で実行されるプログラム5は、上記の図を参照しながら説明された様式で、出力13を通してデータを出力する。

20

【0223】

いくつかの実施形態が説明されたが、これらの実施形態は、単に例として提示されており、本発明の範囲を限定するものではない。実際、本明細書で説明される新規の方法および装置は、様々な他の形態で具体化され得、さらに、本明細書で説明される方法および装置の形態の様々な省略、置換および変更が行われ得る。

30

40

50

【図面】  
【図 1】

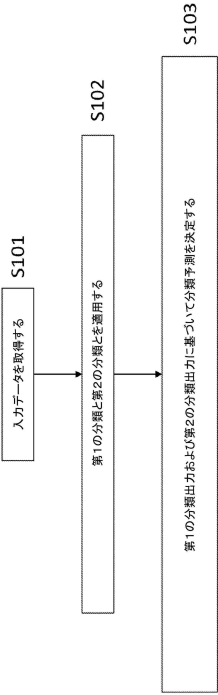


FIG. 1

【図 2 ( a ) 】

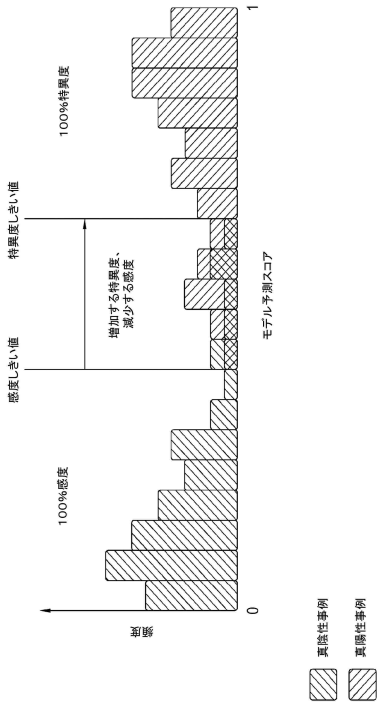


FIG. 2(a)

【図 2 ( b ) 】

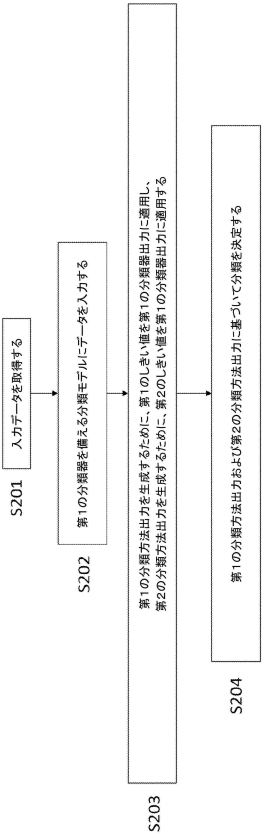


FIG. 2(b)

【図 3】

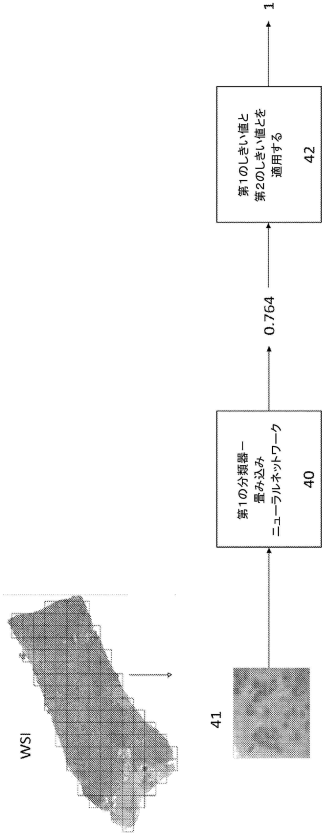


FIG. 3

10

20

30

40

50

【図 4】

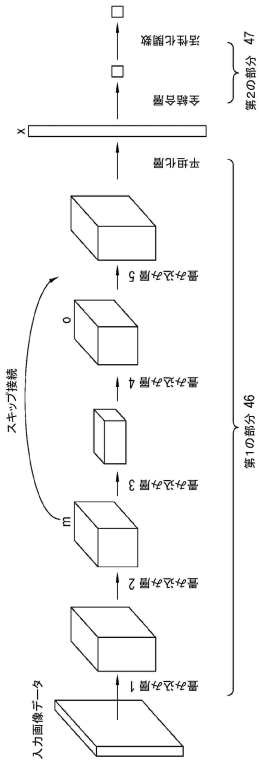


FIG. 4

【図 5】

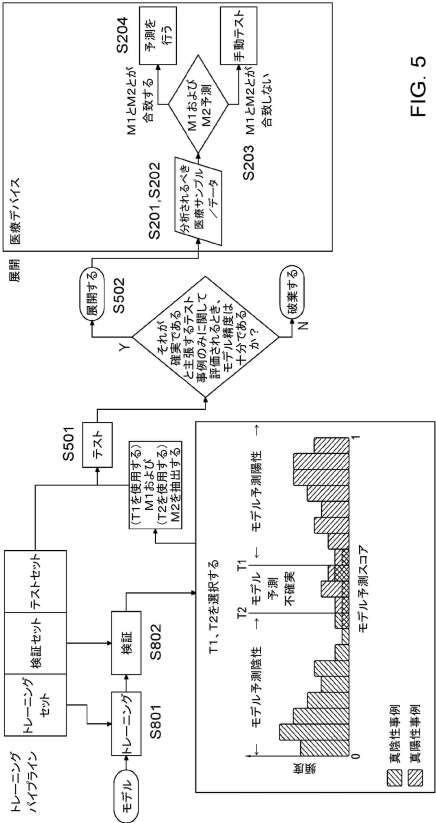


FIG. 5

【図 6】

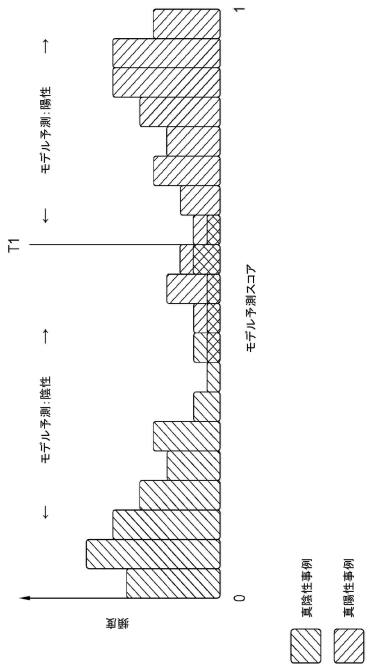


FIG. 6

【図 7】

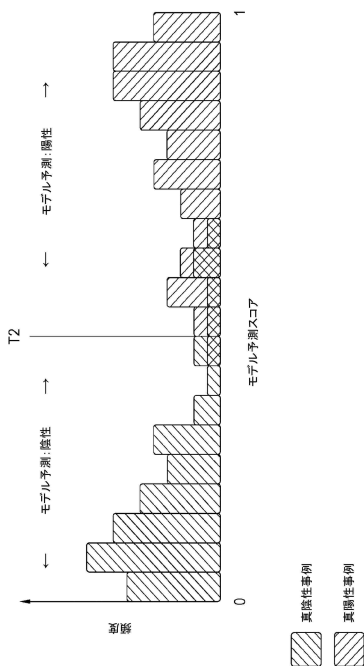


FIG. 7

10

20

30

40

50



【図 10 (b)】

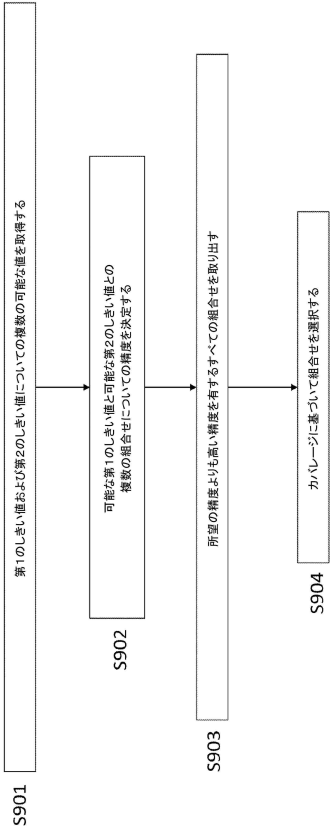


FIG. 10(b)

【図 10 (c)】

| T1 \ T2 | 0.425 | 0.455 | 0.475 | 0.500 | 0.525 | 0.550 | 0.575 | 0.600 | 0.625 | 0.650 |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0.400   | 0.34  | 0.34  | 0.38  | 0.61  | 0.72  | 0.85  | 0.91  | 0.96  | 0.98  | 0.99  |
| 0.425   |       | 0.31  | 0.37  | 0.58  | 0.74  | 0.83  | 0.89  | 0.94  | 0.96  | 0.97  |
| 0.450   |       |       | 0.33  | 0.53  | 0.72  | 0.81  | 0.87  | 0.87  | 0.95  | 0.95  |
| 0.475   |       |       |       | 0.47  | 0.69  | 0.75  | 0.80  | 0.83  | 0.92  | 0.93  |
| 0.500   |       |       |       |       | 0.60  | 0.65  | 0.75  | 0.82  | 0.82  | 0.84  |
| 0.525   |       |       |       |       |       | 0.42  | 0.61  | 0.63  | 0.63  | 0.66  |
| 0.550   |       |       |       |       |       |       | 0.35  | 0.37  | 0.39  | 0.41  |
| 0.575   |       |       |       |       |       |       |       | 0.37  | 0.40  | 0.40  |
| 0.600   |       |       |       |       |       |       |       |       | 0.37  | 0.38  |
| 0.625   |       |       |       |       |       |       |       |       |       | 0.32  |

FIG. 10(c)

【図 11 (a)】

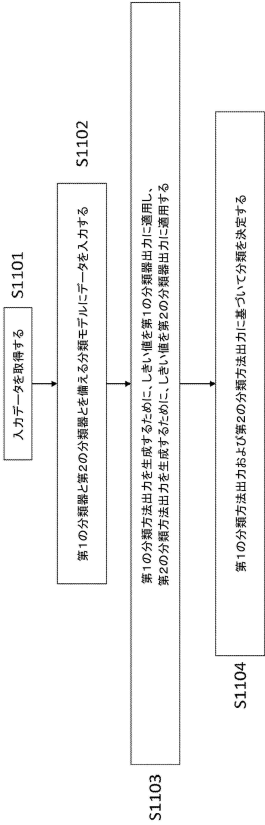


FIG. 11(a)

【図 11 (b)】

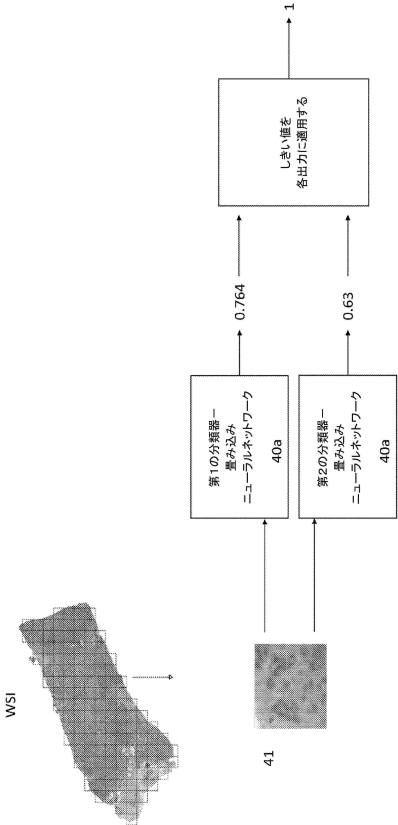


FIG. 11(b)

10

20

30

40

50

【図 1 2 ( a )】

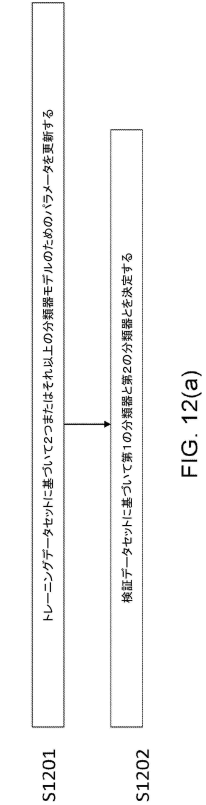


FIG. 12(a)

【図 1 2 ( b )】

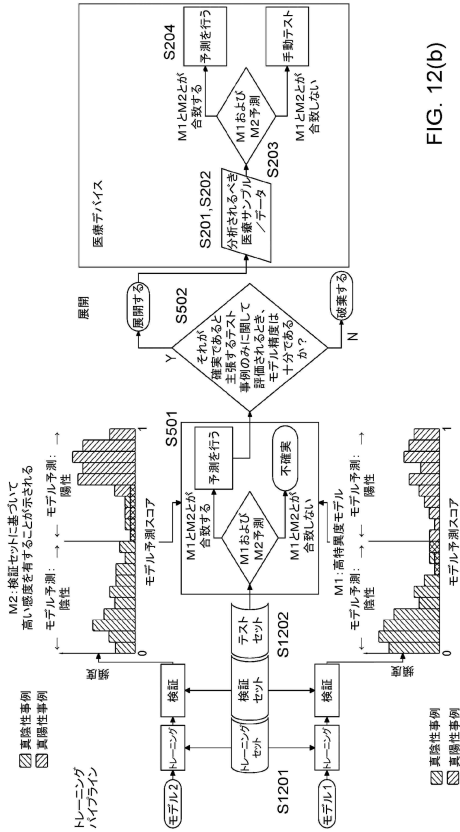


FIG. 12(b)

【図 1 2 ( c )】

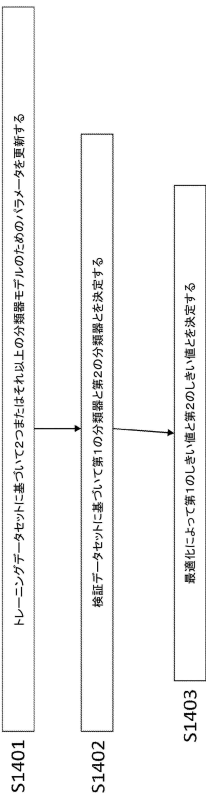


FIG. 12(c)

【図 1 3】

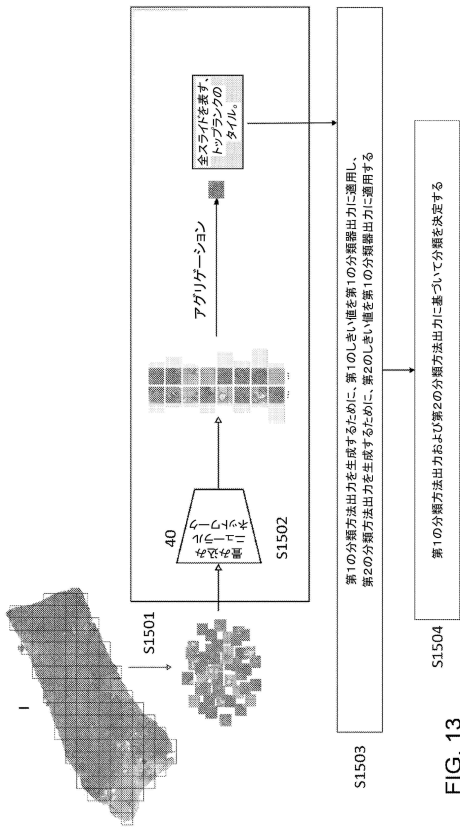


FIG. 13

10

20

30

40

50

【図 1 4】

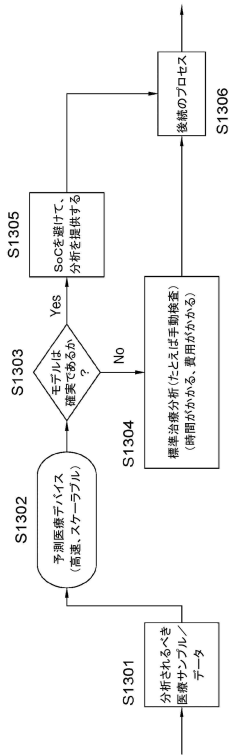


FIG. 14

【図 1 5】

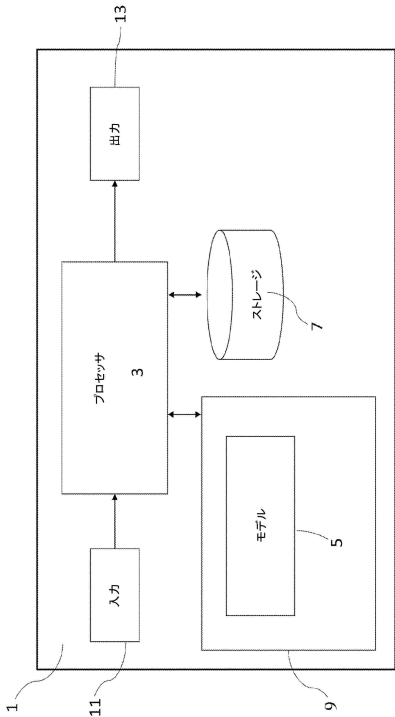


FIG. 15

【図 1 6】

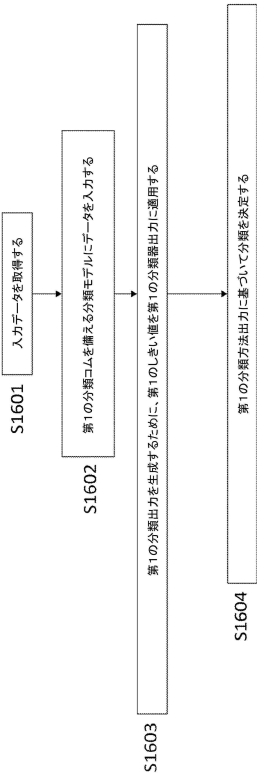


FIG. 16

【図 1 7】

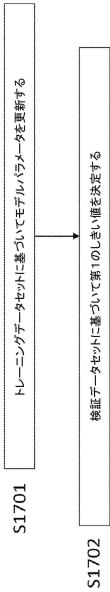


FIG. 17

10

20

30

40

50

## 【国際調査報告】

## INTERNATIONAL SEARCH REPORT

International application No  
**PCT/GB2022/052663**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                         |                                                                             |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| <b>A. CLASSIFICATION OF SUBJECT MATTER</b><br>INV. <b>G06N20/20 G06N3/04 G06N3/08</b><br>ADD.<br><br>According to International Patent Classification (IPC) or to both national classification and IPC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                         |                                                                             |
| <b>B. FIELDS SEARCHED</b><br>Minimum documentation searched (classification system followed by classification symbols)<br><b>G06N</b><br><br>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched<br><br>Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)<br><b>EPO-Internal, WPI Data</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                         |                                                                             |
| <b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                         |                                                                             |
| Category*                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                      | Relevant to claim No.                                                       |
| <b>X</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | <b>US 2018/247107 A1 (MURTHY VENKATESH N [US] ET AL) 30 August 2018 (2018-08-30)</b><br><b>abstract</b><br><b>paragraph [0002] - paragraph [0007]</b><br><b>paragraph [0018] - paragraph [0063];</b><br><b>figures 1-9</b><br><br>----- | <b>1-15</b>                                                                 |
| <b>X</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | <b>US 2011/251989 A1 (KRAAIJ WESSEL [NL] ET AL) 13 October 2011 (2011-10-13)</b><br><b>abstract</b><br><b>paragraph [0001] - paragraph [0013]</b><br><b>paragraph [0018] - paragraph [0082];</b><br><b>figures 1-3</b><br><br>-----     | <b>1-15</b>                                                                 |
| <input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                         |                                                                             |
| * Special categories of cited documents :<br>"A" document defining the general state of the art which is not considered to be of particular relevance<br>"E" earlier application or patent but published on or after the international filing date<br>"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)<br>"O" document referring to an oral disclosure, use, exhibition or other means<br>"P" document published prior to the international filing date but later than the priority date claimed<br>"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<br>"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<br>"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<br>"&" document member of the same patent family |                                                                                                                                                                                                                                         |                                                                             |
| Date of the actual completion of the international search<br><br><b>1 December 2022</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                         | Date of mailing of the international search report<br><br><b>09/12/2022</b> |
| Name and mailing address of the ISA/<br>European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                         | Authorized officer<br><br><b>Hasnas, Sergiu</b>                             |

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

| INTERNATIONAL SEARCH REPORT               |    |                     |                            | International application No |                     |
|-------------------------------------------|----|---------------------|----------------------------|------------------------------|---------------------|
| Information on patent family members      |    |                     |                            | PCT/GB2022/052663            |                     |
| Patent document<br>cited in search report |    | Publication<br>date | Patent family<br>member(s) |                              | Publication<br>date |
| US 2018247107                             | A1 | 30-08-2018          | US 2018247107              | A1                           | 30-08-2018          |
|                                           |    |                     | WO 2017055412              | A1                           | 06-04-2017          |
| -----                                     |    |                     |                            |                              |                     |
| US 2011251989                             | A1 | 13-10-2011          | EP 2182451                 | A1                           | 05-05-2010          |
|                                           |    |                     | EP 2356588                 | A1                           | 17-08-2011          |
|                                           |    |                     | US 2011251989              | A1                           | 13-10-2011          |
|                                           |    |                     | WO 2010050811              | A1                           | 06-05-2010          |
| -----                                     |    |                     |                            |                              |                     |

フロントページの続き

(51)国際特許分類

F I  
G 0 6 N      3/08

テーマコード (参考)

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,N  
E,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,  
CV,CV,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IQ,IR,IS,I  
T,JM,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  
ケンブリッジ、ステーション・ロード、ソールズベリー・ハウス、パナケイア・テクノロジーズ・  
リミテッド気付

F ターム (参考)    5L096   AA02 AA06 BA06 BA13 CA02 DA02 EA18 GA30 GA51 HA11