(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7339337号
(P7339337)

(45)発行日　令和5年9月5日(2023.9.5)

(24)登録日　令和5年8月28日(2023.8.28)

(51)国際特許分類

G01Q 30/04 (2010.01)	G01Q	30/04		
G01Q 60/24 (2010.01)	G01Q	60/24		
G01Q 80/00 (2010.01)	G01Q	80/00	121	
G01N 33/543 (2006.01)	G01N	33/543	595	
G01N 33/50 (2006.01)	G01N	33/50		Z

FI

請求項の数　20　（全37頁）

(21)出願番号　特願2021-525318(P2021-525318)
(86)(22)出願日　令和1年11月7日(2019.11.7)
(65)公表番号　特表2022-507112(P2022-507112A)
(43)公表日　令和4年1月18日(2022.1.18)
(86)国際出願番号　PCT/US2019/060225
(87)国際公開番号　WO2020/097302
(87)国際公開日　令和2年5月14日(2020.5.14)
審査請求日　令和4年9月13日(2022.9.13)
(31)優先権主張番号　62/772,327
(32)優先日　平成30年11月28日(2018.11.28)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　62/756,958
(32)優先日　平成30年11月7日(2018.11.7)
最終頁に続く

(73)特許権者　319009901
トラスティーズ オブ タフツ カレッジ
アメリカ合衆国 マサチューセッツ州 02155 メドフォード バルー ホール
(74)代理人　100117606
弁理士　安部 誠
(74)代理人　100136423
弁理士　大井 道子
(74)代理人　100154449
弁理士　谷 征史
(72)発明者　ソコロフ， イゴール
アメリカ合衆国 02155 マサチューセッツ州， メドフォード， ウィンフォード ウェイ 29
(72)発明者　ミリュコビッチ， ミロス
最終頁に続く

(54)【発明の名称】　表面を識別する原子間力顕微鏡法

(57)【特許請求の範囲】

【請求項１】

原子間力顕微鏡を使用して、複数の異なる表面属性のそれぞれに関連付けられた複数の画像アレイを取得することと、前記複数の画像アレイを結合し多次元画像アレイを生成することと、前記生成された前記多次元画像アレイに適用される機械学習法を使用することと、表面を分類することとを含む方法であって、

前記原子間力顕微鏡を使用することは、複数のチャネルを含むマルチチャネル原子間力顕微鏡を使用することを含み、

前記複数のチャネルの各チャネルは、前記表面の異なる表面属性に対応し、

前記表面を分類することは、トレーニングデータを用いてトレーニングされ、試験データを用いて試験可能である機械学習モジュールの使用を分類することを含み、かつ、検証データを用いて検証可能であり、

前記トレーニングデータは、分類方法を学習するために用いられてきたデータであり、前記試験データは、前記機械学習モジュールによって実行される分類の有効性を検証するために使用可能なデータである、方法。

【請求項２】

前記画像アレイを処理して、表面パラメータを取得することと、機械学習を使用して、少なくとも部分的に前記表面パラメータに基づいて前記表面を分類することとを更に含む、請求項１に記載の方法。

【請求項３】

前記表面を被験者の尿から収集された細胞の表面であるものとして選択することと、癌を示すか、又は癌を示さないものとして前記細胞を分類することとを更に含む、請求項１に記載の方法。

【請求項４】

前記チャネルにより提供される情報を凝縮することと、前記凝縮された情報から凝縮データベースを構築することとを更に含む、請求項１に記載の方法。

【請求項５】

前記チャネルにより提供される情報を凝縮することと、前記凝縮された情報から凝縮データベースを構築することとを更に含み、

前記凝縮データベースを構築することは、

前記チャネルにより提供される前記情報に基づいて第１のデータベースを形成することと、ここで、前記第１のデータベースはインデックスを有する、

前記第１のデータベースから凝縮データベースを導出することと、ここで、前記凝縮データベースは、前記第１のデータベースよりも少数のインデックスを有する、

を含み、

前記凝縮データベースを導出することは、テンソル加算を実行して、前記第１のデータベースからの情報を前記第１のデータベースの１つ又は複数のインデックスに対応する１つ又は複数のスライスと共に結合するテンソル和を生成することと、前記テンソル和を使用して前記凝縮データベースを形成することとを含む、請求項１に記載の方法。

【請求項６】

前記チャネルにより提供される情報を凝縮することと、前記凝縮された情報から凝縮データベースを構築することとを更に含み、

前記凝縮データベースを構築することは、

前記チャネルにより提供される前記情報に基づいて第１のデータベースを形成することと、ここで、前記第１のデータベースはインデックスを有する、

前記第１のデータベースから凝縮データベースを導出することと、ここで、前記凝縮データベースは、前記第１のデータベースよりも少数のインデックスを有する、

を含み、

前記第１のデータベースから凝縮データベースを導出することは、

前記第１のデータベースからの値のサブセットを定義することと、ここで、前記値のそれぞれは前記第１のデータベース内の対応する要素を表す、

前記値のサブセット内の前記値から凝縮値を導出することと、

前記凝縮値を用いて前記第１のデータベースからの前記対応する要素を表すこと

とを含み、

前記凝縮値を導出することは、情報を前記第１のデータベースから表面パラメータ抽出機に通して、表面パラメータセットを取得することを含む、請求項１に記載の方法。

【請求項７】

前記画像アレイは、試料の画像アレイであり、

前記試料の画像アレイを区画に分割することと、各区画の表面パラメータを取得することと、各区画の前記表面パラメータのメジアンまたは平均として前記画像アレイの表面パラメータを定義することとを更に含む、請求項１に記載の方法。

【請求項８】

前記画像アレイを処理して、表面パラメータを取得することと、機械学習を使用して、少なくとも部分的に前記表面パラメータに基づいて外部導出パラメータから前記表面を分類することとを更に含む、請求項１に記載の方法。

【請求項９】

前記表面は、収集された試料から導出されるボディの表面であり、前記試料の少なくとも１つは、ボディを有さないボディフリー試料であり、

前記画像アレイを処理して、表面パラメータを取得し、機械学習を使用して、少なくとも部分的に前記表面パラメータに基づいて外部導出パラメータから前記表面を分類するこ

と、
前記ボディフリー試料からのボディの不在を示すデータを含むものとして前記外部導出パラメータを選択することと
を更に含む、請求項１に記載の方法。
【請求項１０】
前記表面は、患者から取得される試料から導出される細胞の表面であり、
前記画像アレイを処理して、表面パラメータを取得し、機械学習を使用して、少なくとも部分的に前記表面パラメータに基づいて外部導出パラメータから前記表面を分類することと、
前記患者が特定の疾患を有する確率を示すデータを含むものとして前記外部導出パラメータを選択することを更に含む、請求項１に記載の方法。
【請求項１１】
機械学習を使用して、取得された前記画像アレイから少なくとも部分的に表面パラメータに基づいて前記表面を分類することと、
前記表面パラメータのサブセットを定義することと、
前記サブセットに基づいてデータベースを生成することと
を更に含み、
表面パラメータの前記サブセットを定義することは、前記表面パラメータ間の相関を特定することと、前記相関を閾値と比較して、１組の相関パラメータを識別することと、前記１組の相関パラメータのサブセットを表面パラメータの前記サブセットに含めることとを含む、請求項１に記載の方法。
【請求項１２】
前記表面は、患者からの複数の第１の細胞の表面であり、複数の第２の細胞は癌罹患患者からのものとして分類されており、複数の第３の細胞は非癌患者からのものとして分類されており、
前記第１の細胞の数に対する前記第２の細胞の数の比率が所定の閾値を超える場合、前記患者を癌として診断することを更に含む、請求項１に記載の方法。
【請求項１３】
前記原子間力顕微鏡は、カンチレバーと、前記カンチレバーの先端部に配置された探針とを備え、前記カンチレバーは共振周波数を有し、前記原子間力顕微鏡を使用することは、前記探針と前記表面との間の距離を前記共振周波数未満の周波数で振動させることを含む、請求項１に記載の方法。
【請求項１４】
前記画像アレイから画像がアーチファクトを有することを自動的に検出することと、前記アーチファクトを有する前記画像を前記表面の分類に使用されるものから自動的に除外することとを更に含む、請求項１に記載の方法。
【請求項１５】
前記表面を絵画、通貨、株券、身分証明書、出生証明書、パスポートのうちひとつの表面を含むように前記表面を選択することを更に含む、請求項１に記載の方法。
【請求項１６】
前記チャネルにより提供される情報を凝縮することと、
前記凝縮された情報から凝縮データベースを構築することと、
前記チャネルにより提供される前記情報に基づいて第１のデータベースを形成することと
を更に含み、
前記凝縮データベースを構築することは、前記第１のデータベースを前記第１のデータベースよりも低い次元のサブ空間に投影することを含み、前記投影は、前記凝縮データベースを定義し、前記凝縮データベースは、前記第１のデータベースよりも低い次元を有する、請求項１に記載の方法。
【請求項１７】
前記チャネルにより提供される情報を凝縮することと、

前記凝縮された情報から凝縮データベースを構築することと、
前記チャネルにより提供される前記情報に基づいて第１のデータベースを形成することとを更に含み、
前記凝縮データベースを構築することは、前記第１のデータベースを前記第１のデータベースよりも低い次元のサブ空間に投影することを含み、前記投影は、前記凝縮データベースを定義し、前記凝縮データベースは、前記第１のデータベースよりも低い次元を有し、
前記凝縮データベースを構築することは、それぞれのジニインデックスに基づいて最も高いランクを有する表面パラメータのみを含めることを含む、請求項１に記載の方法。
【請求項１８】
前記画像アレイを処理して、表面パラメータを取得することと、
機械学習を使用して、少なくとも部分的に前記表面パラメータに基づいて外部導出パラメータから前記表面を分類することと、ここで、前記表面は、患者から取得される試料から導出される細胞の表面である、
前記患者が特定の疾患を有する確率を示すデータを含むものとして前記外部導出パラメータを選択することと
を更に含み、
前記確率を示す前記データは、前記患者の年齢と、前記患者の喫煙習慣と、前記患者の家族歴とを含む、請求項１に記載の方法。
【請求項１９】
前記チャネルにより提供される情報を凝縮することと、
前記チャネルにより提供される前記情報に基づいて第１のデータベースを形成することと、
前記第１のデータベースを前記第１のデータベースよりも低い次元のサブ空間に投影することと、ここで、前記投影は、前記第１のデータベースよりも低い次元を有する凝縮データベースを定義する、
を更に含み、
前記表面を分類することは、ランダムフォレスト法、エクストリームランダム化フォレスト法、勾配ブースティングツリー法および決定木法からなる群から選択される方法を含む、請求項１に記載の方法。
【請求項２０】
複数の異なる表面属性のそれぞれに関連付けられた複数の画像アレイを取得するマルチチャネル原子間力顕微鏡と、ここで、前記マルチチャネル原子間力顕微鏡は、複数のチャネルを含み、前記複数のチャネルの各チャネルは、表面の異なる表面属性にそれぞれ対応する、
前記マルチチャネル原子間力顕微鏡から、前記画像アレイを表す信号を受信し、前記複数の画像アレイを結合し多次元画像アレイを生成する処理システムと、ここで、前記処理システムは、機械学習モジュールと、前記機械学習モジュールから分類の基礎を学習した後、未知の試料を分類する分類器とを備えている、
を備える装置であって、
前記機械学習モジュールは、分類結果が試験データを用いて検証可能であるように、トレーニングデータを用いて前記表面を分類するためにトレーニングされ、
前記トレーニングデータは、分類方法を学習するために用いられてきたデータであり、前記試験データは、前記分類の有効性を検証するために使用可能なデータである、装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本願は、米国仮特許出願第６２／７５６，９５８号明細書の優先日２０１８年１１月７日及び米国仮特許出願第６２／７７２，３２７号明細書の優先日２０１８年１１月２８日の利益を主張するものであり、これらの内容は参照により本明細書に援用される。
【０００２】

本発明は、原子間力顕微鏡法及び機械学習を表面の特徴を使用して、その表面を分類又は識別することに関する。特に、特徴を使用して生体細胞を識別又は分類することに関する。

【背景技術】

【０００３】

原子間力顕微鏡法では、カンチレバーの先端部に取り付けられた探針が試料の表面を走査する。一動作モードでは、探針は、走査する際に表面を軽く叩く。探針が試料を走査する際、探針が試料に対して及ぼす荷重力に関連付けられた力ベクトルの大きさ及び方向を制御することが可能である。

【０００４】

平衡位置からのカンチレバーの偏向は、大量の情報を抽出することができる信号を提供する。一例として、荷重力又はカンチレバーの偏向を一定に保つことにより、試料の種々の点における試料のトポロジを取得することが可能である。次に、各点で収集された値はアレイに編成され、アレイでは、行及び列が、二次元座標系における点の場所を識別し、行及び列における値が、そのポイントで測定された属性を表す。したがって、その結果生成される数のアレイはマップとして見ることができる。これは、マップの各点がその点における試料表面の何らかの属性を示す試料のマップを作成できるようにする。幾つかの例では、属性は、何らかの基準面よりも上又は下の表面の高さである。

【０００５】

しかしながら、表面高さの画像は、走査時に記録することができる唯一の画像ではない。カンチレバーの偏向は、試料表面の複数の画像を収集するのに使用することができ、各画像は表面の異なる属性のマップである。これらの属性のほんの少数の例には、探針と表面との付着力、表面の剛性、及び粘弾性エネルギー損失がある。

【発明の概要】

【課題を解決するための手段】

【０００６】

本発明は、原子間力顕微鏡により取得された多次元画像を使用して表面を識別し、これらの画像からの情報を使用して、表面を幾つかのクラスの１つに分類する方法を提供する。本発明によれば、空間次元に対応する、空間次元に対応する２つの次元と、２つの空間次元により識別される座標に存在する異なる物理的及び空間的属性に対応する追加の次元とを有する表面の多次元画像を取得することが可能である。幾つかの実施形態では、次元は横寸法である。

【０００７】

生じる問題は、表面の識別及び分類のために、これらの異なる物理的及び空間的属性をいかに選び、いかに使用するかである。本発明によれば、表面の識別及び分類に使用される属性は、予め決められない。表面の識別及び分類に使用される属性は、画像のデータベース及び対応するクラスに適用される機械学習の結果に基づいて計算される。表面の識別及び分類に使用される属性は学習される。特に、表面の識別及び分類に使用される属性は、機械学習により学習される。

【０００８】

本発明の実施形態の中には、原子間力顕微鏡を使用して、表面の異なる属性に対応する異なるマップを取得し、これらのマップ又はこれらのマップから導出されるパラメータの組合せを使用して、試料表面を識別又は分類することを含む実施形態がある。そのような方法は、明確に定義されたクラスに属する表面の例の原子間力顕微鏡画像を記録することと、そのような原子間力顕微鏡マップに、属するクラスが関連付けられるデータベースを形成することと、そうして取得された原子間力顕微鏡マップ及びその組合せを使用して、データベースをトレーニングデータ及びテストデータに分割することにより、表面をいかに分類するかを学習することであって、トレーニングデータは、例えば、決定木、又はニューラルネットワーク、又はそれらの組合せを構築することによりいかに分類するかを学習するのに使用されることと、テストデータを使用して、そうして学習された分類が、所

与の有効性閾値に合格するのに十分効率的であることを確認することとを含む。
【０００９】
別の実施形態は、原子間力顕微鏡により提供されたマップを１組の表面パラメータに縮小することを含み、１組の表面パラメータの値は、それらの属性を入力として使用する数学的関数又はアルゴリズムにより定義される。好ましい実施では、各マップ又は画像は、他の表面パラメータと一緒に、表面の分類又は識別に使用することができる表面パラメータをもたらす。そのような実施形態では、これらの表面パラメータに基づいて分類する分類器が存在する。しかしながら、分類器自体は予め決められない。分類器は、上述したように機械学習手順を通して学習される。
【００１０】
本方法は、表面の性質に依存しない。例えば、本方法は、偽造を見分けるために、絵画、又は通貨、又は出生証明書又はパスポート等のセキュア文書の表面を分類するのに本方法を使用し得る。しかし、種々の疾患を識別するために、生体の細胞又は他の部分の表面を分類するのにも同じ方法を使用し得る。例えば、種々の癌は、特定の表面シグネチャを有する細胞を有する。したがって、本方法は、種々の種類の癌の検出に使用することができる。
【００１１】
調べる細胞を正確に取得することは困難である。幾つかの場合、侵襲的手順が求められる。しかしながら、体から自然に剥がれ落ち、又は最小の侵襲性でもってのみ体から抽出することができる特定の種類の細胞がある。一例は、子宮頸部細胞診での頸部表面を優しく擦るものである。自然に剥がれ落ちる細胞の中には、膀胱を含む尿路からの細胞がある。したがって、本方法は、膀胱鏡検査等の侵襲性で高価な手順の必要性なしでこれらの細胞を検査し、膀胱癌を検出するのに使用することができる。
【００１２】
本発明は、例えば、準共振タッピングモードを使用する場合、物理的属性の多次元アレイを生成することができる原子間力顕微鏡を使用することを特徴とする。幾つかの実施では、１組の画像を取得することは、体液から収集された細胞の表面のナノスケール分解能走査を実行するモードにおいて、原子間力顕微鏡を使用することと、原子間力顕微鏡走査手順から取得されたデータを、試料が、以下「癌罹患患者」と呼ばれる癌を有する患者からのものである確率の指示を提供する機械学習システムに提供することとを含む。本方法は一般に、表面属性に基づいて細胞を分類するのに適用可能である。
【００１３】
膀胱癌の状況で説明されるが、本明細書に開示される方法及びシステムは、細胞又は体液が侵襲的生検を必要とせずに分析に利用可能な他の癌の検出にも適用可能である。例には、上部尿路癌、尿道癌、結腸直腸癌及び他の消化器癌、子宮頸癌、軌道消化管癌、及び同様の属性を有する他の癌がある。
【００１４】
さらに、本明細書に記載される方法は、癌以外の細胞異常の検出及び種々の薬剤への細胞反応のモニタリングにも適用可能である。加えて、本明細書に記載される方法は、導出元が生体であれ、非生体であれ関係なく、あらゆるタイプの表面の分類及び識別に有用である。必要なのは、表面が、原子間力顕微鏡による走査の影響を受けやすいものであるだけである。
【００１５】
例えば、本明細書に記載される方法は、通貨、株券、身分証明書、又は絵画等の芸術品の偽造を含め、偽造の検出に使用することができる。
【００１６】
一態様では、本発明は、原子間力顕微鏡を使用して、患者から取得された複数の細胞のそれぞれの１組の画像を取得することと、画像を処理して、表面パラメータマップを取得することと、画像に適用される機械学習アルゴリズムを使用することと、癌罹患患者又は非癌患者の何れか一方に由来するものとして細胞を分類することとを特徴とする。

【００１７】
これらの実施形態の中には、顕微鏡が準共振タッピングモードで使用される実施形態がある。更に他の実施形態では、顕微鏡はリンギングモードで使用される。
【００１８】
別の態様では、本発明は、原子間力顕微鏡を使用することと、表面に関連付けられた１組の画像を取得することと、画像を処理して表面パラメータマップを取得することと、画像に適用される機械学習アルゴリズムを使用することと、表面を分類することとを特徴とする。
【００１９】
これらの実施の中には、膀胱細胞の表面である表面を選択することと、癌罹患患者又は非癌患者に由来する細胞の表面として表面を分類することとを含む実施がある。
【００２０】
別の態様では、本発明は、原子間力顕微鏡を使用して、表面に関連付けられた１組の画像を取得することと、画像を結合することと、結合された画像に適用される機械学習法を使用することと、表面を分類することとを含む方法を特徴とする。
【００２１】
この方法は、原子間力顕微鏡を実行する必要があり、人の心は機械ではなく、人の心は機械学習法を実行することができないため、紙と鉛筆の有無に関係なく人の心には実行することができない。本方法はまた、技術的効果、すなわち、表面の技術的属性に基づいた表面の分類を達成するために、非抽象的に実行される。抽象的且つ／又は非技術的に本方法をいかに実行するかの説明は、非抽象的実施態様及び技術的実施態様以外のあらゆるものを包含するものとして特許請求の範囲を誤って解釈するのを避けるために、意図的に省かれている。
【００２２】
幾つかの実施では、画像は細胞の画像である。これらの中には、細胞の画像がアーチファクトを有していることを自動的に検出することと、表面の分類に使用されることからその画像を除外することと、及び、シンプルの画像を区画に分割することと、各区画の表面パラメータを取得することと、各区画の表面パラメータのメジアンとして細胞の表面パラメータを定義することとを含む実施とを含む。
【００２３】
幾つかの実施は、画像を処理して、表面パラメータを取得することと、機械学習を使用して、少なくとも部分的に表面パラメータに基づいて表面を分類することとも含む。これらの中には、表面パラメータのサブセットを定義することと、サブセットに基づいてデータベースを生成することとを更に含む実施がある。そのような実施では、表面パラメータのサブセットを定義することは、表面パラメータ間の相関を特定することと、相関を閾値と比較して、１組の相関パラメータを識別することと、１組の相関パラメータのサブセットを表面パラメータのサブセットに含めることとを含む。また、これらの中には、表面パラメータのサブセットを定義することと、サブセットに基づいてデータベースを生成することとを更に含む実施がある。これらの実施では、表面パラメータのサブセットを定義することは、表面パラメータ間の相関行列を特定することを含み、相関行列を特定することは、シミュレートされた表面を生成することを含む。また、これらの実施の中には、表面パラメータのサブセットを定義することと、サブセットに基づいてデータベースを生成することとを含むものがある。これらの実施では、表面パラメータのサブセットを定義することは、同じ試料からの同じ種類の異なる表面パラメータを結合することを含む。
【００２４】
実施は、１組の画像を取得することが、リンギングモードでマルチチャネル原子間力顕微鏡を使用することを含み、原子間力顕微鏡の各チャネルは、表面の対応する表面属性を示す情報を提供するものも含む。
【００２５】
また、本発明の実施の中には、被験者の尿から収集された細胞の表面を選択することと

、癌を示すか、又は癌を示さないものとして細胞を分類することとを含む。
【００２６】
顕微鏡を使用する多種多様な方法が、本発明の範囲から逸脱せずに利用可能な範囲内にある。これらは、各チャネルが表面の表面属性に対応するマルチチャネル原子間力顕微鏡の使用、準共振タッピングモードでの原子間力顕微鏡の使用、及び、それぞれが表面の異なる表面属性に対応する複数のチャネルの情報を取得し、チャネルにより提供された情報を凝縮し、その凝縮された情報から凝縮データベースを構築することと併せた原子間力顕微鏡の使用を含む。
【００２７】
マルチチャネル原子間力顕微鏡に頼る本発明の実施の中には、チャネルにより提供される情報に基づいて第１のデータベースを形成し、多種多様な方法の何れかで凝縮データベースの構築を実行することを更に含む実施がある。これらの中には、第１のデータベースを第１のデータベースよりも低い次元のサブ空間に投影することを含むものがあり、投影は凝縮データベースを定義し、凝縮データベースは、第１のデータベースよりも低い次元を有する。また、これらの中には、第１のデータベースからの凝縮データベースを含むものがあり、凝縮データベースは、第１のデータベースよりも少数のインデックスを有する。これは、例えば、テンソル加算を実行して、第１のデータベースからの情報を第１のデータベースの１つ又は複数のインデックスに対応する１つ又は複数のスライスと共に結合するテンソル和を生成し、テンソル和を使用して凝縮データベースを形成することにより実行することができる。
【００２８】
本発明の幾つかの実施では、第１のデータベースから凝縮データベースを導出することは、第１のデータベースからの値のサブセットを定義することと、値のそれぞれは第１のデータベース内の対応する要素を表すことと、値のサブセット内の値から凝縮値を導出することと、凝縮値を用いて第１のデータベースからの対応する要素を表すこととを含み、凝縮値を導出することは、値のサブセット内の値を総和することを含む。総和することは、テンソル加算を実行して、第１のデータベースからの値を第１のデータベースのインデックスに対応する１つ又は複数のスライスと共に結合するテンソル和を生成することと、テンソル和を使用して凝縮データベースを形成することとによるものを含め、多種多様な方法で実行することができる。
【００２９】
本発明の実施はまた、第１のデータベースからの値のサブセットを定義することと、値のそれぞれは第１のデータベース内の対応する要素を表すことと、値のサブセット内の値から凝縮値を導出することと、凝縮値を用いて第１のデータベースからの対応する要素を表すこととにより、凝縮データベースが第１のデータベースから導出される実施も含み、凝縮値を導出することは、例えば、算出平均又は幾何平均を取得することにより、値のサブセット内の値の平均を計算することを含む。
【００３０】
また、本発明の実施の中には、第１のデータベースから凝縮データベースを導出することは、第１のデータベースからの値のサブセットを定義すること、値のそれぞれは第１のデータベース内の対応する要素を表すことと、値のサブセット内の値から凝縮値を導出することと、凝縮値を用いて第１のデータベースからの対応する要素を表すこととを含むものがあり、凝縮値は、値のサブセット内の値の最大又は最小の一方である。
【００３１】
更に他の実施形態では、第１のデータベースから凝縮データベースを導出することは、第１のデータベースからの値のサブセットを定義することと、値のそれぞれは第１のデータベース内の対応する要素を表すことと、値のサブセット内の値から凝縮値を導出することと、凝縮値を用いて第１のデータベースからの対応する要素を表すこととを含み、凝縮値を導出することは、情報を第１のデータベースから表面パラメータ抽出機に通して、表面パラメータセットを取得することを含む。これらの中には、導出元の画像の表面エリア

から独立するものとして表面パラメータセットを表す表面パラメータを正規化することを含む実施及び表面パラメータを同じ次元の別のパラメータで除することを含む実施がある。

【００３２】

他の実施は、試料の画像がアーチファクトを有することを自動的に検出することと、上記画像を表面の分類に使用されるものから自動的に除外することとを含む。

【００３３】

更に他の実施は、試料の画像を区画に分割することと、各区画の表面パラメータを取得することと、各区画の表面パラメータのメジアンとして細胞の表面パラメータを定義することとを含む。

【００３４】

本発明の幾つかの実施は、画像を処理して、表面パラメータを取得することと、機械学習を使用して、少なくとも部分的に表面パラメータに基づいて外部導出パラメータから表面を分類することとを含む。これらの中には、表面は、収集された試料から導出されるボディの表面であり、試料の少なくとも１つは、ボディを有さないことを意味するボディフリー試料である実施がある。これらの実施では、本方法は、ボディフリー試料からのボディの不在を示すデータを含むものとして外部導出パラメータを選択することを更に含む。ボディフリー試料を含む実施の中には、人工表面パラメータをボディフリー試料に割り当てることを含むものがある。幾つかの実施では、表面は、患者から取得される試料から導出される細胞の表面である。これらの中には、患者が特定の疾患を有する確率を示すデータを含むものとして外部導出パラメータを選択することを含む実施がある。確率を示すそのようなデータの例には、患者の年齢、患者の喫煙習慣、及び患者の家族歴がある。

【００３５】

多種多様な機械学習法が使用可能である。これらには、ランダムフォレスト法、エクストリームランダム化フォレスト法、勾配ブースティングツリー法、ニューラルネットワークの使用、決定木法、及びそれらの組合せがある。

【００３６】

幾つかの実施形態では、表面は、患者からの第１の複数の細胞の表面であり、第２の複数の細胞は癌罹患患者からのものとして分類されており、第３の複数の細胞は非癌患者からのものとして分類されている。これらの方法は、第１の複数に対する第２の複数の比率が所定の閾値を超える場合、患者を癌として診断することを含む。

【００３７】

幾つかの実施では、原子間力顕微鏡は、カンチレバーと、カンチレバーの先端部に配置された探針とを含む。カンチレバーは共振周波数を有する。これらの実施では、原子間力顕微鏡の使用を使用することは、探針と表面との間の距離を共振周波数未満の周波数で振動させることを含む。

【００３８】

幾つかの実施では、原子間力顕微鏡を使用することは、試料表面の異なる物理的属性に対応する複数のチャネルの情報を出力するように構成された顕微鏡を使用することを含む。

【００３９】

他の実施は、画像を処理して、表面パラメータを取得することと、機械学習を使用して、少なくとも部分的に表面パラメータに基づいて外部導出パラメータから表面を分類することとを含む。これらの実施形態では、表面は、患者から取得された試料から導出される細胞の表面であり、試料の少なくとも１つは、患者からの細胞を有さない細胞なし試料である。そのような実施では、本方法は、細胞なし試料からの細胞の不在を示すデータを含むものとして外部導出パラメータを選択することを更に含む。これらの実施の中には、人工表面パラメータを細胞なし試料に割り当てることを更に含むものがある。

【００４０】

別の態様では、本発明は、装置が原子間力顕微鏡及び処理システムを備えることを特徴とする。原子間力顕微鏡は、表面に関連付けられた画像を取得する。処理システムは、原子間力顕微鏡から、画像を表す信号を受信し、画像を結合する。処理システムは、機械学

習モジュールと、機械学習モジュールから分類の基礎を学習した後、未知の試料を分類する分類器とを含む。

【００４１】

幾つかの実施形態では、処理システムは、画像を処理して、表面パラメータを取得し、機械学習モジュールを使用して、少なくとも部分的に表面パラメータに基づいて表面を分類するように構成される。これらの中には、原子間力顕微鏡は、各チャネルが表面の表面属性に対応するマルチチャネル原子間力顕微鏡を含む実施形態がある。これらの中には、チャネルにより提供された情報を凝縮し、凝縮された情報から凝縮データベースを構築する凝縮器も含む実施形態がある。

【００４２】

凝縮データベースを含む実施形態は、分類器が凝縮データベースに基づいて未知の試料を分類する実施形態も含む。

【００４３】

多種多様な凝縮器が、凝縮データベースの構築に利用可能である。これらの中には、第１のデータベースを第１のデータベースよりも低い次元のサブ空間に投影することにより凝縮データベースを構築する凝縮器がある。この投影は、第１のデータベースよりも低い次元を有する凝縮データベースを定義する。

【００４４】

本明細書で使用される場合、「原子間力顕微鏡」、「ＡＦＭ」、「走査探針顕微鏡」、及び「ＳＰＭ」は同義として見なされる。

【００４５】

本明細書に記載される方法は、非抽象的方法のみである。したがって、特許請求の範囲は非抽象的実施態様のみに関することができる。本明細書で使用される場合、「非抽象的」は、本願の出願日時点での米国特許法第１０１条の要件に準拠することを意味すると見なされる。

【００４６】

本発明のこれら及び他の特徴は、以下の詳細な説明及び添付図から明らかになろう。

【図面の簡単な説明】

【００４７】

【図１】原子間力顕微鏡の一例の簡易図を示す。

【図２】図１の処理システムからの更なる詳細を示す。

【図３】図１及び図２に示される原子間力顕微鏡及び処理システムにより実行され得診断方法を示す。

【図４】図１に示される原子間力顕微鏡に内蔵された光学顕微鏡を通る図を示す。

【図５】図１の原子間力顕微鏡により取得される膀胱細胞のマップを示す。

【図６】データベースと図２の処理システム内の機械学習モジュールとの間の対話の詳細を示す。

【図７】初期の大きなデータベースを次元のより小さな凝縮データベースに凝縮することの詳細を示し、凝縮データベースと図２の処理システム内の機械学習モジュールとの間の対話の詳細を示す。

【図８】異なる表面パラメータ間の相関の評価と併用されるシミュレートされる表面の例を示す。

【図９Ａ】表面パラメータの重要係数のヒストグラムプロットを示す。

【図９Ｂ】表面パラメータの重要係数のヒストグラムプロットを示す。

【図９Ｃ】表面パラメータの重要係数のヒストグラムプロットを示す。

【図９Ｄ】表面パラメータの重要係数のヒストグラムプロットを示す。

【図９Ｅ】表面パラメータの重要係数のヒストグラムプロットを示す。

【図９Ｆ】表面パラメータの重要係数のヒストグラムプロットを示す。

【図１０】二分木を示す。

【図１１】分類に必要なデータ構造に適合された機械学習方法を示す。

【図１２】セル表面の汚染により生じ得るアーチファクトの代表的な例を示す。
【図１３】相関閾値への表面パラメータの数の依存性を示す。
【図１４】ランダムフォレスト法内で計算される高さ及び付着属性の表面パラメータの重要性の階層を示す。
【図１５】異なる数の表面パラメータと、高さ及び付着力の結合チャネルについてランダムフォレスト法を使用して計算されたトレーニングデータベース及びテストデータベース間のデータの異なる割り振りでの精度を示す。
【図１６】高さ及び付着力の結合チャネルについてランダムフォレスト法を使用した受信者動作特性を示す。
【図１７】図１６に示されるものと同様であるが、図１６におけるデータの生成に使用される手順の信頼性の確認に使用される人工データを有するプロットを示す。
【図１８】図１７の受信者動作特性下の面積を示す。
【図１９】患者毎に５つの細胞を使用し、癌罹患患者からのものとして識別されるには２つの細胞が必要である（Ｎ＝５、Ｍ＝２）場合、異なる数の表面パラメータ及び高さ及び付着力の結合チャネルについてのランダムフォレスト法を使用したトレーニングデータとテストデータとの間のデータ割り振りの異なる方法での精度を示す。
【図２０】患者毎に５つの細胞を使用し、癌罹患患者からのものとして識別されるには２つの細胞が必要である（Ｎ＝５、Ｍ＝２）場合、高さ及び付着力の結合チャネルでのランダムフォレスト法を使用して計算された受信者動作特性を示す。
【図２１】一方は高さであり、他方は付着力である２つの別個のチャネルでの癌診断に関連付けられた混同行列の統計を示す表である。
【発明を実施するための形態】
【００４８】
図１は、探針１４が取り付けられるカンチレバー１２を支持するスキャナ１０を有する原子間力顕微鏡８を示す。したがって、探針１４はスキャナ１０から片持ち梁のように支持される。スキャナ１０は、試料の表面１６の基準面に平行する走査方向に沿って探針１４を移動させる。そうするに当たり、スキャナ１０は試料の表面１６の領域を走査する。スキャナは、走査方向に探針１４を移動させている間、試料表面１６の基準面に垂直な垂直方向においても探針１４を移動させる。これは、探針１４から表面１６までの距離を変化させる。
【００４９】
探針１４は一般に、カンチレバー１２の反射部に結合される。この反射部は、レーザ２２により提供された照明ビーム２０を反射する。カンチレバー１２のこの反射部は、本明細書ではミラー１８と呼ばれる。反射ビーム２４はミラー１８から光検出器２６に移動し、光検出器２６の出力はプロセッサ２８に接続する。幾つかの実施形態では、プロセッサ２８は、表面の物理的特性又は幾何学的特性に基づいて表面パラメータをリアルタイムで計算できるようにするＦＰＧＡ電子回路を含む。
【００５０】
探針１４の移動はミラー１８の移動に転換され、これにより次に、光検出器２６の異なる部分が反射ビーム２４により照明されることになる。これは、探針の移動を示す探針信号３０を生成する。プロセッサ２８は、後述する方法を使用して探針信号３０に基づいて特定の表面パラメータを計算し、結果３３を記憶媒体３２に出力する。これらの結果３３は、本明細書に記載される任意の表面パラメータを表すデータを含む。
【００５１】
スキャナ１０はプロセッサ２８に接続し、スキャナ位置を示すスキャナ信号３４をプロセッサ２８に提供する。このスキャナ信号３４は、表面パラメータの計算への使用にも利用可能である。
【００５２】
図２は、処理システム２８を詳細に示す。処理システム２８は、インバータ６２に接続されたＡＣ源６０を有する電源５８を特徴とする。電源５８は、後述される種々の構成要

素を動作させる電力を提供する。処理システムは放熱器６４を更に含む。
【００５３】
好ましい実施形態では、処理システム２８は、人が動作を制御できるようにするユーザインターフェース６６を更に含む。
【００５４】
処理システム２８は、探針信号及びスキャナ信号を受信し、バス７２に配置する第１及び第２のＡ／Ｄ変換器６８、７０を更に含む。プログラム記憶セクション７４、ワーキングメモリ７６、及びＣＰＵレジスタ７８もバス７２に接続される。プログラム記憶装置７４からの命令７５を実行するＣＰＵ８０は、レジスタ７８及びＡＬＵ８２の両方に接続する。非一時的コンピュータ可読媒体は、これらの命令７５を記憶する。実行されると、命令７５は、処理システム２８に、第１及び第２のＡ／Ｄ変換器６８、７０を通して受信された入力に基づいて上記パラメータの何れかを計算させる。
【００５５】
処理システム２８は、図６に最もよく見られるように、機械学習モジュール８４と、トレーニングデータ８７及びテストデータ８９を含むデータベース８６とを更に含む。機械学習モジュール８４は、トレーニングデータ８７及びテストデータ８９を本明細書に記載される方法の実施に使用する。
【００５６】
処理システム２８の特定の例は、撮像サービスの属性の値及び／又は本明細書に記載される表面パラメータを特定するように構成された回路を含むＦＰＧＡ電子回路を含み得る。
【００５７】
図３は、原子間力顕微鏡８を使用して、画像を取得し、画像を使用して試料を特徴付ける機械学習モジュール８４に画像を提供するプロセスを示す。図３に示されるプロセスは、患者から尿８８を取得し、尿８８から剥がれ落ちた細胞９０を準備することを含む。走細胞を走査した後、原子間力顕微鏡８は、膀胱細胞９０の画像をデータベース８６への記憶に提供する。
【００５８】
各画像は、各要素が表面１６の属性を表すアレイである。アレイ内の場所は、試料の表面１６上の空間場所に対応する。したがって、画像は、その属性に対応するマップを定義する。そのようなマップは、土壌図が地表上の異なる場所における異なる土質を示すとの略同じように、試料の表面１６上の異なる場所におけるその属性の値を示す。そのような属性を「マップ属性（ｍａｐｐｅｄ　ｐｒｏｐｅｒｔｙ）」と呼ぶことにする。
【００５９】
幾つかの場合、マップ属性は物理的属性である。他の場合、属性は幾何学的属性である。幾何学的属性の一例は、表面１６の高さである。物理的属性の例には、表面の付着力、その剛性、及び表面１６への接触に関連付けられたエネルギー損失がある。
【００６０】
マルチチャネル原子間力顕微鏡８は、異なる属性を同時にマップ形成する能力を有する。各マップ属性は、顕微鏡８の異なる「チャネル」に対応する。したがって、画像は多次元画像アレイＭ（ｋ）と見なすことができ、ここで、チャネルインデックスｋは間隔［１，Ｋ］内の整数であり、ここで、Ｋはチャネルの数である。
【００６１】
準共振タッピングモードで使用される場合、マルチチャネル原子間力顕微鏡８は、以下の属性をマップ形成することができる：高さ、付着力、変形、剛性、粘弾性損失、フィードバックエラー。これは６つのチャネルを生成し、各チャネルは６つのマップ属性の１つに対応する。リンギングモードで使用される場合、原子間力顕微鏡８は、一例として、先の６つの属性に加えて以下の追加の属性の１つ又は複数をマップ形成することができる：回復した付着力、付着力高さ、切断高さ、プルオフネック（ｐｕｌｌ－ｏｆｆ　ｎｅｃｋ）高さ、分離距離、分離エネルギー損失、動的クリープ位相シフト、及びゼロ力高さ。この例では、これは合計で１４のチャネルを生成し、各チャネルは１４のマップ属性の１つ

に対応する。

【００６２】

スキャナ１０は、基準面上の離散ピクセルを定義する。各ピクセルにおいて、顕微鏡の探針１４は測定を行う。便宜上、平面上のピクセルはデカルト座標（x_i，y_j）により定義することができる。そのピクセルで測定される第ｋのチャネルの値は$z_{i,j}(k)$である。これを念頭に置き、第ｋのチャネルのマップ又は画像を表す画像アレイは、

【数１】

$$M^{(k)}=\{x_i, y_j, z_{ij}^{(k)}\} \qquad (1)$$

として正式に表すことができ、式中、「ｉ」及び「ｊ」はそれぞれ間隔［１，Ｎｉ］及び［１，Ｎｊ］内の整数であり、ここで、Ｎｉ及びＮｊは、ｘ及びｙ方向のそれぞれにおいて画像の記録に利用可能なピクセル数である。Ｎｉ及びＮｊの値は異なっていてもよい。しかしながら、本明細書に記載される方法は、そのような違いに有意に依存しない。したがって、考察のために、Ｎｉ＝Ｎｊ＝Ｎである。

【００６３】

試料の画像アレイ内の要素数は、チャネル数とピクセル数との積である。比較的同質の表面１６の場合、表面１６の１つの領域を走査するだけでよい。しかしながら、より異質な表面１６の場合、表面１６上の２つ以上の領域を走査することが好ましい。比喩的に言えば、港の水面を検査したい場合、十中八九、１つの領域を走査するだけでよく、その理由は、他の領域もどのみち同様である可能性が高いためである。他方、その港の町の表面を検査したい場合、複数の領域を走査することが賢明である。

【００６４】

これを念頭に置き、アレイは、走査中の特定の領域を識別する別のインデックスを取得する。これは、アレイの次元を増大させる。したがって、画像アレイの正式な表現は、

【数２】

$$M^{(k,s)}=\{x_i^{(s)}, y_j^{(s)}, z_{ij}^{(k,s)}\} \qquad (2)$$

であり、式中、走査領域インデックスｓは、試料内の特定の走査領域を識別する間隔［１，Ｓ］内の整数である。なお、これは、特定の試料の画像アレイ内の要素数を走査領域数に等しい倍数で増大させる。

【００６５】

好ましくは、そのような走査領域の数は、試料全体を表すのに十分に大きい。適切な走査領域数に収束する一方法は、２つのそのような走査領域間の偏差の分布を比較することである。走査領域数をインクリメントしても、これが統計学的に有意に変わらない場合、走査領域の数は、表面全体を表すのに適切である可能性が高い。別の方法は、妥当なテスト時間と見なされるものを、各走査領域の走査に必要とされる時間量で除し、その商をエリア数として使用することである。

【００６６】

幾つかの場合、各走査領域を区画に分割することが有用である。各走査領域にＰ個のそのような区画がある場合、アレイは、

【数３】

$$M^{(k,s,p)}=\{x_i^{(s,p)}, y_j^{(s,p)}, z_{ij}^{(k,s,p)}\} \qquad (2a)$$

として定義することができ、式中、区画インデックスｐは、間隔［１，Ｐ］内の整数である。正方形の走査エリアの場合、正方形を４つの正方形区画に分割し、したがって、Ｐを４に等しく設定することが好都合である。

【００６７】

走査領域を区画に分割する能力は、画像アーチファクトを除外する有用な方法を提供する。これは、生体細胞９０の検査に特に重要である。これは、検査に向けて細胞９０を準備するプロセスが、アーチファクトを容易に導入し得るためである。これらのアーチファクトは、いかなる分析からも除外されるべきである。これは、ある区画をその他の区画と比較して、もしあれば、除外するのに十分に有意に逸脱するものがどれかを識別できるようにする。

【００６８】

他方、新しいインデックスの追加は、アレイの次元を更に増大させる。

【００６９】

原子間力顕微鏡８により取得された画像アレイＭ（ｋ，ｓ）に基づいて試料が属するクラスを識別するために、機械学習モジュール８４は、特定のクラスＣ（ｉ）に属することが事前に既知である表面の画像を含む適したデータベース８６を構築することに部分的に頼る。そのようなデータベース８６は、

【数４】

$$D_k^{(class)}=\{M_n^{(k,s,p)},C^{(l)}\} \quad (2b)$$

により正式に表すことができ、式中、ｋは、属性又はチャネルを表すチャネルインデックスであり、ｓは、特定の走査領域を識別する走査領域インデックスであり、ｐは、第ｓの走査領域の特定の区画を表す区画インデックスであり、ｎは、特定の試料を識別する試料インデックスであり、ｌは、Ｌクラスの集合からの特定のクラスを識別するクラスインデックスである。したがって、アレイの全体サイズは、クラス数、試料数、走査領域数、走査領域当たりの区画数、及びチャネル数の積である。

【００７０】

図３は、準共振タッピングを使用して動作する原子間力顕微鏡８と、機械学習モジュール８４とを使用して、２つのクラス：癌罹患及び癌なしの一方に患者を分類するために、尿８８から回収された生体細胞９０の表面を検査することを特徴とする診断方法１０を示す。２つのクラスがあるため、Ｌ＝２である。

【００７１】

好ましい実施は、遠心力、重力沈殿、又は濾過を使用し、細胞９０を固定し、次に凍結乾燥又は亜臨界乾燥して細胞９０を収集することを含む。

【００７２】

示される例では、原子間力顕微鏡８は、Ｂｒｕｋｅｒ，Ｉｎｃ．により実施されるＰｅａｋＦｏｒｃｅＱＭＮ等の準共振タッピングモード及び例えば、ＮａｎｏＳｃｉｅｎｃｅＳｏｌｕｔｉｏｎｓ，ＬＬＣにより実施されるリンギングモードの領域を使用して動作した。両モードとも、高さ及び付着力チャネルを記録することができる。しかしながら、リンギングモードは、はるかに高速の画像収集モードである。先に述べたように、これらのモードでは、多くのチャネルが同時に記録することができる。しかしながら、本明細書に記載される実験では２つのみのチャネルが使用される。

【００７３】

図４は、原子間力顕微鏡のカンチレバー１２を患者から取得され、上述したように準備された細胞９０と一緒に示す。図は、原子間力顕微鏡８に結合された光学顕微鏡を通してとられた。

【００７４】

図５は、第１及び第２のマップ対９２、９４を示す。第１のマップ対９２は、非癌患者からの細胞９０のマップを示す。第２のマップ対９４は癌罹患患者からの細胞９０のマップを示す。示されるマップは、両次元で分解能５１２ピクセルを有する一辺が１０μｍの正方形走査エリアのものである。走査速度は、ＰｅａｋＦｏｒｃｅＱＭＮモード等の準共振タッピングモードで走査している場合、０．１Ｈｚであり、リンギングモードで走査している場合、０．４Ｈｚであった。走査中のピーク力は５ｎＮである。

【００７５】
これより図６を参照すると、機械学習モジュール８４は、データベース８６に基づいて候補分類器１００をトレーニングする。特定の機械学習法は、機械学習法の群、例えば、決定木、ニューラルネットワーク、又はそれらの組合せから選ぶことができる。
【００７６】
図６及び図７に示される方法は、データベース８６をトレーニングデータ８７及びテストデータ８９に分割することにより開始される。これは、データベース８６内のどの程度の量のデータがトレーニングデータ８７に行くべきであり、どの程度の量がテストデータ８９に行くべきかという問題を生じさせる。
【００７７】
幾つかの実施形態では、データベース８６の５０％はトレーニングデータ８７に行き、残りの５０％はテストデータ８９に行く。他の実施形態では、データベース８６の６０％はトレーニングデータ８７に行き、残りの４０％はテストデータ８９に行く。更に他の実施形態では、データベース８６の７０％はトレーニングデータ８７に行き、残りの３０％はテストデータ８９に行く。更に他の実施形態では、データベース８６の８０％はトレーニングデータ８７に行き、残りの２０％はテストデータ８９に行く。候補分類器１００は最終的に、分割に使用される比率から独立すべきである。
【００７８】
図３に示される例では、各患者で１０個の膀胱細胞９０を集めた。癌の存在は、侵襲的生検及び病理組織診断を含む標準の臨床方法を使用して識別した。これらの方法は、２つのクラスを明確に定義されていると見なすのに十分な信頼性を有する。その結果、図６に示されるデータベース８６は、
【数５】

$$D_1^{(1,k\times p)}=\{M^{(k\times p)}{}_1,C^{(1)}\},D_2^{(1,k\times p)}=\{M^{(k\times p)}{}_2,C^{(1)}\}\quad\cdots\quad D_{N\mathrm{data}1}^{(1,k\times p)}=\{M^{(k\times p)}{}_{N\mathrm{data}1},C^{(1)}\}$$
$$D_1^{(2,k\times p)}=\{M^{(k\times p)}{}_1,C^{(2)}\},D_2^{(2,k\times p)}=\{M^{(k\times p)}{}_2,C^{(2)}\}\quad\cdots\quad D_{N\mathrm{data}2}^{(2,k\times p)}=\{M^{(k\times p)}{}_{N\mathrm{data}2},C^{(2)}\},\qquad(3)$$

として表すことができ、式中、Ｎｄａｔａ１は、第１のクラス内の患者数であり、Ｎｄａｔａ２はクラス２内の患者数であり、ｓは、１と１０との間の整数（１及び１０を含む）であり、１人の患者から収集された１０個の細胞のうちの特定の１つを識別する。Ｎｄａｔａ１及びＮｄａｔａ２が等しい数である必要はない。
【００７９】
データベース８６をトレーニングデータ８７とテストデータ８９とに分割する際、同じ試料からの異なる走査エリアの画像アレイ
【数６】

$$\{M^{(k,1,p)},M^{(k,2,p)},\ldots M^{(k,S,p)}\}$$

をトレーニングデータ８７とテストデータ８９とに分けるのを回避することが重要である。このルールに違反すると、同じ試料でトレーニング及びテストが行われることになる。これは、分類器１００を独立した新しい試料に適用する場合、再現可能ではないことがあるように、分類器の有効性を人工的に膨らませる。
【００８０】
機械学習モジュール８４は、トレーニングデータ８７を使用して候補分類器１００を構築する。分類器１００のタイプに応じて、トレーニングデータ８７は、学習木、決定木、木のブートストラップ、ニューラルネットワーク、又はそれらの組合せであることができる。分類器１００は、以下「ＡＩ」として表され、特定の試料ｎが特定のクラスＩに属する確率を出力し、
【数７】

$$\mathrm{Prob}_n^{(k;s;p)(i)} = \mathrm{AI}(M_n^{(k;s;p)}|C^{(I)}) \qquad (3a)$$

式中、
【数８】
$\mathrm{Prob}_n^{(k;s;p)(i)}$

は、
【数９】
$M_n^{(k;s;p)}$

により定義された画像又はチャネルがクラスＣ（Ｉ）に属する確率である。
【００８１】
構築された後、検証モジュール１０２はテストデータ８９を使用して、候補分類器１００が実際に十分に有効であるかを検証する。本明細書に記載される実施形態では、検証モジュール１０２は、少なくとも部分的に受信者動作特性及び混同行列に基づいて有効性を評価する。候補分類器１００の堅牢性は、データベース８６のランダム分割を繰り返し、それにより、異なるテストデータ８９及びトレーニングデータ８７を生成し、次に、分類手順を実行して、これが任意の差を生じさせるか否かを調べることにより検証された。
【００８２】
候補分類器１００の有効性が不十分であると分かった場合、機械学習モジュール８４は、トレーニングプロセスのパラメータを変更し、新しい候補分類器１００を生成する。このサイクルは、機械学習モジュール８４が最終的に、所望の有効性閾値を達成する候補分類器１００を提供するまで続けられる。
【００８３】
適した分類器１００を構築するプロセスは、試料ｎに関連付けられた２つ以上の確率値がある場合に生じる計算負荷によりある程度妨げられる。実際には、画像アレイの多次元性により、任意の１つの試料について、プロセスするＫ・Ｓ・Ｐ個の確率Ｐｒｏｂ$_{n}$（ｋ；ｓ；ｐ）（ｉ）がある。必要とされる計算負荷は、そのような大きなデータベースでは非実際的に高くなる。
【００８４】
そのような大きなデータアレイに対処することの別のボトルネックは、分類器の妥当なトレーニングを提供するのに使用される多数の試料である。決定木を構築する場合、大体の目安では、データベースの次元の少なくとも６倍の試料数が必要とされる。原子間力顕微鏡は比較的低速の技法であるため、任意の妥当な分類器を構築するのに十分な試料を取得することは非実際的である。
【００８５】
図７に示されるように、凝縮器１０７は上記問題に対処する。凝縮器１０４は、特定のチャネルにより提供された情報を、そのチャネルについての情報を具現する表面パラメータの空間に凝縮する。凝縮器１０４は、データベース８６を受信し、凝縮データベース１０６を生成する。実際には、これは、かなり高次元の空間である多次元行列のはるかに少ない次元の行列への投影に相当する。
【００８６】
凝縮器１０４は、多種多様なデータベース縮小手順の何れかを実行する。これらの中には、本明細書に記載されるデータベース縮小手順の１つ又は複数を組み合わせた手順がある。これらは共通して、１組のデータから、その組で具現される情報の少なくとも幾つかを具現する表面パラメータを導出することを有する。
【００８７】
幾つかの実施では、凝縮器１０４は第１のデータベース縮小手順を実行する。この第１のデータベース縮小手順は、各画像が最終的に、参加するアレイからの情報の、試料の分

類で有用であるのに十分な側面を保存するという目的を達成するように、そのような他のアレイと結合することができるアレイであるという観測に依拠する。例えば、テンソル加算「

【数１０】

“⊕”

」を使用して、１組の画像M_n（ｋ；ｓ；ｐ）をそのインデックスの１つに対応するスライスと共に結合することができる。

【００８８】

特定の一実施態様では、スライスはインデックスｋに対応する。その場合、画像のテンソル和は

【数１１】

$$M_n^{(1\times p)} \oplus M_n^{(2\times p)} \oplus M_n^{(3\times p)} \oplus \ldots M_n^{(K\times p)}$$

により与えられる。

【００８９】

したがって、機械学習に使用される凝縮データベース１０６の各要素は以下になる。

【数１２】

$$D_n^{(1\times p)} = \{M_n^{(1\times p)} \oplus M_n^{(2\times p)} \oplus M_n^{(3\times p)} \oplus \ldots M_n^{(K\times p)}\} \qquad (3\text{-}1)$$

この特定の例は、データベース８６の次元を１／Ｋに低減する。したがって、分類器１００は以下のように確率を定義する。

【数１３】

$$\mathrm{Prob}_n^{(\times p)(l)} = \mathrm{AI}(M_n^{(1\times p)} \oplus M_n^{(2\times p)} \oplus M_n^{(3\times p)} \oplus \ldots M_n^{(K\times p)} | C^{(l)})$$

残りのインデックスに対して同様の手順を実行することも可能である。最終的には、

【数１４】

$$\mathrm{Prob}_n^{(l)} = \mathrm{AI}(\oplus\oplus\oplus M_n^{(k\times p)} | C^{(l)})$$

であり、式中、「

【数１５】

“⊕⊕⊕”

」は、インデックスｋ、ｓ、ｐにわたるテンソル総和を表す。

【００９０】

他の実施では、凝縮器１０４は代わりに、第２のデータベース縮小手順を実行する。この第２のデータベース縮小手順は、各インデックスｋ、ｓ、ｐに対する別個又は組合せでの幾何平均又は代数平均を計算することに依拠する。第２の手順を実行する具体的な方法の例には、全てのインデックスｋ、ｓ、ｐにわたる以下の平均計算手順がある。

【数１６】

$$\mathrm{Prob}^{(I)}{}_n = \frac{1}{K \times S \times P} \sum_{k,s,p} \mathrm{Prob}^{(k,s,p)(I)}{}_n \quad , \qquad (3\text{-}2)$$

$$\mathrm{Prob}^{(I)}{}_n = \frac{1}{\sqrt[3]{K \times S \times P}} \prod_{k,s,p} \mathrm{Prob}^{(k,s,p)(I)}{}_n \quad , \qquad (3\text{-}3)$$

$$\mathrm{Prob}^{(I)}{}_n = \frac{1}{K \times S \times P} \sum_{k,s,p} (1\text{-}\mathrm{Prob}^{(k,s,p)(I)}{}_n) \quad , \qquad (3\text{-}4)$$

$$\mathrm{Prob}^{(I)}{}_n = \frac{1}{\sqrt[3]{K \times S \times P}} \prod_{k,s,p} (1\text{-}\mathrm{Prob}^{(k,s,p)(I)}{}_n) \quad , \qquad (3\text{-}5)$$

【００９１】

更に他の実施では、凝縮器１０４は代わりに第３のデータベース縮小手順を実行する。この第３のデータベース縮小手順は、シリーズ全体の最高又は最低確率を特定のインデックスに割り当てることに依拠する。例えば、走査領域インデックスｓを考えると、以下の関係式の１つを使用することができる。

【数１７】

$$\mathrm{Prob}^{(k,s,p)(I)}{}_n = \underset{s}{Max}\{\mathrm{Prob}^{(k,s,p)(I)}{}_n\} \quad , \qquad (3\text{-}6)$$

$$\mathrm{Prob}^{(k,s,p)(I)}{}_n = \underset{s}{Min}\{\mathrm{Prob}^{(k,s,p)(I)}{}_n\} \quad , \qquad (3\text{-}7)$$

【００９２】

最終的に、全てのインデックスがこのようにして縮小される場合、

【数１８】

$$\mathrm{Prob}^{(I)}{}_n = \underset{k,s,p}{Max}\{\mathrm{Prob}^{(k,s,p)(I)}{}_n\} \quad \text{or} \qquad (3\text{-}8)$$

$$\mathrm{Prob}^{(I)}{}_n = \underset{k,s,p}{Min}\{\mathrm{Prob}^{(k,s,p)(I)}{}_n\} \quad . \qquad (3\text{-}9)$$

である。

【００９３】

幾つかの実施では、凝縮器１０４は、各画像を表面パラメータ抽出機Ａｍに通して表面パラメータセットＰｎｍ（ｋ；ｓ）を取得することにより、データベースＤｎ（ｋ；ｓ）の次元を低減する。これは

【数１９】

$$P_{nm}^{(k,s)} = A_m\{M_n^{(k,s,p)}\} \quad (4)$$

により正式に表すことができ、式中、表面パラメータインデックスｍは、［１，Ｍ］内の整数であり、チャネルインデックスｋは、マップが表すのが高さであるか、付着力であるか、剛性であるか、それとも何らかの他の物理的又は幾何学的パラメータであるかを識別し、試料インデックスｎは試料を識別し、走査領域インデックスｓは、試料内の特定の走査領域を識別し、区画インデックスｐは、走査領域内の特定の区画を識別する。この手順は、表面パラメータベクトルＰｎｍ（ｋ；ｓ；ｐ）として多次元テンソルＭｎ（ｋ；ｓ；ｐ）

を表すコンパクトな方法を提供する。

【００９４】

表面パラメータベクトルは、分類の基礎として使用可能な、導出元のチャネルに関する十分な残存情報を含む。しかしながら、チャネルにより提供される画像よりもはるかに小さい。したがって、表面パラメータベクトルに頼る分類手順は、はるかに低い計算負荷を維持するが、対応する精度損失はない。

【００９５】

多種多様な表面パラメータをチャネルから抽出することができる。これらは、平均粗さ、二乗平均平方根、表面歪度、表面尖度、ピークツーピーク、十点平均高さ、最大谷深さ、最大ピーク高さ、平均値、平均サミット曲率、テクスチャインデックス、二乗平均平方根勾配、面積二乗平均平方根傾き、表面面積比、投影面積、表面面積、表面ベアリングインデックス、コア溶液保持インデックス、バレー溶液保持インデックス、縮小サミット高さ、コア粗さ深さ、縮小バレー深さ、ベアリング曲線のＩ－ｈ％高さ間隔、サミット密度、テクスチャ方向、テクスチャ方向インデックス、優勢放射状波長、放射状波インデックス、平均半波長、フラクタル次元、２０％時の相関長、３７％時の相関長、２０％時のテクスチャアスペクト比、３７％時のテクスチャアスペクト比を含む。

【００９６】

表面パラメータのリストは、アルゴリズム又は数式を導入することにより更に拡張し得る。例えば、例えば各パラメータを表面面積の関数で除することにより、表面パラメータを、細胞ごとに異なりうる画像の表面面積に正規化することができる。

【００９７】

本明細書に記載される例は、３つの表面パラメータに依拠する：バレー流体保持インデックス（「Ｓｖｉ」）、表面面積比（「Ｓｄｒ」）、及び表面面積（「Ｓ３Ａ」）。

【００９８】

バレー溶液保持インデックスは、バレーゾーンにおける大きなボイドの存在を示す表面パラメータである。これは、

【数２０】

$$Svi = \frac{V(h_{0.80})}{(M-1)(N-1)\delta x \delta y} / Sq \tag{5}$$

により定義され、式中、Ｎは、ｘ方向におけるピクセル数であり、Ｍは、ｙ方向におけるピクセル数であり、Ｖ（h_x）は、水平線h_x下のベアリング面積に対するボイド面積の比率の曲線であり、Ｓｑは平均二乗平方根（ＲＭＳ）であり、以下の式

【数２１】

$$Sq = \sqrt{\frac{1}{MN}\sum_{k=0}^{N-1}\sum_{l=0}^{M-1}\left[h(x_k, y_l)\right]^2} \tag{6}$$

により定義される。

【００９９】

表面面積比（「Ｓｄｒ」）は、投影されたｘ、ｙ平面の面積に対する界面表面面積のインクリメントを表す表面パラメータである。この表面パラメータは、

【数２２】

$$S_{sd} = \frac{\left(\sum_{k=0}^{M-2}\sum_{l=0}^{N-2} A_{kl}\right) - (M-1)(N-1)\delta x \delta y}{(M-1)(N-1)\delta x \delta y} 100\% \tag{7}$$

により定義され、式中、Ｎは、ｘ方向におけるピクセル数であり、Ｍは、ｙ方向におけるピクセル数である。

【０１００】

表面面積（「Ｓ３Ａ」）は

【数２３】

$$S3A=\frac{\left(\sum_{k=0}^{N-2}\sum_{l=0}^{N-2}A_{kl}\right)-(M-1)(N-1)\delta x\delta y}{} \quad (8)$$

により定義される。

【０１０１】

原子間力顕微鏡８により提供される画像からの上述した３つの表面パラメータのそれぞれを計算するために、細胞の各画像はまず、４つの区画に分割される。区画は、この場合、５μｍ辺を有する正方形の象限であった。したがって、各細胞は、各象限に１組ずつ、４組の表面パラメータを生成した。

【０１０２】

細胞におけるアーチファクトの存在は、３つの異なる方法の何れか１つで対処することができる。

【０１０３】

第１の方法は、オペレータにアーチファクトについて細胞を検査させ、１つ又は複数のそのようなアーチファクトを有したあらゆる細胞をそれ以上の処理から除外することである。これは、アーチファクトの識別に人間の介入を必要とする。

【０１０４】

第２の方法は、アーチファクトを認識し、そのアーチファクトを含む細胞を自動的に除外することが可能なアーチファクト認識モジュールを提供することである。これは、手順をオペレータからより独立したものにする。

【０１０５】

第３の方法は、平均値の代わりに各細胞のパラメータのメジアン値を使用することである。本明細書に記載される結果は、平均値の代わりにメジアン値を使用した場合でも略変わらなかった。

【０１０６】

２つのみのクラスの同じ例を使用すると、凝縮データベース１０６は以下のように見える。

【数２４】

$$\begin{matrix} D_1^{(0,0,cp)}=\{P^{(0,0,cp)}{}_1,C^{(1)}\},D_2^{(0,0,cp)}=\{P^{(0,0,cp)}{}_2,C^{(1)}\} & \cdots & D_{Nbins1}^{(0,0,cp)}=\{P^{(0,0,cp)}{}_{Nbins1},C^{(1)}\} \\ D_1^{(2,0,cp)}=\{P^{(2,0,cp)}{}_1,C^{(2)}\},D_2^{(2,0,cp)}=\{P^{(2,0,cp)}{}_2,C^{(2)}\} & \cdots & D_{Nbins2}^{(2,0,cp)}=\{P^{(2,0,cp)}{}_{Nbins2},C^{(2)}\} \end{matrix} \quad (9)$$

【０１０７】

他の実施形態では、原子間力顕微鏡の画像に直接関連しないが、異なるクラスの区別を助けるために追加のパラメータを割り当てることができる。

【０１０８】

例えば、膀胱癌を検出しようとする場合、尿８８の１つ又は複数の試料がいかなる細胞９０も有さないことが十分に生じ得る。そのような結果を考慮した好都合な方法は、真又は偽の何れか一方である新しい「細胞なし」パラメータを追加することである。そのようなパラメータに適合するためのデータ構造を変更する必要性を回避するために、「細胞なし」が「真」に設定された試料は、統計結果の歪ませるのを回避するように選択された表面パラメータの人工値を受け取る。

【０１０９】

別の例として、表面パラメータに関連しないが、それにも関わらず、分類に関連する他のファクタがある。これらには、年齢、喫煙、及び家族歴のような患者の特徴があり、これらは全て、その患者が膀胱癌を有する確率に関連し得る。これらのパラメータは、データ構造を変更する必要性を回避するために、「細胞なし」パラメータと同様にして含むことができる。

【０１１０】

表面パラメータを使用してデータベース８６のサイズを低減する更に他の方法が存在する。

【０１１１】

そのような一手順は、互いに十分に相関する表面パラメータを除外するというものである。幾つかの表面パラメータは、種々の他の表面パラメータに強く依存する。したがって、互いに相関する表面パラメータを包含することにより提供される追加情報はごく僅かである。これらの冗長表面パラメータは、ペナルティが殆どなく削除することができる。

【０１１２】

表面パラメータ間の相関行列を見つける一方法は、シミュレートされた表面を生成することであり、その例を図８に示す。原子間力顕微鏡８を用いて撮像された種々の試料表面は、異なる表面パラメータ間の相関を識別するのにも使用することができる。

【０１１３】

機械学習モジュール８４は、入力の性質に依存しない。したがって、画像アレイで動作して示されているが、代わりに表面パラメータベクトルで動作することが完全に可能である。したがって、同じ機械学習モジュール８４は、特定の表面パラメータベクトルが特定のクラスに属する確率を特定するのに、すなわち、

【数２５】

$$\mathrm{Prob}_n^{(k;s;p)(l)} = \mathrm{AI}(P_n^{(k;s;p)}|C^{(l)})$$

を評価するのに使用可能である。

【０１１４】

したがって、多次元画像アレイ$M_n(k;s;p)$を表面パラメータベクトル$P_{nm}(k;s;p)$に縮小した後、多次元画像アレイ$M_n(k;s;p)$の代わりに表面パラメータベクトル$P_{nm}(k;s;p)$を使用し、次に、機械学習モジュール８４に、分類に重要な表面パラメータがどれか及び細胞の分類に表面パラメータをいかに使用するかを学習させることが可能になる。

【０１１５】

特定の表面パラメータは互いと相関するため、次元を更に低減することが可能である。これは、テンソル総和なしで実行することができる。代わりに、そのような低減は、異なる画像からの同じパラメータの直接操作により実行される。

【０１１６】

（３－１）～（３－９）として先に識別したデータベース縮小手順に頼る方法に加えて、同じ試料からの同じ種類の異なる表面パラメータを結合する分類器１００を使用することも可能である。正式には、このタイプの分類器１００は

【数２６】

$$\mathrm{Prob}_n^{(l)} = \mathrm{AI}(P_n|C^{(l)}) \quad (10)$$

として正式に表すことができ、式中、$P_n = F(P_{nm}(k;s;p))$であり、$F(P_{nm}(k;s;p))$は、試料インデックスｎにより識別された試料に属する、表面パラメータインデックスｍにより識別された異なる表面パラメータの結合である。

【０１１７】

関連する分類器１００は、同じ属性の画像からの同じ試料ｎの同じ種類ｍの異なる表面

パラメータを結合するものである。そのような分類器１００は、
【数２７】

$$\mathrm{Prob}_n^{(k)(l)} = \mathrm{AI}(P_{nm}^{(k)}|C^{(l)}) \qquad (11)$$

として正式に表すことができ、式中、Ｐｎｍ（ｋ）＝Ｆ（Ｐｎｍ（ｋ；ｓ；ｐ））であり、Ｆ（Ｐｎｍ（ｋ；ｓ；ｐ））は、チャネルインデックスｋにより識別されたチャネルからの、同じインデックスｎにより識別される試料の同じ表面パラメータインデックスｍにより識別される異なる表面パラメータの結合である。
【０１１８】
更に別の分類器１００は、パラメータを全ては結合せず、代わりに１つのみのインデックスによる表面パラメータを結合するものである。そのような一分類器１００は、１つの表面パラメータを同じ画像内の区画ｐの全体シリーズに割り当てる。そのような分類器１００は、
【数２８】

$$\mathrm{Prob}_n^{(k,s)(l)} = \mathrm{AI}(P_{nm}^{(k,s)}|C^{(l)}) \quad (12)$$

として正式に表され、式中、Ｐｎｍ（ｋ；ｓ）＝Ｆ（Ｐｎｍ（ｋ；ｓ；ｐ））であり、Ｆ（Ｐｎｍ（ｋ；ｓ；ｐ））は、表面パラメータの結合であり、その例には、区画インデックスにわたるＰｎｍ（ｋ；ｓ；ｐ）の統計分布に関連付けられたパラメータがある。例には、平均：
【数２９】

$$P_n^{(k,s)}{}_m = \frac{1}{N}\sum_{p=1}^{N} P_n^{(k,s,p)}{}_m \qquad (13)$$

及びメジアン
【数３０】

$$P_{nm}^{(k,s)} = \mathrm{median}\{P_n^{(k,s,p)}\} \text{ for } p=1\ldots N \qquad (14)$$

がある。
【０１１９】
各患者からの複数の細胞の膀胱癌検出撮像と併せて使用される場合、分類器１００は平均又はメジアンの何れか一方に頼る。しかしながら、メディアはアーチファクトの影響を受けにくいため、分類器１００が平均よりもメジアンに頼ることが好ましい。
【０１２０】
本明細書に記載される特定の実施形態では、機械学習モジュール８４は、多種多様な機械学習法の何れかを実施する。しかしながら、複数のパラメータに直面する場合、機械学習モジュール８４は容易に過剰トレーニングになり得る。したがって、過剰トレーニングを起こしにくい３つの方法、すなわち、ランダムフォレスト法、エクストリームランダム化フォレスト（Ｅｘｔｒｅｍｅｌｙ　Ｒａｎｄｏｍｉｚｅｄ　Ｆｏｒｅｓｔ）法、及び勾配ブースティングツリー法を使用することが有用である。
【０１２１】
ランダムフォレスト法及びエクストリームランダム化フォレスト法は、ブートストラップ教師なし法である。勾配ブースティングツリー法は、ツリーを構築する教師あり法である。変数ランキング、分類器トレーニング、及び検証は、ＳＣＩＫＩＴ－ＬＥＡＲＮ　Ｐｙｔｈｏｎ機械学習パッケージ（バージョン０．１７．１）からの適切な分類器関数を使用して実行した。
【０１２２】

ランダムフォレスト及びエクストリームランダム化フォレスト法は、多くの分類木の成長に基づく。各分類木は何らかの分類を予測する。しかしながら、全ての木の投票が最終分類を定義する。木はトレーニングデータ８７で成長する。典型的なデータベース８６では、全データの７０％がトレーニングデータ８７であり、残りはテストデータ８９である。本明細書に記載される実験では、トレーニングデータ８７とテストデータ８９との分割はランダムであり、複数回繰り返されて、分類器１００が、データベース８６が分割された様式の影響を受けないことを確認した。

【０１２３】

各分岐ノードは、元の表面パラメータのランダムに選ばれたサブセットに頼る。本明細書に記載される方法では、元の表面パラメータの選ばれたサブセット内の要素数は、元々提供された表面パラメータ数の平方根である。

【０１２４】

次に、学習プロセスは、表面パラメータのランダムに選ばれたサブセットが与えられた木分岐の最良分割を識別することによって進む。機械学習モジュール８４は、分割閾値が分類誤差の推定に基づくことを基礎とする。各パラメータは、トレーニングデータ８７の最も一般的に生じるクラスに関してパラメータ領域に割り当てられる。これらの実施では、機械学習モジュール８４は、その領域内の、最も一般的なクラスに属さないトレーニングデータ８７の割合として分類誤差を定義し、

【数３１】

$$E = 1 - \max_{k}(p_{mk}) \tag{15}$$

式中、p_{mk}は、第mの領域内にあり、且つ第kのクラスにも属するトレーニングデータ８７の割合を表す。しかしながら、実際用途では、式（１）は、木の過成長を回避するのに十分には反応しない。その結果、機械学習モジュール８４は２つの他の尺度に頼る：ジニインデックス及び交差エントロピー。

【０１２５】

ジニインデックスは、K個全てのクラスにわたる分散の尺度であり、以下のように定義される。

【数３２】

$$G = \sum_{k=1}^{K} p_{mk}(1 - p_{mk}) \tag{16}$$

【０１２６】

ジニインデックスは、p_{mk}の全ての値がゼロ又は１に近いままである場合、小さいままである。その結果、ジニインデックスは、特定のノードが主に１つのクラスからの試料を含む程度を測定する。これは「ノード純度」の程度と呼ばれる。したがって、過成長を回避するために、各木は、ジニインデックスがクラスを完全に分離させるまでのみ成長する。これは、２つの子孫ノードが、親ノードよりも小さいジニインデックスを生成する場合、生じる。これらのランダムフォレスト法では成長する分岐のプルーニングはない。

【０１２７】

交差エントロピーは、これもまたノード純度の尺度を提供し、

【数３３】

$$D = -\sum_{k=1}^{K} p_{mk}\log(p_{mk}) \tag{17}$$

として定義される。

【０１２８】

ジニインデックスのように、交差エントロピーは、ｐｍｋの全ての値がゼロに近い場合、小さい。これは純粋なノードを示す。

【０１２９】

ジニインデックスはまた、各表面パラメータの重要度を示す「重要度係数」を取得する方法も提供する。そのような一尺度は、各変数の木ノードにおけるジニインデックスの低下の全ての値を加算し、全ての木にわたる平均を計算することから生じる。

【０１３０】

図９に示されるヒストグラムは、平均から１標準偏差だけずれる程度を示すエラーバーを有する重要度係数の平均値を表す。これらの重要度係数は、特定のチャネルから導出することができる種々の表面パラメータに対応する。したがって、最初の行のヒストグラムは、特徴「高さ」を測定するチャネルから導出することができる表面パラメータを表し、一方、２番目の行の表面パラメータは、特徴「付着力」を測定するチャネルから導出することができる表面パラメータを表す。なお、ニーモニックデバイスを使用して特徴の名称を決定し、「高さ」チャネルから導出可能な全ての表面パラメータは、「ｈ」で始まり、「付着力」チャネルから導出可能な全ての表面パラメータは「ａ」で始まる。

【０１３１】

したがって、最初の行では、最初の列のパネルは、機械学習モジュール８４がランダムフォレスト法を使用する場合の「高さ」チャネルから導出された表面パラメータの重要度係数を示し、２番目の列のパネルは、機械学習モジュール８４がエクストリームランダム化フォレスト法を使用する場合の「高さ」チャネルから導出された表面パラメータの重要度係数を示し、３番目の列のパネルは、機械学習モジュール８４が勾配ブースティングツリー法を使用する場合の「高さ」チャネルから導出された表面パラメータの重要度係数を示す。

【０１３２】

同様に、２番目の行では、最初の列のパネルは、機械学習モジュール８４がランダムフォレスト法を使用する場合の「付着力」チャネルから導出された表面パラメータの重要度係数を示し、２番目の列のパネルは、機械学習モジュール８４がエクストリームランダム化フォレスト法を使用する場合の「付着力」チャネルから導出された表面パラメータの重要度係数を示し、３番目の列のパネルは、機械学習モジュール８４が勾配ブースティングツリー法を使用する場合の「付着力」チャネルから導出された表面パラメータの重要度係数を示す。

【０１３３】

図９のヒストグラムは、試料を正確に分類するのに最も役立つ表面パラメータを選ぶ知的な方法を提供する。例えば、仮に機械学習モジュール８４が、高さを測定する２つのみの表面パラメータを選ぶように強いられた場合、恐らく、「ｈ＿Ｓｙ」及び「ｈ＿Ｓｔｄ」を選ぶのを避けるが、代わりに「ｈ＿Ｓｓｃ」及び「ｈ＿Ｓｆｄ」を選ぶことを好み得る。

【０１３４】

図９の重要度係数は、１００～３００本の木を使用して得られた。元の表面パラメータの選ばれたサブセット内の要素の最大数は、元々提供された表面パラメータ数の平方根であり、ジニインデックスは、分類誤差を評価する基礎を提供した。機械学習手順の選択が、特定の表面パラメータの重要度に大きな差を生じさせないことが、同じ行中のヒストグラムの比較から明らかである。

【０１３５】

図１０は、ブートストラップ法で使用される１００本～３００本の木のアンサンブルからの二分木の一例を示す。最初の分割では、第４の変数「Ｘ［４］」は、分割値１５．０００１で選ばれた。これはジニインデックス０．４９９２をもたらし、７３個の試料をそれぞれ３０個及び４３個の試料を有する２つのビンに分割した。

【０１３６】

第２のレベルの分割では、左側ノードを見ると、第６の変数「Ｘ［６］」は分割値１４

．８０５９で選ばれ、ジニインデックス０．２７７８をもたらし、３０個の試料（クラス１に５個及びクラス２に２５個）を２７個及び３個の試料を有する２つのビンに分割した。分割は、木ノードがジニインデックス０を有し、したがって、２つのクラスの一方のみの存在を示すまで続けられる。

【０１３７】

エクストリームランダム化ツリー法は、分割の選択においてランダムフォレスト法と異なる。ランダムフォレスト法の場合のように最適なパラメータ及びジニインデックスを使用した分割結合を計算する代わりに、エクストリームランダム化フォレスト法を使用した機械学習モジュール８４は、パラメータ経験範囲から各パラメータ値をランダムに選択する。これらのランダム選択が最終的に、ゼロジニインデックスを有する純粋ノードに収束することを保証するために、機械学習モジュール８４は、現在の木が選ばれる１組の選択された変数内のランダムの一様な分割の中から最良の分割のみを選ぶ。

【０１３８】

幾つかの実施では、機械学習モジュール８４は勾配ブースティングツリー法を実施する。この場合、機械学習モジュール８４は、それぞれが何らかの費用関数に関して収束する一連の木を構築する。機械学習モジュール８４は、例えば、平均二乗誤差を最小にすることにより厳密な予測からのずれを最小にするように各後続木を構築する。幾つかの場合、機械学習モジュール８４は、このタイプの回帰にフリードマンプロセスに頼る。この回帰プロセスに適した実装は、「ＳＣＩＫＩＴ－ＬＥＡＲＮ　ＰＹＴＨＯＮ」パッケージで実装されるルーチン「ＴＲＥＥＢＯＯＳＴ」を使用して実行することができる。

【０１３９】

勾配ブースティングツリー法は純粋ノードについての基準がないため、機械学習モジュール８４は木のサイズを予め定義する。代替的には、機械学習モジュール８４は、個々の回帰の数を制限し、したがって、木の最大深さを制限する。

【０１４０】

生じる困難さは、予め定義されたサイズを用いて構築された木が容易に過剰適合し得ることである。この問題の影響を最小にするために、機械学習モジュール８４がブースティング反復回数のような数量に制約を課し、又は例えば、無次元学習率パラメータを使用することにより反復率を弱めることが好ましい。代替の実施では、機械学習モジュール８４は、木の末端ノード又はリーフの最小数を制限する。

【０１４１】

ＳＣＩＫＩＴ－ＬＥＡＲＮ　ＰＹＴＨＯＮパッケージに頼る本明細書に記載される実施態様では、機械学習モジュール８４は、リーフの最小数を１に設定し、最大深さを３に設定する。人間の被験者から収集された膀胱細胞を分類する本明細書に記載される用途では、機械学習モジュール８４は、非常に低い学習率０．０１の選択を検討することにより学習能力を抑制する。その結果生成される低速学習手順は、被験者が少数の人間であり、したがって、試料が少数であることから、生じる分散を低減する。

【０１４２】

トレーニングデータ８７及びテストデータ８９を作成するに当たり、トレーニングデータ８７とテストデータ８９との間で集合｛Ｍ（ｋ；１；ｐ），Ｍ（ｋ；２；ｐ），・・・，Ｍ（ｋ；Ｓ；ｐ）｝を分割するのを回避することが重要である。図１１に開示される手順はこれを回避する。

【０１４３】

膀胱細胞９０を分類する特定の実施態様では、各患者は幾つかの細胞を提供し、各細胞９０の画像は４つの区画に分割される。人間である観測者は、アーチファクトを見つけるために区画を視覚的に検査し、アーチファクトのうちの２つを図１２で見て取ることができる。アーチファクトが区画に存在することが分かった場合、画像を検査した人が誰であれ、その区画を無視すべきものとしてフラグを付ける。

【０１４４】

このプロセスは、多くの細胞９０が関わる場合、退屈になり得る。式（１０）に示され

る分類器１００を使用し、４つの区画のメジアンをとることにより、このプロセスを自動化することができる。これは、アーチファクトの寄与を大幅に下げる。

【０１４５】

機械学習モジュール８４は、データのＳ％がトレーニングデータ８７になり、１００－Ｓ％がテストデータ９８になるように、データベース８６をランダムに分割する。Ｓを５０％、６０％、及び７０％に設定して実験を実行した。機械学習モジュール８４は、同じ個人からのデータを全体的にトレーニングデータ８７又はテストデータ９８の何れか一方に維持して、同じ個人の異なる細胞９０間の相関から生じ得る人工過剰トレーニングを回避するようにデータベース８６を分割する。

【０１４６】

次に、機械学習モジュール８４は、凝縮器１０４に分類に頼る表面パラメータ数を更に低減させる。幾つかの実施では、凝縮器１０４は、各ジニインデックスに基づいて特定のチャネル内の表面パラメータをランク付けし、そのチャネルの何らかの数M_p個の最良パラメータを保持することによりそうする。幾つかの実施では、最良パラメータは、分離力の能力及び他の表面パラメータとの低い相関に基づいて選択される。例えば、パラメータ間相関閾値を変更することにより、分類に頼る表面パラメータ数を変更することが可能になる。

【０１４７】

図１３は、相関係数の閾値の変更が、ランダムフォレスト法を使用して選択される表面パラメータの数にいかに影響するかを示し、左端のパネルは、高さチャネルから利用可能な表面パラメータに対応し、中央のパネルは、付着力チャネルから利用可能は表面パラメータに対応する。縦の目盛りの変更から明らかなように、右端のパネルは、高さチャネル及び付着力チャネルの組合せを表す。図１３はランダムフォレスト法に固有であるが、その他の方法も同様の曲線を有する。

【０１４８】

木がトレーニングされると、テストデータ９８で正確に分類する能力をテストし、又は代替的には、未知の試料の分類に使用することが適切である。分類プロセスは、木投票の結果を取得することと、試料が属するクラスを示す確率の基礎としてその結果を使用することとを含む。次に、この結果は、どの誤差が耐えられるかに基づいて設定された分類器閾値と比較される。この分類器閾値は通常、受信者動作特性の構築の一環として変更される。

【０１４９】

一実験では、２５人の癌罹患患者及び４３人の非癌患者から尿８８の試料が収集された。癌罹患患者のうち、ＴＵＲＢＴにより定義されるように、１４人は低悪性度であり、１１人は高悪性度であった。非癌患者は健康であったか、又は過去に癌に罹患したことがあった。原子間力顕微鏡８に結合された光学顕微鏡を使用して、人間である観測者は、細胞であるように見えた丸い物体をランダムに選択した。

【０１５０】

データベースは、式（１４）で参照されるデータ縮小プロセスを使用することにより更に縮小された。したがって、その結果生成された恐らく生成器１００は、$P_{nm}(k;s)=\mathrm{median}\{P_{nm}(k;s;p)\}$であり、式中、ｐは、各画像の４つの区画に対応する１と４との間の整数（１及び４を含む）である。その結果生成された凝縮データベースは、２つのクラスを有し、

【数３４】

$$D_1^{(1:n)}=\{P_1^{(1:n)}{}_m,C^{(1)}\},D_2^{(1:n)}=\{P_2^{(1:n)}{}_m,C^{(1)}\}\ \ldots\ D_{Num1}{}^{(1:n)}=\{P_{Num1}{}^{(1:n)}{}_m,C^{(1)}\}$$
$$D_1^{(2:n)}=\{P_1^{(2:n)}{}_m,C^{(2)}\},D_2^{(2:n)}=\{P_2^{(2:n)}{}_m,C^{(2)}\}\ \ldots\ D_{Num2}{}^{(2:n)}=\{P_{Num2}{}^{(2:n)}{}_m,C^{(2)}\} \quad (18)$$

として正式に表すことができる。

【０１５１】

患者当たり少なくとも５個の細胞が撮像された。簡潔にするために、２つのみの属性を考慮した：高さ及び付着力。

【０１５２】

図１４は、ランダムフォレスト法内で計算された高さ及び付着力属性の表面パラメータの重要度の階層を示す。図は、平均前後の１標準偏差を示すエラーバーと共に重要度係数の平均を示す。データベース８６は、１０００回、トレーニングデータ８７及びテストデータ８９にランダムに分割された。

【０１５３】

高さ及び付着力のマップ属性は、テンソル加算を通して結合され、テンソル加算は基本的に、表面パラメータのベクトルに適合されたデータ縮小法（３－１）である。関連するテンソル加算演算は、

【数３５】

$$P_{mn}^{(1:s)} \oplus P_{mn}^{(2:s)}$$

により表される。

【０１５４】

図９の場合と同様に、図１４の各表面パラメータは、その名称として、表面パラメータの標準名を有するが、導出元のマップ属性を示す文字が先頭に追加されている。例えば、「ａ＿Ｓｄｓ」は、付着力（ａｄｈｅｓｉｏｎ）属性の画像から導出された「Ｓｄｓ」パラメータを意味する。

【０１５５】

ランダムフォレスト法に適した統計性能尺度は、受信者動作特性及び混同行列を調べることからもたらされる。受信者動作特性は、感度及び特異性の範囲を定義できるようにする。感度の範囲は、細胞を癌罹患患者からのものであると分類する場合の「精度」に対応し、一方、特異性は、細胞が非癌患者からのものであると分類される場合の「精度」に対応する。受信者動作特性は、受信者動作特性を使用して、以下のように特異性の範囲及び感度の範囲を定義できるようにし、

【数３６】

$$\begin{aligned} &\text{sensitivity} = TP/(TP+FN);\\ &\text{specificity} = TN/(TN+FP);\\ &\text{accuracy} = (TN+TP)/(TP+FN+TN+FP), \end{aligned} \qquad (19)$$

式中、ＴＮ、ＴＰ、ＦＰ、ＦＮは、真陰性、真陽性、偽陽性、及び偽陰性をそれぞれ表す。

【０１５６】

図１５は、それぞれが、異なる数の表面パラメータを考慮することにより達成される精度を示す３つの異なる曲線を示し、表面パラメータは、上述したように、異なる自己相関閾値及び重要度係数を選ぶことに基づいて選ばれた。

【０１５７】

図１５における３つの異なる曲線のそれぞれは、トレーニングデータ８７とテストデータ８９との間の１０００回のランダム分割を通して達せられた。曲線は、各セットへのデータの割り振りが異なる。最初の曲線は、データの７０％がトレーニングデータ８７に割り振られ、３０％がテストデータ８９に割り振られたことに対応する。２番目の曲線は、データの６０％のみがトレーニングデータ８７に割り振られ、４０％がテストデータ８９に割り振られたことに対応する。そして３番目の曲線は、トレーニングデータ８７とテストデータ８９とに均等に分割されたことに対応する。

【０１５８】

図１５を調べることから、特定の閾値分割に依存性が略ないことが明らかである。これは、機械学習モジュール８４により実行される手順の堅牢性を示す。

【０１５９】
図１６は、受信者動作特性の群を示す。図１６に示される特性群内の個々の受信者動作特性は、トレーニングデータ８７及びテストデータ８９へのデータベース８６の２００回の異なるランダム分割から生じた。
【０１６０】
各受信者動作特性は、２つのクラス間で分類しようとする場合、異なる閾値での感度及び特異性を示す。図１６のプロットを二分する斜線は、コインを投げることにより分類する分類器に相当する。したがって、受信者動作特性が図１６に示される斜線に近づくほど、分類器は分類が不良である。曲線がこの斜線から遠くにまとまり、個々の曲線間にばらつきが少ないことは、分類器の有効性並びにトレーニングデータ８７及びテストデータ８９の特定の選択の影響の受けにくさの両方を示唆する。
【０１６１】
受信者動作特性を構築するに当たり、特定の確率値が一方のクラスに対応するか、それとも他方のクラスに対応するかを定義する閾値は、自由パラメータである。このパラメータの選択は、特異性及び感度の両方を支配する。各受信者動作特性で、第１のクラスであるべきであった試料を第２のクラスに分類し、またこの逆に分類する最小誤差に対応する点が存在する。これは、１つのチャネルを使用する場合に使用される３つの機械学習法のそれぞれについて図２１に示される。
【０１６２】
図２１に示される表中の各行は、収集された細胞の特定の数（Ｎ）及び診断の閾値として使用されたより小さな数（Ｍ）を特徴とする。各行に、２つのチャネルが考慮された：高さ及び付着力。使用される３つの機械学習法のそれぞれで、表は、データベースの７０％がトレーニングデータに割り振られたトレーニングデータ及びテストデータへのデータベースの１０００回のランダム分割の平均ＡＵＣ並びに精度を示す。精度は、分類の最小誤差に関連付けられたものである。図２１における各行は、感度及び特異性も示す。
【０１６３】
原理上、感度及び特異性は、感度及び特異性が等しい平衡点の前後に定義することもできる。人間の被験者数は限られているため、この平衡点が精密にどこであるかを定義することは難しい。したがって、図２１では、均等要件は緩和され、感度と特異性との間の差の大きさが、選択された値、図２１では５％未満である必要がある平衡範囲が定義された。
【０１６４】
１０のみの表面パラメータが、受信者動作特性の計算に使用された。図１５から明らかなように、表面パラメータをそれ以上追加しても、精度が有意に改善しない収穫逓減点がある。図１５によれば、８～１０のみの賢明に選ばれた表面パラメータの使用が、比較的高い精度である８０％を達成するのに明らかに十分である。上位１０個の表面パラメータは、分類器１００の特異性、感度、及び精度を含め、受信者動作特性の統計学的挙動及び混同行列を特徴付けるのに考慮された。
【０１６５】
非癌患者又は癌罹患患者からのものとして細胞を分類するプロセスは、その確率の取得に使用される手順の全ての反復にわたるその細胞で得られた確率の平均を計算することに頼る。これは、
【数３７】

$$\mathrm{Prob}^{(i)}{}_n=\frac{1}{S}\sum_s \mathrm{Prob}^{(s)(i)}{}_n,\ \text{where}\ \mathrm{Prob}^{(s)(i)}{}_n=\mathrm{AI}(P_s^{(1,s)}{}_n \oplus P_s^{(2,s)}{}_n \mid C^{(i)}), \quad (20)$$

として正式に表され、式中、分類器ＡＩは、トレーニングデータベース８７で展開された機械学習法を使用して作成された。この手順によれば、クラス１が癌細胞を表すと仮定すると、$\mathrm{Prob}_n{}^{(1)}$が特定の閾値を超える場合、細胞は癌罹患患者からのものであると識別され、特定の閾値は、受信者動作特性から取得することができる。

【０１６６】
図１８及び図１９に示されるデータの正確さを確認するために、図１９及び図２０で使用されたものと同じ手順を用いるが、癌細胞及び正常細胞に均等に分かれて分類される試料を用いてコントロール実験を実行した。図１７及び図１８は、分類の１０００回のランダム選択の結果を示す。精度が５３％±１０％に低下したことが明らかであり、これは予想と一貫する。これは、図１９及び図２０に示されるデータの信頼性並びに機械学習法が多すぎるパラメータに対処するようになされた場合に生じる一般的な問題である過剰トレーニングに対する分類器の耐性を示唆する。
【０１６７】
代替の分類方法は、患者の診断の確立に２つ以上の細胞に頼る。これは、高いサンプリングエラーに基づいて堅牢性を欠くことを回避する。さらに、これは、尿８８中で見つけられた細胞が実際に膀胱自体からのものであることを確信することができないため、生じるエラーを回避する。尿路の他の部分も細胞を脱落させることが完全に可能である。加えて、尿８８は、尿路の他の部分から剥脱した上皮細胞等の様々な他の細胞を含むことがある。そのような一分類方法は、分類される細胞の総数Ｎからの癌罹患患者からのものとして分類された細胞数Ｍが予め定義される値以上である場合、患者を癌と診断することを含む。これは、Ｎ＝Ｍ＝１である上述した場合の一般化である。
【０１６８】
Ｎ個の細胞での確率に基づく癌を有する確率は、アルゴリズム（３－２）～（３－９）又は（１０）～（１４）を使用して割り当てることができる。癌患者からのもの（クラス１）としてＮ個のテストされる細胞を分類する確率を定義する好ましい手順として、以下であり、
【数３８】

$$\mathrm{Prob}^{(1)}{}_{n}=\underset{x=1..N}{Max}\{\mathrm{Prob}^{(x1)}{}_{n}\}\ \text{where}\ \mathrm{Prob}^{(x1)}{}_{n}=\mathrm{AI}(P_{n}^{(1,x)}{}_{m}\oplus P_{n}^{(2,x)}{}_{m}\mid C^{(1)}), \qquad (21)$$

式中、分類器ＡＩはトレーニングデータベース８７から作成される。
【０１６９】
図１９及び図２０は、図１５及び図１６におけるものと同様の堅牢性であるが、Ｎ＝５及びＭ＝２である場合での精度及び受信者動作特性を示す。そのような方法の精度が９４％に達することができることを見て取ることができる。上述した無作為化検定は、受信者動作特性曲線下面積で５０±２２％を示す（診断セットの１０００回のランダム選択の結果）。これらは、過剰トレーニングがないことを暗に示す。
【０１７０】
複数のＮ及びＭの混同行列の計算の結果は、２つの単一チャネル（高さ及び付着力）について例示された図２０の表に示される。結合されたチャネルの堅牢性は、単一チャネルに基づく診断よりも良好である。
【０１７１】
上述した手順は、非癌患者の分類に適用することもできる。そのような場合、上述した確率は、細胞が非癌患者に属する確率である。

【図面】

【図１】

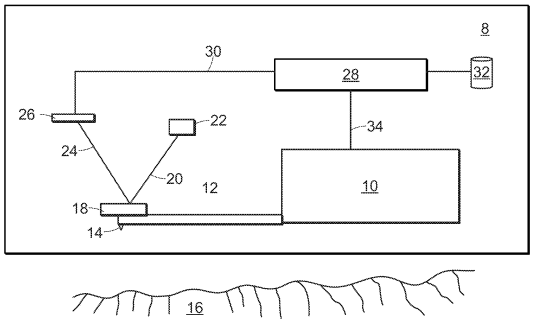


FIG. 1

【図２】

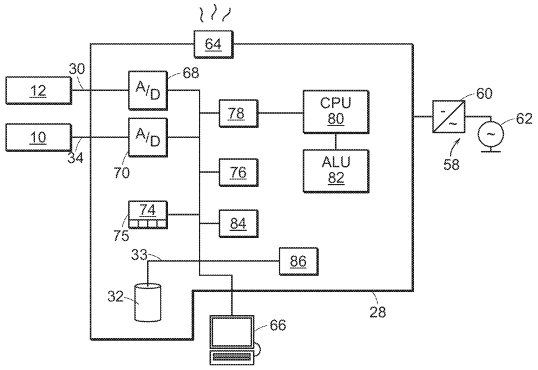


FIG. 2

【図３】

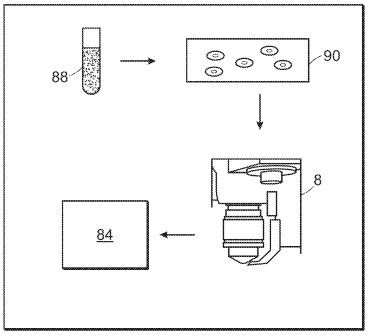


FIG. 3

【図４】

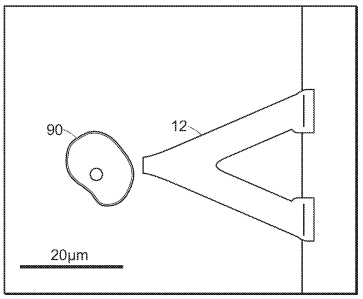


FIG. 4

【図５】

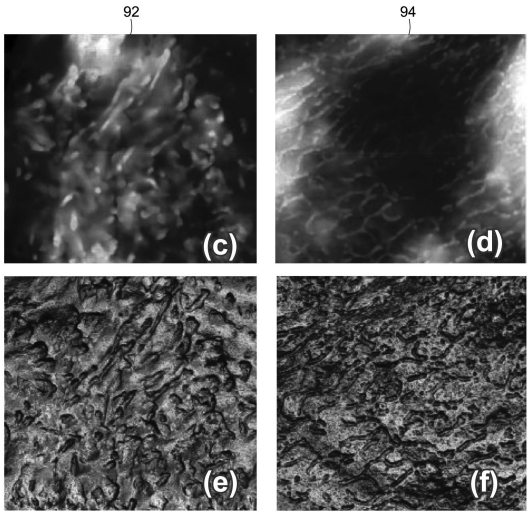


FIG. 5

【図６】

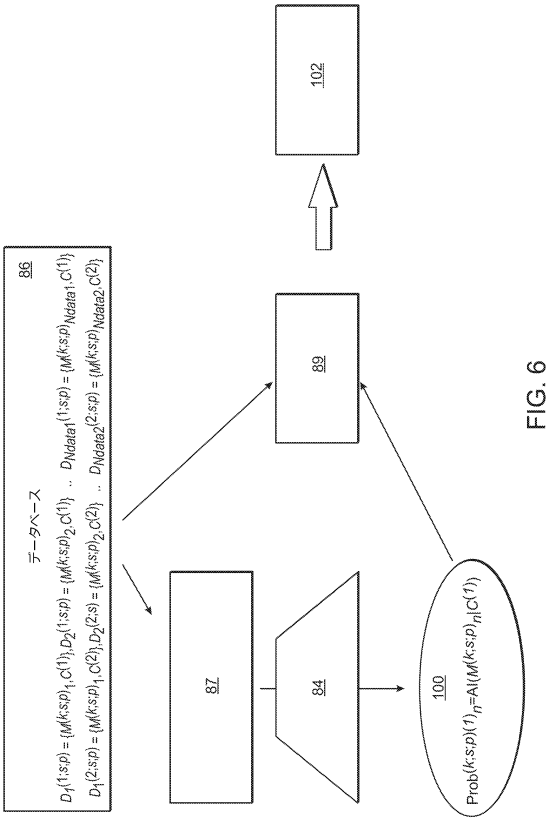


FIG. 6

【図７】

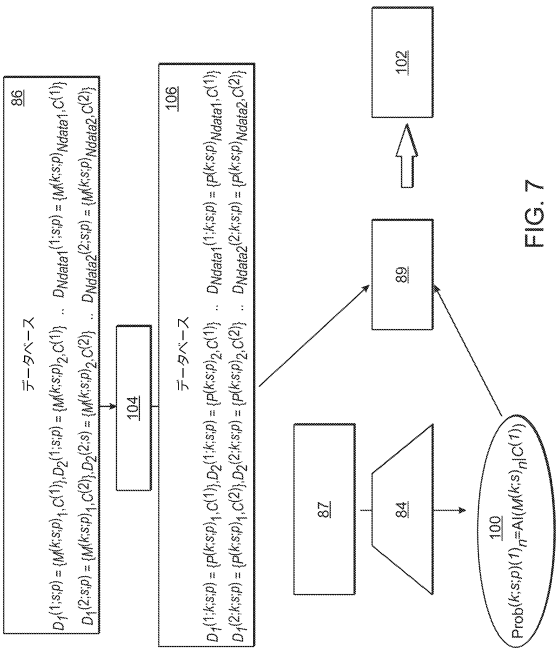


FIG. 7

【図８】

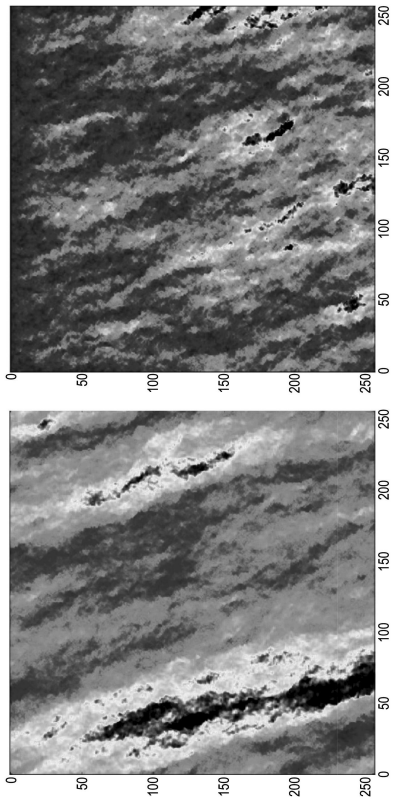
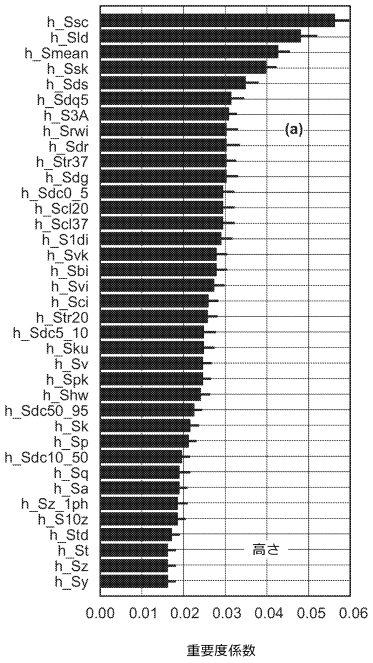


FIG. 8

【図９Ａ】



【図９Ｂ】

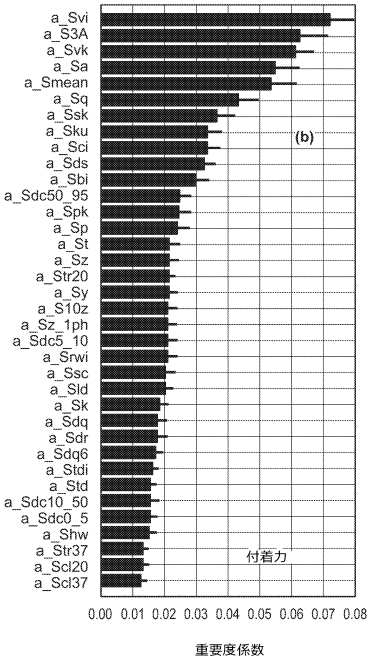
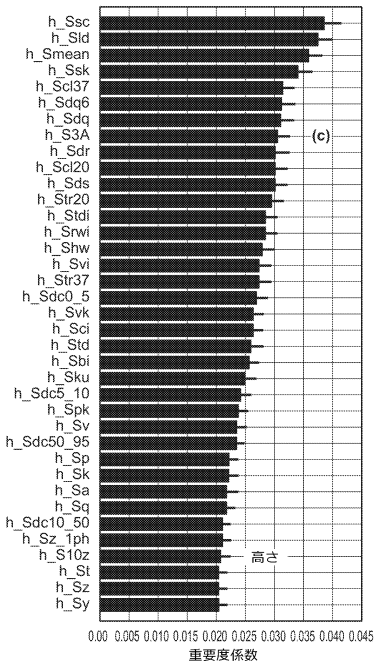
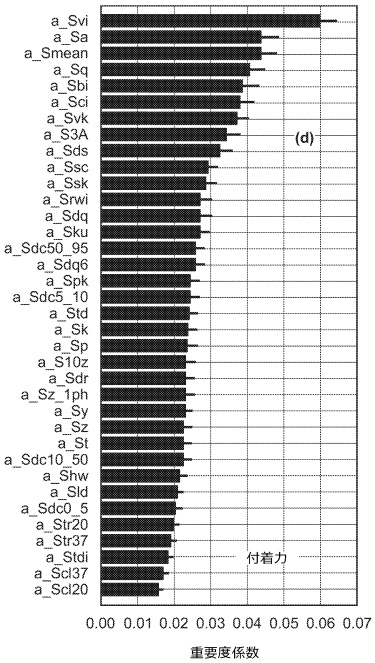


FIG. 9

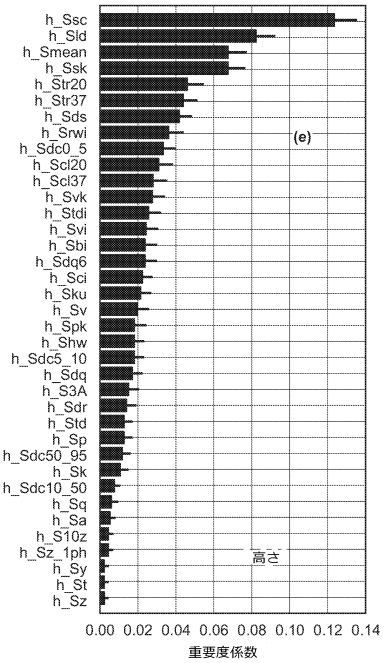
【図９Ｃ】



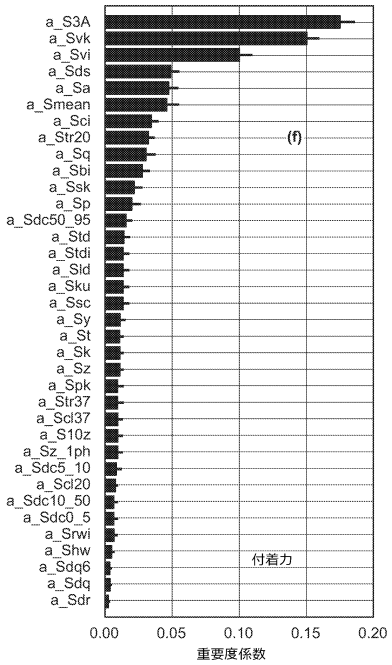
【図９Ｄ】



【図９Ｅ】



【図９Ｆ】



【図１０】

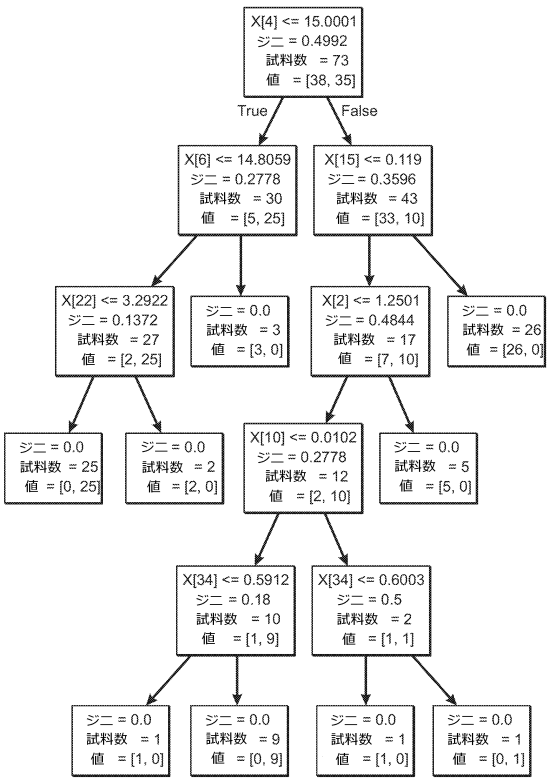


FIG. 10

【図１１】

アルゴリズム　細胞表面パラメータを使用した癌検出

関数 make_daraset　入力：ＣＳＶファイル、統計パラメータ［平均又はメジアン］
　ステップ：
　１）データをＣＳＶファイルからロード
　２）全ての試料が同じ値を有する場合、変数を棄却
　３）データ取得アーチファクトが存在する場合、試料を棄却
　４）事前処理されたデータを有するデータフレームを記憶
　５）細胞がない非癌試料を有するデータフレームを作成
　６）事前処理されたデータフレーム及び非癌試料データフレームを統合
　７）患者ＩＤにより試料をグループ化
　８）全ての患者ＩＤについて細胞象限からの細胞当たりの全ての変数の平均又はメジアン値を取得し、データフレームに記憶
　リターン：細胞試料データフレーム

関数 split_data_train_test()　入力：データフレーム、トレーニング－テスト試料比
　ステップ：
　１）選択可能なトレーニング－テスト試料比を使用して患者ＩＤに基づいてデータフレームをトレーニング試料及びテスト試料に分割
　２）トレーニング試料データフレーム及びテスト試料データフレームを記憶
　リターン：トレーニング試料データフレーム及びテスト試料データフレーム

関数 make_classifier_model()　入力：モデル名、モデルパラメータ
　ステップ：
　１）以下の分類モデルの１つを選択：ランダムフォレスト、エクストラツリー、又は勾配ブーストツリー
　２）選択されたモデルのパラメータを指定し、モデルをインスタンス化
　リターン：分類モデルオブジェクト

関数 classification_performance_metrics()　入力：分類モデルオブジェクト、細胞試料データフレーム、細胞数Ｍ、細胞数Ｎ、反復回数Ｒ
　ステップ：
　for　反復回数Ｒ中のｎ番目の実行
　make_clasiifier_model()を実行
　split_data_train_test()を実行
　１）患者ＩＤ毎に、テストデータフレームの分類性能尺度の測定で使用される細胞数Ｍを選択
　２）患者ＩＤ毎に、患者ＩＤ全体を癌として分類するために、癌として分類される必要がある細胞数Ｎを選択
　３）患者ＩＤにつき利用可能な全ての細胞を使用し、テスト試料中の全ての患者ＩＤのＭ個の細胞の可能な全ての組合せを作成し、データフレームに記憶
　４）モデル分類性能尺度を計算：ＲＯＣ曲線、ＲＯＣ　ＡＵＣスコア、精度、及び混同行列
　リターン：ＲＯＣ曲線、ＲＯＣ　ＡＵＣスコア、精度スコア、混同行列

FIG. 11

【図１２】

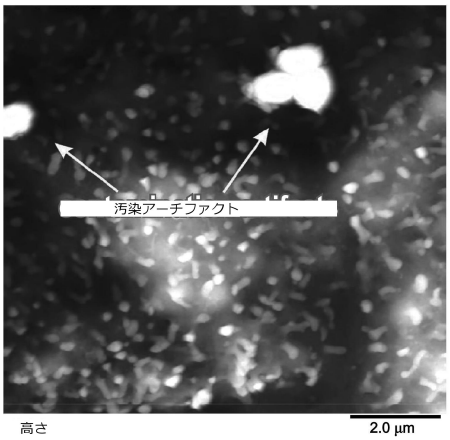


FIG. 12

【図１３】

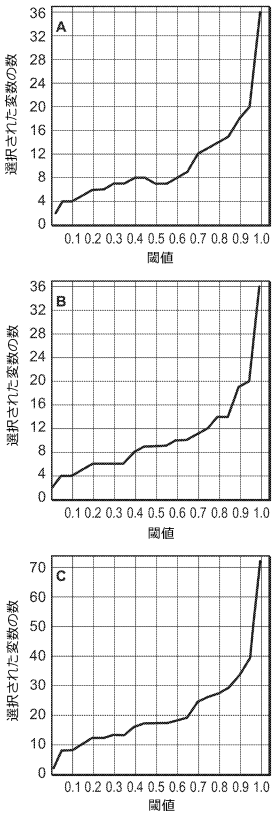
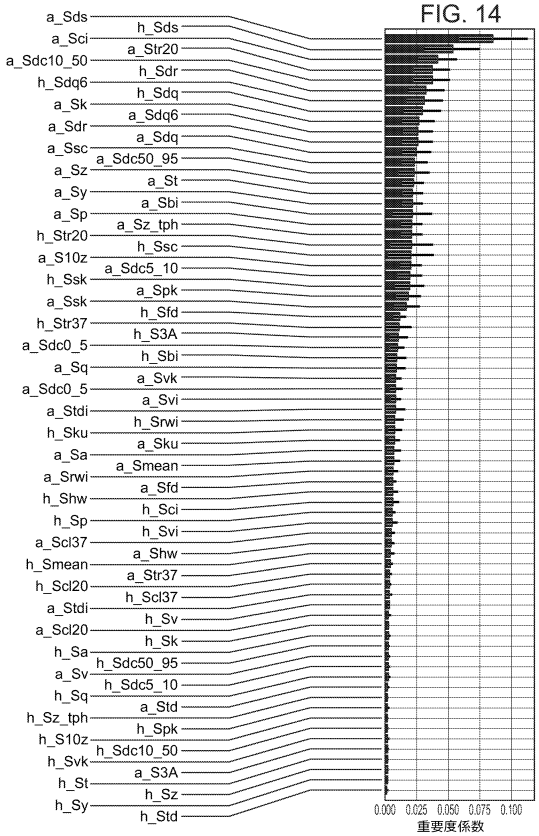


FIG. 13

【図１４】

FIG. 14



【図１５】

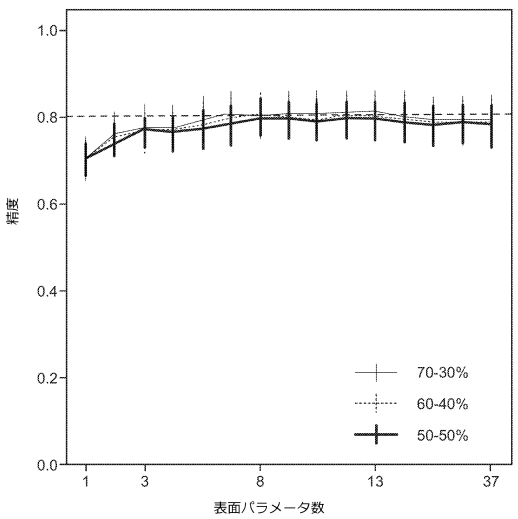


FIG. 15

【図１６】

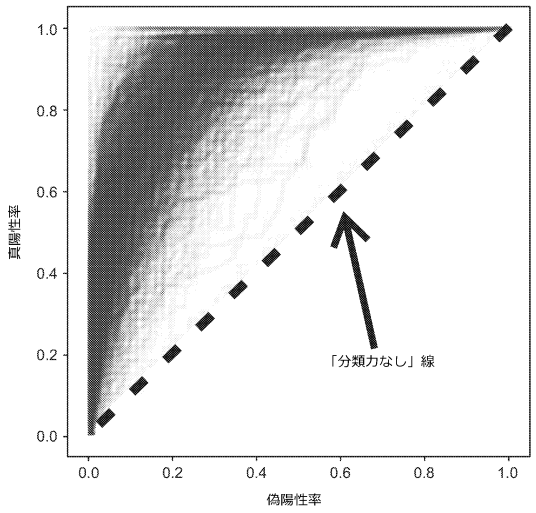


FIG. 16

【図１７】

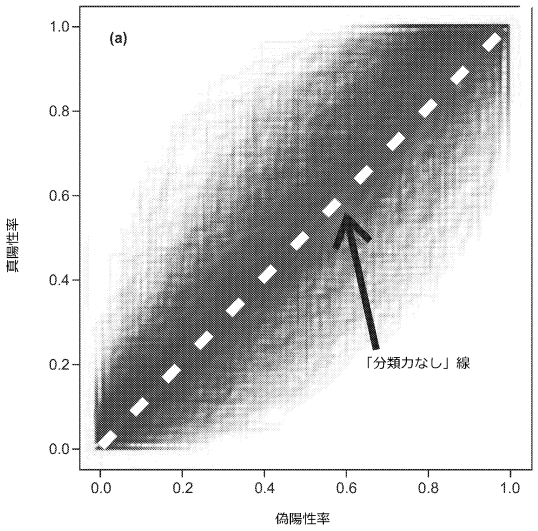


FIG. 17

【図１８】

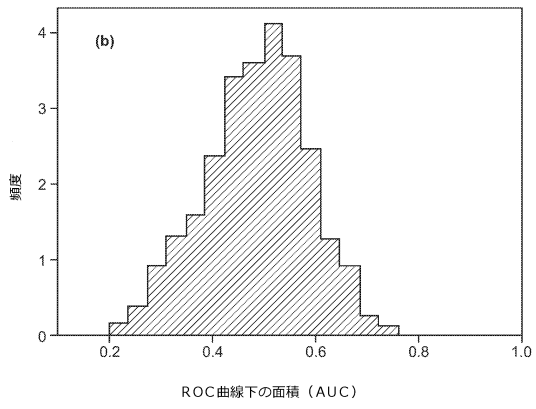


FIG. 18

【図１９】

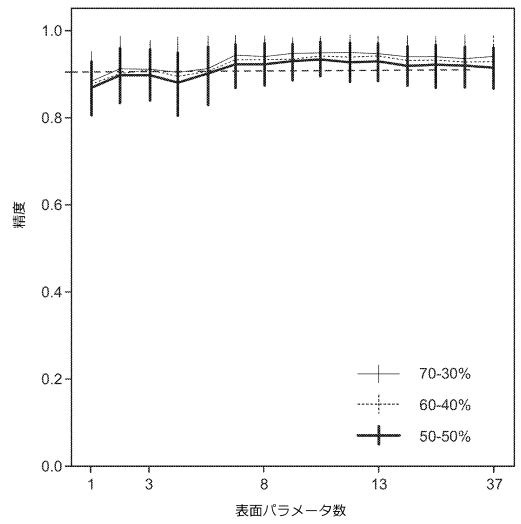


FIG. 19

【図２０】

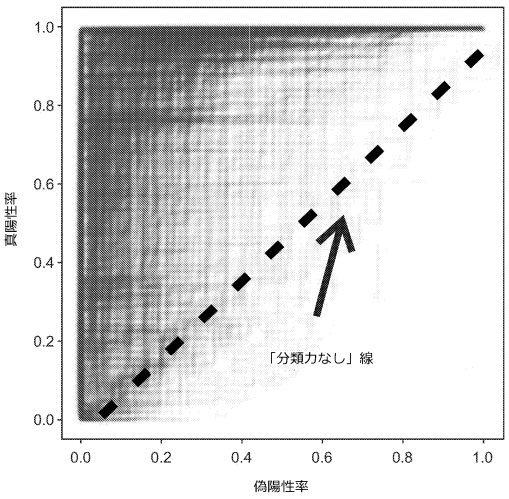


FIG. 20

【図２１】

	データ	ランダムフォレスト			エクストリームランダム化フォレスト			勾配ブースティングツリー		
		AUC/精度	感度/特異度		AUC/精度	感度/特異度		AUC/精度	感度/特異度	
N=1	高さ	75/73	50/84	76/69	77/74	53/84	77/70	74/73	46/86	75/68
M=1	付着力	88/83	68/90	84/77	88/83	69/90	85/78	87/82	69/89	84/77
N=2	高さ	75/77	40/91	76/69	77/78	44/91	77/71	75/77	40/92	75/69
M=1	付着力	89/87	71/93	86/80	89/86	70/93	87/80	89/87	71/93	86/80
N=2	高さ	77/78	43/92	76/69	78/79	48/91	78/71	75/78	38/93	74/68
M=2	付着力	89/86	69/93	86/80	90/87	70/94	86/79	89/87	69/94	85/79
N=3	高さ	73/81	34/95	75/66	74/81	35/95	76/67	73/81	35/95	74/66
M=1	付着力	90/89	70/95	88/81	89/88	69/94	86/79	89/90	70/96	87/80
N=3	高さ	75/82	36/95	75/68	78/82	42/94	77/69	73/81	34/96	73/66
M=2	付着力	91/90	73/96	87/80	90/90	72/96	87/80	89/90	69/96	85/78
N=3	高さ	75/82	37/95	76/67	78/83	42/95	77/69	72/81	32/96	73/66
M=3	付着力	90/90	71/96	87/80	90/90	70/96	87/76	88/89	65/96	85/77
N=4	高さ	69/84	31/97	71/61	70/84	33/97	73/64	69/84	34/97	73/64
M=1	付着力	89/91	75/95	88/80	88/90	71/96	87/79	90/92	74/96	88/80
N=4	高さ	72/84	35/97	73/64	74/85	37/96	76/67	71/84	37/96	72/63
M=2	付着力	91/92	77/97	88/80	90/92	74/96	88/80	90/92	73/97	87/79
N=4	高さ	72/84	34/97	74/64	76/85	41/96	76/67	70/84	32/97	71/62
M=3	付着力	89/92	72/97	88/79	90/92	72/97	87/79	87/91	64/97	84/75
N=4	高さ	72/84	32/97	74/64	75/85	38/96	76/66	70/84	32/97	72/62
M=4	付着力	89/91	71/96	88/79	89/91	71/96	87/78	86/90	64/97	84/75
N=5	高さ	66/86	33/98	69/57	66/85	36/97	72/60	67/86	35/98	70/59
M=1	付着力	88/93	77/96	87/78	87/92	75/96	87/78	89/94	77/97	88/79
N=5	高さ	69/86	34/97	72/61	70/86	39/96	76/64	68/86	36/97	70/61
M=2	付着力	91/94	81/98	91/82	90/94	78/97	89/80	91/94	78/98	88/80
N=5	高さ	68/86	31/98	72/61	72/87	39/97	76/65	67/86	34/97	69/58
M=3	付着力	90/93	77/97	88/78	90/93	77/97	88/78	88/93	70/98	85/75
N=5	高さ	67/86	34/97	72/60	72/87	39/97	76/64	65/85	30/97	69/57
M=4	付着力	89/93	75/97	87/77	88/93	73/97	88/77	85/92	61/98	84/74
N=5	高さ	68/86	30/97	72/59	74/87	39/97	76/63	67/85	32/97	71/59
M=5	付着力	88/93	74/97	88/77	88/93	73/97	88/77	84/91	62/97	84/73

FIG. 21

フロントページの続き

(33)優先権主張国・地域又は機関
米国(US)
早期審査対象出願
アメリカ合衆国 ０２１５５ マサチューセッツ州， メドフォード， バルー ホール， トラスティーズ オブ タフツ カレッジ 内
審査官 前田 敏行
(56)参考文献 米国特許出願公開第２０１５／００３６８８９（ＵＳ，Ａ１）
M. P. Nikiforov, et. al.，Functional Recognition Imaging Using Artificial Neural Networks: Application to Rapid Cellular Identification by Broadband Electromechanical Response，Nanotechnology，2009年
(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ０１Ｑ ３０／０４
Ｇ０１Ｑ ６０／２４
Ｇ０１Ｑ ８０／００
Ｇ０１Ｎ ３３／５４３
Ｇ０１Ｎ ３３／５０