

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2008-520318

(P2008-520318A)

(43) 公表日 平成20年6月19日(2008.6.19)

(51) Int. Cl.	F I	テーマコード (参考)
A 6 1 B 6/03 (2006.01)	A 6 1 B 6/03 3 6 0 D	4 C 0 9 3
G 0 6 F 17/30 (2006.01)	G 0 6 F 17/30 1 7 0 B	4 C 1 1 7
G 0 6 N 3/00 (2006.01)	G 0 6 F 17/30 2 1 0 D	5 B 0 7 5
G 0 6 T 7/00 (2006.01)	G 0 6 N 3/00 5 5 0 C	5 L 0 9 6
A 6 1 B 5/00 (2006.01)	G 0 6 N 3/00 5 6 0 C	

審査請求 未請求 予備審査請求 未請求 (全 14 頁) 最終頁に続く

(21) 出願番号 特願2007-542438 (P2007-542438)
 (86) (22) 出願日 平成17年11月18日 (2005.11.18)
 (85) 翻訳文提出日 平成19年4月26日 (2007.4.26)
 (86) 国際出願番号 PCT/IB2005/053824
 (87) 国際公開番号 W02006/054269
 (87) 国際公開日 平成18年5月26日 (2006.5.26)
 (31) 優先権主張番号 60/629, 751
 (32) 優先日 平成16年11月19日 (2004.11.19)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 60/722, 668
 (32) 優先日 平成17年9月30日 (2005.9.30)
 (33) 優先権主張国 米国 (US)

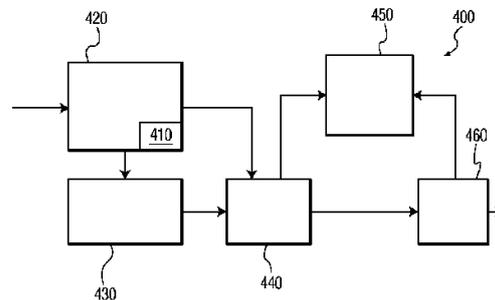
(71) 出願人 590000248
 コーニンクレッカ フィリップス エレクトロニクス エヌ ヴィ
 オランダ国 5 6 2 1 ベーアー アインドーフエン フルーネヴァウツウェッハ 1
 (74) 代理人 100087789
 弁理士 津軽 進
 (74) 代理人 100114753
 弁理士 宮崎 昭彦
 (74) 代理人 100122769
 弁理士 笛田 秀仙

最終頁に続く

(54) 【発明の名称】 サポートベクタマシン (SVM) を用いるコンピュータ支援検出 (CAD) における誤検出低減のためのシステム及び方法

(57) 【要約】

HRCT医療画像データ内で検出される注目領域のコンピュータ支援検出(CAD)及び分類のための方法が、報告される誤検出の数における低減を実現するために、その分類の特異度及び感度を最大化するポスト処理機械学習を含む。その方法は、真及び偽の領域を多数含むよう選択された医療画像トレーニングデータのセットに基づき分類器をトレーニングすることを含む。そこでは、その真及び偽の領域が、CAD処理により識別され、自動的に分割される。そこでは、その分割されたトレーニング領域が、そのグラウンドトゥース、つまり真又は偽に対して各トレーニングを分類するため、少なくとも一人の専門家によりレビューされる。それは本質的に自動分割に適合される。そこでは、各分割された領域から特徴プールが識別され抽出される。そこでは、その特徴プールが、最適な特徴サブセットを識別するため遺伝的アルゴリズムにより処理される。そのサブセットは、サポートベクタマシンをトレーニングするのに使用される。その方法は、非トレーニング医療画像データ内で分類のための候補である領域を検出し、その候補領域を分割し、各分割



【特許請求の範囲】**【請求項 1】**

HRCTデータといった医療画像データ内の領域のコンピュータ支援検出及び分類の間の誤検出低減の方法において、該方法が、分類の特異度及び感度を最大化するポスト処理機械学習ステップを実行し、誤検出低減システムにより報告される偽の陽性の数の低減を実現するものであり、

多数の真及び偽の領域を含むよう選択された医療画像トレーニングデータのセットに基づき分類器をトレーニングするステップであって、前記真及び偽の領域がコンピュータ支援検出処理により特定され自動的に分割され、実質的に前記自動分割に適するよう、真又は偽であるグラウンドトゥースに関して各トレーニング領域を分類する少なくとも1人の専門家によって前記分割されたトレーニング領域がレビューされ、各分割された領域から特徴プールが識別及び抽出され、及びサポートベクタマシンをトレーニングするのに使用される最適な特徴サブセットを識別する遺伝的アルゴリズムにより前記特徴プールが処理される、ステップと、

非トレーニング医療画像データ内で、分類のための候補である領域を検出するステップと、

前記候補領域を分割するステップと、

各分割された候補領域から特徴セットを抽出するステップと、

前記最適な特徴サブセットに基づくトレーニングの後、前記サポートベクタマシンを用いて前記候補領域を分類し、前記候補特徴のセットを処理するステップとを有する、方法。

【請求項 2】

前記トレーニングステップが、最高の適合を与える前記特徴サブセットの大きさ及び前記特徴の同一性を共に決定するステップを更に含む、請求項 1 に記載の方法。

【請求項 3】

前記決定するステップが、2つのフェーズで前記遺伝的アルゴリズムを適用するステップを含み、

a) 前記特徴のセット及び前記特徴の数に関して各染色体を識別するステップと、

b) 異なる染色体の発生数と平均エラー数とに基づき、前記特徴の最適な大きさを決定するため、各染色体に対して、前記識別された特徴セットと前記識別された特徴数とを分析するステップとを含む、請求項 2 に記載の方法。

【請求項 4】

前記トレーニングステップが、染色体として特徴プールを規定するステップを更に含み、各特徴は遺伝子を表すものであり、前記遺伝的アルゴリズムが、まず特徴のランダム選択により前記染色体を母集団とし、各生成に対して評価を繰り返しながらより高い適合度を持つ染色体を繰り返し探し、並びに、変異及び交叉を用いて新たなより適合する染色体を生成する、請求項 1 に記載の方法。

【請求項 5】

汎用コンピュータによる処理のためダウンロードされるコンピュータ可読命令のセットを有するコンピュータ可読媒体であって、

HRCTデータといった医療画像データ内の領域のコンピュータ支援検出及び分類の間の誤検出低減の方法であって、分類の特異度及び感度を最大化するポスト処理機械学習ステップを実行し、誤検出低減システムにより報告される偽の陽性の数の低減を実現するものであり、

多数の真及び偽の領域を含むよう選択された医療画像トレーニングデータのセットに基づき分類器をトレーニングするステップであって、前記真及び偽の領域がコンピュータ支援検出処理により特定され自動的に分割され、実質的に前記自動分割に適するよう、真又は偽であるグラウンドトゥースに関して各トレーニング領域を分類する少なくとも1人の専門家によって前記分割されたトレーニング領域がレビューされ、各分割された領域から特徴プールが識別及び抽出され、前記特徴プールが最適な特徴サブセットを識別する遺伝

10

20

30

40

50

的アルゴリズムにより処理され、前記サブセットがサポートベクタマシンをトレーニングするのに使用される、ステップと、

非トレーニング医療画像データ内で、分類のための候補である領域を検出するステップと、

前記候補領域を分割するステップと、

各分割された候補領域から特徴セットを抽出するステップと、

前記最適な特徴サブセットに基づくトレーニングの後、前記サポートベクタマシンを用いて前記候補領域を分類し、前記候補特徴のセットを処理するステップとを有する方法を実現する、コンピュータ可読媒体。

【請求項 6】

コンピュータ支援検出サブシステム及び誤検出低減のためのシステムを含む医療画像分類システムであって、前記誤検出低減サブシステムが、ポストコンピュータ支援検出でトレーニングされるサポートベクタマシンを有し、報告される偽の陽性を最小化するような特異度と感度とを備える画像化データ内で検出される臨床的に重要な領域を分類するものであり、

前記画像データ内で検出される臨床的に重要な領域を識別及び描写するコンピュータ支援検出サブシステムと、

前記コンピュータ支援検出サブシステムと通信する誤検出低減サブシステムであって、

各コンピュータ支援検出描写による領域から特徴プールを抽出する特徴抽出器と、

前記特徴プールの最適なサブセットを与え、前記特徴抽出器と通信する遺伝的アルゴリズムと、

最小の偽の陽性を備える前記特徴サブセットに基づき各描写された領域を分類し、前記特徴抽出器及び前記遺伝的アルゴリズムと通信するサポートベクタマシンとを有する誤検出低減サブシステムとを有し、

前記システムが、まず真又は偽の陽性のいずれかであることが判明している領域を含む画像のセットに基づきトレーニングされ、前記画像のセットから特徴を抽出し、及び前記サポートベクタマシンが未知の領域を最適に分類するよう、最適なサブセットを特定するため前記遺伝的アルゴリズムを用いる、医療画像分類システム。

【請求項 7】

前記コンピュータ支援検出サブシステムが、前記コンピュータ支援サブシステムにより識別される領域を描写する分割サブシステムを更に含む、請求項 6 に記載の医療画像分類システム。

【発明の詳細な説明】

【技術分野】

【0001】

本願 / 本発明は、本願出願人により2004年11月19日に提出された米国特許仮出願番号 60/629,751号に由来する。本願は、共通して所有される共に係属中のPhilips出願番号PHUS040499、PHUS040500及びPHUS040501と関連する。

【0002】

本発明は、コンピュータ支援による検出システム及び方法に関する。本発明は、より詳細には、ポストCAD機械学習(machine learning)を実現するためのサポートベクタマシン(SVM)を用いて、高分解能の薄層スライスコンピュータ断層撮影(HRCT)画像から肺結節をコンピュータ支援による検出(CAD)する際の誤検出(false positive:偽の陽性)低減のためのシステム及び方法に関する。

【背景技術】

【0003】

現在のコンピュータ関連システムの速度及び精巧さは、高速でより洗練された医療画像化システムの開発をサポートする。処理のために生成されるデータ量の必然的な増加が、その医療画像データを自動的に分析するための多数のアプリケーションプログラムの生成をもたらした。すなわち、医療画像を識別並びに / 又は診断及び評価するため医療画像を

10

20

30

40

50

評価する際に医師、臨床医、放射線技師等を支援するために、様々なデータ処理ソフトウェア及びシステムが開発されてきた。例えば、マルチスライスCT(MSCT)スキャンから疑わしい病巣を自動的に識別するためのコンピュータ支援による検出(CAD)アルゴリズム及びシステムが開発されてきた。異常個所又は病巣だけでなく、生体構造の大きさ、形状及び位置を正確に図示する能力がゆえ、CTつまりコンピュータ断層撮影システムは、画像化を介して病気を診断するのに普通使用される画像化モダリティである。

【0004】

CADシステムは、形態学的に興味ある領域(例えば病巣)、又は臨床的に重要な他の構造的に検出可能な状態を自動的に検出(識別)する。医療画像が描画され表示されると、通常CADシステムは、その調査された領域にマークを付けるか又は特定する。そのマークは、10 マークされた疑わしい領域として注意を引くためのものであり、その病巣(注目領域)の分類又は特徴化を更に提供することができる。すなわち、CAD(及び/又はCADx)システムは、胸部検査における微小石灰化又はMSCTにおける結節を悪性又は良性として特定することができる。CADシステムは、放射線技師の専門的な知識を組み込み、本質的に医療画像データにおける異常個所の検出に関するセカンドオピニオンを提供し、診断的な提案をすることができる。がんの疑いのある病巣の早期検出及び分類をサポートすることより、CADシステムは、患者に対するより良い予後診断をもたらす早期の介入を理論的に可能とする。

【0005】

CAD及び他の機械学習システムに対する現行のほとんどの作業は、教師有学習に対するのと同じ方法論に従う。CADシステムは、グラウンドトリス(ground truth:根拠真理)が判明しているデータの収集で始まり、そのグラウンドトリスを例えば悪性又は良性と区別するのに十分な識別力を持つと信じられる特徴セットを特定するため、トレーニングデータに基づきトレーニングされる。当業者にとっての課題は、カテゴリ間での識別を容易にする特徴を抽出することであり、理想的には特徴プール内で最も重要な特徴を発見することを含む。CADシステムは、異種の情報(例えば、患者データを伴う画像ベースの特徴)を結合することができるか、又はそれらは例示ベースの手法に対する類似性指標を見つけることができる。当業者は、いずれのコンピュータ駆動の決定支援システムの精度が、学習処理に対して既に分類されたパターンセットの利用可能性により(つまりトレーニングセットにより)制限されることを理解する。20

【0006】

不明確な境界に基づきポストCAD処理が行われる場合、不明確な境界描写に基づく結果も同様に不明確である場合がある。すなわち、診断スキャン処理に使用されるいずれのコンピュータ学習システムの出力もアドバイスとなる。そこで、悪性腫瘍の可能性のある候補として医師に提供される各アドバイスにより、医師は、調査することを強いらられる。すなわち、CAD支援による結果が、調査された領域に対して提案される診断としてボトムライントリス(bottom line truth)(例えば真の陽性(true positive))を表す場所で、医師は、まるでその領域をそれ以上詳細に調査しないかのように、無頓着である場合がある。当業者は、医療的な内容において「真の陽性」がしばしば、真に悪性である検出結節を指すものであること、CADの内容においてマーカは、例えそれが良性又は石灰化された結節を指し示すとしても真の陽性のマーカであると考えられることを理解すべきである。当然「真の陰性」は規定されず、正規化された特異度はCADでは与えられないことができない。偽の陽性のマーキングは、全く結節を指し示さない(が、傷跡、気管支壁肥厚、動きアーチファクト、血管分岐等を指し示す)ようなものである。従って、CAD性能は、通常感度(検出率)及び誤検出率(CT検査あたりの偽の陽性マーク数)により制限され(qualified)、そのようなものとして、誤検出を最小化することが当業者にとって非常に望ましい。30

【0007】

(マークあり又はなしの)自動化された検出プロセスの完了後、多くのCADシステムは、ユーザ及びCAD検出による病巣(領域)への適用のため、冗長性を削減するため、解釈的なツールを実現する等のため、1つ又は複数の中間(interception)ツールを自動的に呼び出50

す。そのために、CAD及び診断における誤検出を削減する様々な技術が知られている。例えば、W.A.H.Mousa及びM.A.U.Khanは、タイトル「Lung Nodule Classification Utilizing Support Vector Machines」、Proc.of IEEE ICIP'2002でその技術を開示する。K. Suzuki、S.G.Armato III、F.Li、S.Sone、K.Doiiは、「Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography」、Med. Physics 30(7)、July 2003、pp. 1602-1617において、誤検出を最小化する試みを述べている。Z.Ge、B.Sahiner、H.-p.Chan、L.M.Hadjiski、J.Weii、N.Bogot、P.N.Cascade、E.A.Kazerooni、C.Zhouによる「Computer aided detection of lung nodules: false positive reduction using a 3D gradient field method」、Medical Imaging 2004:Image Processing、pp.1076-1082においても、同様である。

10

【発明の開示】

【発明が解決しようとする課題】

【0008】

上述されたFPRシステムの中には、CADアルゴリズムに埋め込まれているものがある一方、CADアルゴリズムの特異度を改善するのにポスト処理ステップとして使用されるものもある。例えば、R.Wiemkerらは、「COMPUTER-AIDED SEGMENTATION OF PULMONARY NODULES: AUTOMATED VASCULATURE CUTOFF IN THIN- AND THICK-SLICE CT」、2003 Elsevier Science BVにおいて、(部分的なボリューム効果を改善するため)薄層スライスCTにおける結節の周囲の脈管構造から肺結節を効率的に分離するCADアルゴリズムの感度を最大化し、分類誤差を減らす試みを論じている。しかしながら、WeimkerのFPRシステム及び方法は、多くの知られたFPRシステム及び方法同様、しばしば、洗練された機械学習技術を使用することができず、その特徴抽出及び選択方法は最適化されない。例えば、Mousaらは真の肺結節を非結節(FP)から区別するためサポートベクタマシンを利用するが、そのシステムは、特異度を制限する可能性のある非常に単純化した特徴抽出ユニットに基づかれる。

20

【0009】

そこで、本発明の目的は、医療画像データの放射線的に重要な部分の自動検出を正確かつ信頼性高く実行し、非常に良好な特異度及び感度(つまり、最小の誤検出)を実現するような態様でそれらを分類する誤検出低減システムを提供することにある。

【0010】

本発明の別の目的は、医療画像内の形態学的に重要な領域(「候補領域」)を識別して描写するCADサブシステムと、特徴抽出器、遺伝的アルゴリズム(GA)及びサポートベクタマシン(SVM)を含む機械学習サブシステムであって、CADサブシステムによって描写される候補領域に機械学習を適用し、それらを結節及び非結節として分類し、それによりすべての真の陽性が保持される制約条件下でできるだけ多くの誤検出を削減する機械学習サブシステムとを含むFPRシステムを実現することにある。

30

【0011】

本発明の更に別の目的は、実質的に誤検出分類を削減するよう、十分な特異度及び感度を備える医療画像データから、候補結節を検出、抽出及び分類するためのポストCAD機械学習技術を含むことである。候補結節は、まずCADプロセスにより識別され、分類器処理による使用のための理想的な特徴及び特徴数を特定するためGAにより結節特徴が抽出され処理される。それは、ポストCADで決定される特徴のサブセットの機械学習によりサポートされ、誤検出結節の数を効率的に減らすために十分な感度及び特異度ですべての結節を悪性又は良性を識別する。

40

【課題を解決するための手段】

【0012】

1つの実施形態において、誤検出低減(FPR)のための方法が、4つのメインステップのシーケンスとして実現される: 1)(CADによる)画像分割、2)分割されたデータからの特徴抽出、3)GA、ポストCADによる特徴サブセットの最適化、及び4)最適化された特徴サブセットに基づくSVMによる分類であり、結果として、信頼性の高い感度及び特異度、並びに

50

最小の誤検出を生じさせる。そのため、本書で規定される本発明のFPRシステムは、CADサブシステムを有することができる。その場合、そのサブシステムは、解析対象にある領域の「最良の」分割を特定するリコメンダ(recommender)・サブシステムを備える新規な分割器を含むことができる。本発明に関する斯かる変更は、共通して所有され、同時に出願され共に係属中のPhilips出願番号US040505において見出されることができ、その請求項に記載される。

【0013】

本発明のシステム及び方法は、CT又は高分解能CTスキャンデータ(HRCT)に関する処理として説明されるが、当業者は、その説明が、本発明の範囲をCT又はHRCTデータに関する処理に限定することを意味するものではなく、いずれかの取得される画像化データに関して動作するものであり、添付される請求項の範囲によってのみ限定されることを理解されたい。

10

【発明を実施するための最良の形態】

【0014】

画像データセット(例えばCT)における肺結節を検出する際のコンピュータ支援(CAD及びCADx)の基礎をなす目的は、機械による診断を指定するものではなく、放射線科医が決定を行う際に医師をサポートする機械ベースのアルゴリズム又は方法を実現するものである。つまり、全体の感度(検出率)が上昇するよう、疑わしい対象物の位置を指し示すものである。CAD又は他の臨床的な決定支援システムでの主な問題は、偽のマーカ(いわゆる偽の陽性)が、必然的に真の陽性のマーカと共に生じることである。臨床的な研究だけでなく実験によって、放射線科医自身によるものだけでなくCADシステムにより実現される測定された検出率が、共に読影する放射線科医の数にはっきり依存することが示された。共同読影に参加する者が多くなればなるほど、疑わしい病巣が必然的に多く発見されることになり、従って各参加する放射線科医とCADシステムの個別の感度が減少することになる。しかし、例え絶対的な感度の図(figure)が注意深く評価されなければならないとしても、全ての臨床的な研究によれば、かなりの数の結節が追加的なCADソフトウェアのみで検出される一方、共同で読影する放射線科医の全員により見落とされることもある点で一致する。本発明は、斯かる感度を提供する。

20

【0015】

例えばWiemker、Mousaらにより述べられるような誤検出低減処理を含むCADベースのシステムは、1つの大きな課題を抱えており、それは、医療画像データにおいて検出される「すぐに使用可能な」構造体を識別することである。一旦識別される(つまり分割されると、例えば悪性又は良性といったいくつかのグラウンドトゥースに関して、分割された領域を分類するために、重要な特徴の包括的なセットがCADシステムにより得られる。当業者は、コンピュータ駆動の決定支援又はCADシステムの精度が、トレーニングセットとして用いられる、病理学が判明しているパターン又は領域のセットの利用可能性により制限されることを認識するであろう。例えば、Wiemker,R.、T.Blaffertによる「Options to improve the performance of the computer aided detection of lung nodules in thin-slice CT」2003、Philips Research Laboratories: Hamburg、及びWiemker,R.、T.Blaffertによる「Computer Aided Tumor Volumetry in CT Data, Invention disclosure.」2002、Philips Research、Hamburgなどに記載されるように、技術水準のCADアルゴリズムでさえ、リスクを伴い低いユーザ許容性を与える不要な介入をもたらす多数の誤検出を生じさせる可能性がある。更に、現在の誤検出低減アルゴリズムはしばしば、胸部X線画像又は厚層スライスCTスキャンのために開発されたものであり、HRCTから生じるデータに関して必ずしも良好に動作するものではない。

30

40

【0016】

そのため、本書で説明される本発明のFPRシステム及び方法は、分割される候補領域を識別するためのCADサブシステム又は処理を含む。トレーニングの間及びCAD処理の後、トレーニングデータセット内のその分割された領域は、特徴抽出器又は特徴抽出処理を実現するプロセッサへと渡される。特徴抽出は、遺伝的アルゴリズム(GA)サブシステム、又は

50

GAプロセッサに渡される3D及び2D特徴をその検出された構造体から取得する。医療画像における重要な領域の検出に関して当業者である少なくとも1人の臨床医が、トレーニングをサポートするのに必要とされる。そのGAプロセッサは、最適な特徴サブセットを実現するために、(トレーニング画像から)抽出された特徴セットを処理する。最適な特徴サブセットは、FPRと共に、SVMに対する十分な識別力を提供する最適な数の最適な特徴を含む。

【0017】

トレーニングの間、GAによるポストCAD処理が、機械学習処理による使用のための最適な特徴サブセットを決定する。SVMはその機械学習のためその特徴サブセットを使用する。その後、候補領域を識別及び分割するための分割器を用いて又は用いずに、調査下にある画像がCADサブシステムにより処理される。候補領域から抽出された特徴セットは、トレーニングされた分類器(SVM)により処理される。一意なポストCAD機械学習のおかげで、本発明のFPRシステムは正確に、そして十分な特異度及び感度をもって、高分解能で薄層CT(HRCT)画像における小さな肺結節を検出する。当業者は、本発明のFPRシステムが、下位技術(inferior technique)を用いては目に見えない結節又は微小石灰化を正確に検出及び分類することができることを理解するであろう。例えば、スライス厚が1 mm以下のHRCTデータは、非常に小さな結節の検出を可能にするが、そうするためには、信頼性の高い検出及び血管からの識別のため、例えば、本書に記載される発明のような新たな手法を必要とする。

【0018】

本発明のFPRシステム400の好ましい実施形態が、図1を参照して概して説明されることになる。(誤検出低減を備える)FPRシステム400は、特定の基準に合致する領域を識別及び分割するCADサブシステム420を含む。好ましくは、CADサブシステムは、CADプロセッサ410を含み、更に、医療画像データに関する低レベル処理を行う分割ユニット430を含むことができる。CADサブシステム420は、トレーニングデータに関する処理か又は候補領域の調査かに関係なく、CAD処理により識別される候補結節(注目領域)を分割する。そのサブCADシステムは、安定的な分割を実現するためパラメタ調節処理をガイドする。

【0019】

分割データは、FPRサブシステムを有する特徴抽出ユニット440に出力される。特徴プールが、各分割された領域、トレーニング又は候補から抽出され、SVMをトレーニングするのに「最良の」特徴サブセットを識別するため、遺伝的アルゴリズムプロセッサ450により処理される。すなわち、GAプロセッサ450は、特徴プールから、含まれる特徴の選択及び数の両方に関して、最適化された特徴サブセットを生成する。そのサブセットは、候補領域から抽出される特徴セットに関して動作するとき(エラーにおいて)最小の誤検出が特定されるよう、十分良好な感度及び特異度で分類するためのサポートベクタマシン(SVM)460により使用される。つまり、候補領域を調査するとき、トレーニングと区別されるよう、抽出された特徴が分類のためSVMに送られる。

【0020】

上述されたように、CADサブシステム420は、分割ユニット430を有するか有しなにかに関係なく、2値又は3値(trinary)画像を生成する事により、背景から(非結節を含む)候補結節を描写する。そこでは、結節領域、背景領域及び肺壁(又は「カットアウト」)領域がラベル付けされる。グレーレベル及びラベル付けされたVOIを受信すると、特徴抽出器は、例えば、2D及び3D形状特徴、ヒストグラムベースの特徴等といった任意の重要な特徴を計算(抽出)する。トレーニングモードにおいて、それがFPRシステム全体の性能に非常に影響を与えるので、特徴抽出は重要である。特徴の全体のセット又はプールの適切な抽出がなければ、GAは、(オーバーフィッティングを避け、一般化可能性を増加させるため)最良の識別力を持ち最小のサイズの特徴サブセットを決定することはできない。

【0021】

GAベースの特徴選択処理は、共通して所有される係属中のPhilips出願番号US040120(ID disclosure #779446)により教示され、その内容は本書においても参照により含まれる。

10

20

30

40

50

GAによる特徴サブセットの選択は、複数の「遺伝子」から作られる多数の「染色体」を作成する事によって開始される。各遺伝子は、選択された特徴を表す。染色体により表される特徴のセットが、トレーニングデータに関するSVMをトレーニングするのに使用される。染色体の適合性は、結果SVMがいかに好適に動作するかにより評価される。この処理の開始時において、染色体の母集団は、染色体を形成する特徴をランダムに選択することにより生成される。すると、そのアルゴリズム(つまりGA)は、好適に動作する(高い適合度の)染色体を繰り返し探索する。

【0022】

各生成において、GAは母集団における各染色体の適合度を評価し、2つの主要な進化型方法(evolutionary method)である変異(mutation)及び交叉(crossover)を介して、現在の染色体から新たな染色体を作成する。「良性の」染色体における遺伝子は、次世代においても維持されやすく、貧弱な性能のものは切り捨てられやすい。最終的に、最適なソリューション(つまり、特徴の収集)は、最も適合したものが生存するというこの処理を介して発見される。誤って分類されるケースの総数を減らす誤検出低減(FPR)を実現する特徴の最良数を含む、最良の特徴サブセットを知ることによってである。特徴サブセットが決定された後、そのサブセットがSVMをトレーニングするのに使用される。当業者は、SVMが「もとの」特徴空間を、図2に示されるように、トレーニングセットが超平面により分離可能であるような幾つかの高次の特徴空間にマッピングすることを理解すべきである。SVMベースの分類器は、幾つかの内部パラメータを持ち、それは、その性能に影響を与える。斯かるパラメータは、全体の精度を可能性として最良のものとして経験的に最適化される。更に、小さな数値範囲を持つ特徴を大きな数値範囲を持つ特徴が上回り、独占状態になることを避けるため、SVMにより使用される前に特徴値が正規化される。それは、共通して所有される係属中のPhilips出願番号US040499(ID disclosure番号 778965)により教示される本発明のシステム及び処理の焦点である。正規化された特徴値も計算をより簡単なものにする。カーネル値は通常特徴ベクトルの内積に依存するので、大きな属性値が数値問題を生じる場合がある。[0,1]の範囲に対するスケージングは、

$$x' = (x - m_i) / (M_i - m_i)$$

として行われ、ここで x' は「スケール化された」値であり、 x はもとの値であり、 M_i は配列における最大値であり、及び m_i は配列における最小値である。

【0023】

本発明のFPRシステムは、「1つ残し k 畳み込み評価(leave-one-out and k -fold validation)」と呼ばれるものを利用し、トレーニングデータ又は病理が分かっている領域を含んだ肺結節データセットを用いて検証された。その検証が実現され、すべての真の結節を実質的に保持しつつ大多数の偽の結節を削減するため、本発明のFPRシステムが示される。それはCADサブシステムであり、それは、(図1に示されるような)分割器を含む場合も含まない場合もあるが、2値又は3値(trinary)画像を生成する事により、背景から結節及び非結節を描写する。それにより、結節領域、背景領域及び肺壁(又は「カットアウト」)領域がラベル付けされる。グレーレベル及びラベル付けされたVOIを用いて、特徴抽出ユニットを備える機械学習サブシステムが2D及び3D形状特徴、ヒストグラムベースの特徴といった異なる特徴を計算する。

【0024】

図3は、本発明により実現される処理を表すフロー図である。つまり、図3は、本書の発明の適用処理に関する1つの実施形態を説明するフロー図である。ボックス550は、その領域に関する臨床的なグラウンドトゥースが判明している医療画像トレーニングデータのセットに関する分類器のトレーニングを表す。ある実施形態において、そのステップは、多数の真及び偽の領域を含むよう選択される医療画像トレーニングデータのセットに基づき分類器をトレーニングするステップを含むことができる。そのステップでは、真及び偽の領域がCAD処理により識別され、自動的に分割される。そのステップでは、分割されたトレーニング領域がグラウンドトゥース、つまり真又は偽に関して各トレーニング領域を分類する少なくとも一人の専門家によりレビューされる。そのステップでは、特徴

10

20

30

40

50

ールが各分割された領域から識別及び抽出される。そのステップでは、特徴プールが、最適な特徴サブセットを識別する遺伝的アルゴリズムにより処理される。そのサブセットは、サポートベクタマシンをトレーニングするのに使用される。

【0025】

ボックス540は、分類のための候補である領域を非トレーニング医療画像データ内で検出するステップを表わし、ボックス560は、候補領域を分割するステップを表す。ボックス580は、各注目領域に関する完全な特徴セット(プール)を抽出するため分割された領域を更に処理するステップを表す。ボックス600は、サポートベクタマシンをトレーニングするため、最適な特徴サブセットを識別する遺伝的アルゴリズムを用いてそれぞれ知られたトレーニング領域の完全な特徴セットに関する処理を行うステップを表す。トレーニング後、SVMは候補領域から抽出された特徴のセットに関して動作する。トレーニングステップは、分割処理におけるリコメンダを用いることを含むことができる。そのリコメンダは、病理が判明している領域の最良の分割に対する実際の選択をトレーナに提供する。

10

【0026】

本発明の方法を実行する又は本発明のFPR分類器を駆動するのに必要なソフトウェアは、論理的な機能を実現するための実行可能命令の順序付けされたリストを有することができる点に留意することが重要である。そのようなものとして、そのソフトウェアは、コンピュータベースのシステム、プロセッサを含むシステム、若しくは命令実行システム、装置若しくはデバイスから命令をフェッチし、その命令を実行することができる他のシステムといった命令実行システム、装置若しくはデバイスにより、又はそれらに結び付けて使用するいずれかのコンピュータ可読媒体において実現されることができる。本書の文脈において、「コンピュータ可読媒体」は、命令実行システム、装置又はデバイスにより、又はそれらと結び付けて使用するプログラムを含み、格納し、通信し、伝播し、又は運搬するいずれの手段ともすることができる。

20

【0027】

コンピュータ可読媒体は、例えば、電子的、磁氣的、光学的、電磁氣的、赤外線による、又は半導体のシステム、装置、デバイス又は伝播媒体とすることができるが、これらに限定されるものではない。コンピュータ可読媒体のより特定の例(非包括的なりスト)は、以下を含む：1つ又は複数のワイヤを持つ電気接続(電子)、ポータブルコンピュータディスク(磁気)、ランダムアクセスメモリ(RAM)(磁気)、リードオンリメモリ(ROM)(磁気)、消去可能プログラム可能リードオンリメモリ(EPRM又はフラッシュメモリ)(磁気)、光学ファイバ(光学)及びポータブル・コンパクトディスク・リードオンリメモリ(CDROM)(光学)である。コンピュータ可読媒体は、プログラムがプリントされる紙又は他の適切な媒体であってさえよいことに留意されたい。なぜなら、プログラムは、例えば、紙又は他の媒体の光学スキャニングを介して電子的にキャプチャされ、コンパイルされ、インタープリタにかけられ(interpreted)、又は必要なら適切な態様で処理され、その後コンピュータメモリ格納されることができるからである。

30

【0028】

本発明の上述された実施形態は、特に、いずれの「好ましい」実施形態も本発明の原理の明確な理解のために単に記載される実現の可能な例示に過ぎないことを強調しておく。更に、本発明の精神及び原理から実質的に逸脱することなく、多くの変形及び変更が上述した本発明の実施形態に対してなされることができる。斯かる変形及び変更はすべて、本開示により教示されるものとして意図され、本発明の範囲に含まれるものであり、添付の請求項により保護されるものである。

40

【図面の簡単な説明】

【0029】

【図1】サポートベクタマシン(SVM)を用いてコンピュータ断層撮影(CT)医療画像からコンピュータ支援検出(CAD)における誤検出低減(FPR)を実行するためのシステムを表す図である。

50

【図2】サポートベクタマシンの基本的なアイデアを表す図である。
【図3】本発明の例示的な処理を特定する処理フロー図である。

【図1】

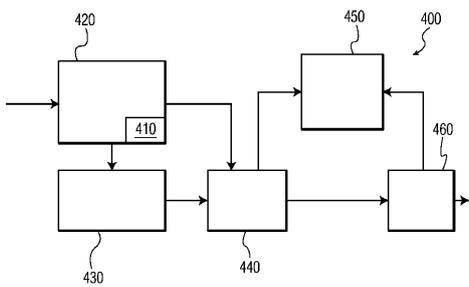


FIG. 1

【図2】

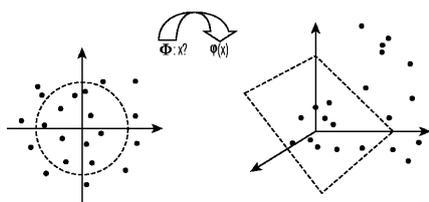


FIG. 2

【図3】

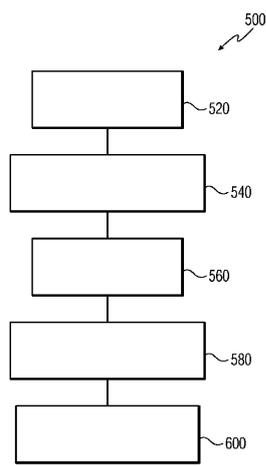


FIG. 3

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2005/053824

A. CLASSIFICATION OF SUBJECT MATTER INV. G06K9/62		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) G06K G06T		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the International search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, PAJ, INSPEC, COMPENDEX		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 136 914 A (UNIVERSITA DEGLI STUDI DI BOLOGNA) 26 September 2001 (2001-09-26) paragraphs [0008] - [0016] paragraphs [0038] - [0074] paragraphs [0157] - [0188] claims 1,4,7,9	1,4-7
Y		2,3
Y	WO 03/070102 A (THE REGENTS OF THE UNIVERSITY OF MICHIGAN; CHAN, HEANG-PING; SAHINER,) 28 August 2003 (2003-08-28) page 3 - page 4 page 27 - page 35 claims 38,39,41,42,45	1-7
	-/--	
<input checked="" 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 : *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. ** document member of the same patent family		
Date of the actual completion of the international search 6 July 2006		Date of mailing of the international search report 17/07/2006
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2200 HV Rijswijk Tel. (+31-70) 940-2040, Tx. 31 651 epo.nl Fax: (+31-70) 940-3016		Authorized officer Deltorn, J-M

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

INTERNATIONAL SEARCH REPORT

International application No PCT/IB2005/053824

G(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	FROHLICH H ET AL: "Feature selection for support vector machines by means of genetic algorithms" PROCEEDINGS 15TH IEEE INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE, ICTAI 2003, SACRAMENTO, CA, NOV. 3 - 5, 2003, IEEE INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE, LOS ALAMITOS, CA, IEEE COMP. SOC, US, vol. CONF. 15, 3 November 2003 (2003-11-03), pages 142-148, XP010672219 ISBN: 0-7695-2038-3 sections 1,3 and 4	1-7
Y	XU J Q ET AL: "A feature selection method based on minimizing generalization bounds of SVM via GA" PROCEEDINGS OF THE 2003 IEEE INTERNATIONAL SYMPOSIUM ON INTELLIGENT CONTROL.(ISIC'03), HOUSTON, TX, OCT. 5 - 8, 2003, IEEE INTERNATIONAL SYMPOSIUM ON INTELLIGENT CONTROL, NEW YORK, NY : IEEE, US, 5 October 2003 (2003-10-05), pages 996-999, XP010674023 ISBN: 0-7803-7891-1 sections I, II and IV	1-7
A	MILLER MT, JEREBKO AK, MALLEY JD, SUMMERS RM: "Feature selection for computer-aided polyp detection using genetic algorithms" PROCEEDINGS OF SPIE - VOLUME 5031 - MEDICAL IMAGING, 2003, PHYSIOLOGY AND FUNCTION: METHODS, SYSTEMS, AND APPLICATIONS, ANNE V. CLOUGH, AMIR A. AMINI, EDITORS, May 2003 (2003-05), pages 102-110, XP002388915 sections I and 2	1-7
A	MICHAEL L RAYMER ET AL: "Dimensionality Reduction Using Genetic Algorithms" IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, IEEE SERVICE CENTER, NEW YORK, NY, US, vol. 4, no. 2, July 2000 (2000-07), XP011020115 ISSN: 1089-778X sections I and II	1-7

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/IB2005/053824

Patent document cited in search report	Publication date	Patent family member(s)	Publication date	
EP 1136914	A	26-09-2001	IT 8020000166 A1	24-09-2001
			US 2001031076 A1	18-10-2001
WO 03070102	A	28-08-2003	AU 2003216295 A1	09-09-2003

フロントページの続き

(51)Int.Cl. F I テーマコード(参考)
 G 0 6 T 7/00 3 5 0 B
 A 6 1 B 5/00 G

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW

(72)発明者 ボロクツキィ リラ
 アメリカ合衆国 ニューヨーク州 1 0 5 1 0 - 8 0 0 1 ブリアクリフ マノアー ピーオー
 ボックス 3 0 0 1

(72)発明者 ツァオ ルイン
 アメリカ合衆国 ニューヨーク州 1 0 5 1 0 - 8 0 0 1 ブリアクリフ マノアー ピーオー
 ボックス 3 0 0 1

(72)発明者 リー クオク プン
 アメリカ合衆国 ニューヨーク州 1 0 5 1 0 - 8 0 0 1 ブリアクリフ マノアー ピーオー
 ボックス 3 0 0 1

Fターム(参考) 4C093 AA21 CA18 FF15 FF18 FF28
 4C117 XB09 XK04 XK05 XR07
 5B075 NR01 UU29
 5L096 AA06 BA06 GA19 JA11 JA22 KA04

【要約の続き】

された候補領域から特徴セットを抽出し、最適な特徴サブセットに基づきトレーニング後サポートベクタマシンを用いてその候補領域を分類し、その候補特徴のセットを処理する。