

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7117246号
(P7117246)

(45)発行日 令和4年8月12日(2022.8.12)

(24)登録日 令和4年8月3日(2022.8.3)

(51)国際特許分類	F I
G 0 6 F 16/906 (2019.01)	G 0 6 F 16/906
G 0 6 F 16/9038(2019.01)	G 0 6 F 16/9038
G 1 6 H 50/70 (2018.01)	G 1 6 H 50/70

請求項の数 9 (全20頁)

(21)出願番号	特願2018-548312(P2018-548312)	(73)特許権者	590000248 コーニンクレッカ フィリップス エヌ ヴェ Koninklijke Philips N.V. オランダ国 5656 アーヘー アイン ドーフェン ハイテック キャンパス 52 High Tech Campus 52, 5656 AG Eindhoven, N etherlands
(86)(22)出願日	平成29年3月8日(2017.3.8)	(74)代理人	100122769 弁理士 笛田 秀仙
(65)公表番号	特表2019-512795(P2019-512795 A)	(74)代理人	100163809 弁理士 五十嵐 貴裕
(43)公表日	令和1年5月16日(2019.5.16)	(72)発明者	アグラワル ヴァルティカ 最終頁に続く
(86)国際出願番号	PCT/IB2017/051345		
(87)国際公開番号	WO2017/158472		
(87)国際公開日	平成29年9月21日(2017.9.21)		
審査請求日	令和2年3月6日(2020.3.6)		
(31)優先権主張番号	62/309,067		
(32)優先日	平成28年3月16日(2016.3.16)		
(33)優先権主張国・地域又は機関	米国(US)		
前置審査			

(54)【発明の名称】 類似するプロファイルを持つ患者を共に分類する分類モデルの性能を改善するための適合性フィードバック

(57)【特許請求の範囲】

【請求項1】

患者コホート識別装置であって、

表示要素及び少なくとも1つのユーザ入力装置を持つコンピュータを有し、前記コンピュータは、患者データベース中の患者についての特徴の値を有する患者データを保存する患者データベースと通信し、前記コンピュータは、患者コホート識別方法を実行するようプログラムされ、前記方法は、

特徴のセットを選択するための患者データに対する自動特徴選択処理、及び前記特徴のセットに依存した患者比較メトリックを用いた前記患者データベースの患者の自動分類を実行するステップ

を含み、前記方法は更に、

前記自動分類に従ってクエリ患者に類似する又は類似しない1人以上のサンプル患者についての情報を前記表示要素上に表示し、前記少なくとも1つのユーザ入力装置を介して前記1人以上のサンプル患者を前記クエリ患者と比較するユーザ入力された比較値を受信するステップと、

前記ユーザ入力された値と前記1人以上のサンプル患者を前記クエリ患者と比較する前記患者比較メトリックによって計算された比較値との間の一致度を上昇させるよう前記患者比較メトリックを調節するステップであって、前記特徴のセット及び前記患者比較メトリックの特徴の重みの少なくとも一方を調節することを含む、ステップと、

前記調節された患者比較メトリックを用いて前記自動分類を繰り返すステップと、

の少なくとも 1 回の繰り返しを実行するステップを含み、前記方法は更に、最後の繰り返しにより生成された前記調節された患者比較メトリックを用いて前記クエリ患者について患者コホートを識別するステップを含み、

前記表示し受信するステップは、

(Ⅰ) 量的なランク付けのスケールで前記クエリ患者に対する少なくとも一人のサンプル患者の類似度をランク付けする要求を表示し、前記量的なランク付けのスケールでの前記サンプル患者の前記受信された類似度ランク付けとして、前記サンプル患者についてのユーザ入力された比較値を受信するステップと、

(Ⅱ) 前記 2 人のサンプル患者のうちどちらが前記クエリ患者に類似しているかを選択する要求を表示し、前記 2 人のサンプル患者のうちどちらが前記クエリ患者に類似しているかの前記受信された選択として、ユーザ入力された比較値を受信するステップと、
10
を有する、

患者コホート識別装置。

【請求項 2】

前記識別するステップは、

前記自動分類の最後の繰り返しにより生成された前記クエリ患者を含むクラスタの少なくとも一部として患者コホートを識別するステップを含む、請求項 1 に記載の患者コホート識別装置。

【請求項 3】

前記表示し受信するステップは、

(Ⅰ) 最も直近に実行された自動分類によって生成された前記クエリ患者を含むクラスタに属する 1 人以上の類似するサンプル患者についての情報を、前記表示要素上に表示するステップと、

(Ⅱ) 最も直近に実行された自動分類によって生成された前記クエリ患者を含むクラスタに属さない 1 人以上の類似しないサンプル患者についての情報を、前記表示要素上に表示するステップと

を有する、請求項 1 又は 2 に記載の患者コホート識別装置。

【請求項 4】

前記表示するステップは、

それぞれが 2 つ以上のモダリティの特徴に対して前記 1 人以上のサンプル患者及び前記クエリ患者をプロットする、2 つ以上のグラフィカルなモダリティ表現を同時に表示するステップであって、前記 2 つ以上のグラフィカルなモダリティ表現は、臨床、放射線、ゲノム、人口統計学的及び生理学的モダリティから成る群から選択されたモダリティについてのグラフィカルなモダリティ表現を含む、ステップ

を有する、請求項 1 乃至 3 のいずれか一項に記載の患者コホート識別装置。

【請求項 5】

前記調節するステップは、

(Ⅰ)

(1) 特徴を追加又は削除して調節された特徴のセットの候補を生成することにより、前記特徴のセットを調節する動作と、

(2) 前記 1 人以上のサンプル患者を前記クエリ患者と比較する、前記調節された特徴のセットの候補を用いた患者比較メトリックを用いて、比較値を計算する動作と、

(3) 前記動作 (2) において計算された計算値が、前記ユーザ入力された比較値に対して一致度が増大したか減少したかに基づいて、前記調節された特徴のセットの候補を許容又は破棄する動作と、

をそれぞれが含む複数の特徴セット調節の繰り返しを実行するステップ、又は

(Ⅱ) 前記特徴のセットにおける特徴の数を減少させる次元低減を実行するステップ
40
を有する、請求項 1 乃至 4 のいずれか一項に記載の患者コホート識別装置。

【請求項 6】

10

20

30

40

50

前記調節するステップは、前記患者比較メトリックの特徴重みを調節するステップを有し、

前記調節するステップは、

(1) 前記患者比較メトリックの少なくとも1つの特徴重みの値を増大又は減少させて、調節された患者比較メトリックの候補を生成することにより、患者比較メトリックを調節する動作と、

(2) 前記1人以上のサンプル患者を前記クエリ患者と比較する、前記調節された患者比較メトリックの候補を用いて、比較値を計算する動作と、

(3) 前記動作(2)において計算された計算値が、前記ユーザ入力された比較値に対して一致度が増大したか減少したかに基づいて、前記調節された患者比較メトリックの候補を許容又は破棄する動作と、

をそれぞれが含む複数の特徴重み調節の繰り返しを実行するステップを有する、

請求項1乃至5のいずれか一項に記載の患者コホート識別装置。

【請求項7】

前記自動特徴選択処理は、主成分分析、情報取得、及び対特徴相關の1つである、請求項1乃至6のいずれか一項に記載の患者コホート識別装置。

【請求項8】

患者コホート識別装置の作動方法であって、前記患者コホート識別装置は、表示要素及び少なくとも1つのユーザ入力装置を持ち、患者データベース中の患者についての特徴の値を有する患者データを保存する患者データベースと通信するコンピュータを有し、前記コンピュータは、前記方法を実行するようプログラムされ、前記方法は、

特徴のセットに依存した患者比較メトリックを用いた前記患者データベースの患者の自動分類を実行するステップ

を含み、前記方法は更に、

前記自動分類に従ってクエリ患者に類似する又は類似しない1人以上のサンプル患者についての情報を前記表示要素上に表示し、前記少なくとも1つのユーザ入力装置を介して前記1人以上のサンプル患者を前記クエリ患者と比較するユーザ入力された比較値を受信するステップと、

調節のない患者比較メトリックと比較して、前記ユーザ入力された値との改善された一致を持つ調節された患者比較メトリックを生成するよう、前記特徴のセット及び前記患者比較メトリックの特徴の重みの少なくとも一方を調節するステップと、

前記調節された患者比較メトリックを用いて前記自動分類を繰り返すステップと、の少なくとも1回の繰り返しを実行するステップ

を含み、前記方法は更に、

最後の繰り返しの前記自動分類の反復によって生成された前記クエリ患者を含むクラスタの少なくとも一部として、前記クエリ患者についての患者コホートを識別するステップを含み、

前記表示し受信するステップは、

量的なランク付けのスケールで前記クエリ患者に対する前記サンプル患者の類似度をランク付けする要求を表示し、前記量的なランク付けのスケールでの前記サンプル患者の前記受信された類似度ランク付けを受信するステップと、

2人のサンプル患者のうちどちらが前記クエリ患者に類似しているかを選択する要求を表示し、前記2人のサンプル患者のうちどちらが前記クエリ患者に類似しているかの選択を受信するステップと、

の少なくとも一方を有する、

方法。

【請求項9】

前記表示するステップは、

それぞれが2つ以上のモダリティの特徴に対して前記1人以上のサンプル患者及び前記クエリ患者をプロットする、2つ以上のグラフィカルなモダリティ表現を同時に表示する

10

20

30

40

50

ステップ

を有する、請求項 8 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般的に医療分野、電子臨床決定支援（CDS）分野、臨床検査分野、ゲノム分野及び関連分野に関する。

【背景技術】

【0002】

種々の医療タスクは、関連する類似性を持つ患者のコホート（cohort）を識別することから利益を得る。例えば、臨床試験を設計する際の重要な最初のステップは、臨床試験の登録のため患者を識別することである。結果の正当性を確実にするため、登録された患者は、他の患者と十分に類似しているべきであり、これにより、異なる患者結果が、年齢、性別、民族性、慢性の医療条件の有無等（これらが臨床試験の目標に関連しない場合）における相違のような外来性の要因による患者結果の相違ではなく、臨床試験の対象（例えば新たな製薬）に信頼性高く起因するものとなるようにするべきである。患者結果は多くの関連する要因によって影響を受け得るため、臨床試験において登録するため適切な患者を識別するタスクは困難である。

10

【発明の概要】

【発明が解決しようとする課題】

20

【0003】

コホート識別はまた、登録の後に、臨床試験の結果の解析の間に実施されても良い。該登録の範囲内で、負の結果を持つか正の結果を持つかで、患者が関連のある2つのコホートを自然に形成する。しかしながら、臨床試験の生のデータ結果に影響を与えるいずれかの外来性の要因を識別し考慮するため、これらのコホートが更に、正及び負のコホート内で類似性及び相違性に基づいて更にセグメント化されても良い。

【0004】

類似コホート識別タスクは、例えば疾患リスク因子を評価するため、又は種々の以前の検査からのデータを組み合わせる「メタ検査」を実行する際のように、他のタイプの医療検査において実行される。

30

【0005】

他の医療タスクは、臨床診断及び患者の処置を含む。斯かるタスクにおいては、臨床医は、類似する過去の患者と現在の患者を比較することから利益を得ることができる。ここでもまた、「類似する」患者を識別するタスクは困難である。2人の患者は同一ではなく、コホート選択タスクは、どの相違が重要であるか重要でないかを評価することを必要とする。

【0006】

以下は、以上に言及された問題及びその他の問題に対処する新規で改善されたシステム及び方法を開示する。

【課題を解決するための手段】

40

【0007】

開示される一態様においては、患者コホート識別装置が開示される。コンピュータは、表示要素及び少なくとも1つのユーザ入力装置を持つ。前記コンピュータは、患者データベース中の患者についての特徴の値を有する患者データを保存する患者データベースと通信する。前記コンピュータは、以下を含む患者コホート識別方法を実行するようプログラムされる。特徴のセットを選択するための患者データに対する自動特徴選択処理が実行され、前記特徴のセットに依存した患者比較メトリックを用いた前記患者データベースの患者の自動分類が実行される。少なくとも1回の繰り返しが実行され、該繰り返しは、前記自動分類に従ってクエリ患者に類似する又は類似しない1人以上のサンプル患者を前記表示要素上に表示するステップと、前記少なくとも1つのユーザ入力装置を介して前記1人

50

以上のサンプル患者を前記クエリ患者と比較するユーザ入力された比較値を受信するステップと、前記ユーザ入力された値と前記1人以上のサンプル患者を前記クエリ患者と比較する前記患者比較メトリックによって計算された比較値との間の一致度を上昇させるよう前記患者比較メトリックを調節するステップであって、前記特徴のセット及び前記患者比較メトリックの特徴の重みの少なくとも一方を調節することを含む、ステップと、前記調節された患者比較メトリックを用いて前記自動分類を繰り返すステップと、を含む。最後の繰り返しにより生成された前記調節された患者比較メトリックを用いて、前記クエリ患者についての患者コホートが識別される。

【0008】

他の開示される態様においては、患者コホート識別装置が開示される。コンピュータは、表示要素及び少なくとも1つのユーザ入力装置を持つ。前記コンピュータは、患者データベース中の患者についての特徴の値を有する患者データを保存する患者データベースと通信する。前記コンピュータは、それぞれが2つ以上のモダリティの座標特徴に対してデータベースの患者をプロットする、2つ以上のグラフィカルなモダリティ表現を、前記表示要素上に同時に表示するステップと、1つのグラフィカルなモダリティ表現における患者のクラスタの選択を受信するステップと、前記選択の受信に応答して、他の同時に表示されたグラフィカルなモダリティ表現における選択された患者のクラスタの患者を強調するステップと、を含む、患者コホート識別方法を実行するようプログラムされる。

【0009】

他の開示される態様においては、患者コホート識別方法が開示され、前記方法は、表示要素及び少なくとも1つのユーザ入力装置を持ち、患者データベース中の患者についての特徴の値を有する患者データを保存する患者データベースと通信するコンピュータとともに実行される。前記患者コホート識別方法は、以下を含む。特徴のセットに依存した患者比較メトリックを用いて、前記患者データベースの患者の自動分類が実行される。前記自動分類に従ってクエリ患者に類似する又は類似しない1人以上のサンプル患者についての情報を前記表示要素上に表示するステップと、前記少なくとも1つのユーザ入力装置を介して前記1人以上のサンプル患者を前記クエリ患者と比較するユーザ入力された比較値を受信するステップと、調節のない患者比較メトリックと比較して、前記ユーザ入力された値との改善された一致を持つ調節された患者比較メトリックを生成するよう、前記特徴のセット及び前記患者比較メトリックの特徴の重みの少なくとも一方を調節するステップと、前記調節された患者比較メトリックを用いて前記自動分類を繰り返すステップと、を含む少なくとも1回の繰り返しが実行される。最後の繰り返しの前記自動分類の反復によって生成された前記クエリ患者を含むクラスタの少なくとも一部として、前記クエリ患者についての患者コホートが識別される。

【0010】

利点のひとつは、改善されたコホート選択についての臨床医からの適合性フィードバック(relevance feedback)を提供する点にある。

【0011】

他の利点は、臨床医による全体的な患者レベルの解析に基づくコホート選択についての適合性フィードバックを提供する点にある。

【0012】

他の利点は、臨床医が特徴レベルの解析を実行することなく、関連する特徴の選択についての臨床医からの適合性フィードバックを提供する点にある。

【0013】

他の利点は、臨床医が種々のモダリティの相互関係(臨床、放射線、ゲノム、人口統計学的、生理学的等)を視覚化することができる、グラフィカルなユーザインターフェースを提供する点にある。

【0014】

所与の実施例は、以上の利点のうちの0個、1つ、2つ、それ以上又は全てを提供し得、及び/又は、本開示を読み理解することにより当業者には明らかとなる他の利点を提供

10

20

30

40

50

し得る。

【0015】

本発明は、種々の構成要素及び構成要素の配置、並びに種々のステップ及びステップの配置の形をとり得る。図面は単に好適な実施例を説明する目的のためのものであり、本発明を限定するものとして解釈されるべきではない。

【図面の簡単な説明】

【0016】

【図1】患者コホート識別装置を図式的に示す。

【図2】図1の患者コホート識別装置により適切に実行される患者コホート識別方法を図式的に示す。

10

【図3】図2の方法の提示動作の適切な実施例の例を図式的に示す。

【図4】図2の方法の提示動作の適切な実施例の例を図式的に示す。

【図5】ここで記載される患者コホート視覚表現及びナビゲーションツールを図式的に示す。

【発明を実施するための形態】

【0017】

ここで、コホートを選択することの複雑さが、患者をコホートにグループ化するための適切な（低減された）患者特徴セットを選択することにより低減させられ得ることが認識される。コホートを選択するために用いられる患者特徴は、当面の医療タスク（例えば臨床試験における登録のため患者を選択すること、又は現在臨床診断中である患者に類似する患者を選択すること、等）に関連のあるものを含むべきであり、当該医療タスクに関連しない患者特徴を含むべきではない。特徴選択は、利用可能な患者特徴の数は非常に多く、例えば人口統計学的データ（年齢、性別、体重、民族性等）、常習的な行動条件（喫煙、大量のアルコール消費、種々の娯楽薬の消費等）の有無、種々の慢性臨床条件（高血圧、糖尿病、喘息、心臓病等）の有無、種々の急性疾患（肺炎又はその他の急性呼吸病、種々の腫瘍状態等）の有無、これらに関する特徴（例えば癌のステージ及びグレード、等を含み得るため、容易なものではない。遺伝子のシーケンシングは、種々の医療条件との変化する既知の又は考えられ得る相関を持つ、多くのゲノム標識を提供し得るため、急速に発展するゲノム分野が、利用可能な患者特徴のリストに迅速に追加している。例えば、幾つかの医療データベースは、数百以上の特徴を定義するデータを含み、ゲノムデータの利用可能性の継続的な拡大は、患者についての特徴の数を数千にまで増大させ得る。斯かる大きな特徴空間は、臨床タスクについてコホートを選択するための「最適な」特徴セットを選択するための、大きな困難さをもたらしている。

20

【0018】

種々の教師なしの（低減された）特徴セット選択手法が知られている。典型的な自動化された特徴選択手法は、特徴の弁別力を測定し、最も弁別的な特徴を選択する。斯かる手法のひとつは、主成分分析（PCA）であり、低減された数の特徴を持つデータセットの分散を捕捉するよう特徴を選択する。例えば特徴毎の情報取得（IG）、又は種々の対毎の特徴相関メトリック（例えば最も高いIGを提供する特徴を選択すること、又は他の特徴と強く相関する特徴を除去すること）といった、他の区別メトリックが利用されても良い。

30

【0019】

強力であるが教師なし自動特徴セット選択手法は、患者コホートを識別するための特徴を選択するために用いられる場合、大きな制限を持つ。臨床タスクと相関しない非常に弁別的な特徴は、あまり弁別力がないが医療タスクと相関している他の特徴より、選択され得る。教師なし特徴セット選択手法は、患者特徴がなぜ根拠となるべきかについての生理学的な根拠を考慮に入れることができない。例えば、特定の代謝経路についての問題によって疑われる臨床状態を考える。当該代謝経路の一部であると知られているゲノム標識は、この場合には関連性のあるものである見込みが高いが、PCA又はその他の教師なし特徴選択手法は、全体的な弁別力が低い場合には、当該ゲノム標識を選択しない可能性があ

40

50

る。

【 0 0 2 0 】

原則的に、これらの問題は、臨床医により実行される手動の特徴選択、又は医師が教師なし自動特徴選択により生成された初期特徴セットを調節する（適合性フィードバック）ハイブリッド型の手法によって、軽減され得る。しかしながら、実際には臨床医は、或る患者がなぜ対象の患者（ここでは「クエリ患者」と呼ばれる）に類似する又は類似しないと考えられるかを、特定の特徴について明瞭に表現できない場合がある。臨床医は、医師の過去の体験及び学習の全体について、患者を全体的にみる傾向がある。従って、臨床医は、どの特徴が効果的に類似性又は非類似性を要約するかを正確に表現することができない状態で、特定の患者がクエリ患者に類似する又は類似しないと認識し得る。更に、所与の臨床タスクについて根拠となる特徴を識別するための数百もの利用可能な特徴の候補を調べるために必要な時間を熟練の臨床医がとることは、実用的ではない。

【 0 0 2 1 】

ここで開示される手法は、自動的に選択される特徴セットを用いて実行された自動分類の閲覧を通して患者レベルにおいて臨床医により提供された後続する適合性フィードバックと教師なし特徴選択を組み合わせることにより、これらの困難さを克服する。これらの手法においては、クエリ患者及びその他のクラスタを含む患者のクラスタを識別するため教師なし自動患者分類を実行するために、初期の自動化された特徴セットが利用される。クエリ患者を含むクラスタは、初期特徴セットによる類似する患者のセットを定義し、他のクラスタは、種々のあまり類似しない患者をグループ化する。臨床医は次いで、これらの分類結果を閲覧し、類似する又は類似しない患者を選択する（適合性フィードバック）。次いで、これらの臨床医の選択に好適に合致するよう特徴セットが自動的に調節され、該調節された特徴セットを用いて分類が繰り返される。この処理は、教師なし自動分類が、臨床医が（少なくとも部分的に）満足できるクラスタを生成するまで、繰り返されても良い。

【 0 0 2 2 】

当該手法は、大きな特徴空間の初期の近似的な選別を提供するための、教師なし特徴セット選択の能力を利用する。PCA又はその他の教師なし特徴選択処理により選択された初期特徴セットを用いて、当該初期特徴セットを用いて測定されたクエリ患者に類似する（又は類似しない）患者を識別するよう患者が分類される。臨床医は、1つ以上の類似する（又は類似しない）サンプル患者を提示され、臨床医が適合性フィードバックを提供することができるユーザインタフェースを提供される。例えば、医師は、診断されている患者（クエリ患者P_Q）に類似するもものとして初期の分類において識別された類似するサンプル患者のセット{P_C}を提示され得る。これらの「類似する」サンプル患者は例えば、分類がクエリ患者P_Qを割り当てた同じクラスタから、又は初期の特徴セットにより定義された距離メトリックを用いて最短距離|P_Q-P_C|を持つクラスタのサブセットから引き出されても良い。医師は次いで、ランクスケール1乃至5を用いて、クエリ患者P_Qに類似する又は類似しないものとして患者をランク付けしても良く、ここで1は最も類似し5が最も類似しないことを示す。その後、考慮される患者についての医師の類似度ランクに最も整合する調節された特徴セットを生成するため、特徴セット調節が実行される。分類が再び繰り返され、クエリ患者P_Q又は幾つかのそのサブセットを含むクラスタが再び、類似度ランク付けのため医師に提示される。当該処理は、クエリ患者P_Qを含むクラスタが当面の医療タスクを実行するための適切なコホートであると医師が満足するまで、繰り返されても良い。

【 0 0 2 3 】

有利にも、適合性フィードバックのための当該手法は、臨床医が特徴空間の抽象化されたレベルにおいて特徴セットを評価することを必要としない。その代わり、臨床医は、個々の患者を比較及び対比するより慣れた設定で操作を行い、それにより臨床医が、医師の過去の体験及び学習の全体を利用して、適合性フィードバックの決定を為すことができる。好適には、ユーザインタフェースは、医師が考察中のそれぞれの提案された類似する患

10

20

30

40

50

者 P C の完全な医療記録を探し、更にクエリ患者 P Q の完全な医療記録を探して、臨床医がアクセスのために利用している情報の同じ供給源を用いて適合性フィードバック評価を為すことを可能とする。

【 0 0 2 4 】

図 1 を参照すると、患者コホート識別装置は、表示要素及び少なくとも 1 つのユーザ入力装置を持つコンピュータを含む。図示されるコンピュータは、2 つのコンピュータ、即ち特徴選択又は分類のような計算が集中する動作を実行するサーバコンピュータ 1 0 と、表示要素 1 4 並びに例えばキーボード 1 6 及びマウス 1 8 (又はトラックボール、トラックパッド、タッチスクリーン若しくはその他のポインティングデバイス) のような少なくとも 1 つのユーザ入力部を含むか又はこれらと動作可能に接続された、デスクトップコンピュータ、ノート型コンピュータ、タブレットコンピュータ等のような、ユーザインタフェースコンピュータ 1 2 と、を含む。コンピュータ 1 0 、 1 2 は、患者データベース中の患者についての特徴の値を有する患者データを保存する患者データベース 2 0 と通信する。患者データベース 2 0 は例えば、電子健康記録 (E H R) 、電子医療記録 (E M R) 、画像保存及び通信システム (放射線画像 / データのために利用される P A C S) 、心臓血管情報システム (C V I S) 、又はこれらの種々の組み合わせ等の 1 つ以上を含んでも良い。種々の要素 1 0 、 1 2 、 2 0 は、病院のローカルエリアネットワーク (L A N) 、無線 L A N (W L A N) 、インターネット、又はこれらの種々の組み合わせ等のような、種々のデータ経路により相互接続されても良い。

10

【 0 0 2 5 】

コンピュータ 1 0 、 1 2 は、種々の処理を実装するようプログラムされる。自動特徴選択処理 2 2 は、患者データベース 2 0 に含まれる利用可能な、又は患者データベース 2 0 に含まれる情報から導出可能な、特徴の典型的に非常に大きなセットから、低減された特徴のセットを選択するよう実行される。特徴選択処理 2 2 は例えば、主成分分析 (P C A) 特徴選択処理、情報取得 (I G) 特徴ランク付け処理、又は対相関特徴除去処理等であっても良い。自動特徴選択処理 2 2 は、特徴 2 4 のセットを識別し、典型的には高い弁別力を持つ特徴を選択する。患者データベース 2 0 は、数十、数百、又はそれ以上の特徴を患者毎に保存し得る (明示的に又は暗黙的に、即ち他の保存された情報から導出可能) ことは理解されるであろう。幾つかの限定するものではない特徴の例は、人口統計学的特徴 (患者の年齢、性別、体重、民族性等) 、常習的な行動条件 (喫煙、大量のアルコール消費、種々の娯楽薬の消費等) の有無を示す特徴、種々の慢性臨床条件 (高血圧、糖尿病、喘息、心臓病等) の有無を示す特徴、種々の急性疾患 (肺炎又はその他の急性呼吸病、種々の腫瘍状態等) の有無を示す特徴、癌のステージ、癌のグレードのような状況に特有な特徴、特定の遺伝子の値、種々のタンパク質発現レベル、又はその他の遺伝標識のようなゲノム特徴、等を含む。患者データセット 2 6 が斯くして生成され、各患者が、特徴のセット 2 4 の特徴について患者データベース 2 0 から引き出された値によって注記付けされるか又は表現される。

20

【 0 0 2 6 】

分類処理 3 0 は、患者データセット 2 6 の患者を、クラスタのセット 3 2 へとグループ化するため、教師なし学習を実行する。一般的に、目標は、クエリ患者 P Q に類似する患者の患者コホートを識別することであり、従って、クラスタのセット 3 2 は、クエリ患者 P Q を含むクラスタ 3 4 (換言すれば、クラスタ 3 4 は、クエリ患者 P Q が属する分類処理 3 0 により生成されたクラスタである) と、分類処理により生成された他のクラスタ 3 6 と、を含む。分類処理は、k 平均分類、連結度ベース又は階層的分類、セントロイドベースの分類、又は期待値最大化 (E M) 分類等のような、いずれの既知の分類手法を利用して良い。該分類は、特徴のセット 2 4 に依存する患者比較メトリックを利用する。2 人の患者 P i 及び P j について、これらの 2 人の患者を比較する患者比較メトリックの値は、ここでは省略表記 | P i - P j | を用いて記述される。限定するものではない例として、患者比較メトリックは、より類似する患者について値が小さくなるような距離メトリックであっても良い。幾つかの適切な距離メトリックは、ユークリッド距離：

30

40

50

【数1】

$$d_{Euclidean}(P_i, P_j) = \sqrt{\sum_{n=1}^N w_n (f_{n,i} - f_{n,j})^2} \quad (1)$$

10 であり、ここで $n = 1, \dots, N$ は、特徴のセット 2 4 の特徴をインデクシングし $f_{n,i}$ 及び $f_{n,j}$ は、それぞれ患者 P_i 及び P_j についての n 番目の特徴の値であり、 w_n は、式 (1) のユークリッド距離における n 番目の特徴についての特徴重みである。他の例としては、患者比較メトリックは、平方根が省略されること以外は式 (1) と同一であるユークリッド二乗距離であっても良い。距離メトリックの代わりに、患者比較メトリックは代替として、より類似する患者について値が大きくなるような類似度メトリックであっても良い。これらは単に例である。一般的に、患者比較メトリックは好適には、特徴のセット 2 4 に機能的に依存し、個々の特徴の寄与が特徴重み (例えは式 (1) のユークリッド距離の例における特徴重み w_n) により制御される。調節可能な特徴重みを含まない患者比較メトリックを利用することも考えられ得る。

【0027】

20 選択された分類処理 3 0 について、分類結果 3 2 の特性は、患者比較メトリックの詳細、特に患者比較メトリックが機能的に依存する特徴のセット 2 4、及び特徴重み (利用可能である場合) に依存する。自動特徴選択処理 2 2 は、弁別能力の評価に基づいて特徴を選択するが、当該手法は、当面の医療タスクにより強く相関する、又は、当面のタスクに関連する幾つかの生理学的な根拠を持つ、低い弁別能力を持つ他の特徴よりも、高い弁別力の特徴を選択し得る。

【0028】

30 図 1 の患者コホート識別装置の例においては、これらの問題は、患者比較メトリックを改善するよう適合性フィードバックを提供することにより、例えは特徴のセット 2 4 及び / 又は特徴重みを調節することにより、対処される。この目的のため、例えは図示された実施例におけるコンピュータ 1 2 において、グラフィカルユーザインターフェース (G U I) 処理 4 0 が実装される。G U I 処理 4 0 は、自動分類に従って、クエリ患者 P_Q に類似する又は類似しない 1 人以上のサンプル患者についての情報を (表示要素 1 4 上に) 提示する。例えは、該サンプル患者は、クエリ患者 P_Q を含むクラスタ 3 4 から (疑似) ランダム的に選択された類似サンプル患者であっても良い。代替としては、類似サンプル患者は、例えは患者比較メトリックにより測定されたクエリ患者 P_Q に最も近い患者を選択することにより、非ランダム的に当該クラスタ 3 4 から選択されても良い。これに加えて又は代替として、類似しないサンプル患者が、例えはランダム的に又は患者比較メトリックにより測定されたクエリ患者 P_Q からのセントロイドが最も遠い他のクラスタから、選択されても良い。サンプル患者は、表示要素 1 4 を介してユーザに提示され、臨床医は、1 人以上のサンプル患者をクエリ患者と比較する比較値を提供するよう要求される。例えは、臨床医は、クエリ患者に対するサンプル患者の類似度を、1 乃至 5 のスケールで (又は 1 乃至 1 0 のスケール等で) ランク付けするよう要求されても良い。これに加えて又は代替として、臨床医は、2 人のサンプル患者のどちらのほうがクエリ患者に類似するかを選択するよう要求されても良い。斯かる手法は、特徴レベルで類似度を評価するよう臨床医に (少なくとも直接的には) 要求しないものであり、患者レベルにおける評価を要求するものであることは、留意されるであろう。このことは、一般に臨床医の教育及び経験とともに患者記録におけるあらゆる利用可能な情報に基づいて患者を分析するよう訓練された一般的な臨床医の強みを利用するものである。この手法は、臨床医の自然な動作範囲ではない特徴レベルの分析を医師が実行することを要求することを回避する。

【0029】

40 G U I 処理 4 0 は、少なくとも 1 つのユーザ入力装置 1 6、1 8 を介して、1 人以上の

10

20

30

40

50

サンプル患者をクエリ患者と比較するユーザ入力された比較値を受信する。このことは、「適合性フィードバック」を構成する。次いで、患者比較メトリック調節処理 42 が、特徴のセット 24 を調節し、及び / 又は特徴重み w_n を調節し、ユーザ入力された比較値と、1人以上のサンプル患者をクエリ患者 P_Q と比較する患者比較メトリックにより計算された比較値と、の間の一一致度を上昇させるようにする。

【 0 0 3 0 】

一手法においては、患者比較メトリック調節処理 42 は、それぞれが以下のように実行される、特徴セット調節の繰り返しを実行する。該繰り返しの第 1 のステップにおいては、特徴のセット 24 は、該セットに特徴を追加し、又は該セットから特徴を除去し、特徴の調節されたセットの候補を生成することにより調節される。次いで、1人以上のサンプル患者をクエリ患者 P_Q と比較する特徴の調節されたセットの候補を持つ患者比較メトリックを用いて、比較値が計算される。特徴の調節されたセットの候補は、計算された比較値が、ユーザ入力された比較値に対して、それぞれ増大した一致度を呈するか、減少した一致度を呈するかに基づいて、許容されるか又は棄却される。棄却された場合には、特徴の調節されたセットの候補は破棄される。許容された場合、特徴の調節されたセットの候補は、新たな（即ち更新された）特徴のセット 24 となる。この処理は、一定の回数だけ繰り返されても良いし、又は、幾つかの連続する反復の結果が棄却されるまで繰り返されても良いし、若しくは他の何らかの基準が用いられても良い。

10

【 0 0 3 1 】

他の手法においては、患者比較メトリック調節処理 42 は、それぞれが以下のように実行される、特徴重み調節の繰り返しを実行する。該繰り返しの第 1 のステップにおいては、特徴比較メトリックの少なくとも 1 つの特徴重みの値を増大又は減少させて、調節された特徴比較メトリックの候補を生成することにより、患者比較メトリックが調節される。1人以上のサンプル患者をクエリ患者と比較する調節された患者比較メトリックを用いて、比較値が計算される。調節された患者比較メトリックの候補は、ユーザ入力された比較値に対して、それぞれ増大した一致度を呈するか、減少した一致度を呈するかに基づいて、許容されるか又は棄却される。許容された場合には、新たな特徴重みが利用され、棄却された場合には破棄される。

20

【 0 0 3 2 】

ここで図 2 を参照しながら、図 1 の患者コホート識別装置を用いて実行される処理が説明される。動作 50 において、特徴選択処理 22 が実行され、特徴の（初期の）セット 24 を選択する。動作 52 において、分類処理 30 が実行され、（初期の）クラスタ 32 を生成する。動作 54 において、臨床医は、1人以上の類似する及び / 又は類似しないサンプル患者を提示され、患者比較メトリックを用いて、クエリ患者 P_Q に対する類似度 / 非類似度が測定される。より詳細には、好適には、クエリ患者に対するサンプル患者の類似度のランク付けのような、臨床医が慣れた態様で形式化された情報に対する要求、又は2人のサンプル患者のうちどちらがクエリ患者に類似するかを識別する要求の形をとる、サンプル患者についての情報が提示される。動作 56 において、ユーザ入力された比較値が受信される（例えばサンプル患者のランク付け、又は2人のサンプル患者のセットのうち類似するほうのサンプル患者の選択）。動作 60 において、サンプル患者に適用された患者比較メトリックの、ユーザ入力された比較値との一致度を増大させるよう、特徴のセット 24 及び / 又は特徴重み w_n が調節される。例えば、ユーザがサンプル患者をクエリ患者と非常に類似するとランク付けした場合、（調節された）患者比較メトリックにより測定される、より短いサンプル患者 - クエリ患者間距離に帰着する調節が許容され、当該距離を増大させる調節は棄却される。動作 62 において、調節された患者比較メトリックを用いて分類処理 30 が繰り返される。フローは次いで動作 54 に戻り、該動作により、臨床医は更新された分類に従って類似する及び / 又は類似しない患者を提示される。このループは、動作 64 において、最後の分類結果を閲覧した臨床医が、満足する結果であると結論するまで、いずれの回数だけ繰り返されても良い。

30

【 0 0 3 3 】

40

50

以下、臨床専門家による関連する特徴（動作 5 6 からの）が短い距離を呈する、元の空間から新たな空間への特徴の自動化されたマッピングとして、動作 6 0 を実装するための、幾つかの手法の例が開示される。第 1 の手法の例は、次元低減方法を利用し、第 2 の手法の例は、特徴重み調節方法を利用する。

【 0 0 3 4 】

次元低減を用いる第 1 の手法の例においては、患者 $P = \{ p_1, \dots, p_m \}$ についての特徴 $F = \{ f_1, \dots, f_n \}$ を含む患者データ (V) が表される。次いで、患者間の距離が計算されて、距離行列 (S_m 、サイズ $m \times m$ の正方な対称の行列) が得られ、古典的な多次元尺度構成法 (MDS) が利用されて、当該データのより低い次元の射影が得られる。代替の手法においては、MDS 解析は、2 から ($m - 1$) の次元の数を特定し、それぞれの低次元の投射について患者 p_1, \dots, p_m の間の対のヨークリッド距離を計算して、距離行列 $D(2), \dots, D(m - 1)$ を得ることにより、実行される。動作 5 6 において、特定の患者（グループ又は別個の対）がより類似すると予測されると考える場合、当該グループにおける全てのとり得る対の間の対の距離が最小化される。ここで、当該メトリックが最小となる $\{ 2, \dots, (m - 1) \}$ における K を特定する。行列記述：

【 数 2 】

$$V = \begin{bmatrix} v_{p_1 f_1} & \cdots & \cdots & v_{p_1 f_n} \\ v_{p_2 f_1} & \cdots & \cdots & v_{p_2 f_n} \\ \vdots & \vdots & \vdots & \vdots \\ v_{p_m f_1} & \cdots & \cdots & v_{p_m f_n} \end{bmatrix}$$

10

20

及び

【 数 3 】

$$S_m = \begin{bmatrix} d_{m_{(1,1)}} & \cdots & \cdots & d_{m_{(1,m)}} \\ d_{m_{(2,1)}} & \cdots & \cdots & d_{m_{(2,m)}} \\ \vdots & \vdots & \vdots & \vdots \\ d_{m_{(m,1)}} & \cdots & \cdots & d_{m_{(m,m)}} \end{bmatrix}$$

30

を用い、ここで行列 S_m は対称であり ($d_m(p, q) = d_m(q, p)$; $p \neq q$; $p = \{1, \dots, m\}$; $q = \{1, \dots, m\}$) 、MDS 関数は距離行列（サイズ $m \times m$ ）及び次元の数 (l ; $1 < l < m$) をとる。範囲 $\{2, \dots, (m - 1)\}$ における l について、 m 個の点の対毎の距離を計算して、対称な距離行列 $D(l)$ を得る。医師のフィードバックに基づく類似する患者のグループは、 $G = \{g_1, g_2, \dots\}$ と示され、 g_i は P からの患者のセットである。このとき、

【 数 4 】

$$D^{(l)} = \begin{bmatrix} d_{(1,1)} & \cdots & \cdots & d_{(1,l)} \\ d_{(2,1)} & \cdots & \cdots & d_{(2,l)} \\ \vdots & \vdots & \vdots & \vdots \\ d_{(l,1)} & \cdots & \cdots & d_{(l,l)} \end{bmatrix}$$

40

及び

【 数 5 】

$$k = \min_{l \in \{2, \dots, (m-1)\}} \sum_{\substack{g_i \in G; \\ p_a, p_b \in g_i \\ a \neq b; a < b}} D^{(l)}[a, b]$$

50

であり、ここで k は、 G における患者グループを最も近くに配置する最も小さな次元を表す $\{2, \dots, (m-1)\}$ における整数である。主成分分析 (PCA) 又はその他の特徴低減アルゴリズムが、上位 k 個の重要な特徴を特定するために利用される。これら k 個の特徴は、動作 62 において新たな患者を分類するために利用される。医師が情報入力したグループ G は任意に、相互検証を得るためにオーバーフィッティング問題を回避するため分割される。

【0035】

動作 60 を実装するための第 2 の手法の例は、新たな空間における特徴値を、これら特徴の重要度の重みを調節することにより表す。例として、3人の患者の例が以下のとおりであるとする：

10

特徴値 (3, 2, 4, 7) を持つ患者 P_1

特徴値 (3, 3, 3, 3) を持つ患者 P_2

特徴値 (4, 4, 4, 7) を持つ患者 P_3

この表記において、各患者 P_i は、値 (f_1, f_2, f_3, f_4) を持つ列 1 乃至 4 に特徴を持つ。説明のため、以下の距離が仮定される：

患者距離 $D(P_1, P_3) = 3$

患者距離 $D(P_1, P_2) = 6$

患者距離 $D(P_2, P_3) = 5$

マンハッタン距離を用いた初期の分類動作 52 において、第 1 のクラスタは患者 P_1 及び P_2 を含み、患者 P_2 は第 2 のクラスタにある。しかしながら、ことによると医師が特徴 f_2 及び f_3 がより重要であると考えるために、動作 56 において該医師が、患者 P_2 と P_3 とはより類似すると判断し、それにより分類が P_2 及び P_3 が同じクラスタに割り当てられ、 P_1 が別個のクラスタに属するよう更新される。

20

【0036】

新たなクラスタのセントロイドは、クラスタにおける特徴値の平均として算出され、 $P_c = (3.5, 3, 3, 5)$ である。次に、元のサンプルが新たな空間にマッピングされ、新たな空間における 2 つのサンプルのセントロイドに対する距離が最小である（前もって規定されていても良いし、又はユーザによって規定されても良い）。新たな空間の座標を調節するため、元の座標が各特徴についての調節された重みと乗算される（新たな空間における座標）。

30

【0037】

これを解くため、一次方程式のセットが適切に用いられる。しかしながら、患者の数 n と特徴の数 m とは、通常等しくはない。従って、選択された患者の数 p ($p < n$) について、新たな空間にマッピングされるべき最も変動する特徴のセットが導出される。表記すると、

$$w_1 * f_{11} + w_2 * f_{12} + \dots + w_p * f_{1p} = d_1$$

$$w_2 * f_{21} + w_2 * f_{22} + \dots + w_p * f_{2p} = d_2$$

...

$$w_p * f_{p1} + w_2 * f_{p2} + \dots + w_p * f_{2p} = d_2$$

この目的のため、全ての特徴についての分散が算出され、上位 p 個の変動する特徴が選択される。新たな行列は、 $p \times p$ の次元を持つ。該新たな行列について、適切な重みを見出すために、一次方程式のセットが解かれる。重みが決定されると、ユーザによって選択されなかった患者に対して、新たな空間へと、同じ重みが適用される。

40

【0038】

以上の例においては、これは以下のように変形される：

$$w_1 * 3 + w_4 * 3 = d_1$$

$$w_1 * 4 + w_4 * 7 = d_2$$

ここで w_1 及び w_4 は重みであり、列 1 及び列 4 における特徴は（患者 P_1 及び P_2 について）最も変動する特徴であると仮定される。

【0039】

50

以上は単に例であり、動作 6 0 を実行するための他の手法も考えられる。また、例えば重み調節（第 2 の手法の例）により後続される次元低減（第 1 の手法の例）を実行する、又はその逆のような、調節の組み合わせを利用することも考えられる。

【 0 0 4 0 】

図 3 を参照すると、表示要素 1 4 における類似するサンプル患者の表現の例が示されており（即ち図 2 の動作 5 4 ）、ここでの例においては、クエリ患者 P Q は「John Smith」であり、最後の分類の繰り返しにより特定された 2 人の類似するサンプル患者は「Bob Brown」及び「Mickey Red」である。2 つの適合性フィードバック応答が要求される。第 1 のものは、「Bob Brown」の「John Smith」に対する類似度を 1 乃至 5 のスケールでランク付けする要求の形をとり、ここで「1」が最も類似し「5」が最も類似しないとするものである。臨床医は、この要求に回答するため、マウスポインタを用いて、「1」乃至「5」とラベル付けされたボタンの 1 つを選択することができる。第 2 の要求は、2 人の患者「Bob Brown」及び「Mickey Red」のどちらがクエリ患者「John Smith」に類似するかを選択するものである。臨床医は、「Bob Brown」ボタン又は「Mickey Red」ボタンのいずれかを選択するようマウスポインタを用いることによりこれに回答することができる。

【 0 0 4 1 】

これら要求に有意に回答するため、臨床医は、クエリ患者「John Smith」について並びに各サンプル患者「Bob Brown」及び「Mickey Red」についての医療記録又はその他の患者情報を閲覧したいと欲する見込みが高いであろうことは理解されよう。この目的のため、これらの患者の 1 人に対する各参照は、図 3 の表示におけるハイパーリンクとして示され（患者名に下線を引くことにより示されている）、表示は「注：患者の医療記録をポップアップウィンドウで閲覧するには上の患者名のいずれかをクリックしてください」と説明している。斯くして、臨床医がマウスポインタを用いて「John Smith」をクリックすることに応答して、John Smith についての患者記録情報が表示されたポップアップウィンドウ（図示されていない）が表示され、好適には John Smith の医療記録を臨床医が閲覧することを可能とする適切なナビゲーションツールが伴われる。同様に、患者名「Bob Brown」又は「Mickey Red」の表示がマウスでクリックされる。斯かるポップアップの表示は、患者の特徴情報を含んでも良いが、臨床医は、患者記録の全体を閲覧することが可能であり、1 つの患者の特徴又は患者の特徴の小さな群に基づいて患者の類似度を評価するようには要求されない。他の閲覧ツールの枠組みが用いられても良く、例えばポップアップウィンドウではなく、別個の表示要素（利用可能である場合；図 1 には示されていない）に患者記録が表示されても良いことは、理解されるであろう。

【 0 0 4 2 】

図 4 を参照すると、他の実施例においては、サンプル患者についての情報は他の態様で表示されても良い。例えば図 4 は、2 つ以上のグラフィカルなモダリティ表現が表示要素 1 4 に同時に表示される視覚化ツールを示している。図 4 の例は、3 つの同時に表示されるグラフィカルなモダリティ表現を含み、それぞれは、ゲノムモダリティについてのグラフィカルなモダリティ表現 7 0 、放射線モダリティについてのグラフィカルなモダリティ表現 7 2 、及び臨床モダリティについてのグラフィカルなモダリティ表現 7 4 、である。各グラフィカルなモダリティ表現 7 0 、7 2 、7 4 は、2 つ以上のモダリティの特徴に対して、ウォーターフォール（waterfall）プロットにおける 1 人以上のサンプル患者及びクエリ患者（図 4 の例においては、2 人の患者 Bob Brown 及び Mickey Red 並びにクエリ患者 John Smith）をプロットする。図 2 の例においては、ゲノムのモダリティ表現 7 0 は、患者を E R 、H E R 2 及び P R のゲノムマーカ特徴に対してプロットしている。放射線のモダリティ表現 7 2 は、患者をテクスチャ（粗さ）、体積及び形態の画像特徴に対してプロットしている。臨床のモダリティ表現 7 4 は、腫瘍のステージのモダリティについてのものであり、患者を腫瘍サイズ（T）、結節状態（N）及び転移値（M）の腫瘍特徴に対してプロットしている。図 4 から、臨床医は、提示された特徴について、サンプル患者 Mickey Red に比べてサンプル患者 Bob Brown のほうがクエリ患者 John Smith に類似

10

20

30

40

50

しているように考えられることを、容易に観察することができる。

【0043】

図5を参照しながら、図4の視覚表現がより一般的に適用可能であり、インタラクティブなグラフィカルな視覚化によって患者コホートを識別するため患者データベース20を閲覧するために用いられることに注意されたい。図5の例においては、図4のゲノム、放射線及び臨床のモダリティ表現70、72、74が再び表示されている。ゲノムのモダリティ表現70においては、G U I処理40(図1参照)が、クラスタの各患者を個別にクリックすることのような適切な選択方法によって、又は本例においては、少なくとも1つのユーザ入力装置(例えばマウス18、トラックボール、トラックパッド、タッチスクリーン又はその他のポインティングデバイス)を介して患者のクラスタ{1, 2, 4, 8}の包含80を受信することによって、患者のクラスタ(任意に患者{1, 2, 4, 8}と示されている)の選択を受信している。他の同時に表示されているモダリティのグラフィカル表現72、74のそれぞれにおいては、選択80に応答して、選択された患者のクラスタ{1, 2, 4, 8}の患者が同様に強調されている。図5の例においては、当該強調は、他の全ての患者の表示を削除し、患者1、2、4及び8のみが表示されるようすることにより、他のモダリティのグラフィカル表現72、74において為される。該強調は、選択されたクラスタの患者を赤色で表示し、他の患者の全てを黒色で表示し続けることのような、他の手法を用いても良い。

10

【0044】

図5に示されるように、患者1、4及び8は、放射線のグラフィカル表現72において好適に分類されているが、患者2は当該モダリティのビューにおいて分類から外れている。臨床のモダリティのグラフィカル表現74においては、患者1及び8のみがともに分類されており、患者2及び4は外れている。これらの結果に基づいて、臨床医は、種々の結論を導くことが可能であり得る。例えば、クエリ患者が患者1である場合、患者8がクエリ患者1に最も近い患者であると決定され得(患者8が3つのビューのいずれにおいても患者1に近いため)、患者2は最も類似しないものであるように見える。他のビューにおいては、臨床医は、2つ以上の座標特徴を更新することにより、表現72、74の1つ以上を変更することを試みるよう決定し得、これに応答して、G U I処理40は、更新されたグラフィカルなモダリティ表現を再描画し、患者を更新された2つ以上のモダリティの座標特徴に対してプロットする。

20

【0045】

図4及び5は、ゲノム、放射線及び臨床についての視覚化の例を示すが、種々のモダリティ表現はより一般的に、臨床、放射線、ゲノム、人口統計学、及び/又は生理学的なモダリティのようなモダリティを含んでも良い。一般的に、ユーザは、利用可能なモダリティのパレット又はリストから、どのモダリティを表示するかを選択し、更に各モダリティのどの特徴に対してプロットするかを選択しても良い。

30

【0046】

以下、図4及び5の表現70、72、74のような2つ以上のグラフィカルなモダリティ表現を用いて図1のG U I処理40により適切に実行された視覚化表現及びナビゲーション処理の更に詳細な例が記載される。該手法は、例えば名前又は電子医療記録(E M R)番号の検索を介して、クエリ患者を選択することにより開始する。名前、年齢、主治医、疾患等のような、当該患者についての基本情報が表示されても良い。グラフィカルなモダリティ表現70、72、74を含む、ユーザ作業領域が表示される。図4及び5の例においては、各モダリティ表現は、円として表され、患者がプロットされるモダリティ特徴が、該円のまわりに等間隔で配置される(例えばゲノムのモダリティの場合には特定の疾患に関連する生物マーカーの存在)。他の患者が、当該視覚化に自動的に統合され、いずれかの利用可能なコホート(例えば図1の患者コホート識別装置を用いて生成されたもの)から引き抜かれ、該円に配置される。このことは、より大きな患者の集団という観点から、対象の患者を配置する。デフォルトでは、全てのモダリティが同時に表示されるが、それぞれが個々に調べるためにズームインされる能够である。円の周辺に配置された特徴

40

50

(即ちプロットされる特徴)は選択されることがある。値が割り当てられた患者に結合されたいずれの値も、全ての利用可能なモダリティに亘って後続して強調される。後続する解析が、これらの選択された患者に対して実行されても良い。

【0047】

任意に、ユーザが患者を選択すると、選択された患者の顕著な特性を強調する統計的な概要が、画面に表示される。当該概要是、患者の選択が更新されると、動的に更新される。該概要の内容は、離散的又は連続的といった変数の特性に基づいて記述されても良い。

【0048】

多量の利用可能な人口統計学的、病理学的、臨床的及びゲノムの特徴(例えば幾つかの患者データベースにおいては200以上)に対して、対象の特定のコンテキストに関連するデータの効果的な要約及び視覚化のため、生物マーカ、署名、予後スコア等及びコホートのサンプルのような特徴の選択を支援するナビゲーションツールが提供される。任意に、G U Iツールもが、臨床医がカスタマイズされた選択を定義及び保存し容易に或るコンテキストから他のコンテキストに切り換えることができるようとする。

10

【0049】

図5の例においては、臨床医は、データベース又は選択されたコホートにおける他の患者のコンテキストにおける選択された患者のゲノムのグラフィカルなビュー70において、E R、H E R 2及びP R受容体の状態を閲覧し、同じ患者を他のモダリティのグラフィカルなビュー72、74において閲覧する柔軟性を持つ。臨床医は、例えば包含80を用いて、ゲノムのビューから関心のある患者のサブセットを選択/強調することができ、これら患者は、他のビュー72、74において強調される。斯くて、例えば、選択された患者{1, 2, 4, 8}のT、N、Mのステージは、臨床の(癌のステージの)グラフィカルなビュー74において閲覧され、選択された患者のT(腫瘍サイズ)、N(結節状態)及びM(転移状態)の分布が評価される(選択された患者の数が4人の選択された患者の例よりも多い場合には、より良く定義された分布が得られる見込みが高いことは、理解されるであろう)。同様に、撮像のグラフィカルなビュー72においては、例えば体積、捩じり上げ、捩じり下げ特性、テクスチャ、及び形態特徴のようなM R I特徴が表示される。他のモダリティのビューが、臨床医により選択されても良い(図示されていない)。このようにして、臨床医は、異なるモダリティ間で、選択された患者の群について、1つのモダリティにおける特徴又は特徴の群の関連について、インタラクティブに試験することができる。

20

【0050】

更なる例として、ゲノムのグラフィカルなビュー70が更に詳細に説明される。図5に示されるように、ゲノムの層が円に表示され、乳癌を評価するタスクに適切な例が特に示されている。このタスクのため、対象の特徴は、乳癌の診断及び予後についての臨床的な有用性を示したE R、P R及びH E R 2活動レベルを含む(当然ながら、他のタスクについては、他の顕著なゲノム特徴がプロットのため選択される)。臨床医がアプリケーションを開くと、該アプリケーションは選択されたコホート(例えば図1のコホート識別装置を用いて生成されたもの)における他の患者とともに対象の患者(クエリ患者)を表示する。図5の例においては、3つのウォーターフォールプロット(降順でプロットされたバー)が示され、それぞれがE R / P R / H E R 2活動度の1つを表し、ゲノムの円70に均等に配置される。ナビゲーション手法の1つにおいては、クエリ患者が、コホートの残りのものに対して、円における各生物マーカについてのクエリ患者の活動レベルを強調するよう、自動的に選択される。任意に、患者を表すこれら3本のバーから、円の中心における領域まで、線(図示されていない)が描画され、コホートのE R / P R / H E R 2データから算出された機械学習アルゴリズム(例えば主成分解析)が、適切に視覚化される。該線は、コホートに対して対象の患者が位置する場合へと正確に描画される。この全体的な視覚化から、例えば包含80を用いて、類似するE R / P R / H E R 2発現レベルを持つ更なる患者が選択される。機械学習空間において選択されたいずれの更なる患者も、円を囲むそれぞれのE R / P R / H E R 2活動度レベルに対して引かれた線を持つことと

30

40

50

なる。

【0051】

選択されたグループ（例えば図5におけるグループ{1, 2, 4, 8}）について、統計的な概要が提供されても良い。年齢及び遺伝子の発現値のような連続的な変数については、選択されたグループの平均が算出されても良い。性別及びER状態のような離散的な変数については、超幾何試験のような試験及びp値の降順に保存された特性を用いるエンリッチメント解析が実行されても良い。乳癌のデータセットについて選択された患者の典型的な概要が、表1に示される。

【表1】

表1：乳癌データセットについての選択された患者の概要

平均年齢	46
P53の平均発現値	2.3 FPKM
優位な性別	女性 (p値0.001)
優位なER状態	正 (p値0.003)
優位なPR状態	負 (p値0.005)
優位なHER2状態	正 (p値0.007)
優位な乳癌サブタイプ	基底細胞型 (p値0.009)

10

20

【0052】

この表においては、FPKM (fragments per kilobase of exon per million reads mapped) が、RNAシーケンシングデータに基づいてp53遺伝子の発現値を示す。これら変数の有意な数は、乳癌の診断のタスクの例に特有であり、統計的な概要の要素は、各疾患又は臨床タスクについて概要形式で適切に事前に記述される。

【0053】

図4及び5のグラフィカルな視覚化及びナビゲーションツールは、例である。例示した円の幾何以外の幾何が利用されても良い。円の幾何の利点は、プロットされるべきモダリティの特徴の適切な数（即ち円の周りに好適にフィッティングされ得る特徴のいずれかの数）に容易に更新される点であり、正方形の幾何は、2つのモダリティの特徴に対するプロットにのみ適している。

30

【0054】

1つのグラフィカルなモダリティ表現において患者のクラスタを選択する動作は、臨床医がユーザ入力装置16、18を操作する（例えば図5におけるような包含選択80を為す）以外のエンティティ/機構によって実行されても良いことも想到される。例えば、図4の例においては、例えば分類処理30のような、コンピュータ10、12上で動作する他のプログラムが、クエリ患者及び1人以上のサンプル患者（図4の例においてはサンプル患者Bob Brown及びMickey Red並びにクエリ患者John Smith）のセットとして患者のクラスタを選択する。

40

【0055】

図1及び2に戻り参照しながら、患者レベルの適合性フィードバックを用いた患者コード識別喜悦の実装の他の例がテキストで説明され、以下のステップを含む。

【0056】

ステップ1：大量のデータセット（第2の実施例においては百万以上のサンプル）における全ての患者及び選択された患者の特徴に対する階層的な分類をもちいて、教師なし学習を実行する。

【0057】

ステップ2：クラスタの数を決定し、クラスタのセントロイドを算出する。

【0058】

50

ステップ3：全ての特徴に基づいてクエリ患者P Qを含むクラスタを選択し、同じクラスタから付加的なシード(seed)を選択する。

【0059】

ステップ4：各シードについて、患者比較メトリックを用いて測定された全ての異なるクラスタのセントロイドに対する当該患者の距離に基づいて、最も類似する患者を見つける。

【0060】

ステップ5：(单一のクラスタに属するもののうち)類似する患者の優先順位付けされたリストに基づいて、サンプル及び現在のサンプルに類似するサンプルを選択する。

【0061】

ステップ6：一度に1つの特徴を削除することにより、どの特徴が患者の当該類似度を重要とするかを決定する。

【0062】

ステップ7：患者比較メトリックを用いて、現在の患者と全ての選択された患者との間の距離を見出す。

【0063】

ステップ8：中央値が0に近い列を見つける。大きな数を持つ列を破棄する。

【0064】

ステップ9：選択された特徴に基づいて、選択された分類のみを用いて、データセット全体に対して教師なし分類を実行する。

20

【0065】

ステップ10：元のクエリ患者P Qのクラスタ又は選択された患者の大部分が出現するクラスタにおける患者を最後に提示する。

【0066】

最後に、サンプルのセットが全て臨床医に関連するものとなるまで、以上のステップ1乃至10を反復的に繰り返す。

【0067】

本発明は、好適な実施例を参照しながら説明された。以上の詳細な説明を読むことにより、他への変更及び変形が生じ得る。本発明は、添付される請求の範囲又はそれと同等のものの範囲内である限り、斯かる変更及び変形の全てを含むものとして解釈されることを意図される。

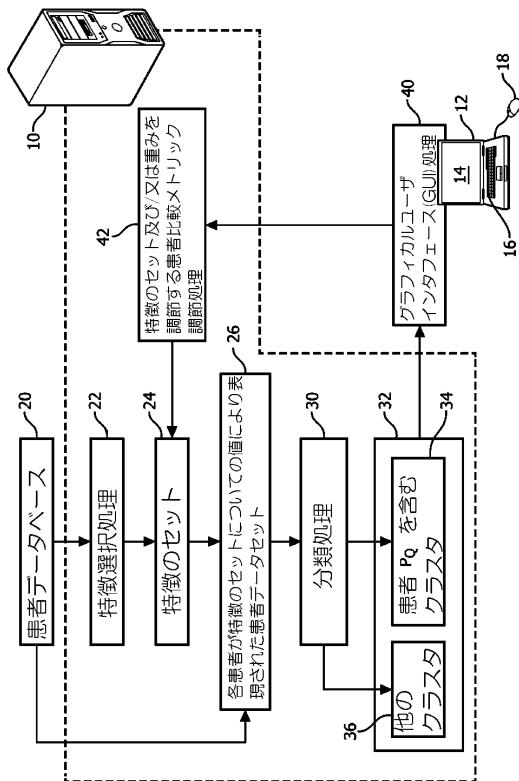
30

40

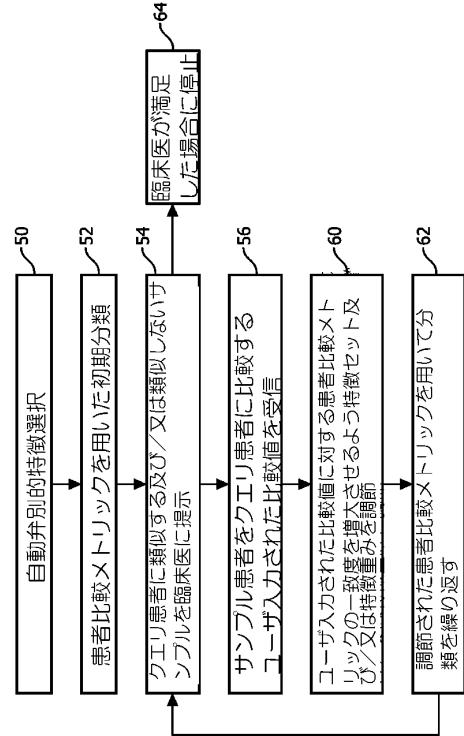
50

【図面】

【図 1】



【図 2】



10

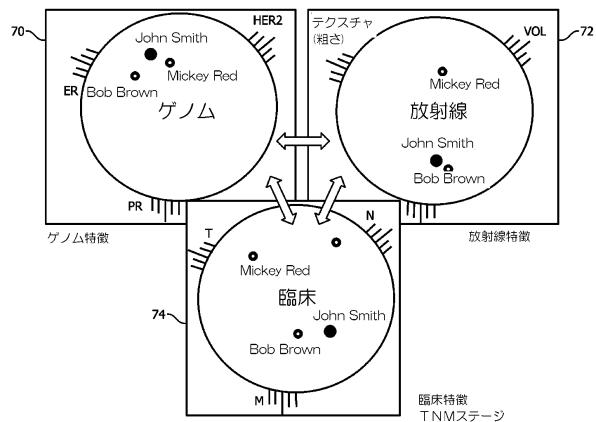
20

30

【図 3】



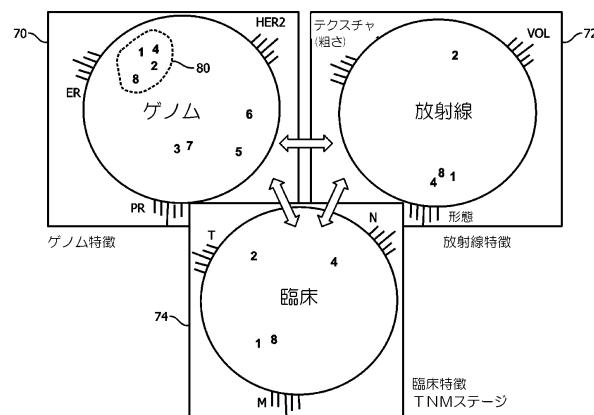
【図 4】



40

50

【図 5】



10

20

30

40

50

フロントページの続き

オランダ国 5656 アーエー アインドーフェン ハイ テック キャンパス ビルディング 5
(72)発明者 マンコヴィッチ アレクサンデル ライアン
オランダ国 5656 アーエー アインドーフェン ハイ テック キャンパス ビルディング 5
(72)発明者 ディミトロヴァ ネヴェンカ
オランダ国 5656 アーエー アインドーフェン ハイ テック キャンパス ビルディング 5
(72)発明者 バネルジー ニランジャナ
オランダ国 5656 アーエー アインドーフェン ハイ テック キャンパス ビルディング 5
(72)発明者 チューン イー ヒム
オランダ国 5656 アーエー アインドーフェン ハイ テック キャンパス ビルディング 5
(72)発明者 デ ボント ヨハンナ マリア
オランダ国 5656 アーエー アインドーフェン ハイ テック キャンパス ビルディング 5
(72)発明者 ライマケルス ヨゼフ ヒエロニムス マリア
オランダ国 5656 アーエー アインドーフェン ハイ テック キャンパス ビルディング 5
審査官 木村 大吾
(56)参考文献 特表2012-523877 (JP, A)
米国特許出願公開第2011/0270628 (US, A1)
特表2011-520206 (JP, A)
(58)調査した分野 (Int.Cl., DB名)
G 06 F 16/00 - 16/958
G 16 H 50/70