(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2024年10月3日(03.10.2024)

WIPO | PCT

(10) 国際公開番号
WO 2024/203701 A1

(51) 国際特許分類:
G01N 27/62 (2021.01) *G01N 33/50* (2006.01)

(21) 国際出願番号： PCT/JP2024/010963

(22) 国際出願日： 2024年3月21日(21.03.2024)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

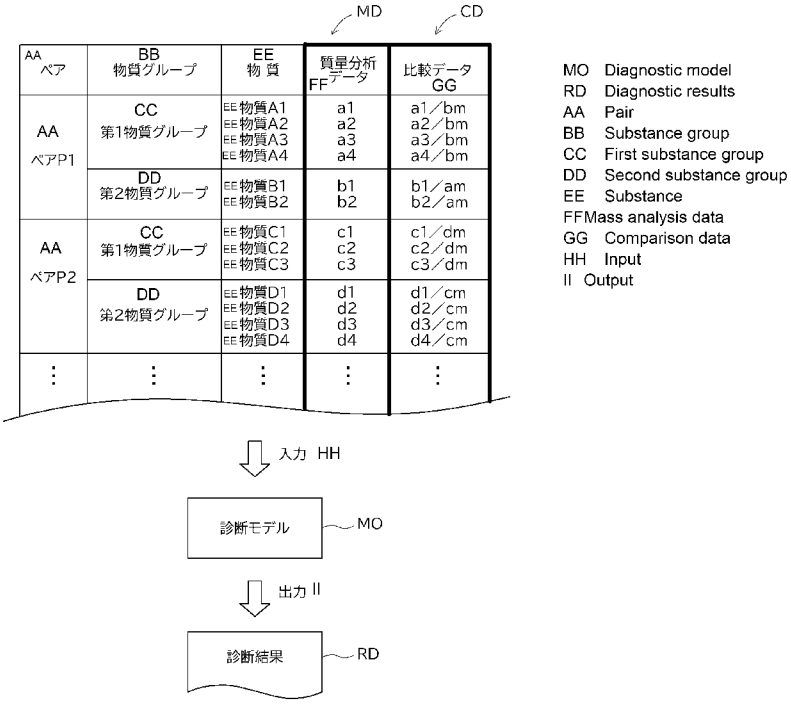
(30) 優先権データ：
特願 2023-054539 2023年3月30日(30.03.2023) JP

(71) 出願人: 国立大学法人東海国立大学機構 (**NATIONAL UNIVERSITY CORPORATION TOKAI NATIONAL HIGHER EDUCATION AND RESEARCH SYSTEM**) [JP/JP]; 〒4648601 愛知県名古屋市千種区不老町1番 Aichi (JP).

(72) 発明者: 馬場 泰輔 (**BABA Taisuke**); 〒4648601 愛知県名古屋市千種区不老町1番 国立大学法人東海国立大学機構内 Aichi (JP). 砂川 真輝 (**SUNAGAWA Masaki**); 〒4648601 愛知県名古屋市千種区不老町1番 国立大学法人東海国立大学機構内 Aichi (JP). 國料 俊男 (**KOKURYO Toshio**); 〒4648601 愛知県名古屋市千種区不老町1番 国立大学法人東海国立大学機構内 Aichi (JP). 水野 隆史 (**MIZUNO Takashi**); 〒4648601 愛知県名古屋市千種区不老町1番 国立大学法人東海国立大学機構内 Aichi (JP).

(54) **Title:** INFORMATION PROCESSING DEVICE, INFORMATION PROCESSING METHOD, AND COMPUTER PROGRAM

(54) 発明の名称： 情報処理装置、情報処理方法、および、コンピュータプログラム



(57) **Abstract:** The present invention improves the accuracy of sample classification from reference to a mass analysis result. This information processing device comprises a model acquisition unit, an object data acquisition unit, and a classification execution unit. The model acquisition unit acquires a trained model using comparison data as input and a sample classification result as output. The comparison data include data indicating a magnitude relationship between a peak intensity of a substance belonging to one substance group and a peak intensity of a substance belonging to the other sub-

WO 2024/203701 A1

[続葉有]

(74) 代理人: 弁理士法人アルファ国際特許事務所 (ALPHA INTERNATIONAL PATENT FIRM); 〒4600003 愛知県名古屋市中区錦一丁目１１番２０号 Aichi (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類：

— 国際調査報告（条約第21条(3)）

stance group for at least one substance group pair which is a pair of two substance groups each constituted from at least one substance in a mass analysis result for a plurality of substances included in a sample. The object data acquisition unit acquires comparison data about an object sample. The classification execution unit executes classification of the object sample by inputting comparison data about the object sample to the trained model, and outputs a classification result.

(57) 要約：質量分析結果を参照した試料の分類の精度を向上させる。 情報処理装置は、モデル取得部と、対象データ取得部と、分類実行部とを備える。モデル取得部は、比較データを入力とし、試料の分類結果を出力とする学習済モデルを取得する。比較データは、試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の物質グループに属する物質のピーク強度と、他方の物質グループに属する物質のピーク強度と、の大小関係を示すデータを含む。対象データ取得部は、対象試料についての比較データを取得する。分類実行部は、対象試料についての比較データを学習済モデルに入力することにより、対象試料の分類を実行し、分類の結果を出力する。

明　細　書

発明の名称 ：

情報処理装置、情報処理方法、および、コンピュータプログラム

技術分野

[0001]　本明細書に開示される技術は、試料の分類を行う情報処理に関する。

背景技術

[0002]　疾病の診断のために、種々のマーカーが利用されている。例えば膵管がんや胆管がんの腫瘍マーカーとして、ＣＡ１９－９が利用されている（例えば、特許文献１参照）。

先行技術文献

特許文献

[0003] 特許文献1：特開２０２３－３７００３号公報

発明の概要

発明が解決しようとする課題

[0004]　診断対象者から採取された試料（血液、尿等）を質量分析にかけることにより得られる特定の物質のピーク面積を参照して、診断対象者が特定の疾病に罹患しているか否かの分類を行うことが考えられる。しかしながら、質量分析では様々な要因によりその結果にバラツキが生じ得る。そのため、特定の物質の定量的質量分析では標準物質を用いるが、このような定量的質量分析では、同時に測定されている上記特定の物質以外の他の物質の含有量が診断に利用されておらず、診断精度向上の余地がある。なお、このような課題は、特定の疾病の診断を目的とする場合に限らず、試料の質量分析結果を参照して何らかの分類を行う場合に共通の課題である。

[0005]　本明細書では、上述した課題を解決することが可能な技術を開示する。

課題を解決するための手段

[0006]　本明細書に開示される技術は、例えば、以下の形態として実現することが

可能である。

[0007] （１）本明細書に開示される情報処理装置は、モデル取得部と、対象データ取得部と、分類実行部とを備える。モデル取得部は、比較データを入力とし、前記試料の分類結果を出力とする学習済モデルを取得する。比較データは、試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の前記物質グループに属する物質のピーク強度と、他方の前記物質グループに属する物質のピーク強度と、の大小関係を示すデータを含むデータである。対象データ取得部は、対象試料についての前記比較データを取得する。分類実行部は、前記対象試料についての前記比較データを前記学習済モデルに入力することにより、前記対象試料の分類を実行し、前記分類の結果を出力する。

[0008] このように、本情報処理装置では、比較データを入力とし試料の分類結果を出力とする学習済モデルを用いて、対象試料の分類を実行する。比較データは、試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の物質グループに属する物質のピーク強度と、他方の物質グループに属する物質のピーク強度と、の大小関係を示すデータを含む。そのため、比較データを利用することにより、様々な要因に起因する試料間での値のバラツキの影響を相対的に小さくすることができ、質量分析結果を参照した試料の分類の精度を向上させることができる。また、比較データは、試料に含まれる複数の物質についての質量分析結果を利用して作成されたデータであるため、特定の物質のみについての質量分析結果を利用する構成と比較して、質量分析結果を参照した試料の分類の精度を向上させることができる。

[0009] （２）上記情報処理装置において、前記比較データは、前記他方の物質グループに属する物質のピーク強度に対する、前記一方の物質グループに属する物質のピーク強度の比を示すデータを含む構成としてもよい。本構成を採用

すれば、試料におけるピーク強度が同程度であって、属する分類により大小関係が反転するような物質のペアに限られず、任意の物質の組合せにより物質グループペアを構成することができる。これにより、物質グループペアの設定の自由度が高まるため、より好ましい比較データを得ることができ、質量分析結果を参照した試料の分類の精度を効果的に向上させることができる。

[0010] （３）上記情報処理装置において、前記比較データは、前記他方の物質グループに属するすべての物質のピーク強度の平均値に対する、前記一方の物質グループに属する各物質のピーク強度の比を示すデータを含む構成としてもよい。本構成を採用すれば、それぞれの物質について比較データの値が定まることとなり、情報量の減少を抑制することができる。これにより、より情報量の多い比較データを得ることができ、質量分析結果を参照した試料の分類の精度を効果的に向上させることができる。

[0011] （４）上記情報処理装置において、前記比較データは、複数の前記物質グループペアについてのデータを含む構成としてもよい。本構成を採用すれば、より情報量の多い比較データを得ることができ、質量分析結果を参照した試料の分類の精度を効果的に向上させることができる。

[0012] （５）上記情報処理装置において、さらに、前記複数の物質の互いの相関に基づき、前記少なくとも１つの物質グループペアを設定するペア設定部を備える構成としてもよい。本構成を採用すれば、より適切な分類に資する物質グループペアを設定することができ、質量分析結果を参照した試料の分類の精度を効果的に向上させることができる。

[0013] （６）上記情報処理装置において、前記試料は、ヒトから採取された生体試料である構成としてもよい。本構成を採用すれば、生体試料の質量分析結果を参照した生体試料の分類の精度を向上させることができる。

[0014] （７）上記情報処理装置において、前記生体試料は、尿と、胆汁と、汗と、唾液と、血液と、髄液と、腹水と、胸水と、膵液と、胃液と、腸液と、の少なくとも１つである構成としてもよい。本構成を採用すれば、それらの生体

試料の質量分析結果を参照した生体試料の分類の精度を向上させることができる。

[0015] （８）上記情報処理装置において、前記分類実行部は、特定の疾病についての前記分類を行う構成としてもよい。本構成を採用すれば、質量分析結果を参照した特定の疾病の診断の精度を向上させることができる。

[0016] （９）上記情報処理装置において、前記特定の疾病は、がんである構成としてもよい。本構成を採用すれば、質量分析結果を参照したがんの診断の精度を向上させることができる。

[0017] なお、本明細書に開示される技術は、種々の形態で実現することが可能であり、例えば、情報処理装置、情報処理方法、それらの方法を実現するコンピュータプログラム、そのコンピュータプログラムを記録した一時的でない記録媒体等の形態で実現することができる。

図面の簡単な説明

[0018] [図1]本実施形態における診断モデルＭＯを概略的に示す説明図

[図2]本実施形態における診断モデルＭＯを概略的に示す説明図

[図3]質量分析データＭＤおよび比較データＣＤを概念的に示す説明図

[図4]情報処理装置１００の概略構成を示す説明図

[図5]本実施形態における診断モデル取得処理を示すフローチャート

[図6]本実施形態における診断処理を示すフローチャート

[図7]質量分析データＭＤの主成分分析結果を示す説明図

[図8]がんに特異的な物質に絞り込んだ質量分析データＭＤの主成分分析結果を示す説明図

[図9]比較データＣＤの主成分分析結果を示す説明図

[図10]本実施形態の診断モデルＭＯの精度を示す説明図

[図11]本実施形態の診断モデルＭＯを膵がんのデータに適用した場合の分類精度を示す説明図

[図12]クレアチニン補正を用いて分類を行った場合の主成分分析結果および分類精度を示す説明図

[図13]本実施形態の診断モデルＭＯを用いて分類を行った場合の主成分分析結果および分類精度を示す説明図

発明を実施するための形態

[0019] Ａ．実施形態：

Ａ－１．診断モデルＭＯの概要：

図１および図２は、本実施形態における診断モデルＭＯを概略的に示す説明図である。診断モデルＭＯは、特定の疾病についての診断を行うための学習済モデルであり、例えば、診断対象者が特定の疾病に罹患しているか否かの分類を行うために利用される。特定の疾病としては、例えば、がん（胆道がん、膵がん、胃がん、大腸がん等）やがん以外の疾病（炎症性腸疾患等）が挙げられる。以下の説明では、特定の疾病は、がん（より具体的には胆道がん）であるものとする。

[0020] 図１および図２に示すように、診断モデルＭＯへの入力は、診断対象者から採取された生体試料を対象とした質量分析データＭＤに基づき特定される比較データＣＤである。生体試料としては、例えば、尿、胆汁、汗、唾液、血液、髄液、腹水、胸水、膵液、胃液、腸液等が用いられる。以下の説明では、生体試料は尿であるものとし、質量分析の対象は、尿に含まれる複数の物質（代謝物（メタボローム））であるものとする。尿を対象とした質量分析では、例えば１５００～２０００個程度の多数の物質のピーク面積（ピーク強度）が測定される。

[0021] 質量分析では、試料の採取条件や分析前の試料処理条件、日ごとの分析機器のセットアップ（バッチ間誤差）、試料内に存在するイオン化を妨げるような物質（マトリック効果）などの様々な要因により、結果にバラツキが生じ得る。また、尿等の生体試料は濃度のバラツキも加わり、生体試料を対象とした質量分析では、分析結果のバラツキが比較的大きくなりやすい。そのため、本実施形態では、以下に説明する様に、質量分析結果を示す質量分析データＭＤに基づき特定される比較データＣＤを用いている。

[0022] 図１および図２に示すように、比較データＣＤの生成のために、試料に含

まれる複数の物質を対象として、複数の物質グループペア（以下、単に「ペア」ともいう。）が予め設定される。物質グループペアは、それぞれ少なくとも１つの物質から構成される２つの物質グループ（第１物質グループおよび第２物質グループ）のペアである。図１に示す例では、各物質グループペアを構成する２つの物質グループは、共に、１つの物質から構成されている。例えば、１番目の物質グループペアＰ１は、１つの物質Ａ１から構成される第１物質グループと、他の１つの物質Ｂ１から構成される第２物質グループと、のペアである。このように、本明細書では、単一の要素から構成される単位も「グループ」という。

[0023] また、図２に示す例では、各物質グループペアを構成する２つの物質グループは、１つまたは複数の物質から構成されている。例えば、１番目の物質グループペアＰ１は、４つの物質Ａ１，Ａ２，Ａ３，Ａ４から構成される第１物質グループと、２つの物質Ｂ１，Ｂ２から構成される第２物質グループと、のペアである。

[0024] なお、複数の物質グループペアは、試料に含まれる複数の物質の互いの相関に基づき、設定される。この点については、後に詳述する。

[0025] 各物質グループペアについての比較データＣＤは、一方の物質グループ（第１物質グループまたは第２物質グループ）に属する物質のピーク面積と、他方の物質グループ（第２物質グループまたは第１物質グループ）に属する物質のピーク面積と、の間の大小関係を示すデータを含む。例えば、図１の例では、各物質グループペアについての比較データＣＤは、第１物質グループに属する物質のピーク面積が第２物質グループに属する物質のピーク面積より小さければ値「０」をとり、反対に第１物質グループに属する物質のピーク面積が第２物質グループに属する物質のピーク面積以上であれば値「１」をとる、「０／１」の２値データである。例えば、１番目の物質グループペアＰ１については、第１物質グループを構成する物質Ａ１のピーク面積ａ１が第２物質グループを構成する物質Ｂ１のピーク面積ｂ１より小さいため、比較データＣＤは値「０」を取っており、２番目の物質グループペアＰ２

については、第１物質グループを構成する物質Ｃ１のピーク面積ｃ１が第２物質グループを構成する物質Ｄ１のピーク面積ｄ１以上であるため、比較データＣＤは値「１」を取っている。

[0026] また、各物質グループペアについての比較データＣＤは、他方の物質グループに属する物質のピーク面積に対する、一方の物質グループに属する物質のピーク面積の比を示すデータを含んでいてもよい。例えば、図２の例では、各物質グループペアを構成する一方の物質グループに属する各物質についての比較データＣＤは、当該一方の物質グループに属する各物質のピーク面積を、他方の物質グループに属するすべての物質のピーク面積の平均値で除した値である。例えば、１番目の物質グループペアＰ１の第１物質グループに属する物質Ａ１については、物質Ａ１のピーク面積ａ１を、当該ペアの第２物質グループに属するすべての物質（物質Ｂ１，Ｂ２）のピーク面積の平均値ｂｍ（＝（ｂ１＋ｂ２）／２）で除した値である。また、当該ペアの第２物質グループに属する物質Ｂ１については、物質Ｂ１のピーク面積ｂ１を、当該ペアの第１物質グループに属するすべての物質（物質Ａ１，Ａ２，Ａ３，Ａ４）のピーク面積の平均値ａｍ（＝（ａ１＋ａ２＋ａ３＋ａ４）／４）で除した値である。なお、各物質グループペアにおいて、一方の物質グループに属する物質について算出された比較データＣＤの値が１未満であれば、当該物質グループに属する物質のピーク面積が他方の物質グループに属する物質のピーク面積より小さいと言え、反対に、一方の物質グループに属する物質について算出された比較データＣＤの値が１以上であれば、当該物質グループに属する物質のピーク面積が他方の物質グループに属する物質のピーク面積以上であると言える。そのため、図２の例でも、各物質グループペアについての比較データＣＤは、一方の物質グループに属する物質のピーク面積と、他方の物質グループに属する物質のピーク面積と、の間の大小関係を示すデータであると言える。

[0027] このように、各物質グループペアについての比較データＣＤは、一方の物質グループに属する物質と他方の物質グループに属する物質とについて、ど

ちらのピーク面積が大きいかを示す情報である。また、この比較データＣＤは、一方の物質グループに属する物質と他方の物質グループに属する物質とについて、どちらがどれだけピーク面積が大きいかを示す情報を含んでいてもよい。このような比較データＣＤを用いることにより、様々な要因に起因する試料間での値のバラツキの影響を相対的に小さくすることができる。

[0028] 図３は、質量分析データＭＤおよび比較データＣＤを概念的に示す説明図である。図３において、「バッチ１」、「バッチ２」および「バッチ３」は、試料の採取単位を表しており、バッチが異なれば試料の採取条件が異なり得る。２つある「バッチ２」のうちの右側のものは、健常者から採取された試料についてのデータであり、残りのものは、胆道がん患者から採取された試料についてのデータである。

[0029] 図３の上段には、質量分析データＭＤの一例を概念的に示している。質量分析データＭＤは、試料の質量分析により特定される、試料に含まれる複数の物質（例えば、guanine、dopamine_3_o_sulfate、glycochenodeoxycholate_3_sulfate等）のピーク面積を示すデータである。図３では、各物質のピーク面積を濃度で表現している。図示するように、胆道がん患者から採取された試料でも、質量分析の計測日が異なるだけで、尿中の各物質のピーク面積の傾向が大きく異なり得る。

[0030] 図３の下段には、比較データＣＤの一例を概念的に示している。図３に示す例では、図１に示す例と同様に、各物質グループペアを構成する２つの物質グループは、共に１つの物質から構成されており、各物質グループペアについての比較データＣＤは、第１物質グループに属する物質のピーク面積が第２物質グループに属する物質のピーク面積より小さければ値「０」をとり、反対に第１物質グループに属する物質のピーク面積が第２物質グループに属する物質のピーク面積以上であれば値「１」をとる、「０／１」の２値データである。胆道がん患者について、各物質グループペアの比較データＣＤの傾向は、バッチにかかわらず互いに類似し、健常者についての比較データＣＤの傾向と大きく異なっている。このように、各物質グループペアについ

ての比較データＣＤを用いることにより、様々な要因に起因する試料間での値のバラツキの影響を相対的に小さくすることができる。

[0031] 図１および図２に示すように、診断モデルＭＯの出力は、診断結果を示す診断結果データＲＤである。診断結果は、対象試料の分類結果であり、より具体的には、対象試料である尿の採取元の診断対象者が胆道がんに罹患しているか否かの分類結果である。

[0032] このように、試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである物質グループペアについて、一方の物質グループに属する物質のピーク面積と、他方の物質グループに属する物質のピーク面積と、の大小関係を示すデータを含む比較データＣＤを入力とし、試料の分類結果を出力とする学習済モデルである診断モデルＭＯを利用することにより、様々な要因に起因する試料間での値のバラツキの影響を相対的に小さくすることができ、より精度の高い試料の分類、すなわち、より精度の高い疾病の診断を実現することができる。なお、本実施形態における分類の手法を、Ｉｎｖｅｒｓｅ Ｐａｉｒ Ｂｏｏｓｔｉｎｇ（略して、「ＩＰＢ」）ともいう。

[0033] Ａ－２．情報処理装置３の構成：

次に、診断モデルＭＯの作成や診断モデルＭＯを用いた診断を実行するための情報処理装置１００の構成を説明する。図４は、情報処理装置１００の概略構成を示す説明図である。情報処理装置１００は、コンピュータ（ＰＣ、サーバ等）により構成されている。

[0034] 情報処理装置１００は、制御部１１０と、記憶部１２０と、表示部１３０と、操作入力部１４０と、インターフェース部１５０とを備える。これらの各部は、バス１９０を介して互いに通信可能に接続されている。なお、情報処理装置１００が出力手段としてのスピーカを備えていてもよい。

[0035] 情報処理装置１００の表示部１３０は、例えば液晶ディスプレイ等により構成され、各種の画像や情報を表示する。操作入力部１４０は、例えばキーボードやマウス、ボタン、マイク、トラックパッド等により構成され、管理

者の操作や指示を受け付ける。なお、表示部１３０がタッチパネルを備えることにより、操作入力部１４０として機能するとしてもよい。インターフェース部１５０は、例えばＬＡＮインターフェースやＵＳＢインターフェース等により構成され、有線または無線により他の装置との通信を行う。

[0036] 情報処理装置１００の記憶部１２０は、例えばＲＯＭやＲＡＭ、ハードディスクドライブ（ＨＤＤ）等により構成され、各種のプログラムやデータを記憶したり、各種のプログラムを実行する際の作業領域やデータの一時的な記憶領域として利用されたりする。例えば、記憶部１２０には、後述する診断モデル取得処理や診断処理を実行するためのコンピュータプログラムである診断プログラムＣＰが格納されている。診断プログラムＣＰは、例えば、ＣＤ－ＲＯＭやＤＶＤ－ＲＯＭ、ＵＳＢメモリ等のコンピュータ読み取り可能な記録媒体（不図示）に格納された状態で提供され、あるいは、インターフェース部１５０を介して外部装置（ネットワーク上のサーバや他の端末装置）から取得可能な状態で提供され、情報処理装置１００上で動作可能な状態で記憶部１２０に格納される。

[0037] また、情報処理装置１００の記憶部１２０には、予め、または、後述する各種処理中に、学習用データＬＤと、診断モデルＭＯと、診断結果データＲＤとが格納される。これらの情報やデータの内容については、後述する各種処理の説明に合わせて説明する。

[0038] 情報処理装置１００の制御部１１０は、例えばＣＰＵ等により構成され、記憶部１２０から読み出したコンピュータプログラムを実行することにより、情報処理装置１００の動作を制御する。例えば、制御部１１０は、記憶部１２０から診断プログラムＣＰを読み出して実行することにより、後述の各種処理を実行する。このとき、制御部１１０は、原データ取得部１１１と、ペア設定部１１２と、学習用データ取得部１１３と、モデル取得部１１４と、対象データ取得部１１５と、分類実行部１１６として機能する。これら各部の機能については、後述の各種処理の説明に合わせて説明する。

[0039] Ａ－３．診断モデル取得処理：

次に、本実施形態の情報処理装置１００により実行される診断モデル取得処理について説明する。図５は、本実施形態における診断モデル取得処理を示すフローチャートである。診断モデル取得処理は、上述した診断モデルＭＯを取得するための処理である。本実施形態では、情報処理装置１００は、自ら所定の機械学習により診断モデルＭＯを作成することにより、診断モデルＭＯを取得する。診断モデル取得処理は、ユーザが情報処理装置１００の操作入力部１４０を操作して開始指示を入力したことに応じて開始される。

[0040] はじめに、情報処理装置１００の原データ取得部１１１（図４）が、診断モデルＭＯの作成に用いられる原データを取得する（Ｓ１１０）。原データは、複数の胆道がん患者および健常者（コントロール）から採取された生体試料としての尿の質量分析結果を示す質量分析データＭＤ（図１および図２参照）である。原データにおいては、尿中に含まれる多数の物質（代謝物）のピーク面積が特定されている。原データは、インターフェース部１５０を介してあるいは操作入力部１４０を介して取得される。

[0041] 次に、情報処理装置１００のペア設定部１１２（図４）が、複数の物質グループペアを設定する（Ｓ１２０）。上述したように、物質グループペアは、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである（図１および図２参照）。ペア設定部１１２は、試料に含まれる複数の物質の互いの相関に基づき、複数の物質グループペアを設定する。

[0042] ペア設定部１１２による複数の物質グループペアの具体的な設定方法は、例えば以下の通りである。まず、健常者の９５％よりも外にある物質（代謝物）について、例えばＷｉｌｃｏｘｏｎ　ｒａｎｋ－ｓｕｍ　ｔｅｓｔを行い、統計学的有意差が十分に大きな物質を選択する。選択された物質は、がんに特異的な物質であると言える。選択された物質のうち、健常者に対して低下する（Ｌｏｇ２　ｆｏｌｄ　ｃｈａｎｇｅ（Ｌ２ＦＣ）＜０の関係を満たす）物質のピーク面積データを－１倍して正負反転させ、ピアソンの相関係数を計算する。この相関係数を距離とした教師なしクラスタリング（例えば、Ｒパッケージのｍ３Ｃを用いる）をｍａｘＫ＝１０まで行い、ｐ値が最

も低くなるクラスタ数に複数の物質を分類する。以上のように正負反転させたデータを混ぜてクラスタリングすることで、正負反転した逆相関を示す物質グループが同じクラスタ内に分類される。これら同じクラスタ内の逆相関した物質グループ同士でペアを作成する。クラスタ内に逆相関を示す物質が十分（例えば５以上）あるか否かを確認し、これが十分でないクラスタでは、ピアソンの相関係数の近い他のクラスタとペアを組む。以上のようにして設定された各クラスタが、各物質グループペアとなる。癌では代謝リプログラミングによって様々な代謝経路に異常が生じているが、代謝経路のある一点で代謝が進まなくなると代謝前の物質グループが増加し、逆に代謝後の物質グループが減少する。このようにして生じる代謝経路上の複数の物質グループの逆相関を網羅的に抽出することで、胆道がん患者と健常者の代謝の違いを浮き彫りにする２つの物質グループから構成された物質グループペアを設定することができる。

[0043] 次に、情報処理装置１００の学習用データ取得部１１３（図４）が、設定された複数の物質グループペアについて、上述した比較データＣＤを取得する（Ｓ１３０）。例えば、学習用データ取得部１１３は、図１に示すように、複数の物質グループペアについて、比較データＣＤとして、第１物質グループに属する物質のピーク面積が第２物質グループに属する物質のピーク面積より小さければ値「０」をとり、反対に第１物質グループに属する物質のピーク面積が第２物質グループに属する物質のピーク面積以上であれば値「１」をとる、「０／１」の２値データを取得する。あるいは、学習用データ取得部１１３は、図２に示すように、複数の物質グループペアについて、比較データＣＤとして、一方の物質グループに属する各物質のピーク面積を、他方の物質グループに属するすべての物質のピーク面積の平均値で除した値を算出する。学習用データ取得部１１３は、取得された比較データＣＤを用いて学習用データＬＤを作成する。学習用データＬＤは、比較データＣＤと、胆道がん患者および健常者の別と、が対応付けられたデータである。

[0044] 次に、情報処理装置１００のモデル取得部１１４（図４）が、学習用デー

タＬＤを用いた機械学習により診断モデルＭＯを作成する（Ｓ１４０）。診断モデルＭＯの作成のための機械学習には、公知の種々の機械学習アルゴリズムを利用可能である。例えば、診断モデルＭＯの作成のために、ランダムフォレストやサポートベクターマシンが利用されてもよい。機械学習により作成された診断モデルＭＯは、情報処理装置１００の記憶部１２０に格納される。以上により、診断モデルＭＯの取得処理（図５）が完了する。

[0045] Ａ－４．診断処理：

次に、本実施形態の情報処理装置１００により実行される診断処理について説明する。図６は、本実施形態における診断処理を示すフローチャートである。診断処理は、診断モデルＭＯを用いて、対象試料の分類を行う、より具体的には、対象試料の採取元の診断対象者が胆道がんに罹患しているか否かの分類を行う処理である。診断処理は、ユーザが情報処理装置１００の操作入力部１４０を操作して開始指示を入力したことに応じて開始される。

[0046] はじめに情報処理装置１００の対象データ取得部１１５（図４）が、対象試料についての比較データＣＤを取得する（Ｓ２１０）。対象試料は、診断対象者から採取された尿である。また、上述したように、比較データＣＤは、各物質グループペアについて、一方の物質グループに属する物質のピーク面積と、他方の物質グループに属する物質のピーク面積と、の大小関係を示すデータや、他方の物質グループに属する物質のピーク面積に対する、一方の物質グループに属する物質のピーク面積の比を示すデータを含む。対象データ取得部１１５は、インターフェース部１５０を介して対象試料の質量分析データＭＤを取得し、該質量分析データＭＤに基づき比較データＣＤを作成する。あるいは、対象データ取得部１１５は、質量分析データＭＤに基づき予め作成された比較データＣＤを、インターフェース部１５０を介して取得する。

[0047] 次に、情報処理装置１００の分類実行部１１６（図４）が、対象試料についての比較データＣＤと診断モデルＭＯとを用いて、診断を実行する（Ｓ２２０）。具体的には、分類実行部１１６は、対象試料についての比較データ

ＣＤを診断モデルＭＯに入力することにより、診断モデルＭＯから出力される分類結果（対象者が胆道がんに罹患しているか否かの分類結果）を取得する。分類実行部１１６は、診断結果を示す診断結果データＲＤを生成し、情報処理装置１００の記憶部１２０に格納する。

[0048] 次に、分類実行部１１６は、診断結果データＲＤに基づき、診断結果を出力する（Ｓ２３０）。例えば、分類実行部１１６は、診断結果を表示部１３０に表示させる。以上により、診断処理が完了する。

[0049] Ａ－５．実施例：

上述した診断モデルＭＯの実施例について、以下説明する。図７は、質量分析データＭＤの主成分分析結果を示す説明図である。図７では、Ｘ軸に第１主成分（ＰＣ１）を取り、Ｙ軸に第２主成分（ＰＣ２）を取っている。各軸に各成分の寄与率を示している。また、図７では、各プロットのハッチングにより、対象データの分類を示している。図中、「ＣＣＡ」は胆道がん患者のサンプルであり、「Ｎｏ＿ＰＤＡＣ」は膵がんを疑って手術をしたが病理診断で膵がんではないことが判明した症例のサンプルであり、「ＣＴＲ」は健常者（コントロール）のサンプルであり、「ＩＰＭＣ」は膵管内乳頭状粘液性癌(Intraductal Papillary Mucinous Carcinoma)前癌病変の一種(癌化後)のサンプルであり、「ＩＰＭＮ」は膵管内乳頭状粘液性腫(Intraductal Papillary Mucinous Neoplasm)前癌病変の一種(癌化前)のサンプルであり、「ＰＤＡＣ」は膵がん患者のサンプルである。また、図７では、各プロットの形状により、対象データの属するバッチを示している。これらの点は、後述する図８および図９においても同様である。

[0050] 図７に示すように、質量分析データＭＤの主成分分析結果では、グラフエリアにおける右側に、胆道がん患者サンプル（ＣＣＡ、黒）および健常者サンプル（ＣＴＲ、白抜き）が集まり、グラフエリアにおける左側に、膵がん患者サンプル（ＰＤＡＣ、縦ハッチング）が集まり、両者が左右に大きく分離されている。この実施例では、胆道がん患者サンプルおよび健常者サンプルは朝一番に採取したサンプルである一方、膵がん患者サンプルは麻酔がか

かった後に採取したサンプルである。そのため、サンプル採取条件の相違に起因するバラツキが最も大きなバラツキのベクトル(第１主成分)となったものと考えられる。また、この主成分分析結果では、グラフエリアにおける上側に、バッチ４のサンプル（菱形）が集まり、グラフエリアにおける下側に、それ以外のバッチのサンプルが集まり、両者が上下に大きく分離されている。そのため、第２主成分はバッチ間の相違であると考えられる。一方、胆道がん患者サンプル（ＣＣＡ、黒）の分布と健常者サンプル（ＣＴＲ、白抜き）の分布は、グラフエリアにおいて完全に重なっており、胆道がん患者サンプルと健常者サンプルとの相違は、第３主成分以下である。

[0051] 図８は、上述したように、がんに特異的な物質（特徴量）に絞り込んだ質量分析データＭＤの主成分分析結果を示す説明図である。図８に示すように、「がんに特異的な物質」を選択することにより、麻酔下の影響に関係のない物質を集めることができる。これにより、図７に示す主成分分析結果では第１主成分(サンプル採取の条件の相違に起因するバラツキ)方向に分離していた胆道がん患者サンプル（ＣＣＡ、黒）および健常者サンプル（ＣＴＲ、白抜き）と、膵がん患者サンプル（ＰＤＡＣ、縦ハッチング）との間の距離が、かなり縮まっている。また、胆道がん患者サンプル（ＣＣＡ、黒）と健常者サンプル（ＣＴＲ、白抜き）とが、第２主成分の方向において互いに分離されている。しかしながら、図８に示す主成分分析結果では、第１主成分の方向においては、胆道がん患者サンプル（ＣＣＡ、黒）と健常者サンプル（ＣＴＲ、白抜き）とは分離されていない。すなわち、第１主成分は、あいかわらず、サンプル採取条件の相違に起因するバラツキであると言える。

[0052] 図９は、質量分析データＭＤに基づき生成された比較データＣＤの主成分分析結果を示す説明図である。図９に示す主成分分析結果では、胆道がん患者サンプル（ＣＣＡ、黒）と健常者サンプル（ＣＴＲ、白抜き）とがしっかりと分離されている。具体的には、第１主成分の値が正であれば胆道がん患者、第１主成分の値が負であれば健常者として分類しても、高い精度で診断を行うことができる。すなわち、この主成分分析結果では、胆道がんに罹患

しているか否かが一番大きなバラツキの要因となっており、胆道がんについての診断に資する状態となっている。また、健常者サンプル（ＣＴＲ、白抜き）と膵がん患者サンプル（ＰＤＡＣ、縦ハッチング）とが、上下方向に大きく分離されていることから、第２主成分は膵がんか否かのバラツキであると言える。バッチ間の相違に起因するバラツキは、第３主成分以下となっている。第１主成分の寄与率が２２％、第２主成分の寄与率が１８．９％と生成されたデータの４０．９％%が胆道癌・膵癌の診断に寄与する形にブーストされた一方で、第３主成分の寄与率は６％まで低下し、バッチ間の相違に起因するバラツキは相対的に縮小した。

[0053] 図７から図９に示す主成分分析の結果から、比較データＣＤを用いて学習することにより作成した診断モデルＭＯを利用することにより、様々な要因に起因する試料間での値のバラツキの影響を相対的に小さくすることができ、より精度の高い疾病の診断を実現することができると言える。

[0054] 図１０は、本実施形態の診断モデルＭＯの精度を示す説明図である。図１０には、２０２１年に提出されたデータを用いた学習により診断モデルＭＯを作成し、その後の２０２２年に提出されたデータを用いて検証を行ったときの（すなわち前向き研究を行ったときの）、特異度（ｓｐｅｃｉｆｉｃｉｔｙ）および感度（ｓｅｎｓｉｔｉｖｉｔｙ）を示している。図１０に示すように、診断モデルＭＯを用いた胆道がんの診断（胆道がんに罹患しているか否かの分類）において、感度：８２．８％、特異度：１００％、ＲＯＣ－ＡＵＣ：０．９９４を示し、精度の良い診断が実現されていると言える。

[0055] 図１１は、本実施形態の診断モデルＭＯの精度を示す説明図である。図１１には、本実施形態の診断モデルＭＯを膵がんのデータに適用した場合の、膵がんのステージ（ステージ１～４）毎の特異度および感度を示している。図１１に示すように、本実施形態の診断モデルＭＯを用いた膵がんの診断において、ステージ２～４についてはもちろんのこと、初期段階であるために診断が比較的難しいステージ１についても、精度の良い診断が実現されている。

[0056] 図１２および図１３は、比較例としてのクレアチニン補正と本実施形態の診断モデルＭＯとの比較結果を示す説明図である。図１２および図１３には、それぞれ、クレアチニン補正および本実施形態の診断モデルＭＯを用いて２つのデータセット（バッチ１およびバッチ２）の分類を行った場合について、主成分分析結果および分類精度を示している。図１２に示すように、クレアチニン補正では、バッチ１とバッチ２とが分離しており、バッチ間の相違に起因するバラツキが相対的に大きくなっており、その結果、分類精度も比較的低くなっている。一方、図１３に示すように、本実施形態の診断モデルＭＯを用いた分類（ＩＰＢ）では、バッチ１とバッチ２とが分離しておらず、バッチ間の相違に起因するバラツキが相対的に小さくなっており、その結果、分類精度も比較的高くなっている。このように、本実施形態の診断モデルＭＯを用いた分類によれば、クレアチニン補正を用いた分類と比較して、分類の精度を向上させることができる。

[0057] Ａ－６．本実施形態の効果：

以上説明したように、本実施形態の情報処理装置１００は、試料の分類を行うための装置であって、モデル取得部１１４と、対象データ取得部１１５と、分類実行部１１６とを備える。モデル取得部１１４は、比較データＣＤを入力とし、試料の分類結果を出力とする学習済モデルである診断モデルＭＯを取得する。比較データＣＤは、試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の物質グループに属する物質のピーク面積と、他方の物質グループに属する物質のピーク面積と、の大小関係を示すデータを含む。対象データ取得部１１５は、対象試料についての比較データＣＤを取得する。分類実行部１１６は、対象試料についての比較データＣＤを診断モデルＭＯに入力することにより、対象試料の分類を実行し、分類の結果を出力する。

[0058] このように、本実施形態の情報処理装置１００では、比較データＣＤを入力とし試料の分類結果を出力とする学習済モデルである診断モデルＭＯを用

いて、対象試料の分類を実行する。比較データＣＤは、試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の物質グループに属する物質のピーク面積と、他方の物質グループに属する物質のピーク面積と、の大小関係を示すデータを含む。そのため、比較データＣＤを利用することにより、様々な要因に起因する試料間での値のバラツキの影響を相対的に小さくすることができ、質量分析結果を参照した試料の分類の精度を向上させることができる。また、比較データＣＤは、試料に含まれる複数の物質についての質量分析結果を利用して作成されたデータであるため、特定の物質のみについての質量分析結果を利用する構成と比較して、質量分析結果を参照した試料の分類の精度を向上させることができる。

[0059] また、本実施形態の情報処理装置１００において、比較データＣＤは、他方の物質グループに属する物質のピーク面積に対する、一方の物質グループに属する物質のピーク面積の比を示すデータを含むとしてもよい。このような構成とすれば、試料におけるピーク面積が同程度であって、属する分類により大小関係が反転するような物質のペアに限られず、任意の物質の組合せにより物質グループペアを構成することができる。これにより、物質グループペアの設定の自由度が高まるため、より好ましい比較データＣＤを得ることができ、質量分析結果を参照した試料の分類の精度を効果的に向上させることができる。

[0060] また、本実施形態の情報処理装置１００において、比較データＣＤは、他方の物質グループに属するすべての物質のピーク面積の平均値に対する、一方の物質グループに属する各物質のピーク面積の比を示すデータを含むとしてもよい。このような構成とすれば、それぞれの物質について比較データＣＤの値が定まることとなり、情報量の減少を抑制することができる。これにより、より情報量の多い比較データＣＤを得ることができ、質量分析結果を参照した試料の分類の精度を効果的に向上させることができる。

[0061] また、本実施形態の情報処理装置１００において、比較データＣＤは、複数の物質グループペアについてのデータを含む。そのため、より情報量の多い比較データＣＤを得ることができ、質量分析結果を参照した試料の分類の精度を効果的に向上させることができる。

[0062] また、本実施形態の情報処理装置１００は、さらに、複数の物質の互いの相関に基づき少なくとも１つの物質グループペアを設定するペア設定部１１２を備える。そのため、より適切な分類に資する物質グループペアを設定することができ、質量分析結果を参照した試料の分類の精度を効果的に向上させることができる。

[0063] また、本実施形態の情報処理装置１００において、対象の試料は、ヒトから採取された生体試料であるとしてもよい。このような構成とすれば、生体試料の質量分析結果を参照した生体試料の分類の精度を向上させることができる。また、生体試料は、尿と、胆汁と、汗と、唾液と、血液と、髄液と、腹水と、胸水と、膵液と、胃液と、腸液と、の少なくとも１つであってもよい。このような構成とすれば、それらの生体試料の質量分析結果を参照した生体試料の分類の精度を向上させることができる。

[0064] また、本実施形態の情報処理装置１００において、分類実行部１１６は、特定の疾病についての分類を行ってもよい。このような構成とすれば、質量分析結果を参照した特定の疾病の診断の精度を向上させることができる。また、特定の疾病は、がんであるとしてもよい。このような構成とすれば、質量分析結果を参照したがんの診断の精度を向上させることができる。

[0065] Ｂ．変形例：

本明細書で開示される技術は、上述の実施形態に限られるものではなく、その要旨を逸脱しない範囲において種々の形態に変形することができ、例えば次のような変形も可能である。

[0066] 上記実施形態における情報処理装置１００の構成は、あくまで一例であり、種々変形可能である。また、上記実施形態における診断モデル取得処理および診断処理の内容は、あくまで一例であり、種々変形可能である。例えば

、上記実施形態では、情報処理装置１００が、診断モデルＭＯを自ら生成することによって診断モデルＭＯを取得しているが、情報処理装置１００が、他の装置により生成された診断モデルＭＯを取得するとしてもよい。

[0067] 上記実施形態における診断モデルＭＯの作成に用いられるデータの構成や機械学習アルゴリズムは、あくまで一例であり、種々変形可能である。例えば、診断モデルＭＯの作成に用いられる比較データＣＤは、単一の物質グループペアについてのデータから構成されていてもよいし、複数の物質グループペアについてのデータから構成されていてもよい。また、図１の例において、各物質グループペアについての比較データＣＤが、一方の物質グループを構成する物質のピーク面積を、他方の物質グループを構成する物質のピーク面積で除した値であるとしてもよい。また、図２の例において、各物質グループペアについての比較データＣＤが、一方の物質グループに属するすべての物質のピーク面積の平均値を、他方の物質グループに属するすべての物質のピーク面積の平均値で除した値であるとしてもよい。また、図２の例において、平均値の代わりに、他の代表値（例えば中央値）を用いてもよい。また、質量分析データＭＤや比較データＣＤは、具体的な物質名が特定されたデータである必要はなく、例えば質量分析により得られるマススペクトルのピークの位置（例えば、ｍ／ｚ値）が特定されたデータであってもよい。

[0068] 比較データＣＤを診断モデルＭＯに入力する前に主成分分析を行い、第１主成分・第２主成分などを診断モデルＭＯの入力に用いてもよい。

[0069] 上記実施形態では、試料の採取元の対象者が胆道がんに罹患しているか否かの分類を行うための情報処理を例示しているが、本明細書に開示される技術は、胆道がんに限らず、特定の疾病に罹患しているか否かの分類にも同様に適用可能である。さらに、本明細書に開示される技術は、診断に限らず、試料を何らかの観点で分類する場合にも同様に適用可能である。

[0070] 上記実施形態において、情報処理装置１００の制御部１１０に含まれる各機能部の少なくとも１つが、情報処理装置１００の制御部１１０ではなく、他の装置に含まれていてもよい。上記実施形態における診断モデル取得処理

および診断処理は、必ずしも単一の装置により実行される必要はなく、それぞれ別の装置により実行されてもよい。上記実施形態における診断モデル取得処理の各ステップは、必ずしも単一の装置により実行される必要はなく、互いに異なる装置により実行されてもよい。同様に、上記実施形態における診断処理の各ステップは、必ずしも単一の装置により実行される必要はなく、互いに異なる装置により実行されてもよい。上記実施形態において、ハードウェアによって実現されている構成の一部をソフトウェアに置き換えるようにしてもよく、反対に、ソフトウェアによって実現されている構成の一部をハードウェアに置き換えるようにしてもよい。

[0071] 機械学習手法は、ランダムフォレストやサポートベクターマシンなどの単体でもよいし、これらのいくつかを組み合わせたアンサンブルモデル（スタックモデルまたはブレンドモデル）でもよい。

[0072] 質量分析の結果は、各物質のピーク面積に限られず、何らかのピーク強度の値であればよい。

符号の説明

[0073] １００：情報処理装置　１１０：制御部　１１１：原データ取得部　１１２：ペア設定部　１１３：学習用データ取得部　１１４：モデル取得部　１１５：対象データ取得部　１１６：分類実行部　１２０：記憶部　１３０：表示部　１４０：操作入力部　１５０：インターフェース部　１９０：バス　ＣＰ：診断プログラム　ＬＤ：学習用データ　ＭＤ：質量分析データ　ＭＯ：診断モデル　ＲＤ：診断結果データ

請求の範囲

[請求項1] 試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の前記物質グループに属する物質のピーク強度と、他方の前記物質グループに属する物質のピーク強度と、の大小関係を示すデータを含む比較データを入力とし、前記試料の分類結果を出力とする学習済モデルを取得するモデル取得部と、

対象試料についての前記比較データを取得する対象データ取得部と、

前記対象試料についての前記比較データを前記学習済モデルに入力することにより、前記対象試料の分類を実行し、前記分類の結果を出力する分類実行部と、

を備える、情報処理装置。

[請求項2] 請求項１に記載の情報処理装置であって、

前記比較データは、前記他方の物質グループに属する物質のピーク強度に対する、前記一方の物質グループに属する物質のピーク強度の比を示すデータを含む、情報処理装置。

[請求項3] 請求項２に記載の情報処理装置であって、

前記比較データは、前記他方の物質グループに属するすべての物質のピーク強度の平均値に対する、前記一方の物質グループに属する各物質のピーク強度の比を示すデータを含む、情報処理装置。

[請求項4] 請求項１から請求項３までのいずれか一項に記載の情報処理装置であって、

前記比較データは、複数の前記物質グループペアについてのデータを含む、情報処理装置。

[請求項5] 請求項１から請求項３までのいずれか一項に記載の情報処理装置であって、さらに、

前記複数の物質の互いの相関に基づき、前記少なくとも１つの物質グループペアを設定するペア設定部を備える、情報処理装置。

[請求項6] 請求項１から請求項３までのいずれか一項に記載の情報処理装置であって、

前記試料は、ヒトから採取された生体試料である、情報処理装置。

[請求項7] 請求項６に記載の情報処理装置であって、

前記生体試料は、尿と、胆汁と、汗と、唾液と、血液と、髄液と、腹水と、胸水と、膵液と、胃液と、腸液と、の少なくとも１つである、情報処理装置。

[請求項8] 請求項６に記載の情報処理装置であって、

前記分類実行部は、特定の疾病についての前記分類を行う、情報処理装置。

[請求項9] 請求項８に記載の情報処理装置であって、

前記特定の疾病は、がんである、情報処理装置。

[請求項10] 試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の前記物質グループに属する物質のピーク強度と、他方の前記物質グループに属する物質のピーク強度と、の大小関係を示すデータを含む比較データを入力とし、前記試料の分類結果を出力とする学習済モデルを、学習を実行することにより生成するモデル取得部を備える、情報処理装置。

[請求項11] 試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の前記物質グループに属する物質のピーク強度と、他方の前記物質グループに属する物質のピーク強度と、の大小関係を示すデータを含む比較データを入力とし、前記試料の分類結果を出力とする学習済モデルを取得す

る工程と、

対象試料についての前記比較データを取得する工程と、

前記対象試料についての前記比較データを前記学習済モデルに入力することにより、前記対象試料の分類を実行し、前記分類の結果を出力する工程と、

を備える、情報処理方法。

[請求項12] 試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の前記物質グループに属する物質のピーク強度と、他方の前記物質グループに属する物質のピーク強度と、の大小関係を示すデータを含む比較データを入力とし、前記試料の分類結果を出力とする学習済モデルを、学習を実行することにより生成する工程を備える、情報処理方法。

[請求項13] コンピュータに、

試料に含まれる複数の物質についての質量分析結果において、それぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の前記物質グループに属する物質のピーク強度と、他方の前記物質グループに属する物質のピーク強度と、の大小関係を示すデータを含む比較データを入力とし、前記試料の分類結果を出力とする学習済モデルを取得する処理と、

対象試料についての前記比較データを取得する処理と、

前記対象試料についての前記比較データを前記学習済モデルに入力することにより、前記対象試料の分類を実行し、前記分類の結果を出力する処理と、

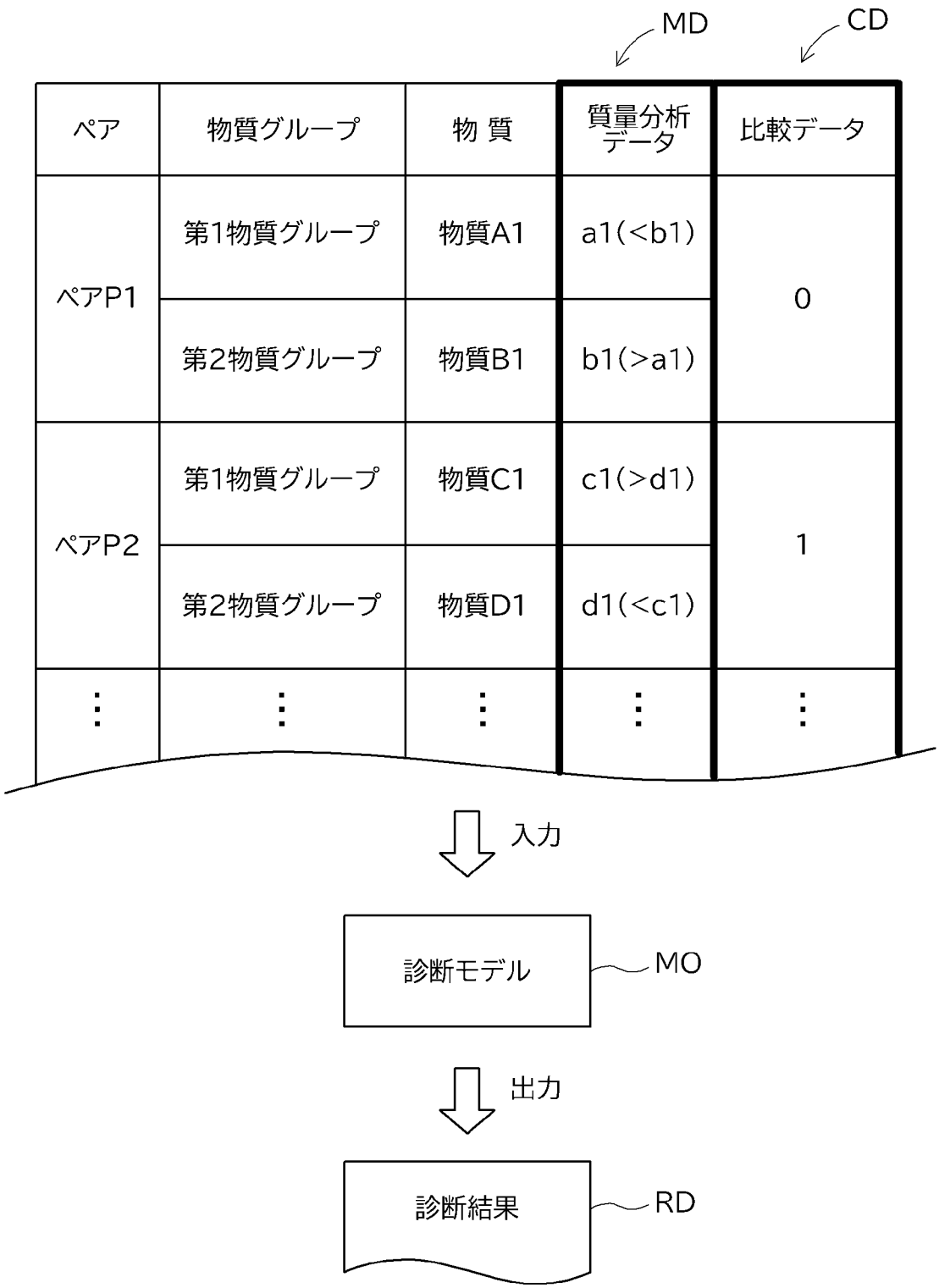
を実行させる、コンピュータプログラム。

[請求項14] コンピュータに、

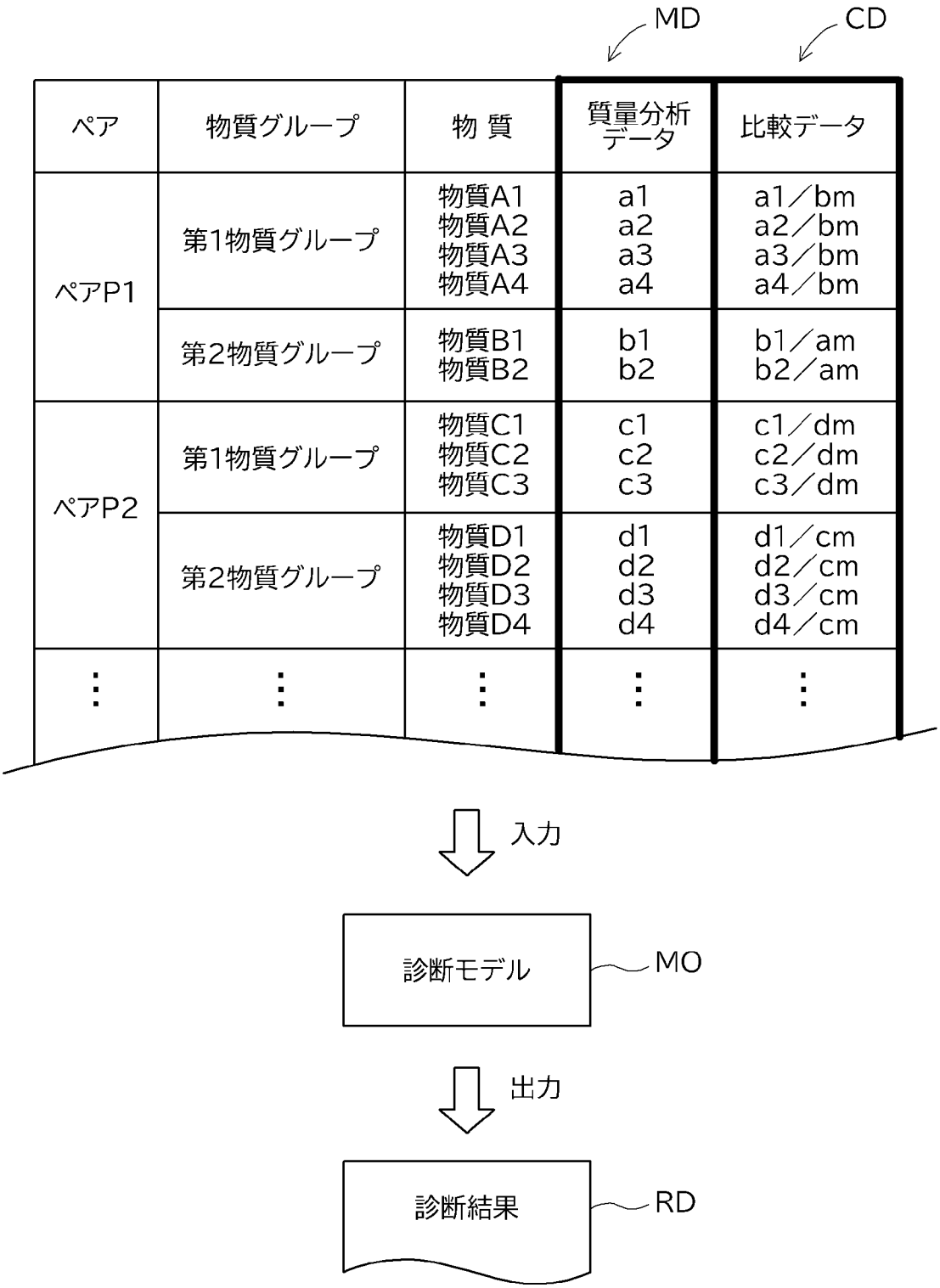
試料に含まれる複数の物質についての質量分析結果において、それ

ぞれ少なくとも１つの物質から構成される２つの物質グループのペアである少なくとも１つの物質グループペアについて、一方の前記物質グループに属する物質のピーク強度と、他方の前記物質グループに属する物質のピーク強度と、の大小関係を示すデータを含む比較データを入力とし、前記試料の分類結果を出力とする学習済モデルを、学習を実行することにより生成する処理を実行させる、コンピュータプログラム。

[図1]


MD
CD
ペア
物質グループ
物 質
質量分析データ
比較データ
ペアP1
第1物質グループ
物質A1
a1(<b1)
第2物質グループ
物質B1
b1(>a1)
0
ペアP2
第1物質グループ
物質C1
c1(>d1)
第2物質グループ
物質D1
d1(<c1)
1
入力
診断モデル
MO
出力
診断結果
RD

[図2]


MD
CD
ペア
物質グループ
物 質
質量分析データ
比較データ
ペアP1
第1物質グループ
物質A1
物質A2
物質A3
物質A4
a1
a2
a3
a4
a1／bm
a2／bm
a3／bm
a4／bm
第2物質グループ
物質B1
物質B2
b1
b2
b1／am
b2／am
ペアP2
第1物質グループ
物質C1
物質C2
物質C3
c1
c2
c3
c1／dm
c2／dm
c3／dm
第2物質グループ
物質D1
物質D2
物質D3
物質D4
d1
d2
d3
d4
d1／cm
d2／cm
d3／cm
d4／cm
入力
診断モデル
MO
出力
診断結果
RD

[図3]

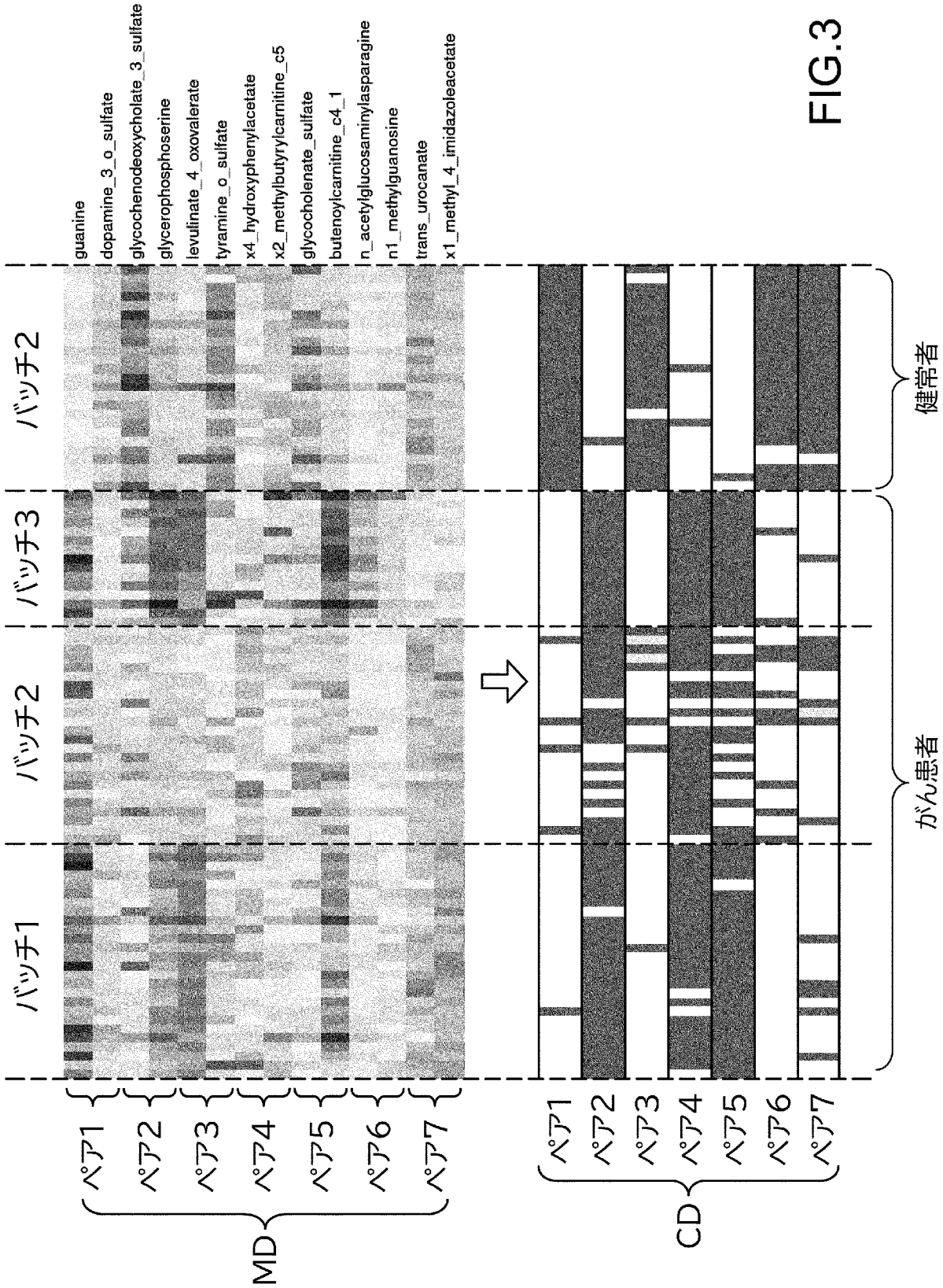
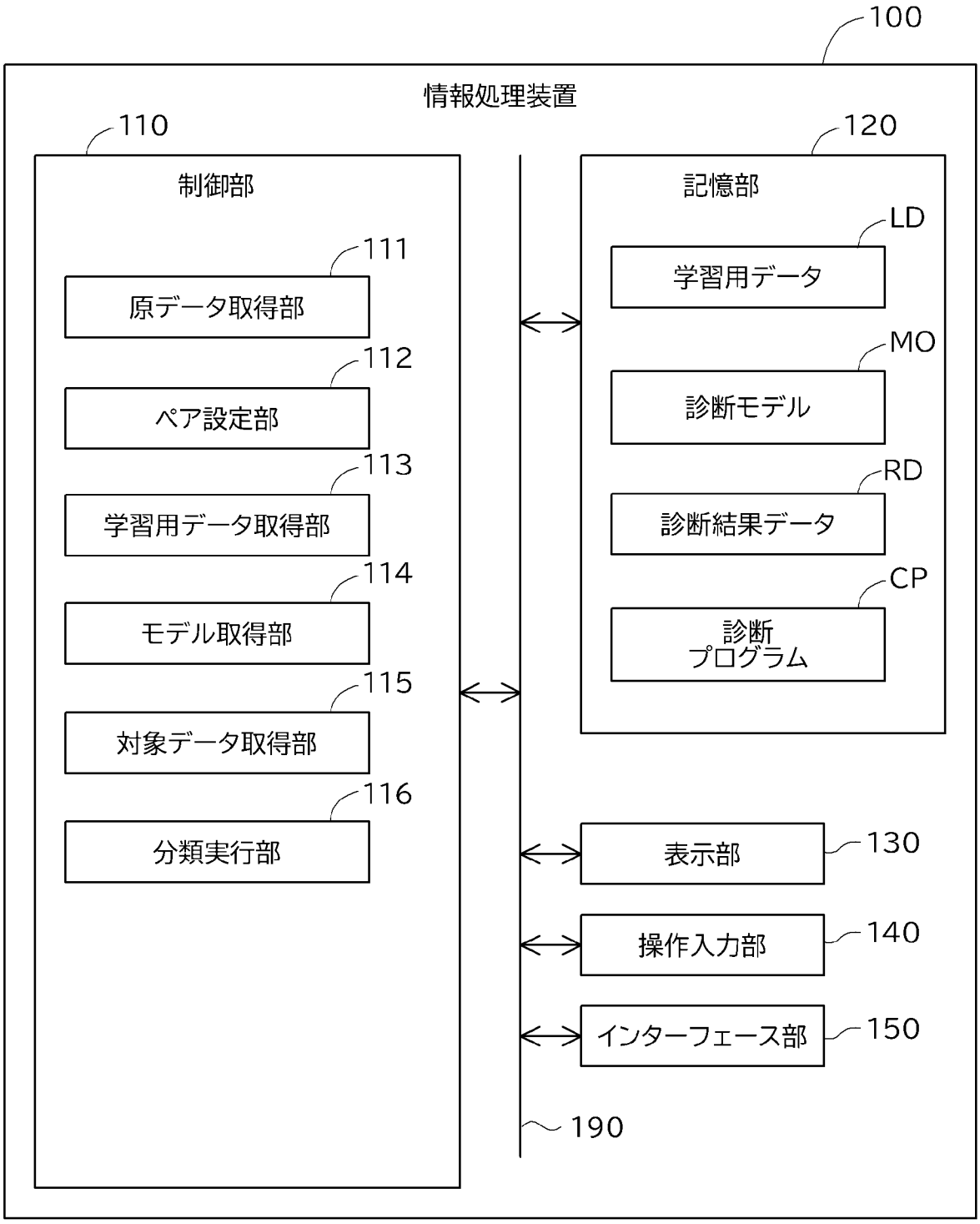
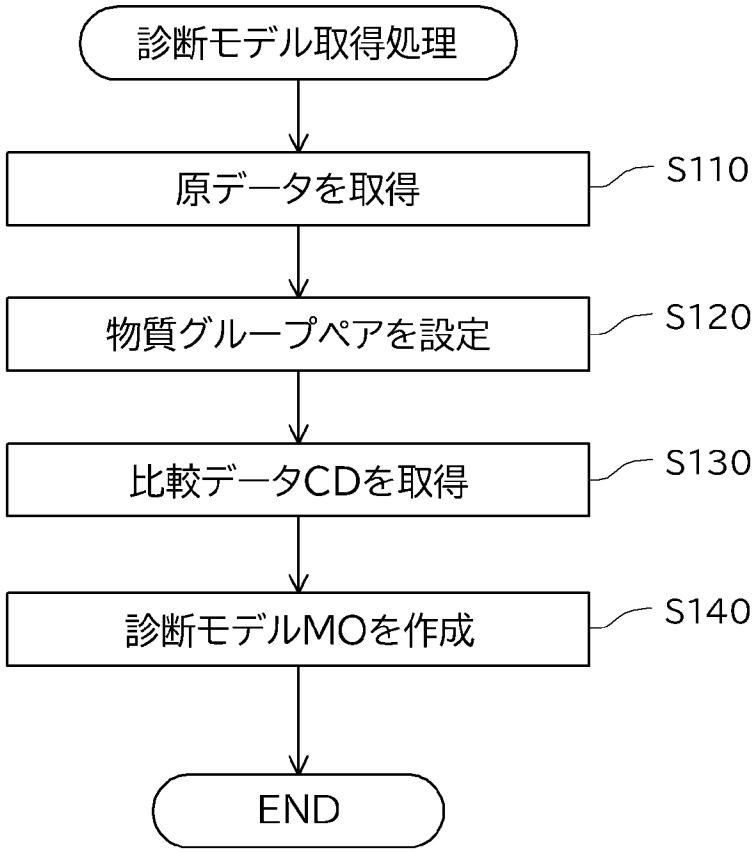

バッチ1
バッチ2
バッチ3
バッチ2
guanine
dopamine_3_o_sulfate
glycochenodeoxycholate_3_sulfate
glycerophosphoserine
levulinate_4_oxovalerate
tyramine_o_sulfate
x4_hydroxyphenylacetate
x2_methylbutyrylcarnitine_c5
glycocholenate_sulfate
butenoylcarnitine_c4_1
n_acetylglucosaminylasparagine
n1_methylguanosine
trans_urocanate
x1_methyl_4_imidazoleacetate
ペア1
ペア2
ペア3
ペア4
ペア5
ペア6
ペア7
MD
CD
がん患者
健常者

FIG.3

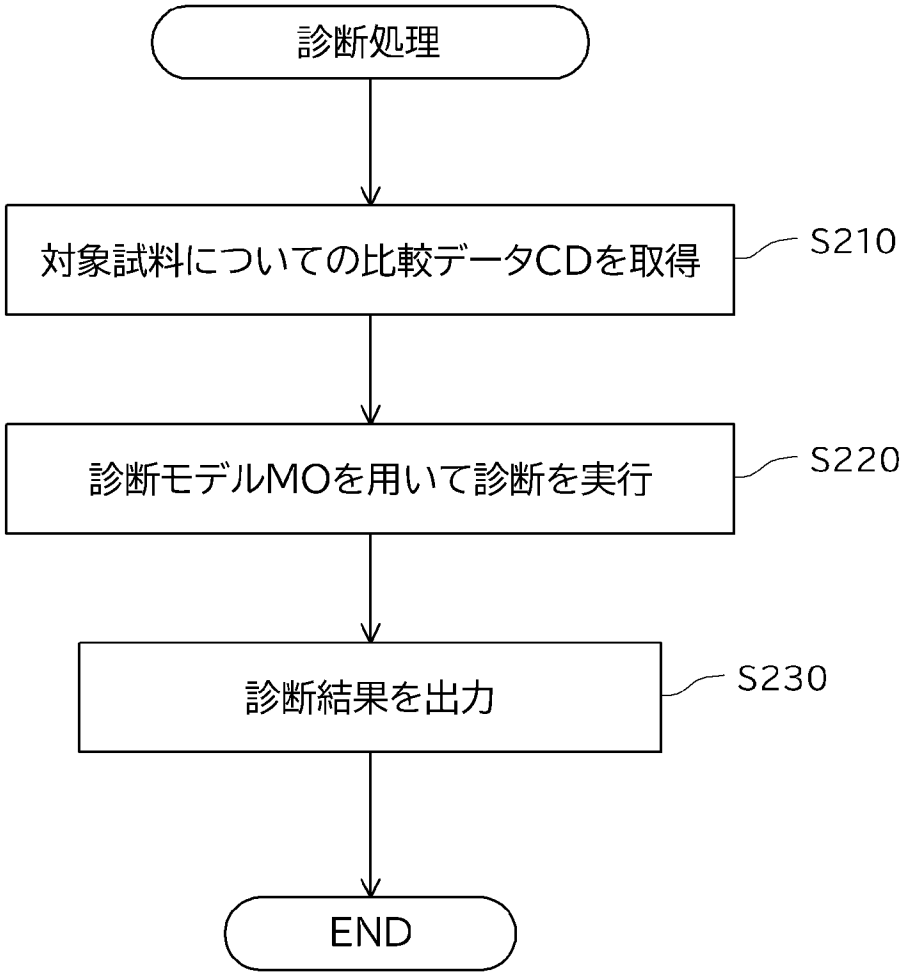
[図4]


100
情報処理装置
110
制御部
111
原データ取得部
112
ペア設定部
113
学習用データ取得部
114
モデル取得部
115
対象データ取得部
116
分類実行部
120
記憶部
LD
学習用データ
MO
診断モデル
RD
診断結果データ
CP
診断
プログラム
表示部
130
操作入力部
140
インターフェース部
150
190

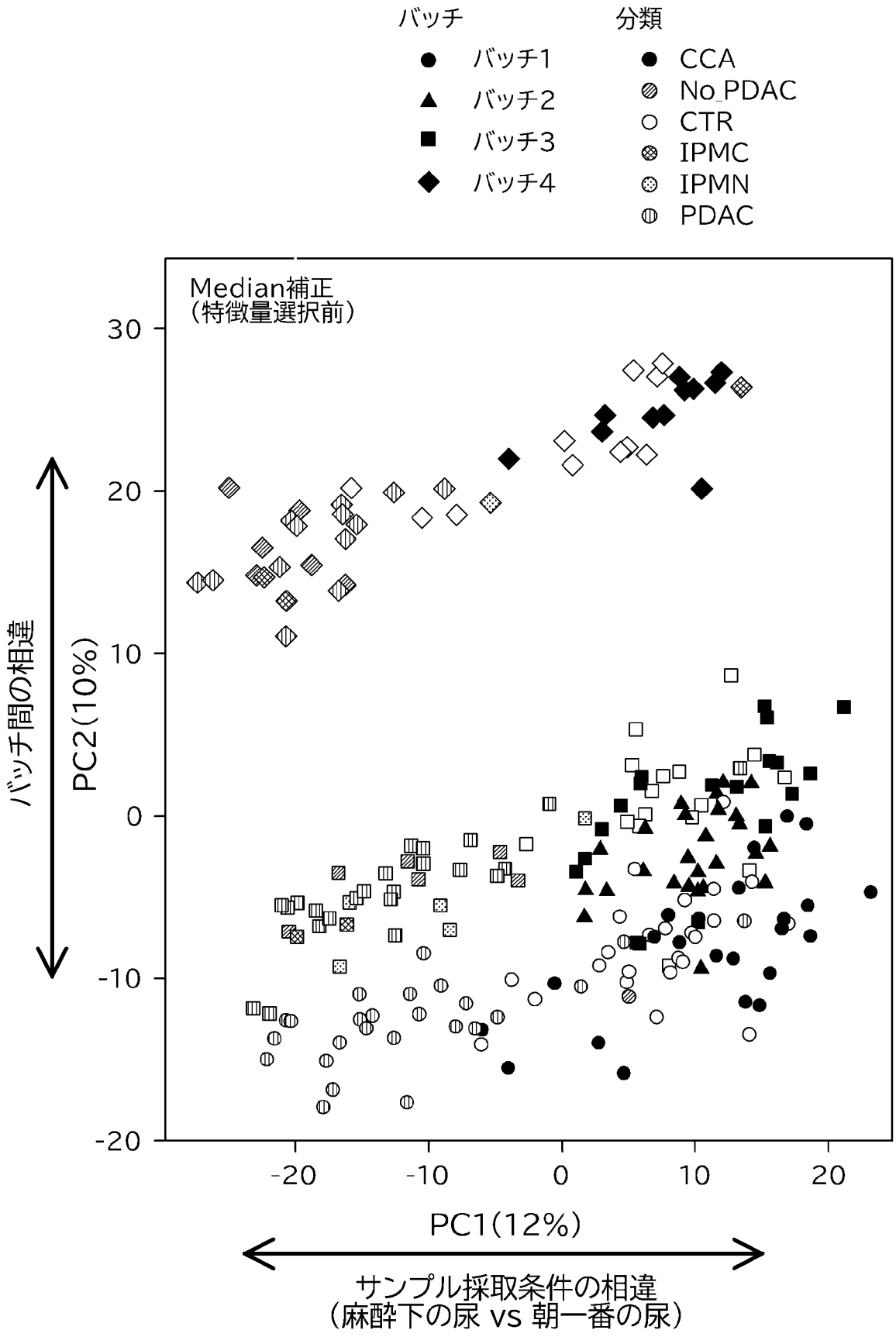
[図5]


診断モデル取得処理
原データを取得
S110
物質グループペアを設定
S120
比較データCDを取得
S130
診断モデルMOを作成
S140
END

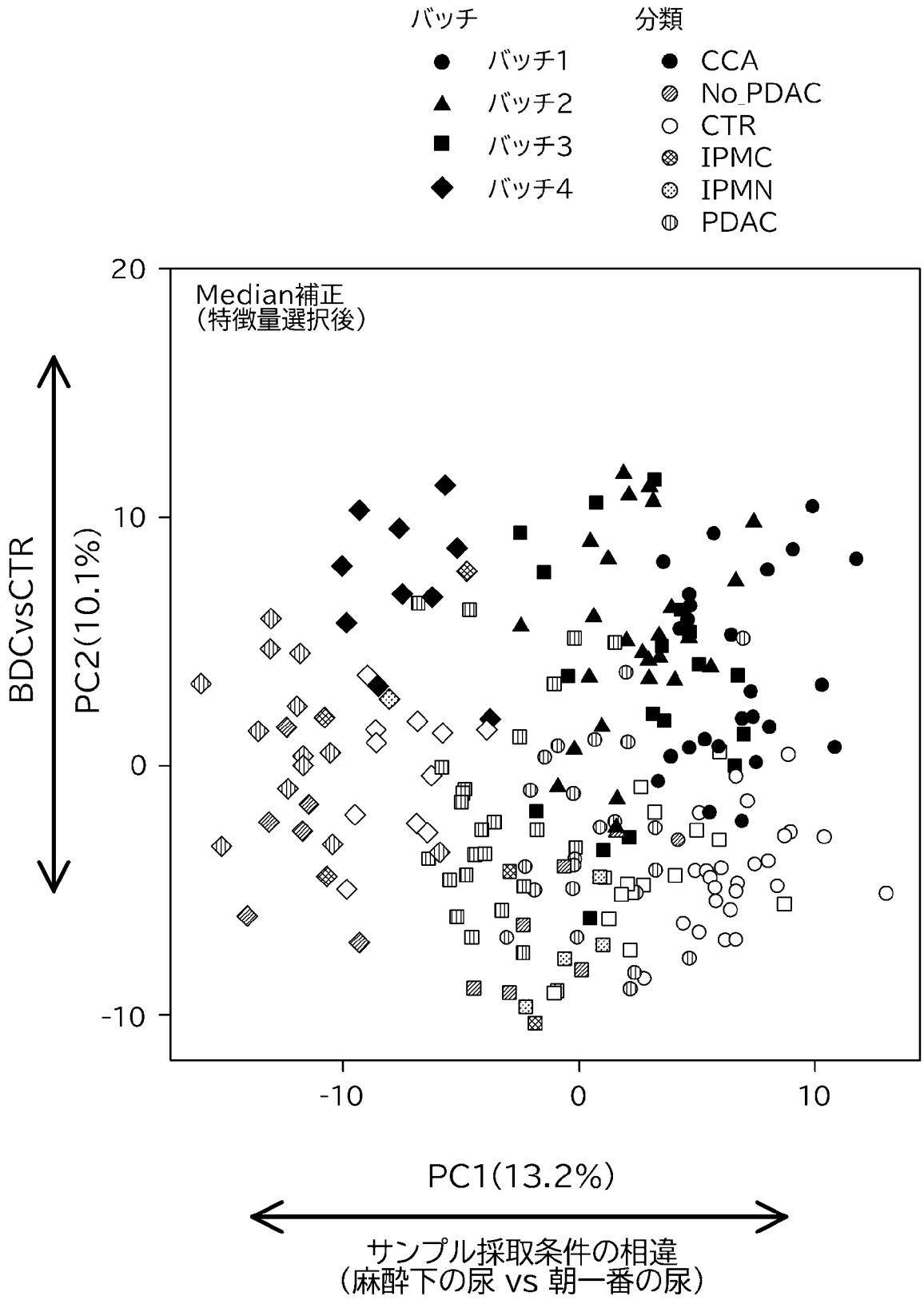
[図6]


診断処理
対象試料についての比較データCDを取得
S210
診断モデルMOを用いて診断を実行
S220
診断結果を出力
S230
END

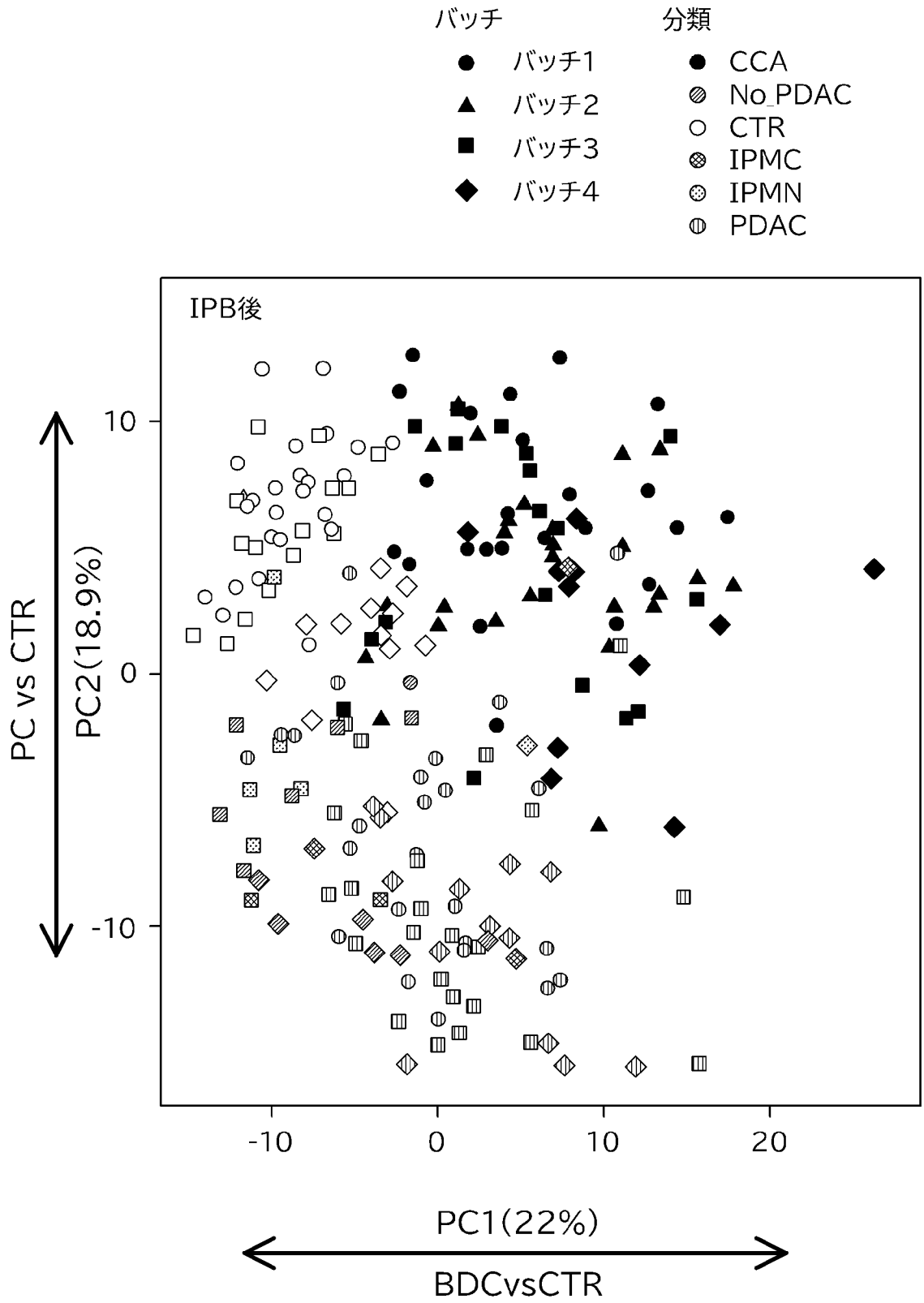
[図7]


バッチ
バッチ1
バッチ2
バッチ3
バッチ4
分類
CCA
No_PDAC
CTR
IPMC
IPMN
PDAC
Median補正
（特徴量選択前）
30
20
10
0
-10
-20
PC2(10%)
バッチ間の相違
-20
-10
0
10
20
PC1(12%)
サンプル採取条件の相違
（麻酔下の尿 vs 朝一番の尿）

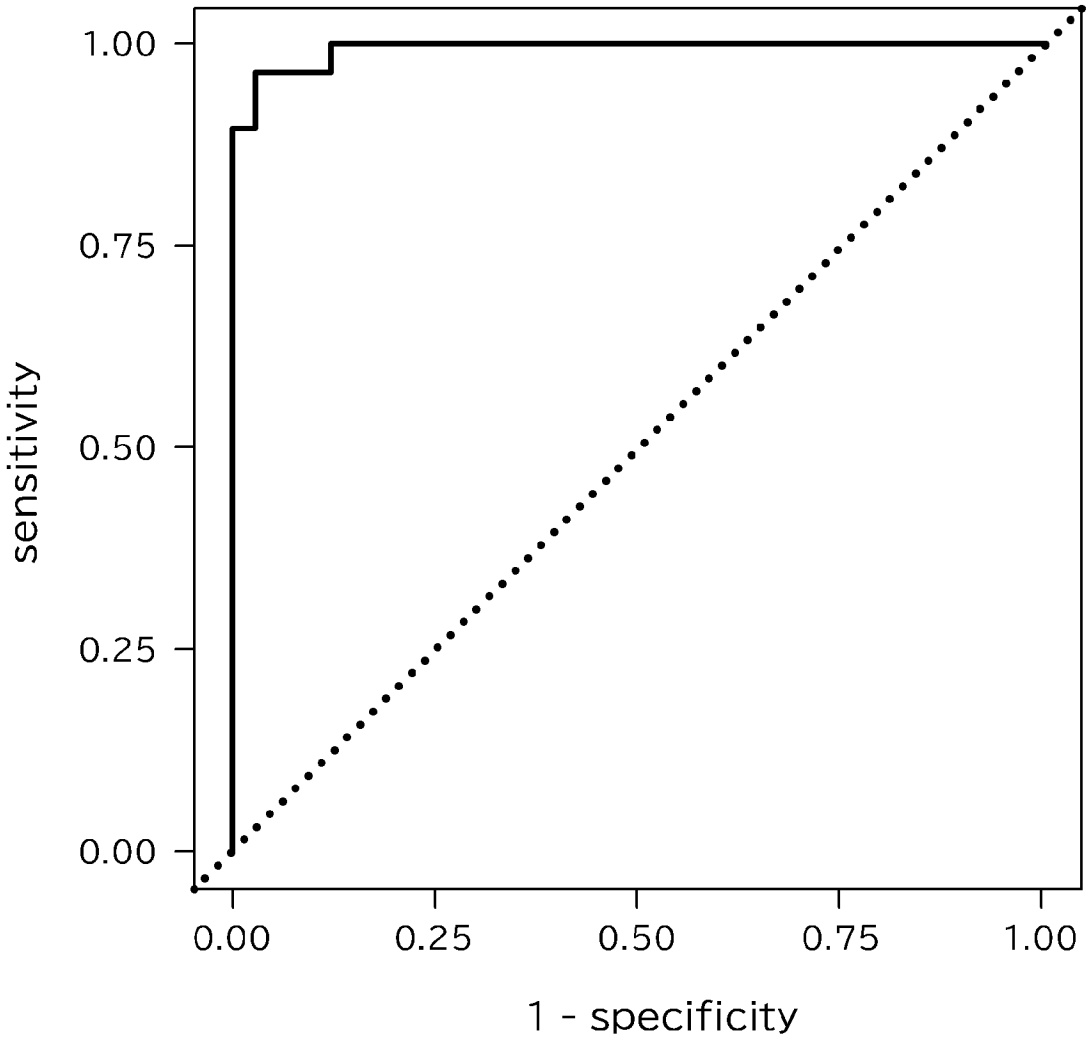
[図8]


バッチ
バッチ1
バッチ2
バッチ3
バッチ4
分類
CCA
No_PDAC
CTR
IPMC
IPMN
PDAC
Median補正
(特徴量選択後)
20
10
0
-10
-10
0
10
BDCvsCTR
PC2(10.1%)
PC1(13.2%)
サンプル採取条件の相違
(麻酔下の尿 vs 朝一番の尿)

[図9]


バッチ
バッチ1
バッチ2
バッチ3
バッチ4
分類
CCA
No_PDAC
CTR
IPMC
IPMN
PDAC
IPB後
10
0
-10
PC vs CTR
PC2(18.9%)
-10
0
10
20
PC1(22%)
BDCvsCTR

[図10]

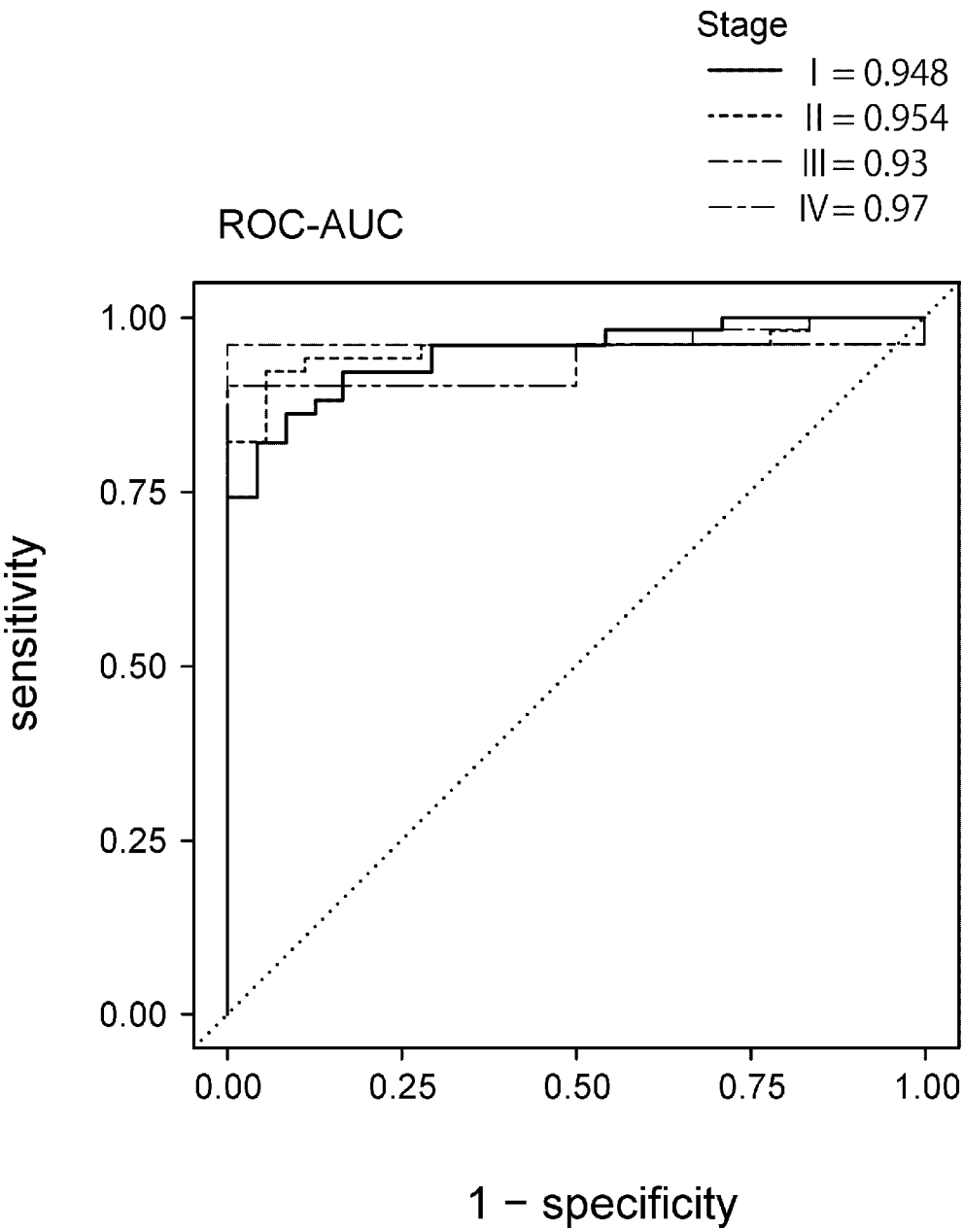


sensitivity 82.8%

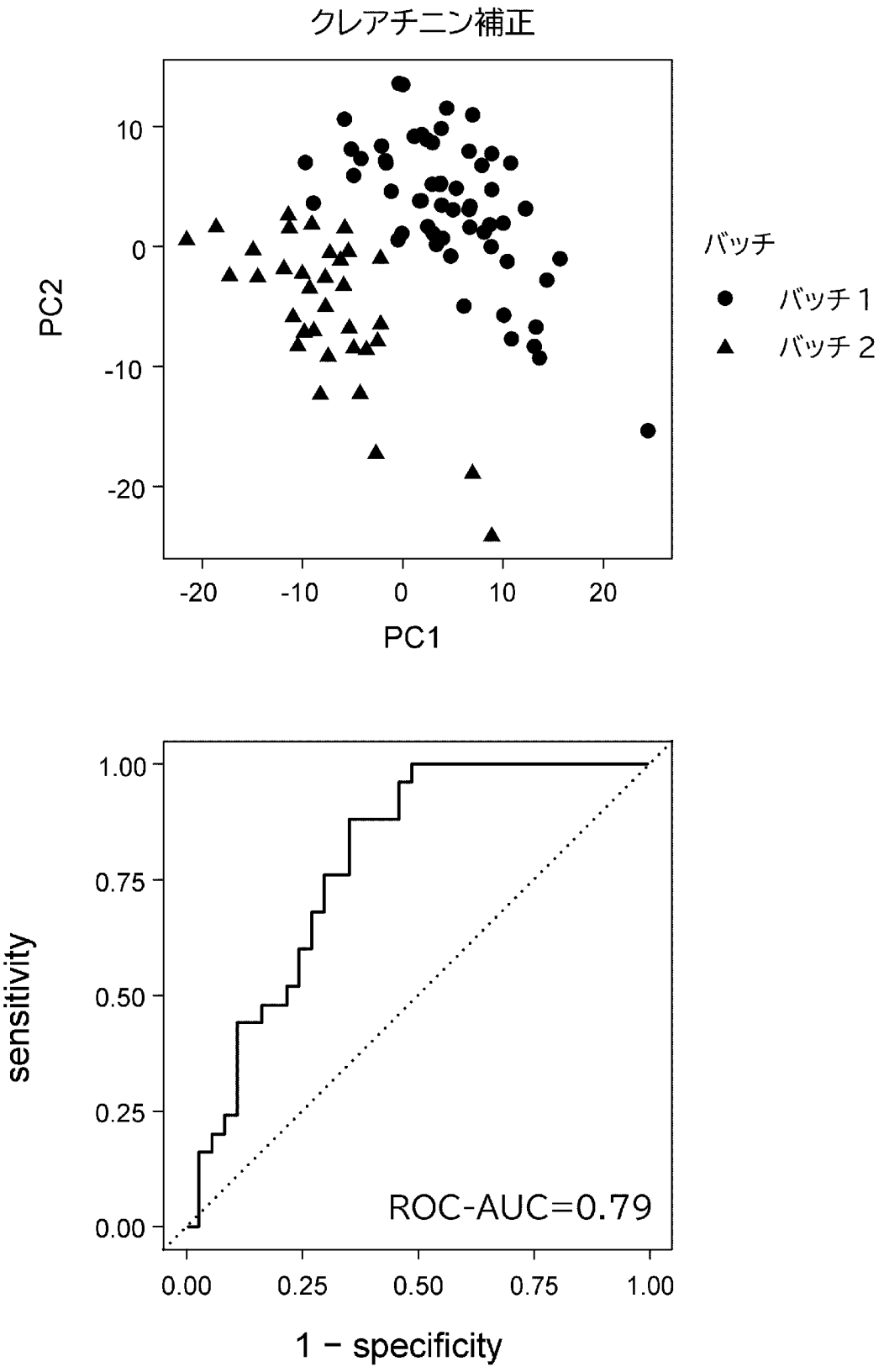
specificity 100%

ROC-AUC 0.994

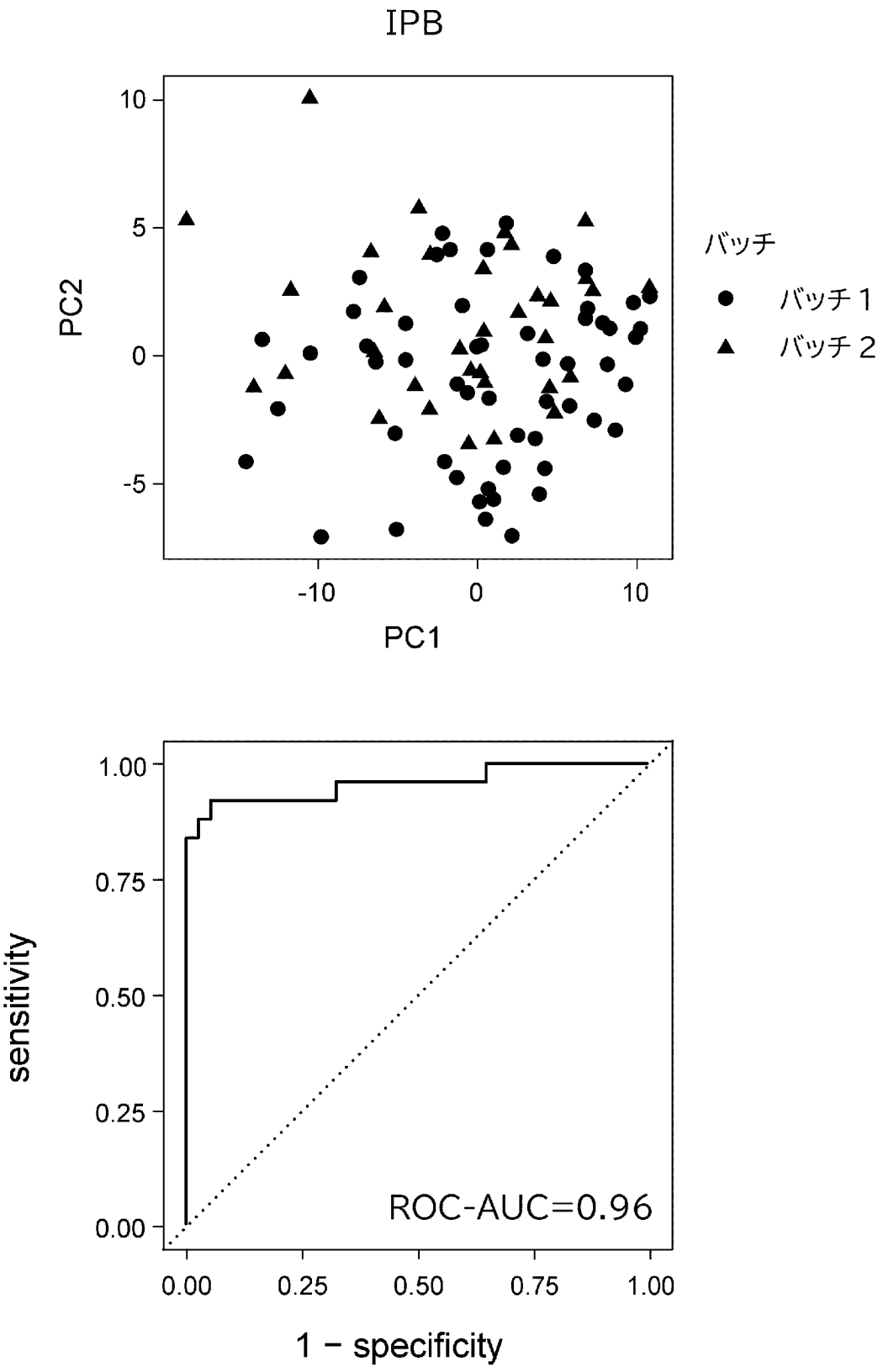
[図11]


Stage
I = 0.948
II = 0.954
III = 0.93
IV = 0.97
ROC-AUC
sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 − specificity

[図12]


クレアチニン補正
PC2
PC1
10
0
-10
-20
-20
-10
0
10
20
バッチ
バッチ１
バッチ２
sensitivity
1 − specificity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
ROC-AUC=0.79

[図13]


IPB
10
5
0
-5
PC2
-10
0
10
PC1
バッチ
バッチ１
バッチ２
1.00
0.75
0.50
0.25
0.00
sensitivity
ROC-AUC=0.96
0.00
0.25
0.50
0.75
1.00
1 − specificity

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2024/010963

A. CLASSIFICATION OF SUBJECT MATTER

G01N 27/62(2021.01)i; ***G01N 33/50***(2006.01)i
FI: G01N27/62 D; G01N33/50 Z

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)

G01N27/62; G01N33/50

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

Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2024
Registered utility model specifications of Japan 1996-2024
Published registered utility model applications of Japan 1994-2024

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2007-263641 A (SHIMADZU CORPORATION) 11 October 2007 (2007-10-11) paragraphs [0015]-[0049], fig. 1-8	1-7, 10-14
Y		8-9
Y	WO 2009/025080 A1 (SCHOOL JURIDICAL PERSON KITASATO INSTITUTE) 26 February 2009 (2009-02-26) paragraph [0042], fig. 17-18	8-9
A		1-7, 10-14
Y	JP 2008-519261 A (TARGET DISCOVERY, INC.) 05 June 2008 (2008-06-05) paragraph [0124]	8-9
A		1-7, 10-14

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
08 April 2024	**23 April 2024**

Name and mailing address of the ISA/JP	Authorized officer
Japan Patent Office (ISA/JP) **3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915** **Japan**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2024/010963

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP 2007-263641 A	11 October 2007	(Family: none)	
WO 2009/025080 A1	26 February 2009	(Family: none)	
JP 2008-519261 A	05 June 2008	US 2006/0120961 A1 paragraph [0148] WO 2006/050130 A2 EP 1824516 A2 CA 2588260 A1 IL 182775 A AU 2005302417 A	

国際調査報告

国際出願番号
PCT/JP2024/010963

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G01N 27/62(2021.01)i; G01N 33/50(2006.01)i
FI: G01N27/62 D; G01N33/50 Z

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

G01N27/62; G01N33/50

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０２４年
日本国実用新案登録公報	１９９６－２０２４年
日本国登録実用新案公報	１９９４－２０２４年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献のカテゴリー*	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
X	JP 2007-263641 A (株式会社島津製作所) 11.10.2007 (2007-10-11) [0015]-[0049], 図1-8	1-7, 10-14
Y		8-9
Y	WO 2009/025080 A1 (学校法人　北里研究所) 26.02.2009 (2009-02-26) [0042], 図17-18	8-9
A		1-7, 10-14
Y	JP 2008-519261 A (ターゲット ディスカバリー インク.) 05.06.2008 (2008-06-05) [0124]	8-9
A		1-7, 10-14

☐ Ｃ欄の続きにも文献が列挙されている。

☑ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー
"A" 特に関連のある文献ではなく、一般的技術水準を示すもの
"D" 国際出願で出願人が先行技術文献として記載した文献
"E" 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
"L" 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
"O" 口頭による開示、使用、展示等に言及する文献
"P" 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
"T" 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの
"X" 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
"Y" 特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
"&" 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
08.04.2024	23.04.2024

名称及びあて先

日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）

三宅　克馬 2W 1571

電話番号 03-3581-1101 内線 3258

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2024/010963

引用文献	公表日	パテントファミリー文献	公表日
JP 2007-263641 A	11.10.2007	（ファミリーなし）	
WO 2009/025080 A1	26.02.2009	（ファミリーなし）	
JP 2008-519261 A	05.06.2008	US 2006/0120961 A1 [0148] WO 2006/050130 A2 EP 1824516 A2 CA 2588260 A1 IL 182775 A AU 2005302417 A	