



(12) 发明专利申请

(10) 申请公布号 CN 114678122 A

(43) 申请公布日 2022.06.28

(21) 申请号 202210149664.9

G06N 3/08 (2006.01)

(22) 申请日 2022.02.18

G01N 21/65 (2006.01)

(66) 本国优先权数据

G01N 1/34 (2006.01)

202210144807.7 2022.02.17 CN

G01N 1/40 (2006.01)

(71) 申请人 上海市第十人民医院

地址 200072 上海市静安区延长中路301号

(72) 发明人 吕中伟 曲伸 陈贺昌 贾成友

丛宪玲 于博 鄢阳 张俊峰

杨哲健 刘铁 刘瑾 刘凡新

(74) 专利代理机构 上海领誉知识产权代理有限

公司 31383

专利代理师 王琰

(51) Int.Cl.

G16H 50/30 (2018.01)

G16H 50/50 (2018.01)

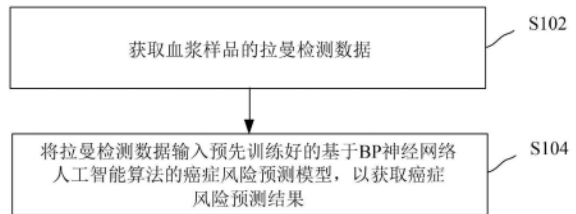
权利要求书5页 说明书28页 附图4页

(54) 发明名称

一种癌症风险预测方法、系统、设备及介质

(57) 摘要

本发明涉及一种癌症风险预测方法、系统、设备及介质,方法包括获取血浆样品的拉曼检测数据;将上述数据输入基于BP神经网络预先训练好的癌症风险预测模型以获取预测结果。其优点在于,利用拉曼检测数据与人工智能算法结合建立癌症风险预测模型,大幅度缩短检测时间,能够在15分钟左右获取癌症风险预测结果,进而决定是否进行后续精确检测;本检测方法操作简单,一次可同时检测多种物质,癌症风险预测结果准确率高;检测时间快;费用低廉;医疗废弃物少;该预测方法特异性高、敏感性高、准确率高;可实现小分子代谢物快速的大规模检测,具有高通量、高准确性的优势,检测成本相对较低。该方法如果结合临床检测结果会达到更好的预测效果。



1. 一种癌症风险预测方法,其特征在于,包括:

获取血浆样品的拉曼检测数据;

将所述拉曼检测数据输入预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型,以获取癌症风险预测结果;

其中,所述癌症风险预测结果的准确率大于91.48%、敏感度大于88.70%、特异性大于95.98%。

2. 根据权利要求1所述的癌症风险预测方法,其特征在于,所述拉曼检测数据包括通用类拉曼检测数据,其至少包括:

	ID	Name
1	M365T412_2	Lactose
2	M487T404	Bloodgroupbtrisaccharide
3	M203T272	DL-tryptophan
4	M173T309	Gly-Val
5	M437T96	Hctoxin
6	M291T33	Dl-norleucinemethylester
7	M697T235	Izenamidec

3. 根据权利要求2所述的癌症风险预测方法,其特征在于,所述拉曼检测数据还包括:肾癌拉曼检测数据,其至少包括:

	ID	Name
1	M751T38	1-myristoyl-2-palmitoyl-sn-glycero-3-phosphocholine
2	M141T292	Kojicacid
3	M865T137	Pc (16:1e/17-hdohe)
4	M791T137	Pe (18:1e/10-hdohe)
5	M671T141_2	1-Hexadecanoyl-2- (9Z,12Z-octadecadienoyl) -sn-glycero-3-phosphoricacid
6	M309T25_2	Mestranol
7	M841T140	Pc (16:0e/8-hepe)
8	M580T180	1-behenoyl-2-hydroxy-sn-glycero-3-phosphocholine
9	M538T185	1-Palmitoyllysophosphatidylcholine
10	M193T82_1	Trans-3'-hydroxycotinine

;和/或

胃癌拉曼检测数据,其至少包括:

	ID	Name
1	M747T38	(2-aminoethoxy)[2-[docosa-4.7.10.13.16.19-hexaenoyloxy]-3-[hexadec-1-en-1-yloxy]propoxy]phosphinic acid
2	M147T122	3-hydroxyglutaric acid
3	M103T300	D(-)-beta-hydroxy butyric acid
4	M119T122	Tartronate
5	M283T37	Octadecanoic acid
6	M293T115	[6]-gingerol
7	M141T34_1	Ethyl imidazole-2-carboxylate
8	M751T135	1-palmitoyl-2-thiopalmitoyl phosphatidylcholine

;和/或

宫颈癌拉曼检测数据,其至少包括:

	ID	Name
1	M113T186	Glutaric acid
2	M716T34	1-Stearoyl-sn-glycerol
3	M481T211	11-(6-{{2-(3,4-dimethoxyphenyl)ethyl}amino}-4-chloro-1,3,5-triazin-2-yl)-7,11-diazatricyclo[7.3.1.0<2,7>]trideca-2,4-dien-6-one
4	M163T32	2-tert-butyl-6-methylphenol
5	M212T29	Indoxyl sulfate
6	M130T90	Paxilline
7	M158T433	Trans-4-(aminomethyl)cyclohexanecarboxylic acid
8	M100T72	Cyclohexylamine
9	M104T411	Choline

;和/或

直肠癌拉曼检测数据,其至少包括:

	ID	Name
1	M179T369	D- (+) -mannose
2	M341T370	Sucrose
3	M665T501	Stachyose
4	M179T453	D-psicose
5	M281T75_3	Oleicacid
6	M297T102	Enterolactone
7	M321T72	Deoxythymidine5'-phosphate (dTMP)
8	M858T196	Pi36:4
9	M361T157_2	Phe-Glu
10	M511T461	3'-fucosyllactose
11	M107T82	(-) -perillyl alcohol
12	M74T71	N,n-dimethylformamide
13	M171T188	Glycyl-1-4-hydroxyproline

;和/或

前列腺癌拉曼检测数据,其至少包括:

	ID	Name
1	M173T442	Cis-aconitate
2	M923T492	Disialyllactose
3	M283T29_2	3,4'-dimethoxy-2'-hydroxychalcone

4	M151T218	Xanthine
5	M131T474	Asparagine
6	M791T38	Pe(18:1e/7-hdohe)
7	M341T489	Cellobiose
8	M116T312	Guanidoacetic acid
9	M118T412	Threonine
10	M173T505	Isocitrate
11	M473T26	Amastatin
12	M186T237_1	2-hydroxy-6-methylquinoline-3-carbaldehyde
13	M167T333	Uric acid
14	M180T330	Tyrosine
15	M179T309_2	D-fructose
16	M277T36	C14-sas (tentative)
17	M311T28_2	Thymol-beta-d-glucoside
18	M325T27	Hydroquinidine
19	M223T216	D-Quinovose
20	M339T27_2	Gly-His-Lys
21	M107T25	3-hydroxyphenylacetic acid
22	M130T279_2	Guanidinopropionic acid
23	M135T33_1	Diglyme
24	M167T57	1-methyl-3-isobutylxanthine
25	M139T482	Trans-aconitic acid
26	M102T87	4-hydroxy-l-isoleucine
27	M429T299_2	Carfentrazone-ethyl
28	M230T574	Lauryldimethylamine oxide
29	M869T102	Solamargine
30	M230T233	Arg-Gln-Arg
31	M212T458	Phosphocreatine
32	M229T447	Pro-hyp
33	M118T278	Betaine
34	M167T116_1	Spectinomycin
35	M88T321	N-methylpropionamide
36	M295T483_2	Arg-Cys
37	M317T503	.gamma.-L-Glu-.epsilon.-L-Lys
38	M116T32	Sulfallate
39	M138T222	6-aminopyridine-3-carboxamide

;和/或

肺癌拉曼检测数据,其至少包括:

	ID	Name
1	M113T186	Glutaric acid
2	M716T34	1-Stearoyl-sn-glycerol
3	M481T211	11-(6-{[2-(3,4-dimethoxyphenyl)ethyl]amino}-4-chloro-1,3,5-triazin-2-yl)-7,11-diazatricyclo[7.3.1.0<2,7>]trideca-2,4-dien-6-one
4	M163T32	2-tert-butyl-6-methylphenol
5	M212T29	Indoxyl sulfate
6	M130T90	Paxilline
7	M158T433	Trans-4-(aminomethyl)cyclohexanecarboxylic acid
8	M100T72	Cyclohexylamine
9	M104T411	Choline

;和/或

卵巢癌拉曼检测数据,其至少包括:

	ID	Name
1	M187T44	1-hydroxy-2-naphthoicacid
2	M355T37_1	Fumarprotocetraricacid
3	M159T119_2	3-hydroxyoctanoicacid
4	M498T44	Taurochenodeoxycholate
5	M182T40	4-pyridoxicacid
6	M300T204	N-Acetyl-D-Glucosamine6-Phosphate
7	M204T47	N,n'-diacetylchitobiose
8	M191T255	5-methyl-5-phenylhydantoin
9	M174T570	Ala-Thr-Arg
10	M130T582	D-Pipecolinicacid
11	M196T258	1-deoxy-1-(methylamino)-d-galactitol
12	M283T72	Hexaethyleneglycol

;和/或

乳腺癌拉曼检测数据,其至少包括:

	ID	Name
1	M129T86	5-methyl-5,6-dihydrouracil
2	M144T245	L-glutamic acid, dimethyl ester
3	M188T328_2	Leu-Ser-Arg
4	M210T532	Pro-Phe-Arg
5	M211T74_1	Cyclo(proline-leucine)
6	M213T166	m-Chlorohippuric acid
7	M282T33	Oleamide
8	M283T51	Swietenocoumarin b
9	M283T70	Stearic acid
10	M343T81	2-[1-hydroxy-1-(4-methoxyphenyl)propan-2-yl]oxy-6-(hydroxymethyl)oxane-3,4,5-triol

4. 根据权利要求1~3任一所述的癌症风险预测方法,其特征在于,在获取血浆样品的拉曼检测数据之前,所述方法还包括:

获取血液样品;

对所述血液样品进行离心处理,以获取初始血浆样品;

对所述初始血浆样品进行基于膜过滤的离心处理,以获取血浆样品;

其中,对所述初始血浆样品进行基于膜过滤的离心处理的时间小于等于5min,所述初始血浆样品的体积小于等于450 $\mu$ l,所述血浆样品的体积小于等于5 $\mu$ l。

5. 根据权利要求1~4任一所述的癌症风险预测方法,其特征在于,将所述拉曼检测数据输入预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型,以获取癌症风险预测结果包括:

处理所述拉曼检测数据,将所述拉曼检测数据映射为1024维的初始特征向量;

将1024维的初始特征向量输入至预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型,以获取4维的最终特征向量;

将4维的最终特征向量输入至分类函数进行处理,以获取癌症风险预测结果。

6. 根据权利要求1~5任一所述的癌症风险预测方法,其特征在于,基于BP神经网络人工智能算法的癌症风险预测模型的训练方法包括:

构建癌症风险预测模型,其中,所述癌症风险预测模型包括输入层、隐藏层和输出层,所述输入层的输入为1024、输出为512,所述隐藏层的输入为512、输出为128,所述输出层的输入为128、输出为4;

根据拉曼检测数据构建训练集和测试集,其中,所述训练集与所述测试集的比例为3:1;

将所述训练集输入至所述癌症风险预测模型进行训练,以获取输出结果;

将所述输出结果输入至分类函数进行处理,以获取癌症风险预测结果;

对所述癌症风险预测结果进行测试,以检测训练的有效性,并根据有效性检测结果,对所述癌症风险预测模型进行迭代,直至训练完成。

7. 一种癌症风险预测系统,包括:

血浆样品获取装置,用于对血液样品进行离心处理以获取初始血浆样品,以及对所述初始血浆样品进行基于膜过滤的离心处理以获取血浆样品;其中,对所述初始血浆样品进行基于膜过滤的离心处理的时间小于等于5min;

拉曼检测装置,用于对所述血浆样品进行拉曼检测,以获取拉曼检测数据;

癌症风险预测装置,用于获取血浆样品的拉曼检测数据,将所述拉曼检测数据输入预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型,以获取癌症风险预测结果;

其中,所述癌症风险预测结果的准确率大于91.48%、敏感度大于88.70%、特异性大于95.98%。

8. 根据权利要求7所述的癌症风险预测系统,其特征在于,所述血浆样品获取装置包括:

抗凝取血单元,用于放置血液样品;

离心单元,用于对所述抗凝取血单元的所述血液样品进行离心处理以获取初始血浆样品,以及对所述初始血浆样品进行离心处理以获取血浆样品;

膜过滤单元,用于在所述离心单元对所述初始血浆样品进行离心处理的同时,对所述初始血浆样品进行膜处理,以获取所述血浆样品。

9. 一种计算机设备,包括存储器、处理器以及存储在所述存储器上并可在所述处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现如权利要求1至6中任一项所述的癌症风险预测方法。

10. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,该程序被处理器执行时实现如权利要求1至6任一项所述的癌症风险预测方法。

## 一种癌症风险预测方法、系统、设备及介质

### 技术领域

[0001] 本发明涉及医疗数据管理技术领域,尤其涉及一种癌症风险预测方法、系统、计算机设备及计算机可读存储介质。

### 背景技术

[0002] 癌细胞代谢的一个共同特征是能够从经常缺乏营养的环境中获取必要的营养,并利用这些营养来维持生存能力并建立新的生物质。伴随癌症相关代谢重编程的细胞内和细胞外代谢物的改变对基因表达、细胞分化和肿瘤微环境具有深远的影响。从代谢物角度可将癌症相关代谢变化总结为六个标志:(1)对葡萄糖和氨基酸的摄取失控;(2)使用投机模式获取营养,(3)使用糖酵解/TCA循环中间体进行生物合成和生产NADPH;(4)增加对氮的需求;(5)代谢物驱动的基因调控的改变;(6)与微环境的代谢相互作用。虽然很少有肿瘤同时具备所有六个标志,但大多数显示其中的几个特征,所以单个肿瘤表现出的特定特征可能最终有助于更好的肿瘤分类并有助于指导治疗。

[0003] 虽然对肿瘤特征性代谢改变的首次观察是在近一个世纪前首次进行的,但在过去十年中,癌症代谢领域已成为重新引起人们关注的话题。借助新的生化和分子生物学工具,对癌细胞代谢的研究已经扩展了我们对肿瘤发生各个阶段肿瘤相关代谢改变的机制和功能后果的理解。虽然基因水平的改变如基因敲除(Deletion)或敲入(Insertion),导致基因拷贝数改变(Gene copy number alteration GCNA),基因突变(Mutation),非编码RNA及转录后修饰如甲基化(Methylation),乙酰化(Acetylation)等标志物已经被研究得非常广泛,但这些基因水平,蛋白水平及转录,转录后水平的调节,必然会以某种形式以代谢物的形式展示出来。因此,代谢物才是所有基因表达后的共同汇聚之处。

[0004] 支持哺乳动物细胞生存和生物合成的两种主要营养素是葡萄糖和谷氨酰胺。通过葡萄糖和谷氨酰胺的分解代谢,细胞维持多种碳中间体,被用作组装各种大分子的结构基础,同时它们介导电子传递链以促进ATP的生成,或以相关辅因子NADPH的形式,它为各种生物合成反应提供还原能力以维持细胞氧化还原能力。与非增殖的正常组织相比,肿瘤对葡萄糖的消耗显著增加,这种现象在90多年前由德国生理学家Otto Warburg首次描述。在各种肿瘤环境中得到相继证实,并显示出与肿瘤预后不良相关。基于正电子发射断层扫描(PET)的放射性氟标记葡萄糖类似物<sup>18</sup>F-氟脱氧葡萄糖(<sup>18</sup>F-FDG)摄取的成像已取得成功完全用于临床肿瘤诊断和分期,以及监测对治疗的反应。

[0005] 谷氨酰胺是第二个主要的支持肿瘤生长的代谢物。它不仅为许多不同的含氮化合物的从头生物合成提供碳,还提供还原的氮。因此,谷氨酰胺提供嘌呤和嘧啶核苷酸、葡萄糖胺6-磷酸和非必需氨基酸的生物合成所需的氮。据报道,谷氨酰胺也在必需氨基酸的吸收中发挥作用。虽然非必需氨基酸可以由哺乳动物细胞从头产生,但必需氨基酸必须从外部来源获得。有趣的是,必需氨基酸亮氨酸通过质膜定位的中性氨基酸逆向转运蛋白(LAT1)的转运进入细胞与谷氨酰胺的同时胞外运输有关。以这种方式,细胞内谷氨酰胺可以促进多种LAT1底物的输入,包括亮氨酸、异亮氨酸、缬氨酸、蛋氨酸、酪氨酸、色氨酸和苯

丙氨酸。

[0006] 有综述文章对31篇包括血清中靶向和非靶向代谢物研究的综述显示：多种代谢物，包括葡萄糖、果糖、半乳糖、甘露糖、丙二酸和肌苷，胆固醇和花生四烯酸，糖基化（免疫球蛋白G[IgG]Fc-糖基化），胆碱、胆碱衍生物、乳酸、脂肪酸等均是甲状腺癌发生过程中显著改变，此中柠檬酸盐被认为是第一个最重要的生物标志物，其次是乳酸。

[0007] 由此可见，代谢物在肿瘤改变方面研究由来已久，其与肿瘤的相关性具有非常广泛和扎实的研究基础。只是目前并未见到基于上述代谢物预测癌症肿瘤的方法。

[0008] 现有技术存在以下缺陷：

[0009] 1) 检测过程复杂：目前基于蛋白的标志物基本上都要靠酶联免疫反应，最后显色来判断物质的浓度；

[0010] 2) 检测成分单一：目前临床血清或血浆检测不同指标经常要抽多管血，分别检测不同的肿瘤标志物；

[0011] 3) 检测时间长：不同的血清标志物检测时间从15分钟到12小时不等，甚至更长时间；

[0012] 4) 检测费用高：每种费用几十到上百元不等，多项加起来费用就聚少成多；

[0013] 5) 污染环境：检测过程会产生大量的试剂盒的废弃物，即医疗垃圾，处理费时，且容易污染环境。

[0014] 目前，针对肿瘤的血清学标志物检测相关技术中存在的检测过程复杂、检测成分单一、检测项目多、检测费用高、检测时间长、以及污染环境的问题，目前尚无有效的整体解决方案。

## 发明内容

[0015] 本申请的目的是针对现有技术中的不足，基于血清中的代谢物检测，提供一种癌症风险预测方法、系统、设备及存储介质，以至少解决相关技术中存在的检测过程复杂、检测成分单一、检测项目多、检测费用高、检测时间长以及污染环境的问题。

[0016] 为实现上述目的，本申请采取的技术方案是：

[0017] 第一方面，本发明提供一种癌症风险预测方法，包括：

[0018] 获取血浆样品的拉曼检测数据；

[0019] 将所述拉曼检测数据输入预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型，以获取癌症风险预测结果；

[0020] 其中，癌症风险预测结果的准确率大于91.48%、敏感度大于88.70%、特异性大于95.98%。

[0021] 在其中的一些实施例中，所述癌症包括肾癌、胃癌、宫颈癌、直肠癌、前列腺癌、肺癌、卵巢癌、乳腺癌。

[0022] 在其中的一些实施例中，所述拉曼检测数据包括通用类拉曼检测数据，其至少包括：

[0023]

	ID	Name
1	M365T412_2	Lactose
2	M487T404	Blood group b trisaccharide

3	M203T272	DL-tryptophan
4	M173T309	Gly-Val
5	M437T96	Hc toxin
6	M291T33	Dl-norleucine methyl ester
7	M697T235	Izenamide c

[0024] 在其中的一些实施例中,所述拉曼检测数据还包括:

[0025] 肾癌拉曼检测数据,其至少包括:

[0026]

	ID	Name
1	M751T38	1-myristoyl-2-palmitoyl-sn-glycero-3-phosphocholine
2	M141T292	Kojic acid
3	M865T137	Pc (16:1e/17-hdohe)
4	M791T137	Pe (18:1e/10-hdohe)
5	M671T141_2	1-Hexadecanoyl-2- (9Z,12Z-octadecadienoyl) -sn-glycero-3-phosphoric acid
6	M309T25_2	Mestranol
7	M841T140	Pc (16:0e/8-hepe)
8	M580T180	1-behenoyl-2-hydroxy-sn-glycero-3-phosphocholine
9	M538T185	1-Palmitoyllysophosphatidylcholine
10	M193T82_1	Trans-3'-hydroxycotinine

[0027] 在其中的一些实施例中,所述拉曼检测数据还包括:

[0028] 胃癌拉曼检测数据,其至少包括:

[0029]

	ID	Name
1	M747T38	(2-aminoethoxy)[2-[docosa-4.7.10.13.16.19-hexaenoyloxy]-3-[hexadec-1-en-1-yloxy]propoxy]phosphinic acid
2	M147T122	3-hydroxyglutaric acid
3	M103T300	D(-)-beta-hydroxy butyric acid
4	M119T122	Tartronate
5	M283T37	Octadecanoic acid
6	M293T115	[6]-gingerol
7	M141T34_1	Ethyl imidazole-2-carboxylate
8	M751T135	1-palmitoyl-2-thiopalmityl phosphatidylcholine

[0030] 在其中的一些实施例中,所述拉曼检测数据还包括:

[0031] 宫颈癌拉曼检测数据,其至少包括:

	ID	Name	
[0032]	1	M113T186	Glutaric acid
	2	M716T34	1-Stearoyl-sn-glycerol
	3	M481T211	11-(6-{{2-(3,4-dimethoxyphenyl)ethyl}amino}-4-chloro-1,3,5-triazin-2-yl)-7,11-diazatricyclo[7.3.1.0<2,7>]trideca-2,4-dien-6-one
	4	M163T32	2-tert-butyl-6-methylphenol
	5	M212T29	Indoxyl sulfate
	6	M130T90	Paxilline
	7	M158T433	Trans-4-(aminomethyl)cyclohexanecarboxylic acid
	8	M100T72	Cyclohexylamine
	9	M104T411	Choline

[0033] 在其中的一些实施例中,所述拉曼检测数据还包括:

[0034] 直肠癌拉曼检测数据,其至少包括:

	ID	Name	
[0035]	1	M179T369	D-(+)-mannose
	2	M341T370	Sucrose
	3	M665T501	Stachyose
	4	M179T453	D-psicose
	5	M281T75_3	Oleic acid
	6	M297T102	Enterolactone
	7	M321T72	Deoxythymidine 5'-phosphate (dTMP)
	8	M858T196	Pi 36:4
	9	M361T157_2	Phe-Glu
	10	M511T461	3'-fucosyllactose
	11	M107T82	(-)-perillyl alcohol

[0036]	12	M74T71	N,n-dimethylformamide
	13	M171T188	Glycyl-l-4-hydroxyproline

[0037] 在其中的一些实施例中,所述拉曼检测数据还包括:

[0038] 前列腺癌拉曼检测数据,其至少包括:

	ID	Name	
	1	M173T442	Cis-aconitate
	2	M923T492	Disialyllactose
	3	M283T29_2	3,4'-dimethoxy-2'-hydroxychalcone
	4	M151T218	Xanthine
	5	M131T474	Asparagine
	6	M791T38	Pe(18:1e/7-hdohe)
	7	M341T489	Cellobiose
	8	M116T312	Guanidoacetic acid
	9	M118T412	Threonine
	10	M173T505	Isocitrate
	11	M473T26	Amastatin
	12	M186T237_1	2-hydroxy-6-methylquinoline-3-carbaldehyde
	13	M167T333	Uric acid
	14	M180T330	Tyrosine
	15	M179T309_2	D-fructose
	16	M277T36	C14-sas (tentative)
[0039]	17	M311T28_2	Thymol-beta-d-glucoside
	18	M325T27	Hydroquinidine
	19	M223T216	D-Quinovose
	20	M339T27_2	Gly-His-Lys
	21	M107T25	3-hydroxyphenylacetic acid
	22	M130T279_2	Guanidinopropionic acid
	23	M135T33_1	Diglyme
	24	M167T57	1-methyl-3-isobutylxanthine
	25	M139T482	Trans-aconitic acid
	26	M102T87	4-hydroxy-l-isoleucine
	27	M429T299_2	Carfentrazone-ethyl
	28	M230T574	Lauryldimethylamine oxide
	29	M869T102	Solamargine
	30	M230T233	Arg-Gln-Arg
	31	M212T458	Phosphocreatine
	32	M229T447	Pro-hyp
	33	M118T278	Betaine
	34	M167T116_1	Spectinomycin
	35	M88T321	N-methylpropionamide
	36	M295T483_2	Arg-Cys
[0040]	37	M317T503	.gamma.-L-Glu-.epsilon.-L-Lys
	38	M116T32	Sulfallate
	39	M138T222	6-aminopyridine-3-carboxamide

[0041] 在其中的一些实施例中,所述拉曼检测数据还包括:

[0042] 肺癌拉曼检测数据,其至少包括:

[0043]

	ID	Name
1	M113T186	Glutaric acid
2	M716T34	1-Stearoyl-sn-glycerol
3	M481T211	11-(6-{{2-(3,4-dimethoxyphenyl)ethyl}amino}-4-chloro-1,3,5-triazin-2-yl)-7,11-diazatricyclo[7.3.1.0<2,7>]trideca-2,4-dien-6-one
4	M163T32	2-tert-butyl-6-methylphenol
5	M212T29	Indoxyl sulfate
6	M130T90	Paxilline
7	M158T433	Trans-4-(aminomethyl)cyclohexanecarboxylic acid
8	M100T72	Cyclohexylamine
9	M104T411	Choline

[0044] 在其中的一些实施例中,所述拉曼检测数据还包括:

[0045] 卵巢癌拉曼检测数据,其至少包括:

[0046]

	ID	Name
1	M187T44	1-hydroxy-2-naphthoic acid
2	M355T37_1	Fumarprotocetraric acid
3	M159T119_2	3-hydroxyoctanoic acid
4	M498T44	Taurochenodeoxycholate
5	M182T40	4-pyridoxic acid
6	M300T204	N-Acetyl-D-Glucosamine 6-Phosphate
7	M204T47	N,n'-diacetylchitobiose
8	M191T255	5-methyl-5-phenylhydantoin
9	M174T570	Ala-Thr-Arg
10	M130T582	D-Pipecolinic acid
11	M196T258	1-deoxy-1-(methylamino)-d-galactitol
12	M283T72	Hexaethylene glycol

[0047] 在其中的一些实施例中,所述拉曼检测数据还包括:

[0048] 乳腺癌拉曼检测数据,其至少包括:

[0049]

	ID	Name
1	M129T86	5-methyl-5,6-dihydrouracil
2	M144T245	L-glutamic acid, dimethyl ester
3	M188T328_2	Leu-Ser-Arg
4	M210T532	Pro-Phe-Arg
5	M211T74_1	Cyclo(proline-leucine)
6	M213T166	m-Chlorohippuric acid
7	M282T33	Oleamide
8	M283T51	Swietenocoumarin b
9	M283T70	Stearic acid
10	M343T81	2-[1-hydroxy-1-(4-methoxyphenyl)propan-2-yl]oxy-6-(hydroxymethyl)oxane-3,4,5-triol

[0051] 在其中的一些实施例中,在获取血浆样品的拉曼检测数据之前,所述方法还包括:

[0052] 获取血液样品;

- [0053] 对所述血液样品进行离心处理,以获取血浆超滤样品;
- [0054] 对所述初始血浆样品进行基于膜过滤的离心处理,以获取血浆样品;
- [0055] 其中,对所述初始血浆样品进行基于膜过滤的离心处理的时间小于等于5min,所述初始血浆样品的体积小于等于450 $\mu$ l,所述血浆样品的体积小于等于5 $\mu$ l。
- [0056] 在其中的一些实施例中,对所述血液样品进行离心处理,以获取初始血浆样品的工作参数为2000~3000rpm、3~5min。
- [0057] 在其中的一些实施例中,对所述血液样品进行离心处理,以获取初始血浆样品的工作参数为3000rpm、2min。
- [0058] 在其中的一些实施例中,对所述初始血浆样品基于膜过滤的离心处理,以获取血浆样品的工作参数为10000~15000rpm、3~6min。
- [0059] 在其中的一些实施例中,对所述初始血浆样品基于膜过滤的离心处理,以获取血浆超滤后样品的工作参数为12000rpm、3min。
- [0060] 在其中的一些实施例中,所述血浆样品的体积为2~5 $\mu$ l。
- [0061] 在其中的一些实施例中,所述膜处理为超滤处理。
- [0062] 在其中的一些实施例中,在获取血浆样品之后,所述方法还包括:
- [0063] 将所述血浆样品放置于镀铝膜的载玻片进行干燥。
- [0064] 在其中的一些实施例中,所述载玻片预先放置于37 $^{\circ}$ C的环境下。
- [0065] 在其中的一些实施例中,将所述血浆样品放置于镀铝膜的载玻片进行干燥的时间为30s。
- [0066] 在其中的一些实施例中,将所述血浆样品放置于镀铝膜的载玻片进行干燥的温度为37 $^{\circ}$ C。
- [0067] 在其中的一些实施例中,获取血浆样品的拉曼检测数据的拉曼检测参数为:
- [0068] 激发波长:532nm;
- [0069] 功率:10-14mw;
- [0070] 光栅:1200g/mm;
- [0071] 单谱积分时间:16-20s;
- [0072] 物镜:100x/0.9。
- [0073] 在其中的一些实施例中,获取血浆样品的拉曼检测数据的拉曼检测参数还包括:
- [0074] 每个所述血浆样品采集5-8张图谱。
- [0075] 在其中的一些实施例中,将所述拉曼检测数据输入预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型,以获取癌症风险预测结果包括:
- [0076] 处理所述拉曼检测数据,将所述拉曼检测数据映射为1024维的初始特征向量;
- [0077] 将1024维的初始特征向量输入至预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型,以获取4维的最终特征向量;
- [0078] 将4维的最终特征向量输入至分类函数进行处理,以获取癌症风险预测结果。
- [0079] 在其中的一些实施例中,所述分类函数为softmax分类函数。
- [0080] 在其中的一些实施例中,癌症风险预测模型的训练方法包括:
- [0081] 构建癌症风险预测模型,其中,所述癌症风险预测模型包括输入层、隐藏层和输出层,所述输入层的输入为1024、输出为512,所述隐藏层的输入为512、输出为128,所述输出

层的输入为128、输出为4；

[0082] 根据拉曼检测数据构建训练集和测试集,其中,所述训练集与所述测试集的比例为3:1;

[0083] 将所述训练集输入至所述癌症风险预测模型进行训练,以获取输出结果;

[0084] 将所述输出结果输入至分类函数进行处理,以获取癌症风险预测结果;

[0085] 对所述癌症风险预测结果进行测试,以检测训练的有效性,并根据有效性检测结果,对所述癌症风险预测模型进行迭代,直至训练完成。

[0086] 在其中的一些实施例中,对所述癌症风险预测结果进行测试为:

[0087] 每训练10轮进行一次测试。

[0088] 在其中的一些实施例中,癌症风险预测模型的训练方法还包括:

[0089] 使用交叉熵目标函数计算损失数值;

[0090] 使用随机梯度下降法和反向传播算法对所述癌症风险预测模型的权重进行优化。

[0091] 第二方面,本发明提供一种癌症风险预测系统,包括:

[0092] 血浆样品获取装置,用于对血液样品进行离心处理以获取初始血浆样品,以及对所述初始血浆样品进行基于膜过滤的离心处理以获取血浆样品;其中,对所述初始血浆样品进行基于膜过滤的离心处理的时间小于等于8min;

[0093] 拉曼检测装置,用于对所述血浆样品进行拉曼检测,以获取拉曼检测数据;

[0094] 癌症风险预测装置,用于获取血浆样品的拉曼检测数据,将所述拉曼检测数据输入预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型,以获取癌症风险预测结果,如果根据不同的肿瘤结合临床信息,包括患者病史及临床血清学指标,设置不同的权重,将有助于更精准地诊断不同的肿瘤;

[0095] 其中,所述癌症风险预测结果的准确率大于91.48%、敏感度大于88.70%、特异性大于95.98%。

[0096] 在其中的一些实施例中,所述血浆样品获取装置包括:

[0097] 抗凝取血单元,用于放置血液样品;

[0098] 离心单元,用于对所述抗凝取血单元的所述血液样品进行离心处理以获取初始血浆样品,以及对所述初始血浆样品进行离心处理以获取血浆样品;

[0099] 膜过滤单元,用于在所述离心单元对所述初始血浆样品进行离心处理的同时,对所述初始血浆样品进行膜处理,以获取所述血浆样品。

[0100] 在其中的一些实施例中,所述抗凝取血单元为EDTA-K2抗凝取血单元。

[0101] 在其中的一些实施例中,所述膜过滤单元为超滤离心管。

[0102] 在其中的一些实施例中,还包括:

[0103] 在其中的一些实施例中,所述拉曼检测装置采用激光共聚焦拉曼技术、表面增强拉曼散射技术。

[0104] 第三方面,本发明提供一种计算机设备,包括存储器、处理器以及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如上所述的癌症风险预测方法。

[0105] 第四方面,本发明提供一种计算机可读存储介质,其上存储有计算机程序,该程序被处理器执行时实现如上所述的癌症风险预测方法。

[0106] 相比于相关技术,本申请实施例提供的一种癌症风险预测方法、系统、设备及存储介质,利用拉曼检测与基于BP神经网络的人工智能(AI)癌症风险预测模型的结合,大幅度缩短检测时间,能够在15分钟左右获取癌症风险预测结果,可以根据该预测结果决定是否进行后续精确检测;本检测方法简单易操作,一次测试即可同时检测多种物质,癌症风险预测结果准确率高;检测费用低廉,不会产生大量医疗废弃物,避免污染环境;该预测方法特异性高、敏感性高、准确率高;拉曼检测可以实现小分子代谢物快速的大规模检测,具有高通量、高准确性的优势,如果能够结合临床血清学和影像检测等信息,并根据不同癌症指标或影像学变化在癌症发生或转移过程中的权重进行调整,则基于此法在肿瘤的精准诊断和(或)预测转移复发中将具有非常广阔的应用前景,具备大规模推广应用的优点。

## 附图说明

[0107] 此处所说明的附图用来提供对本申请的进一步理解,构成本申请的一部分,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。在附图中:

[0108] 图1是根据本申请实施例的癌症风险预测方法的流程图(一);

[0109] 图2是根据本申请实施例的癌症风险预测方法的流程图(二);

[0110] 图3是根据本申请实施例的癌症风险预测方法的流程图(三);

[0111] 图4是根据本申请实施例的癌症风险预测模型的训练方法的流程图;

[0112] 图5是根据本申请实施例的癌症风险预测系统的框架图;

[0113] 图6是根据本申请实施例的血浆样品获取装置的框架图;

[0114] 图7是七种泛癌血清代谢物的AUC曲线。

[0115] 其中的附图标记为:

[0116] 500、癌症风险预测系统;

[0117] 510、癌症风险预测装置;

[0118] 520、血浆样品获取装置;

[0119] 521、抗凝取血单元;

[0120] 522、离心单元;

[0121] 523、膜过滤单元;

[0122] 530、拉曼检测装置。

## 具体实施方式

[0123] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行描述和说明。应当理解,此处所描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。基于本申请提供的实施例,本领域普通技术人员在没有作出创造性劳动的前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0124] 显而易见地,下面描述中的附图仅仅是本申请的一些示例或实施例,对于本领域的普通技术人员而言,在不付出创造性劳动的前提下,还可以根据这些附图将本申请应用于其他类似情景。此外,还可以理解的是,虽然这种开发过程中所作出的努力可能是复杂并且冗长的,然而对于与本申请公开的内容相关的本领域的普通技术人员而言,在本申请揭露的技术内容的基础上进行的一些设计,制造或者生产等变更只是常规的技术手段,不应

当理解为本申请公开的内容不充分。

[0125] 在本申请中提及“实施例”意味着,结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例,也不是与其它实施例互斥的独立的或备选的实施例。本领域普通技术人员显式地和隐式地理解的是,本申请所描述的实施例在不冲突的情况下,可以与其它实施例相结合。

[0126] 除非另作定义,本申请所涉及的技术术语或者科学术语应当为本申请所属技术领域内具有一般技能的人士所理解的通常意义。本申请所涉及的“一”、“一个”、“一种”、“该”等类似词语并不表示数量限制,可表示单数或复数。本申请所涉及的术语“包括”、“包含”、“具有”以及它们任何变形,意图在于覆盖不排他的包含;例如包含了一系列步骤或单元(单元)的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元,而是可以还包括没有列出的步骤或单元,或可以还包括对于这些过程、方法、产品或设备固有的其它步骤或单元。本申请所涉及的“连接”、“相连”、“耦接”等类似的词语并非限于物理的或者机械的连接,而是可以包括电气的连接,不管是直接的还是间接的。本申请所涉及的“多个”是指两个或两个以上。“和/或”描述关联对象的关联关系,表示可以存在三种关系,例如,“A和/或B”可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。本申请所涉及的术语“第一”、“第二”、“第三”等仅仅是区别类似的对象,不代表针对对象的特定排序。

[0127] 实施例1

[0128] 图1是根据本发明实施例的癌症风险预测方法的流程图(一)。如图1所示,一种癌症风险预测方法,包括以下步骤:

[0129] S102、获取血浆样品的拉曼检测数据;

[0130] S104、将拉曼检测数据输入预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型,以获取癌症风险预测结果。

[0131] 其中,所述癌症风险预测结果的准确率大于91.48%、敏感度大于88.70%、特异性大于95.98%。

[0132] 在步骤S102中,拉曼检测数据的拉曼检测参数为:

[0133] 激发波长:532nm;

[0134] 功率:10-14mw;

[0135] 光栅:1200g/mm;

[0136] 单谱积分时间:16-20s;

[0137] 物镜:100x/0.9。

[0138] 其中,结合BP神经网络算法的拉曼检测数据的时间为1~2min。

[0139] 进一步地,获取血浆样品的拉曼检测数据的拉曼检测参数还包括:

[0140] 每个血浆样品采集5-8张图谱。

[0141] 在步骤S104中,获取癌症风险预测结果的时间为0~2min。

[0142] 优选地,获取癌症风险预测结果的时间为30s-1min。

[0143] 在其中的一些实施例中,拉曼检测数据包括通用类拉曼检测数据,其至少包括:

	ID	Name	AUC	95%CI	
[0144]	1	M173T309	Gly-Val	0.9637	0.9303 to 0.9971
	2	M203T272	DL-tryptophan	0.9357	0.8712 to 1.000
	3	M365T412_2	Lactose	0.9077	0.8514 to 0.9640
	4	M437T96	Hc toxin	0.9018	0.8443 to 0.9593
	5	M697T235	Izenamide c	0.8988	0.8389 to 0.9588
	6	M291T33	DI-norleucine methyl ester	0.8476	0.7732 to 0.9221
	7	M487T404	Blood group b trisaccharide	0.6363	0.5373 to 0.7353

[0145] 通过上述7种代谢物,可以预测是否有癌症发生风险。

[0146] 在其中的一些实施例中,拉曼检测数据还包括:

[0147] 肾癌拉曼检测数据,其至少包括:

	ID	Name	
[0148]	1	M751T38	1-myristoyl-2-palmitoyl-sn-glycero-3-phosphocholine
	2	M141T292	Kojic acid
	3	M865T137	Pc(16:1e/17-hdohe)
	4	M791T137	Pc(18:1e/10-hdohe)
	5	M671T141_2	1-Hexadecanoyl-2-(9Z,12Z-octadecadienoyl)-sn-glycero-3-phosphoric acid
	6	M309T25_2	Mestranol
	7	M841T140	Pc(16:0e/8-hepe)
	8	M580T180	1-behenoyl-2-hydroxy-sn-glycero-3-phosphocholine
	9	M538T185	1-Palmitoyllysophosphatidylcholine

[0149] 10 M193T82\_1 Trans-3'-hydroxycotinine

[0150] 通过上述10种代谢物,可以辅助预测是否有肾癌发生风险。

[0151] 在其中的一些实施例中,拉曼检测数据还包括:

[0152] 胃癌拉曼检测数据,其至少包括:

	ID	Name	
[0153]	1	M747T38	(2-aminoethoxy)[2-[docosa-4.7.10.13.16.19-hexaenoyloxy]-3-[hexadec-1-en-1-yloxy]propoxy]phosphinic acid
	2	M147T122	3-hydroxyglutaric acid
	3	M103T300	D(-)-beta-hydroxy butyric acid
	4	M119T122	Tartronate
	5	M283T37	Octadecanoic acid
	6	M293T115	[6]-gingerol
	7	M141T34_1	Ethyl imidazole-2-carboxylate
	8	M751T135	1-palmitoyl-2-thiopalmitoyl phosphatidylcholine

[0154] 通过上述8种代谢物,可以辅助预测是否有胃癌发生风险。

[0155] 在其中的一些实施例中,拉曼检测数据还包括:

[0156] 宫颈癌拉曼检测数据,其至少包括:

	ID	Name
	M113T186	Glutaric acid
	M716T34	1-Stearoyl-sn-glycerol
	M481T211	11-(6-([2-(3,4-dimethoxyphenyl)ethyl]amino)-4-chloro-1,3,5-triazin-2-yl)-7,11-diazatricyclo[7.3.1.0<2,7>]trideca-2,4-dien-6-one
[0157]	M163T32	2-tert-butyl-6-methylphenol
	M212T29	Indoxyl sulfate
	M130T90	Paxilline
	M158T433	Trans-4-(aminomethyl)cyclohexanecarboxylic acid
	M100T72	Cyclohexylamine
	M104T411	Choline

[0158] 通过上述9种代谢物,可以辅助预测是否有宫颈癌发生风险。

[0159] 在其中的一些实施例中,拉曼检测数据还包括:

[0160] 直肠癌拉曼检测数据,其至少包括:

	ID	Name
[0161]	M179T369	D-(+)-mannose
	M341T370	Sucrose

	M665T501	Stachyose
	M179T453	D-psicose
	M281T75_3	Oleic acid
	M297T102	Enterolactone
	M321T72	Deoxythymidine 5'-phosphate (dTMP)
[0162]	M858T196	Pi 36:4
	M361T157_2	Phe-Glu
	M511T461	3'-fucosyllactose
	M107T82	(-)-perillyl alcohol
	M74T71	N,n-dimethylformamide
	M171T188	Glycyl-l-4-hydroxyproline

[0163] 通过上述13种代谢物,可以辅助预测是否有直肠癌发生风险。

[0164] 在其中的一些实施例中,拉曼检测数据还包括:

[0165] 前列腺癌拉曼检测数据,其至少包括:

	ID	Name	
	1	M173T442	Cis-aconitate
	2	M923T492	Disialyllactose
	3	M283T29_2	3,4'-dimethoxy-2'-hydroxychalcone
	4	M151T218	Xanthine
	5	M131T474	Asparagine
	6	M791T38	Pe(18:1e/7-hdohe)
	7	M341T489	Cellobiose
	8	M116T312	Guanidoacetic acid
	9	M118T412	Threonine
	10	M173T505	Isocitrate
	11	M473T26	Amastatin
[0166]	12	M186T237_1	2-hydroxy-6-methylquinoline-3-carbaldehyde
	13	M167T333	Uric acid
	14	M180T330	Tyrosine
	15	M179T309_2	D-fructose
	16	M277T36	C14-sas (tentative)
	17	M311T28_2	Thymol-beta-d-glucoside
	18	M325T27	Hydroquinidine
	19	M223T216	D-Quinovose
	20	M339T27_2	Gly-His-Lys
	21	M107T25	3-hydroxyphenylacetic acid
	22	M130T279_2	Guanidinopropionic acid
	23	M135T33_1	Diglyme
	24	M167T57	1-methyl-3-isobutylxanthine
	25	M139T482	Trans-aconitic acid
	26	M102T87	4-hydroxy-l-isoleucine
	27	M429T299_2	Carfentrazone-ethyl
	28	M230T574	Lauryldimethylamine oxide
	29	M869T102	Solamargine
	30	M230T233	Arg-Gln-Arg
	31	M212T458	Phosphocreatine
[0167]	32	M229T447	Pro-hyp
	33	M118T278	Betaine
	34	M167T116_1	Spectinomycin
	35	M88T321	N-methylpropionamide
	36	M295T483_2	Arg-Cys
	37	M317T503	.gamma.-L-Glu-.epsilon.-L-Lys
	38	M116T32	Sulfallate
	39	M138T222	6-aminopyridine-3-carboxamide

[0168] 通过上述39种代谢物,可以辅助预测是否有前列腺癌发生风险。

[0169] 在其中的一些实施例中,拉曼检测数据还包括:

[0170] 肺癌拉曼检测数据,其至少包括:

	ID	Name
	M113T186	Glutaric acid
	M716T34	1-Stearoyl-sn-glycerol
	M481T211	11-(6-{{2-(3,4-dimethoxyphenyl)ethyl}amino}-4-chloro-1,3,5-triazin-2-yl)-7,11-diazatricyclo[7.3.1.0<2,7>]trideca-2,4-dien-6-one
[0171]	M163T32	2-tert-butyl-6-methylphenol
	M212T29	Indoxyl sulfate
	M130T90	Paxilline
	M158T433	Trans-4-(aminomethyl)cyclohexanecarboxylic acid
	M100T72	Cyclohexylamine
	M104T411	Choline

[0172] 通过上述9种代谢物,可以辅助预测是否有肺癌发生风险。

[0173] 在其中的一些实施例中,拉曼检测数据还包括:

[0174] 卵巢癌拉曼检测数据,其至少包括:

	ID	Name
[0175]	M187T44	1-hydroxy-2-naphthoic acid
	M355T37_1	Fumarprotocetraric acid
	M159T119_2	3-hydroxyoctanoic acid

	M498T44	Taurochenodeoxycholate
	M182T40	4-pyridoxic acid
	M300T204	N-Acetyl-D-Glucosamine 6-Phosphate
	M204T47	N,n'-diacetylchitobiose
[0176]	M191T255	5-methyl-5-phenylhydantoin
	M174T570	Ala-Thr-Arg
	M130T582	D-Pipecolinic acid
	M196T258	1-deoxy-1-(methylamino)-d-galactitol
	M283T72	Hexaethylene glycol

[0177] 通过上述12种代谢物,可以辅助预测是否有卵巢癌发生风险。

[0178] 在其中的一些实施例中,拉曼检测数据还包括:

[0179] 乳腺癌拉曼检测数据,其至少包括:

	ID	Name
	M129T86	5-methyl-5,6-dihydrouracil
	M144T245	L-glutamic acid, dimethyl ester
	M188T328_2	Leu-Ser-Arg
	M210T532	Pro-Phe-Arg
	M211T74_1	Cyclo(proline-leucine)
[0180]	M213T166	m-Chlorohippuric acid
	M282T33	Oleamide
	M283T51	Swietenocoumarin b
	M283T70	Stearic acid
	M343T81	2-[1-hydroxy-1-(4-methoxyphenyl)propan-2-yl]oxy-6-(hydroxymethyl)oxane-3,4,5-triol

[0181] 通过上述10种代谢物,可以辅助预测是否有乳腺癌发生风险。

[0182] 图2是根据本发明实施例的癌症风险预测方法的流程图(二)。如图2所示,在获取血浆样品的拉曼检测数据之前,方法还包括以下步骤:

- [0183] 步骤S202、获取血液样品；
- [0184] 步骤S204、对血液样品进行离心处理，以获取初始血浆样品；
- [0185] 步骤S206、对初始血浆样品进行基于膜过滤的离心处理，以获取血浆样品。
- [0186] 其中，对初始血浆样品进行基于膜过滤的离心处理的时间小于等于8min，初始血浆样品的体积小于等于450 $\mu$ l，血浆样品的体积小于等于5 $\mu$ l。
- [0187] 在步骤S202中，获取血液样品为将采集的血液放置入抗凝取血管。
- [0188] 在步骤S204中，对血液样品进行离心处理，以获取初始血浆样品的工作参数为2000~5000rpm、3~5min。
- [0189] 优选地，对血液样品进行离心处理，以获取初始血浆样品的工作参数为3000rpm、2min。
- [0190] 在步骤S206中，将初始血浆样品放置入超滤离心管，膜处理为超滤处理。
- [0191] 在步骤S206中，对初始血浆样品同时进行膜处理和离心处理，以获取血浆样品的工作参数为10000~15000rpm、3~6min。
- [0192] 优选地，对初始血浆样品同时进行膜处理和离心处理，以获取血浆样品的工作参数为12000rpm、5min。
- [0193] 优选地，血浆样品的体积为2~5 $\mu$ l。
- [0194] 优选地，血浆样品的体积为2 $\mu$ l。
- [0195] 进一步地，在步骤S206之后，方法还包括以下步骤：
- [0196] 步骤S208、将血浆样品放置于镀铝膜的载玻片进行干燥。
- [0197] 其中，将血浆样品放置于镀铝膜的载玻片进行干燥的工作参数为37 $^{\circ}$ C。
- [0198] 其中，干燥时间为30s~2min。
- [0199] 通过步骤S202~S206，可以在8min内（一般为7~12min，优选为7~8min）获取血浆样品以进行后续检测，大大缩短获取样品的时间。
- [0200] 图3是根据本发明实施例的癌症风险预测方法的流程图（三）。如图3所示，将拉曼检测数据输入预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型，以获取癌症风险预测结果包括：
- [0201] 步骤S302、处理拉曼检测数据，将拉曼检测数据映射为1024维的初始特征向量；
- [0202] 步骤S304、将1024维的初始特征向量输入至预先训练好的基于BP神经网络人工智能算法的癌症风险预测模型，以获取4维的最终特征向量；
- [0203] 步骤S306、将4维的最终特征向量输入至分类函数进行处理，以获取癌症风险预测结果。
- [0204] 在步骤S302中，拉曼检测数据包括波数和强度，其中，波数的范围为279-2186 $\text{cm}^{-1}$ ，一共1024个点。
- [0205] 在步骤S306中，分类函数为softmax分类函数。
- [0206] 图4是根据本发明实施例的癌症风险预测模型的训练方法的流程图。如图4所示，癌症风险预测模型的训练方法包括：
- [0207] 步骤S402、构建癌症风险预测模型，其中，癌症风险预测模型包括输入层、隐藏层和输出层，输入层的输入为1024、输出为512，隐藏层的输入为512、输出为128，输出层的输入为128、输出为4；

- [0208] 步骤S404、根据拉曼检测数据构建训练集和测试集,其中,训练集与测试集的比例为3:1;
- [0209] 步骤S406、将训练集输入至癌症风险预测模型进行训练,以获取输出结果;
- [0210] 步骤S408、将输出结果输入至分类函数进行处理,以获取癌症风险预测结果;
- [0211] 步骤S410、对癌症风险预测结果进行测试,以检测训练的有效性,并根据有效性检测结果,对癌症风险预测模型进行迭代,直至训练完成。
- [0212] 在步骤S408中,分类函数为softmax分类函数。
- [0213] 在步骤S410中,每训练10轮进行一次测试。
- [0214] 在步骤S410中,使用测试集对癌症风险预测模型进行测试,以检测训练的有效性。
- [0215] 进一步地,癌症风险预测模型的训练方法还包括:
- [0216] 步骤S412、使用交叉熵目标函数计算损失数值;
- [0217] 步骤S414、使用随机梯度下降法和反向传播算法对癌症风险预测模型的权重进行优化。
- [0218] 通过步骤S402~步骤S414,可以训练得到所述癌症风险预测结果的准确率大于91.48%、敏感度大于88.70%、特异性大于95.98%的癌症风险预测模型,在减少预测时间的情况下,大幅度提高准确率、敏感度和特异性。
- [0219] 图5是根据本发明实施例的癌症风险预测系统的框架图。如图5所示,癌症风险预测系统500包括癌症风险预测装置510,用于获取血浆样品的拉曼检测数据,将拉曼检测数据输入预先训练好的癌症风险预测模型,以获取癌症风险预测结果。
- [0220] 在其中的一些实施例中,癌症风险预测装置510包括但不限于移动终端、云端服务器、本地服务器、计算机、笔记本等。
- [0221] 进一步地,癌症风险预测系统500还包括血浆样品获取装置520和拉曼检测装置530。其中,血浆样品获取装置520用于对血液样品进行离心处理以获取初始血浆样品,以及对初始血浆样品同时进行膜处理和离心处理以获取血浆样品;拉曼检测装置530用于对血浆样品进行拉曼检测,以获取拉曼检测数据。
- [0222] 在其中的一些实施例中,拉曼检测装置530采用采用激光共聚焦拉曼技术、表面增强拉曼散射技术。
- [0223] 图6是根据本发明实施例的血浆样品获取装置的框架图。如图6所示,血浆样品获取装置620包括抗凝取血单元521、离心单元522和膜过滤单元523。其中,抗凝取血单元521用于放置血液样品;离心单元522用于对抗凝取血单元521的血液样品进行离心处理以获取初始血浆样品,以及对初始血浆样品进行离心处理以获取血浆样品;膜过滤单元523用于在离心单元522对初始血浆样品进行离心处理的同时,对初始血浆样品进行膜处理,以获取所述血浆样品。
- [0224] 在其中的一些实施例中,抗凝取血单元521为抗凝取血管,其内部装载有EDTA-K2抗凝。
- [0225] 在其中的一些实施例中,膜过滤单元523为超滤离心管。
- [0226] 优选地,膜过滤单元523为Millipore UFC503096 Amicon Ultra-30K超滤离心管。
- [0227] 另外,本申请实施例的癌症风险预测方法可以由计算机设备来实现。计算机设备的组件可以包括但不限于处理器以及存储有计算机程序指令的存储器。

[0228] 在一些实施例中,处理器可以包括中央处理器(CPU),或者特定集成电路(Application Specific Integrated Circuit,简称为ASIC),或者可以被配置成实施本申请实施例的一个或多个集成电路。

[0229] 在一些实施例中,存储器可以包括用于数据或指令的大容量存储器。举例来说而非限制,存储器可包括硬盘驱动器(Hard Disk Drive,简称为HDD)、软盘驱动器、固态驱动器(Solid State Drive,简称为SSD)、闪存、光盘、磁光盘、磁带或通用串行总线(Universal Serial Bus,简称为USB)驱动器或者两个或更多个以上这些的组合。在合适的情况下,存储器可包括可移除或不可移除(或固定)的介质。在合适的情况下,存储器可在数据处理装置的内部或外部。在特定实施例中,存储器是非易失性(Non-Volatile)存储器。在特定实施例中,存储器包括只读存储器(Read-Only Memory,简称为ROM)和随机存取存储器(Random Access Memory,简称为RAM)。在合适的情况下,该ROM可以是掩模编程的ROM、可编程ROM(Programmable Read-Only Memory,简称为PROM)、可擦除PROM(Erasable Programmable Read-Only Memory,简称为EPROM)、电可擦除PROM(Electrically Erasable Programmable Read-Only Memory,简称为EEPROM)、电可改写ROM(Electrically Alterable Read-Only Memory,简称为EAROM)或闪存(FLASH)或者两个或更多个以上这些的组合。在合适的情况下,该RAM可以是静态随机存取存储器(Static Random-Access Memory,简称为SRAM)或动态随机存取存储器(Dynamic Random Access Memory,简称为DRAM),其中,DRAM可以是快速页模式动态随机存取存储器(Fast Page Mode Dynamic Random Access Memory,简称为FPMDRAM)、扩展数据输出动态随机存取存储器(Extended Date Out Dynamic Random Access Memory,简称为EDODRAM)、同步动态随机存取内存(Synchronous Dynamic Random-Access Memory,简称SDRAM)等。

[0230] 存储器可以用来存储或者缓存需要处理和/或通信使用的各种数据文件,以及处理器所执行的可能的计算机程序指令。

[0231] 处理器通过读取并执行存储器中存储的计算机程序指令,以实现上述实施例中的任意一种癌症风险预测方法。

[0232] 在其中一些实施例中,计算机设备还可包括通信接口和总线。其中,处理器、存储器、通信接口通过总线连接并完成相互间的通信。

[0233] 通信接口用于实现本申请实施例中各单元、装置、单元和/或设备之间的通信。通信接口还可以实现与其他部件例如:外接设备、图像/数据采集设备、数据库、外部存储以及图像/数据处理工作站等之间进行数据通信。

[0234] 总线包括硬件、软件或两者,将计算机设备的部件彼此耦接在一起。总线包括但不限于以下至少之一:数据总线(Data Bus)、地址总线(Address Bus)、控制总线(Control Bus)、扩展总线(Expansion Bus)、局部总线(Local Bus)。举例来说而非限制,总线可包括图形加速接口(Accelerated Graphics Port,简称为AGP)或其他图形总线、增强工业标准架构(Extended Industry Standard Architecture,简称为EISA)总线、前端总线(Front Side Bus,简称为FSB)、超传输(Hyper Transport,简称为HT)互连、工业标准架构(Industry Standard Architecture,简称为ISA)总线、无线带宽(InfiniBand)互连、低引脚数(Low Pin Count,简称为LPC)总线、存储器总线、微信道架构(Micro Channel Architecture,简称为MCA)总线、外围组件互连(Peripheral Component Interconnect,简

称为PCI)总线、PCI-Express (PCI-X)总线、串行高级技术附件 (Serial Advanced Technology Attachment, 简称为SATA)总线、视频电子标准协会局部 (Video Electronics Standards Association Local Bus, 简称为VLB)总线或其他合适的总线或者两个或更多个以上这些的组合。在合适的情况下,总线可包括一个或多个总线。尽管本申请实施例描述和示出了特定的总线,但本申请考虑任何合适的总线或互连。

[0235] 该计算机设备可以执行本申请实施例中的癌症风险预测方法。

[0236] 另外,结合上述实施例中的癌症风险预测方法,本申请实施例可提供一种计算机可读存储介质来实现。该计算机可读存储介质上存储有计算机程序指令;该计算机程序指令被处理器执行时实现上述实施例中的任意一种癌症风险预测方法。

[0237] 本发明的癌症风险预测方法、系统、计算机设备及计算机存储介质,具有以下优点:

[0238] 1) 样本前处理简单:与常规的血清离心方法相比,本发明仅增加一个膜处理过程,可以在离心过程中,快速分离得到血浆过滤物,以便进行后续的拉曼检测,该离心处理时间仅需7-12分钟,比常规的血清离心方法的时间短;

[0239] 2) 检测时间短:拉曼单点检测最多需要20秒左右,即使多点检测(如3-5点检测),其耗时最多不超过2分钟,远低于当前血清学检测时间——各种指标从半小时到半天不等;

[0240] 3) 检测成分全面:能够检测血清中的全部代谢物小分子,成分全面,且结果与质谱检测结果一致性非常高,结果判断可靠;

[0241] 4) 检测时效性好:通过拉曼检测可以快速分析出代谢物的整体变化,对癌症提前进行早期和快速预警,便于医生针对性地进行后续精准检测;

[0242] 5) 检测成本低廉:本发明的检测仅需一个样本处理管和一个检测铝片或表面增强衬底即可完成,费用十分低廉;

[0243] 6) 低碳少污染:在检测过程中只会产生极少医疗废弃物,远比当前检测环保低碳。

[0244] 实施例2

[0245] 本实施例为本发明的具体应用。

[0246] 1、标准品标定

[0247] 1.1样品处理

[0248] 获取血液样品,对上述血液样品分别进行离心,以获取对应的初始血浆样品,并取1.5ml转置于EP管中,其中,离心条件为3000rpm、2min;

[0249] 对上述初始血浆样品进行基于膜过滤的离心处理,以获取对应的血浆样品,其中,取初始血浆样品450 $\mu$ l放入Millipore UFC503096 Amicon Ultra-30K超滤离心管的内管,离心条件为12000rpm、2.5-3min。

[0250] 1.2拉曼检测

[0251] 取上述血浆样品2 $\mu$ l放置于镀铝膜的载玻片(事先置于37 $^{\circ}$ C进行加热)30s-1分钟,此段时间调整好拉曼仪的聚焦,之后立即进行拉曼检测,其中,拉曼检测的条件为:Witech alpha300R激光共聚焦拉曼仪、激发波长:532nm、功率:10-14mw、光栅:1200g/mm、单谱积分时间:16-20s、物镜:100x/0.9、每个血浆样品采集5-8张图谱;

[0252] 对所有图谱进行极限校准和归一化处理,使用R语言对每个血浆样品的所有图谱求均值并做PCA分析,并对差异峰进行统计学分析。

[0253] 1.3质谱检测作为方法学标准进行校准

[0254] 取上述血浆样品进行质谱检测,并获取质谱检测结果。

[0255] 其中,质谱检测主要检测非靶代谢物,并对上述血浆样品整体进行PCA分析和LDA分析、差异代谢物的信号通路分析。在总共126个样本中,以25%(即32例)数据进行准确率训练时,拉曼与质谱分析结果的准确率完全一致,均是90.6%(29/32)。充分说明基于拉曼检测血清结果具有与质谱相同的准确性,本方法学结果非常可靠。

[0256] 共同发生的及各别发生的血清代谢物在肿瘤发生机制分析

[0257] 1乳糖

[0258] 乳糖是人类和哺乳动物乳汁中特有的碳水化合物,是由葡萄糖和半乳糖组成的双糖。目前研究都集中在乳糖不耐受方面,但对于乳糖和癌症的相关性,目前研究极少涉及。在本研究中,乳糖的浓度在8种肿瘤中平均升高15.82倍,在宫颈癌中升高倍数为29.40,在卵巢癌中则为4.99倍,充分说明了其在肿瘤中的普遍性。乳糖的最主要来源是牛奶,当然,不能下此结论“乳糖致癌”,但至少说明血清中乳糖升高与致癌有显著的相关性,也可能是其中和牛奶一起摄入的某些代谢物所引起,但肯定和包含有密切关系。在被纽约时报称为“21世纪流行病学巅峰之作”的《中国健康调查报告》中,被称为营养学界“爱因斯坦”的美国康奈尔大学的教授,柯林·坎贝尔先生和中科院陈君石院士在1970年代开始的,跨度30多年,横跨40多个国家,被称为“营养学有史以来最全面的调查”的研究显示:动物蛋白,包括牛奶,是包括肿瘤在内的多种慢性疾病的启动因素,尤其是牛奶中的酪蛋白,是启动和活化肿瘤基因表达的重要血清代谢物。

[0259] 2.B血型三糖(Blood group b trisaccharide)。

[0260] 三糖由三分子单糖以糖苷键连结而组成的化合物之总称。B血型三糖在Pubmed上搜索时,从2001年到2021年,仅8篇文献提及,且无有提及与肿瘤的相关性。在糖代谢过程中产生的多糖,种类比较多,比如甘露三糖,龙胆三糖,车三糖等,目前在文献中基本是空白。

[0261] 3.色氨酸

[0262] 研究表明:犬尿氨酸/色氨酸之比值可作为血清中前列腺癌,卵巢癌发生标志物;血清组氨酸和血浆色氨酸的组合可用作检测肾透明细胞癌(ccRCC)的生物标志物。在乳腺癌患者血清中,色氨酸水平显著升高,在乳腺癌细胞株培育过程中,培养基中加入色氨酸(100 $\mu$ M)可以显著抑制CD4<sup>+</sup>T细胞分泌IL-10。说明色氨酸通过CD4<sup>+</sup>T细胞抑制IL-10分泌是乳腺癌的潜在发病机制。本研究中,肾癌,乳腺癌和卵巢癌血清中色氨酸的浓度与正常人相比较,分别是4.32倍,4.13倍和3.791倍。

[0263] 4.HC毒素

[0264] HC毒素是一种首先从碳长丝孢菌的次级代谢产物中分离出来的环状四肽,其抗肿瘤活性优于其他组蛋白脱乙酰基化酶(HDAC)抑制剂。HDAC的功能是通过表观遗传学修饰,调节组蛋白乙酰化的程度,进而调控基因的表达。HDAC活性过高会使核小体结构过于紧密,抑制基因的表达,这也是多种癌症的成因。组蛋白脱乙酰基化酶抑制剂(HDACi)可以有效抑制组蛋白活性,是潜在的癌症治疗药物。HC毒素,是针对I、IIa和IIb类HDAC的HDAC抑制剂,在脑膜瘤生物学和作为靶向机制中的重要作用,在其他肿瘤中亦具有非常广谱的作用。研究表明:HC毒素导致神经母细胞瘤(NB)细胞向分化的良性表型转变,这与视网膜母细胞瘤(RB)抑制网络的激活有关。本研究的8种癌症中,HC毒素显著降低到0.106倍,其对HDAC的抑

制作用显著降低。

[0265] 5. 甘氨酸-缬氨酸二聚体

[0266] 20种氨基酸是构成人体蛋白质的最重要的组成基本单位,而甘氨酸作为非必需氨基酸的一种,对正常组织和肿瘤的生长都是非常必要的。肿瘤生长过程中新血管是给肿瘤提供能量的重要结构,其血管壁的形成也需要大量由甘氨酸、缬氨酸等合成的弹性蛋白,尤其是甘氨酸,几乎占到弹性蛋白组成的1/4。因此血清中甘氨酸升高,恰好为肿瘤的血管壁的生长,提供了重要的物质基础。在去年的《自然》杂志上,英国比特森癌症研究所的研究人员在结直肠癌,黑色素瘤和淋巴瘤小鼠模型中就发现:不含丝氨酸和甘氨酸的饮食将小鼠的生存率提高了一半,充分证明了甘氨酸和缬氨酸在促进癌症发展过程中的重要作用。而两个 $\alpha$ -氨基酸在体内可环化衍生成2,6-二酮哌嗪(DKP),被认为是一种重要的化合物,因为它们具有治疗多种疾病(包括癌症)的生物学特性,新型咪喃-二酮哌嗪类衍生物有效抑制微管聚合,可被视为开发抗癌药物的潜在靶点,苯氧基二酮哌嗪类衍生物,可显著示非常显著的抗微管剂细胞活性,具有显著的搞肿瘤应用前景。本研究结果显示:8种肿瘤中甘氨酸-缬氨酸二聚体显著下降到平均值为0.174倍。

[0267] 6. D1-正亮氨酸甲酯:

[0268] 这方面文献研究几乎空白。输入“norleucine methyl ester”只能找到1969-1979年共10篇文献,比如其中一篇:重氮乙酰-DL-正亮氨酸甲酯和胃酶抑素对来自脑、肾、骨骼肌和食虫植物的酸性蛋白酶的特异性抑制作用。本研究结果显示:D1-正亮氨酸甲酯在8种肿瘤中的平均倍数为0.02倍。

[0269] 7. 伊泽酰胺c (Izenamide c)

[0270] Izenamide是从海洋蓝藻中提取的线性缩肽,Izenamides A和B是组织蛋白酶D (cathepsin D) 的新型抑制剂,文献中2019年首次合成完成了组织蛋白酶D的缩肽抑制剂 izenamides A、B和C。组织蛋白酶是在各种动物组织的细胞内(特别是溶酶体部分)发现的一类蛋白酶,是半胱氨酸蛋白酶家族的主要成员,在生物界发现的组织蛋白酶有20余种,人体中主要存在至少10余种,与人类多种重大疾病密切相关,如肿瘤如结肠癌、骨质疏松、关节炎等。研究表明 $\alpha$ 1抗胰蛋白酶(A1AT)和组织蛋白酶D(CTSD)在结直肠癌标本在组织和血清水平上,与正常对照组相比较,分别是显著降低和升高,通过对组织免疫组织化学分析,A1AT和CTSD共同作用可使96.77%的CRC组织与正常组织区分开来( $P < 0.0001$ ) 本研究中发现Izenamide c的浓度显著降低到0.008倍。

[0271] 综上所述:本次共七个代谢物,作为共同预测肿瘤的发生的预测标志物。其AUC曲线分别是0.9637,0.9357,0.9077,0.9018,0.8988,0.8476和0.6363。整体而言,有的可归于肿瘤发生过程提供代谢物基础,为肿瘤的生长和进展提供原料;有的代谢物起抑制剂作用,显著降低。两类物质形成的代谢物失衡的存在状态为肿瘤的发生的精准判断提供了非常高的可能性。

[0272] 具体肿瘤中:

[0273] 肾癌中共有10种特异改变的代谢物,其中1-棕榈酰溶血磷脂酰胆碱升高17.692倍,1-behenoyl-2-羟基-sn-甘油-3-磷酸胆碱显著升高70.962倍,而磷脂酰胆碱Pc(16:1e/17-hdohe),Pe(18:1e/10-hdohe)Pc(16:0e/8-hepe)显著降低,分别是0.014倍,0.029倍和0.016倍。

[0274] 胃癌中共有8种特异改变的代谢物,其中咪唑-2-羧酸乙酯(Ethyl imidazole-2-carboxylate) 7.184倍,而1-棕榈酰-2-硫代棕榈酰磷脂酰胆碱(1-palmitoyl-2-thiopalmityl phosphatidylcholine)显著降低到0.071倍。

[0275] 宫颈癌中共有9种特异改变的代谢物,其中环己胺(Cyclohexylamine) 硫酸吲哚酚(Indoxyl sulfate) 分别是0.119和0.193倍。2-叔丁基-6-甲基苯酚(2-tert-butyl-6-methylphenol) 反式-4-(氨基甲基)环己烷甲酸(Trans-4-(aminomethyl)cyclohexanecarboxylic acid) 分别升高30.653和32.896倍。

[0276] 直肠癌共有13种特异改变的代谢物,其中脱氧胸苷5'-磷酸(Deoxythymidine 5'-phosphate (dTMP), 水苏糖(Stachyose), 3'-岩藻糖基乳糖(3'-fucosyllactose) 分别升高87.35倍,70.57倍和66.217倍。

[0277] 肺癌共有9种特异改变的代谢物,其中β-羟乙基茶碱(beta-hydroxyethyltheophylline) 3-乙基苯酚分别升高108.637倍和11.187倍,另外两种氨基酸的二肽Glu-Pro与Ala-Pro分别升高14.76倍和11.646倍。

[0278] 卵巢癌共有12种特异改变的代谢物,其中牛磺鹅去氧胆酸盐(Taurochenodeoxycholate) N,n'-二乙酰壳二糖(N,n'-diacetylchitobiose), 富马原十六酸(Fumarprotocetraric acid) 分别是153.807倍,162.187倍和17.183倍。

[0279] 乳腺癌共有10种特异改变的代谢物,其中甜香豆素b(Swietenocoumarin b) 环(脯氨酸-亮氨酸) 和Pro-Phe-Arg三肽分别升高倍数是116.380,4.490和3.222倍。

[0280] 前列腺癌中共有39种特异改变的代谢物,其中纤维二糖(Cellobiose) 二唾液酸乳糖(Disialyllactose) 和顺乌头酸分别升高243.043倍,96.584倍和16.253倍。

[0281] 具体代谢物改变情况如图7以及下表1~表9所示:

[0282] 表1通用类拉曼检测数据的7种代谢物改变情况

	ID	英文名	中文名	前列腺癌	乳腺癌	卵巢癌	肺癌	肝癌	宫颈癌	胃癌	肾癌	平均倍数	标准差	
[0283]	1	M365T412_2	Lactose	乳糖	1.891	24.712	4.999	7.282	18.982	29.405	26.656	NA	16.275	11.355
	2	M487T404	Blood	B型血三	NA	19.604	NA	7.273	14.086	20.767	20.948	NA	16.536	5.886

		group b	糖											
		trisaccharide												
[0284]	3	M203T272	DL-tryptophan	DL-色氨酸	NA	4.131	3.791	5.750	5.806	4.541	5.253	4.317	4.798	0.805
	4	M173T309	Gly-Val	甘氨酸-缬氨酸二聚体	0.093	0.135	0.367	0.199	0.215	0.110	0.174	0.174	0.183	0.085
	5	M437T96	Hc toxin	HC毒素	0.143	0.059	0.067	0.126	0.149	0.127	0.112	0.066	0.106	0.037
	6	M291T33	Di-norleucine methyl ester	DI-正亮氨酸甲酯	0.006	0.009	0.049	0.050	0.011	0.008	0.008	0.016	0.020	0.019
	7	M697T235	Izenamide c	伊泽酰胺c	0.005	0.007	0.013	0.009	0.009	0.011	0.006	0.006	0.008	0.003

[0285] NA:表示不涉及

[0286] 表2肾癌拉曼检测数据的10种代谢物改变情况

	ID	英文名字	VIP	Fold change	p-value	
	1	M751T38	1-myristoyl-2-palmitoyl-sn-glycero-3-phosphocholine	1.345	5.012	0.001
	2	M141T292	Kojic acid	1.510	1.418	0.007
	3	M865T137	Pc(16:1e/17-hdohe)	1.494	0.014	0.040
	4	M791T137	Pe(18:1e/10-hdohe)	1.224	0.029	0.043
[0287]	5	M671T141_2	1-Hexadecanoyl-2-(9Z,12Z-octadecadienoyl)-sn-glycero-3-phosphoric acid	1.252	0.008	0.046
	6	M309T25_2	Mestranol	1.348	0.388	0.049
	7	M841T140	Pc(16:0e/8-hepe)	2.405	0.016	0.050
	8	M580T180	1-behenoyl-2-hydroxy-sn-glycero-3-phosphocholine	1.326	70.962	0.000
	9	M538T185	1-Palmitoyllysophosphatidylcholine	1.358	17.692	0.000
	10	M193T82_1	Trans-3'-hydroxycotinine	1.655	7.483	0.016

[0288] 表3胃癌拉曼检测数据的8种代谢物改变情况

	ID	英文名字	VIP	Fold change	p-value	
	1	M747T38	(2-aminoethoxy)[2-[docosa-4.7.10.13.16.19-hexaenoyloxy]-3-[hexadec-1-en-1-yloxy]propoxy]phosphinic acid	1.843	7.334	0.000
	2	M147T122	3-hydroxyglutaric acid	1.183	2.059	0.000
[0289]	3	M103T300	D(-)-beta-hydroxy butyric acid	1.098	5.299	0.014
	4	M119T122	Tartronate	1.152	1.444	0.023
	5	M283T37	Octadecanoic acid	5.704	1.522	0.032
	6	M293T115	[6]-gingerol	2.329	1.283	0.036
	7	M141T34_1	Ethyl imidazole-2-carboxylate	1.484	7.184	0.000
	8	M751T135	1-palmitoyl-2-thiopalmitoyl phosphatidylcholine	1.209	0.071	0.042

[0290] 表4宫颈癌拉曼检测数据的9种代谢物改变情况

	ID	英文名字	VIP	Fold change	p-value	
	1	M113T186	Glutaric acid	1.060	9.950	0.000
	2	M716T34	1-Stearoyl-sn-glycerol	1.685	96.320	0.000
[0291]	3	M481T211	11-(6-{{2-(3,4-dimethoxyphenyl)ethyl}amino}-4-chloro-1,3,5-triazin-2-yl)-7,11-diazatricyclo[7.3.1.0<2,7>]trideca-2,4-dien-6-one	1.126	245.217	0.001
	4	M163T32	2-tert-butyl-6-methylphenol	1.683	30.653	0.009
	5	M212T29	Indoxyl sulfate	2.880	0.193	0.041
	6	M130T90	Paxilline	1.129	2.360	0.001
	7	M158T433	Trans-4-(aminomethyl)cyclohexanecarboxylic acid	2.852	32.896	0.006
	8	M100T72	Cyclohexylamine	2.485	0.119	0.032
	9	M104T411	Choline	1.614	0.642	0.037

[0292] 表5直肠癌拉曼检测数据的13种代谢物改变情况

[0293]

	ID	英文名字	VIP	Fold change	p-value
1	M179T369	D-(+)-mannose	1.490	3.637	0.000
2	M341T370	Sucrose	1.588	7.452	0.000
3	M665T501	Stachyose	1.560	66.217	0.000
4	M179T453	D-psicose	1.600	7.198	0.000
5	M281T75_3	Oleic acid	1.514	2.829	0.011
6	M297T102	Enterolactone	1.194	12.103	0.017
7	M321T72	Deoxythymidine 5'-phosphate (dTMP)	2.252	87.350	0.022
8	M858T196	Pi 36:4	1.094	5.415	0.038
9	M361T157_2	Phe-Glu	1.015	20.498	0.000
10	M511T461	3'-fucosyllactose	1.314	70.570	0.000
11	M107T82	(-)-perillyl alcohol	1.018	5.578	0.009
12	M74T71	N,n-dimethylformamide	1.350	1.692	0.014
13	M171T188	Glycyl-L-4-hydroxyproline	1.505	28.834	0.022

[0294]

表6前列腺癌拉曼检测数据的39种代谢物改变情况

[0295]

	ID	英文名字	VIP	Fold change	p-value
1	M173T442	Cis-aconitate	1.899	16.253	0.000
2	M923T492	Disialyllactose	2.402	243.043	0.000
3	M283T29_2	3,4'-dimethoxy-2'-hydroxychalcone	1.093	0.245	0.000
4	M151T218	Xanthine	2.028	0.222	0.000
5	M131T474	Asparagine	3.627	27.887	0.000
6	M791T38	Pe(18:1e/7-hdohe)	1.803	6.949	0.000
7	M341T489	Cellobiose	1.847	96.584	0.000
8	M116T312	Guanidoacetic acid	2.751	1.636	0.000
9	M118T412	Threonine	1.581	0.443	0.000

[0296]

10	M173T505	Isocitrate	3.978	5.308	0.001
11	M473T26	Amastatin	1.102	0.544	0.001
12	M186T237_1	2-hydroxy-6-methylquinoline-3-carbaldehyde	1.537	0.417	0.002
13	M167T333	Uric acid	2.162	1.847	0.002
14	M180T330	Tyrosine	1.645	0.630	0.003
15	M179T309_2	D-fructose	4.425	0.750	0.005
16	M277T36	C14-sas (tentative)	1.093	0.648	0.009
17	M311T28_2	Thymol-beta-d-glucoside	1.586	0.526	0.013
18	M325T27	Hydroquinidine	1.837	0.492	0.021
19	M223T216	D-Quinovose	1.979	0.440	0.022
20	M339T27_2	Gly-His-Lys	2.540	0.413	0.028
21	M107T25	3-hydroxyphenylacetic acid	1.045	2.302	0.037
22	M130T279_2	Guanidinopropionic acid	2.632	1.361	0.039
23	M135T33_1	Diglyme	7.165	50.058	0.000
24	M167T57	1-methyl-3-isobutylxanthine	7.098	35.866	0.000
25	M139T482	Trans-aconitic acid	2.262	57.340	0.000
26	M102T87	4-hydroxy-l-isoleucine	1.318	2.876	0.000
27	M429T299_2	Carfentrazone-ethyl	1.357	2.269	0.000
28	M230T574	Lauryldimethylamine oxide	1.273	2.694	0.000
29	M869T102	Solamargine	1.493	3.970	0.000
30	M230T233	Arg-Gln-Arg	1.047	2.051	0.000
31	M212T458	Phosphocreatine	1.031	29.453	0.000
32	M229T447	Pro-hyp	1.042	2.962	0.000
33	M118T278	Betaine	6.122	0.686	0.000
34	M167T116_1	Spectinomycin	1.893	17.881	0.000
35	M88T321	N-methylpropionamide	1.706	6.292	0.000
36	M295T483_2	Arg-Cys	2.482	17.771	0.001
37	M317T503	.gamma.-L-Glu.-epsilon.-L-Lys	1.129	2.692	0.003
38	M116T32	Sulfallate	1.416	1.675	0.006
39	M138T222	6-aminopyridine-3-carboxamide	1.549	5.569	0.035

[0297] 表7肺癌拉曼检测数据的9种代谢物改变情况

[0298]

	ID	英文名字	VIP	Fold change	p-value
1	M299T57	(r)-2-hydroxystearic acid	1.078	4.497	0.000
2	M187T196_2	Gabapentin related compound e	2.145	19.468	0.000
3	M227T219	Glu-Pro	1.163	14.760	0.000
4	M123T140	3-ethylphenol	1.534	11.187	0.000
5	M187T141	Ala-Pro	1.983	11.646	0.000
6	M95T140	1,2,3,6-tetrahydrobenzylalcohol	1.109	14.598	0.000
7	M225T52	.beta.-hydroxyethyltheophylline	5.300	108.637	0.000
8	M226T55	6-benzylaminopurine	1.202	6.255	0.000
9	M229T35	Bisphenol a	1.042	3.963	0.000

[0299] 表8卵巢癌拉曼检测数据的12种代谢物改变情况

	ID	英文名字	VIP	Fold change	p-value	
[0300]	1	M187T44	1-hydroxy-2-naphthoic acid	1.505	8.541	0.000
	2	M355T37_1	Fumarprotocetraric acid	1.452	17.183	0.000
	3	M159T119_2	3-hydroxyoctanoic acid	1.411	1.856	0.004
	4	M498T44	Taurochenodeoxycholate	1.989	153.807	0.005
	5	M182T40	4-pyridoxic acid	1.584	4.872	0.016
	6	M300T204	N-Acetyl-D-Glucosamine 6-Phosphate	1.028	0.627	0.020
	7	M204T47	N,n'-diacetylchitobiose	2.092	162.187	0.005
	8	M191T255	5-methyl-5-phenylhydantoin	1.300	3.029	0.005
	9	M174T570	Ala-Thr-Arg	1.167	1.600	0.006
	10	M130T582	D-Pipecolinic acid	1.993	0.681	0.009
	11	M196T258	1-deoxy-1-(methylamino)-d-galactitol	1.061	7.710	0.019
	12	M283T72	Hexaethylene glycol	6.955	0.653	0.028

[0301] 表9乳腺癌拉曼检测数据的10种代谢物改变情况

	ID	英文名字	VIP	Fold change	p-value	
[0302]	1	M129T86	5-methyl-5,6-dihydrouracil	1.150	2.485	0.008
	2	M144T245	L-glutamic acid, dimethyl ester	1.113	2.158	0.007
	3	M188T328_2	Leu-Ser-Arg	1.103	1.349	0.006
	4	M210T532	Pro-Phe-Arg	1.126	3.222	0.001
	5	M211T74_1	Cyclo(proline-leucine)	1.810	4.490	0.003
	6	M213T166	m-Chlorohippuric acid	1.020	0.847	0.017
	7	M282T33	Oleamide	1.199	2.343	0.001
	8	M283T51	Swietenocoumarin b	2.116	116.380	0.008
	9	M283T70	Stearic acid	1.541	0.502	0.007
	10	M343T81	2-[1-hydroxy-1-(4-methoxyphenyl)propan-2-yl]oxy-6-(hydroxymethyl)oxane-3,4,5-triol	2.300	14.602	0.001

[0303] 1.4标准品标定

[0304] 对上述代谢物的标准品进行拉曼检测,并形成标准拉曼峰谱合集,用于判断多种肿瘤。

[0305] 2、癌症风险预测模型

[0306] 2.1构建癌症风险预测模型

[0307] 采用BP神经网络构建癌症风险预测模型,其包括输入层、隐藏层和输出层,其中,输入层的输入为1024、输出为512,隐藏层的输入为512、输出为128,输出层的输入为128、输出为4。

[0308] 2.2构建数据集

[0309] 将不同血浆样品以及标准品的拉曼检测数据统一映射为1024维的初始特征向量,并存储于txt文件中,其包括两列数据,一列为波数、一列为强度,所述数据的波数一致,均为279-2186 $\text{cm}^{-1}$ ,一共1024个点;

[0310] 将上述拉曼检测数据划分为训练集和测试集,其中,训练集和测试集的比例为3:1。

[0311] 其中,上述拉曼检测数据可以分成4个类别,分别为肿瘤组和Con(对照组)。

[0312] 2.3训练癌症风险预测模型

[0313] 将训练集输入至癌症风险预测模型,获得输出结果;

- [0314] 将输出结果输入至softmax分类函数进行处理,以获取癌症风险预测结果。
- [0315] 2.4迭代癌症风险预测模型
- [0316] 将测试集输入至训练后的癌症风险预测模型,获得输出结果;
- [0317] 将输出结果输入至softmax分类函数进行处理,以获取癌症风险预测结果;
- [0318] 将癌症风险预测结果与测试集的原有结果进行比对,以检验训练的有效性。
- [0319] 使用训练集对癌症风险预测模型训练300轮,每10轮进行一次测试用以检验训练的有效性。
- [0320] 2.5训练好的癌症风险预测模型
- [0321] 经过不断训练,癌症风险预测模型的实验数据如下:
- [0322] 预测模型算法简要说明

[0323]

	某癌	对照	敏感性	特异性	阳性预测值	阴性预测值	预测准确率
某癌	A	C	A/(A+B)	D/(C+D)	A/(A+C)	D/(B+D)	#
正常	B	D					
	<b>AI 算法整合</b>		(基于神经网络算法对上述基于拉曼的结果进行算法优化)				
ABCD 代表基于临床样本的拉曼信息数量; #基于 75%的拉曼数据建立模型,对另 25%的数据进行判断准确性得出的结果。							

[0324]

	肾癌	对照	敏感性	特异性	阳性预测值	阴性预测值	预测准确率
肾癌	25	13	71.43%	92.44%	65.79%	94.08%	79.47%
对照	10	159					
	<b>AI 算法整合</b>		82.93%	82.93%	93.41%	73.24%	95.34%

[0325]

	胃癌	对照	敏感性	特异性	阳性预测值	阴性预测值	预测准确率
胃癌	36	12	76.60%	92.50%	75.00%	93.08%	84.29%
对照	11	148					
	<b>AI 算法整合</b>		85.52%	85.52%	94.52%	87.45%	94.31%

[0326]

	宫颈癌	对照	敏感性	特异性	阳性预测值	阴性预测值	预测准确率
宫颈癌	28	4	90.32%	97.78%	87.50%	98.32%	93.48%
对照	3	176					
	<b>AI 算法整合</b>		0.9212	92.12%	98.23%	92.15%	98.44%

[0328]

	直肠癌	对照	敏感性	特异性	阳性预测值	阴性预测值	预测准确率
直肠癌	16	10	72.73%	93.75%	61.54%	96.15%	81.04%
对照	6	150					
	<b>AI 算法整合</b>		87.42%	87.42%	94.53%	77.36%	97.32%

[0329]

	肺癌	对照	敏感性	特异性	阳性预测值	阴性预测值	预测准确率
肺癌	25	12	73.53%	92.86%	67.57%	94.55%	82.12%
对照	9	156					
	<b>AI 算法整合</b>		85.71%	94.53%	78.43%	96.54%	90.37%

[0330]

	卵巢癌	对照	敏感性	特异性	阳性预测值	阴性预测值	预测准确率
卵巢癌	33	1	94.29%	99.44%	97.06%	98.88%	97.42%
对照	2	177					
	<b>AI 算法整合</b>		95.01%	99.32%	97.22%	99.01%	98.37%

	乳腺癌	对照	敏感性	特异性	阳性预测值	阴性预测值	预测准确率	
[0331]	乳腺癌	47	5	77.05%	96.99%	90.38%	92.00%	89.11%
	对照	14	161					
	<b>AI 算法整合</b>		89.15%	97.01%	92.51%	93.45%	92.51%	
	前列腺癌	前列腺癌	对照	敏感性	特异性	阳性预测值	阴性预测值	预测准确率
	对照	56	6	77.05%	96.99%	90.38%	92.00%	89.11%
[0332]		16	142					
	<b>AI 算法整合</b>		89.15%	97.01%	92.51%	93.45%	92.51%	
	<b>拉曼算法平均值</b>		79.21%	95.21%	79.40%	94.62%	86.93%	
	<b>AI 算法整合平均值</b>		89.70%	95.98%	86.47%	95.72%	91.48%	

[0333] 由此可见：八肿瘤平均的敏感性，特异性和准确率在基于拉曼的检测分别是79.21%，95.21%和86.93%，在经BP神经网络的AI算法优化后，敏感性和特异性和准确率分别是89.70%，95.98%和91.48%。由此可见，一方面，基于拉曼的血清学检测与质谱检测在模型训练过程中结果一致，方法学上完全可靠，且相对比较全面和准确地判断癌症的发生；另一方面说明基于AI的神经网络算法可显著提高基于拉曼血清学检测结果的敏感性和特异性。

[0334] 关于本发明，其主要研究过程如下：

[0335] 对每一份样本都进行质谱和拉曼检测，并经基于AI的BP神经网络算法融合提高，将拉曼、质谱及人工智能算法整合在一起，扣除了临床用药的影响，使用75%数据建立癌症的检测模型，此模型最终以AI算法形式整合到拉曼检测过程，最终达到基于拉曼检测血清学快速检测并预测癌症发生或转移/复发概率的效果。

[0336] 本发明以正常对照组作为癌症组，将患者或正常对照组的血清检测后拉曼检测数据导入预先训练好的癌症患者风险预测模型，即以快速(2分钟之内)获取癌症风险预测结果；在总共126个样本中，以25% (即32例) 数据进行准确率训练时，拉曼与质谱分析结果的准确率完全一致，均是90.6% (29/32)。

[0337] 因此，使用拉曼直接检测血清中代谢物与检测金标准质谱 (Mass spectrometry) 相比较几乎没有差异，证实本发明预测癌症发生方法的可靠性。

[0338] 本发明先以临床明确诊断的多种癌症患者数据建立泛癌模型，并基于BP神经网络的AI算法优化，以7种血清代谢物为共同代谢物，其他癌症的预测过程中可分别结合显著差异表达的代谢物作为预测癌症的辅助代谢物。本方法癌症风险预测结果的准确，可靠，应用前景广阔。

[0339] 本发明的最重要的优点在于：

[0340] 拉曼检测能够十分全面地检测到癌症样本中关于代谢物的改变，对血清代谢产物进行定性分析，有助于实现癌症的早筛和早诊。在我们的实验中，对同一癌症患者的血清样本分别进行拉曼和质谱的检测，并利用深度神经网络模型进行比较，结果显示拉曼检测结果与质谱检测结果的准确率完全一致，充分反映了拉曼检测血清的准确性与可靠性。此外，拉曼检测可以实现小分子代谢物快速的大规模检测，具有高通量、高准确性的优势，而且由于检测成本相对较低，因此在多种癌症的精准诊断/转移/复发预测中的应用前景十分光明，具备大规模推广应用的的优势。如果能够扩大样本量，结合不同癌症患者的病史和不同癌症特异的血清学标志物进一步优化拉曼数据与临床信息的权重，则会更精准地识别不同癌

症发生/转移/复发等的可能性,为更多患者早期发生预警,肿瘤转移复发等进展提供便捷,快速和精准的预测系统。

[0341] 以上所述实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0342] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请专利的保护范围应以所附权利要求为准。

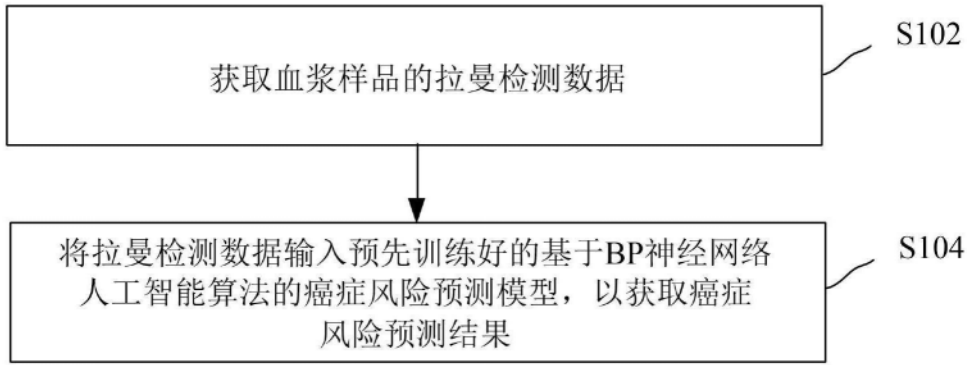


图1

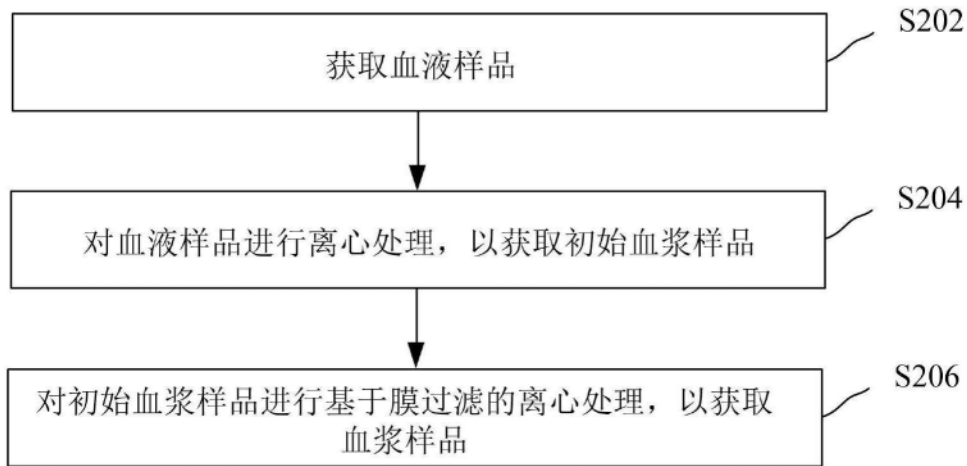


图2

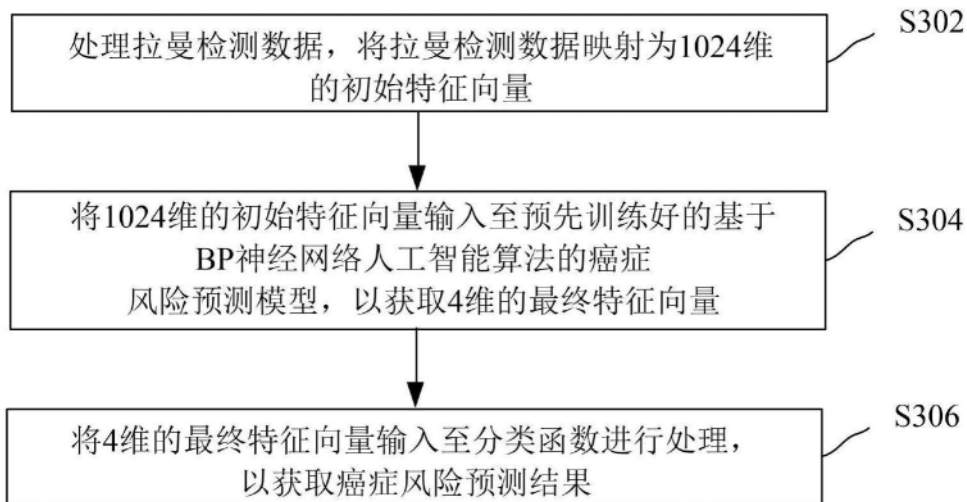


图3

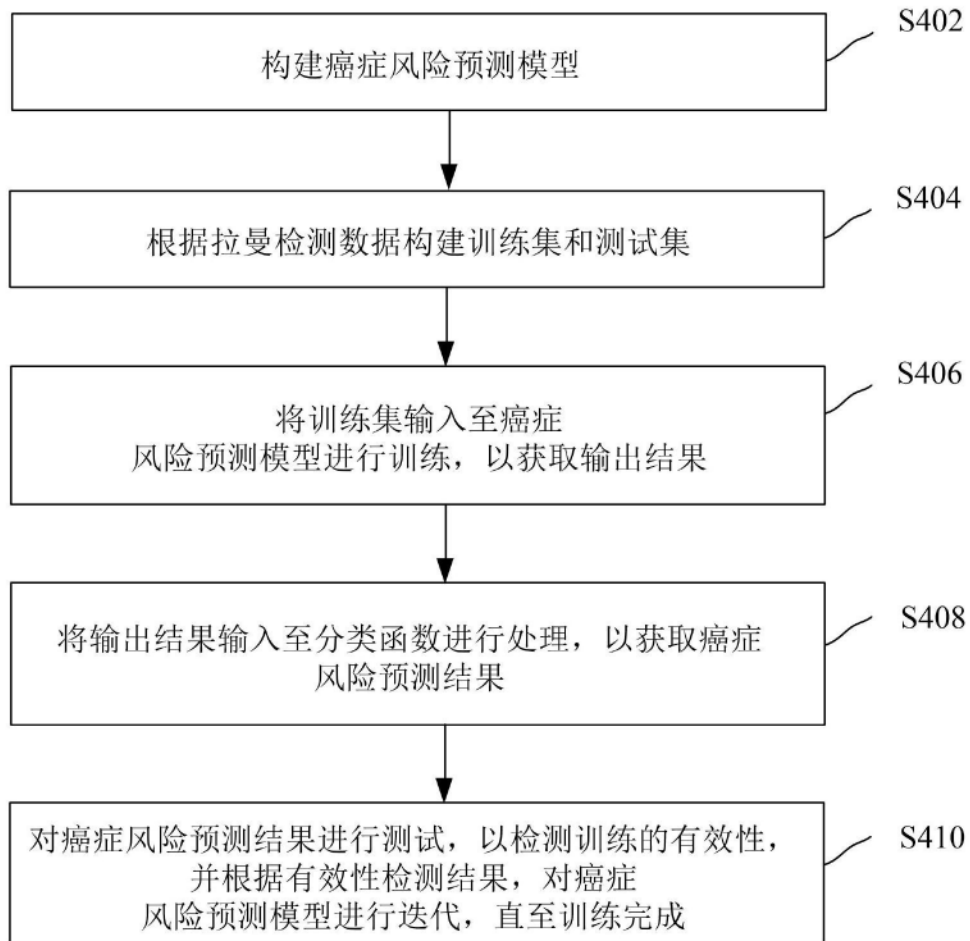


图4

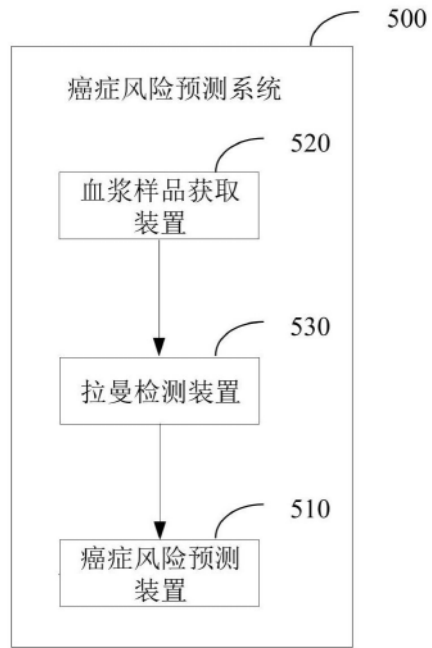


图5

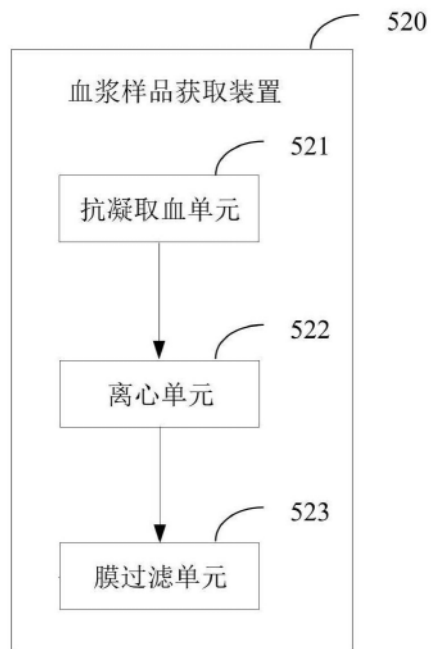


图6

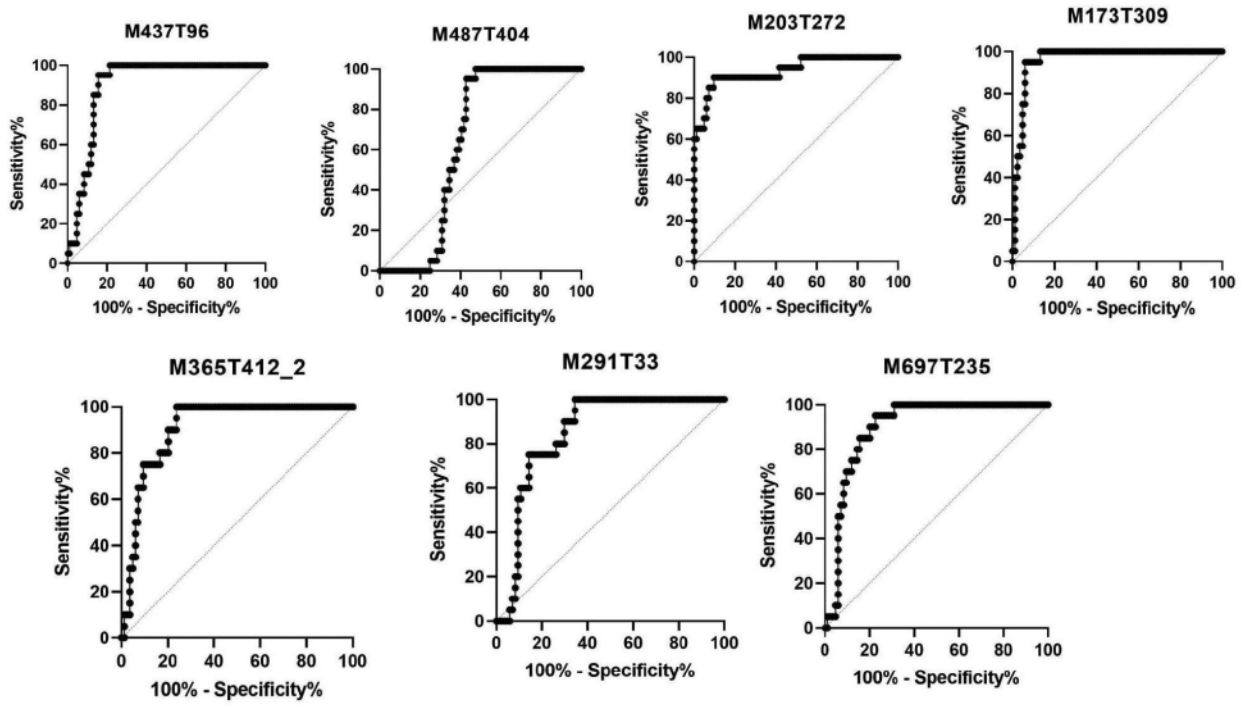


图7