



(12) 发明专利

(10) 授权公告号 CN 116794245 B

(45) 授权公告日 2025. 10. 31

(21) 申请号 202210685683.3
 (22) 申请日 2022.06.16
 (65) 同一申请的已公布的文献号
 申请公布号 CN 116794245 A
 (43) 申请公布日 2023.09.22
 (30) 优先权数据
 17/654,603 2022.03.14 US
 (73) 专利权人 香港理工大学
 地址 中国香港九龙
 (72) 发明人 刘乐庭 曹建农 张磊 林昕昕
 李文根 何晨杭 章焕 谭慧珊
 钟之桦
 (74) 专利代理机构 北京世峰知识产权代理有限公司 11713
 专利代理师 卓霖 张春媛

(51) Int.Cl.
 G01N 33/02 (2006.01)
 G01N 33/03 (2006.01)
 G01N 33/04 (2006.01)
 G01N 33/14 (2006.01)
 (56) 对比文件
 WO 2020044495 A1, 2020.03.05
 CN 104777115 A, 2015.07.15
 审查员 冯洁婷

权利要求书3页 说明书14页 附图17页

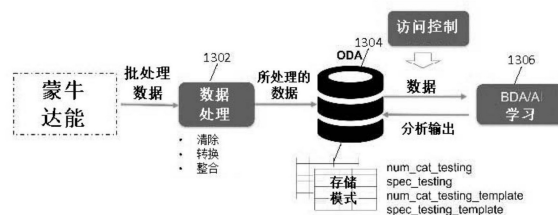
(54) 发明名称

用于物质的异常检测的方法、装置和计算机可读介质

(57) 摘要

在物质的异常检测中,获得第一组化学指纹。第一组中的每个指纹指示物质的正常样品的多个物理化学特性。第一组在多维主成分分析(PCA)图中被转换为数据点群集。该图的每个维度均基于对应于物理化学特性中的一个的主成分(PC)。将数据点群集的轮廓图案构建为预测模型,以将具有在轮廓图案之外的化学指纹的新样品识别为异常。使用第二组化学指纹优化预测模型。第二组中的每个化学指纹均指示包括正常检验样品和物质的异常样品的检验样品的多个物理化学特性。

数据流框架



ODA: Oracle 数据库设备 BDA: 大数据分析, AI: 人工智能



1300

1. 一种用于物质的异常检测的方法,该方法包括:

获得第一组化学指纹,其中,所述第一组化学指纹的每个化学指纹均指示所述物质的一组正常样品中的每个样品的多个物理化学特性;

将所述第一组化学指纹转换为多维主成分分析(PCA)图中的数据点群集,其中,多维PCA图的每个维度均基于主成分(PC),每个PC均对应于所述多个物理化学特性中的一个物理化学特性;

将所述数据点群集的轮廓图案构建为预测模型,该预测模型被配置为将具有落在所述轮廓图案的外部的化学指纹的新样品识别为异常;以及

使用第二组化学指纹优化所述预测模型,其中,所述第二组化学指纹指示包括所述物质的多个正常检验样品和多个异常检验样品的一组检验样品的多个物理化学特性,

其中,所述轮廓图案的构建包括:

对于群集中的每个数据点,计算所有平方PC总和的平方根;

基于数据点的所计算的总和的平方根的相应值,将数据点群集排列成预定数量的间隔;以及

通过从所述群集中去除落入具有超过预定数量的数据点的一个或多个排列中的数据点,来获得所述PCA图中的轮廓图案。

2. 根据权利要求1所述的方法,其特征在于,间隔的预定数量为20,数据点的预定数量为1000。

3. 根据权利要求1所述的方法,其特征在于,所述预测模型的优化包括:

对于所述一组检验样品,将所述第二组化学指纹转换为所述PCA图中的新数据点;以及

对于所述一组检验样品中的每个样品,计算所述样品与所述PCA图中的轮廓图案的质心之间的平方马氏距离(MD)分数,并且确定阈值平方MD分数以区分所述样品是落在所述轮廓图案的内部还是外部。

4. 根据权利要求3所述的方法,其特征在于,确定所述阈值平方MD分数包括:

对所述一组检验样品的平方MD分数进行排列;

迭代一系列平方MD分数作为灵敏度计算中的阈值;以及

将在所迭代的一系列平方MD分数中具有最高灵敏度值且总体准确度超过95%的平方MD分数确定为阈值平方MD分数。

5. 根据权利要求4所述的方法,其特征在于,所述灵敏度值是基于以下等式计算的:

$$\text{灵敏度值} = \frac{\text{正确识别为异常的样品数}}{\text{平方 MD 分数大于阈值的样品数}} \times 100\%。$$

6. 根据权利要求4所述的方法,其特征在于,所述总体准确度是基于以下等式计算的:

$$\text{总体准确度} = \frac{\text{正确识别为真实正常样品和异常的样品数}}{\text{该组检验样品中所有样品数}} \times 100\%。$$

7. 根据权利要求4所述的方法,其特征在于,还包括:

基于所述一组检验样品的光谱数据的层次群集学习,来训练所述预测模型以定义真实正常样品。

8. 根据权利要求4所述的方法,其特征在于,还包括:

基于对所述一组检验样品的多个物理化学特性的Extratree或XGBoost学习,来训练所

述预测模型以优化所述总体准确度。

9. 根据权利要求1所述的方法,其特征在於,所述物质是牛奶。

10. 一种用于物质的异常检测的装置,该装置包括:

至少一个处理器;以及

存储器,该存储器包括用于由所述至少一个处理器执行的计算机程序代码,其中,所述计算机程序代码指示所述至少一个处理器,以:

获得第一组化学指纹,其中,所述第一组化学指纹的每个化学指纹均指示所述物质的一组正常样品中的每个样品的多个物理化学特性;

将所述第一组化学指纹转换为多维主成分分析(PCA)图中的数据点群集,其中,多维PCA图的每个维度均基于主成分(PC),每个PC均对应于所述多个物理化学特性中的一个物理化学特性;

将所述数据点群集的轮廓图案构建为预测模型,该预测模型被配置为将具有落在所述轮廓图案的外部的化学指纹的新样品识别为异常;以及

使用第二组化学指纹优化所述预测模型,其中,所述第二组化学指纹指示包括所述物质的多个正常检验样品和多个异常检验样品的一组检验样品的多个物理化学特性,

其中,在构建所述轮廓图案期间,所述计算机程序代码进一步指示所述至少一个处理器,以:

对于群集中的每个数据点,计算所有平方PC总和的平方根;

基于数据点的所计算的总和的平方根的相应值,将数据点群集排列成预定数量的间隔;以及

通过从所述群集中去除落入具有超过预定数量的数据点的一个或多个排列中的数据点,来获得所述PCA图中的轮廓图案。

11. 根据权利要求10所述的装置,其特征在於,间隔的预定数量为20,数据点的预定数量为1000。

12. 根据权利要求10所述的装置,其特征在於,在所述预测模型的优化期间,所述计算机程序代码进一步指示所述至少一个处理器,以:

对于所述一组检验样品,将所述第二组化学指纹转换为所述PCA图中的新数据点;以及

对于所述一组检验样品中的每个样品,计算所述样品与所述PCA图中的轮廓图案的质心之间的平方马氏距离(MD)分数,并且确定阈值平方MD分数以区分所述样品是落在所述轮廓图案的内部还是外部。

13. 根据权利要求12所述的装置,其特征在於,在确定所述阈值平方MD分数期间,所述计算机程序代码进一步指示所述至少一个处理器,以:

对所述一组检验样品的平方MD分数进行排列;

迭代一系列平方MD分数作为灵敏度计算中的阈值;以及

将在所迭代的一系列平方MD分数中具有最高灵敏度值且总体准确度超过95%的平方MD分数确定为阈值平方MD分数。

14. 根据权利要求13所述的装置,其特征在於,所述灵敏度值是基于以下等式计算的:

$$\text{灵敏度值} = \frac{\text{正确识别为异常的样品数}}{\text{平方 MD 分数大于阈值的样品数}} \times 100\%。$$

15. 根据权利要求13所述的装置,其特征在於,所述总体准确度是基于以下等式计算的:

$$\text{总体准确度} = \frac{\text{正确识别为真实正常样品和异常的样品数}}{\text{该组检验样品中所有样品数}} \times 100\%。$$

16. 根据权利要求13所述的装置,其特征在於,所述计算机程序代码进一步指示所述至少一个处理器,以:

基于对所述一组检验样品的光谱数据的层次群集学习,来训练所述预测模型以定义真实正常样品。

17. 根据权利要求10所述的装置,其特征在於,所述物质是牛奶。

18. 一种其上编码有指令的非暂时性计算机可读存储介质,当所述指令由处理器执行时,使所述处理器执行根据权利要求1所述的用于物质的异常检测的方法中的一个或多个步骤。

用于物质的异常检测的方法、装置和计算机可读介质

技术领域

[0001] 本说明书广泛但非排他地涉及用于物质的异常检测的方法、装置和计算机可读介质。

背景技术

[0002] 食品掺假对于公共健康来说有重大风险。遭遇食品掺假的一些常见食品包括橄榄油、牛奶、蜂蜜、藏红花、橙汁、咖啡、苹果汁、葡萄酒、香草精和枫糖浆。例如,在牛奶蛋白质中掺入富含氮的化合物,或者宣传为含有牛肉的食品被发现为含有未声明的马肉,或者销售假冒的橄榄油,等等。因此,期望一个有效的协议来检测以前未遇到的掺杂物。

[0003] 尽管有全球合作,但是目前的检测方法实际上是如各个地方法律部门所规定的目标导向的。例如,由于所有分析都在寻求特定的化学物质和浓度,因此新设计的和以前未知的掺杂物可以规避现有目标导向的(或有针对性的,如在本申请中可互换使用的)检验方法,从而构成严重威胁。然而,不可能利用所有可用的检验方法来检验产品,更不用说目前食品工业中使用的质量评估方法通常很昂贵、需要专门的基础设施、而且是劳动密集型的。

[0004] 因此,期望一种有效的非目标导向(或非针对性的,如在本申请中可互换使用的)协议来检测以前未遇到的掺杂物(即物质中的异常)。

发明内容

[0005] 根据一个方面,提供了一种用于物质的异常检测的方法。该方法包括:获得第一组化学指纹,其中第一组化学指纹的每个化学指纹均指示该物质的一组正常样品中的每个样品的多个物理化学特性;将第一组化学指纹转换为多维主成分分析(PCA)图中的数据点群集,其中多维PCA图的每个维度均基于主成分(PC),每个PC均对应于多个物理化学特性中的一个物理化学特性;将数据点群集的轮廓图案构建为预测模型,该预测模型被配置为将具有落在轮廓图案的外部的化学指纹的新样品识别为异常;以及使用第二组化学指纹优化预测模型,其中第二组化学指纹指示包括物质的多个正常检验样品和多个异常检验样品的一组检验样品的多个物理化学特性。

[0006] 根据另一方面,提供了一种用于物质的异常检测的装置。该装置包括:至少一个处理器;以及存储器,该存储器包括用于由至少一个处理器执行的计算机程序代码,计算机程序代码指示该至少一个处理器来:获得第一组化学指纹,其中第一组化学指纹的每个化学指纹均指示该物质的一组正常样品中的每个样品的多个物理化学特性;将第一组化学指纹转换为多维主成分分析(PCA)图中的数据点群集,其中多维PCA图的每个维度均基于主成分(PC),每个PC均对应于多个物理化学特性中的一个物理化学特性;将数据点群集的轮廓图案构建为预测模型,该预测模型被配置为将具有落在轮廓图案的外部的化学指纹的新样品识别为异常;以及使用第二组化学指纹优化预测模型,其中第二组化学指纹指示包括该物质的多个正常检验样品和多个异常检验样品的一组检验样品的多个物理化学特性。

附图说明

[0007] 实施例和实现方式仅以示例的方式提供,并且通过结合附图阅读以下书面描述,本领域普通技术人员将更好地理解并且很容易明白这些实施例和实现方式。

[0008] 图1是根据一个实施例的用于物质的异常检测的装置100的示意图。

[0009] 图2是示出根据一个实施例的用于物质的异常检测的方法200的流程图。

[0010] 图3A是数据点群集的主成分分析 (PCA) 图300。数据点群集是从物质的一组正常样品的第一组化学指纹转换而来的。

[0011] 图3B是直方图,示出了基于数据点的所计算的所有平方主成分 (PC) 总和的平方根的相应值,数据点群集以预定数量的间隔被排列。

[0012] 图3C示出了数据点群集的轮廓图案350。轮廓图案350示出数据点群集的“核心”和“皮肤”。

[0013] 图4A和图4B是示出PCA图300中的新数据点的替代视角。新数据点是从包括物质的多个正常检验样品和多个异常检验样品的一组检验样品的第二组化学指纹转换而来的。

[0014] 特别地,图4A示出了轮廓图案(以形成包络形状402的点描绘)围绕来自检验样品的正常检验样品(以被包络形状402包裹的点404描绘)包裹,而图4B示出了在轮廓图案之外发现的样品(以诸如点406之类的点描绘)是异常检验样品。

[0015] 图4C示出了具有不同数量(在该示例中从1到8)的主成分 (PC) 的轮廓图案的累积百分比方差的陡坡图。

[0016] 图5示出了图表500,其描绘了一组检验样品中的每个样品的平方马氏距离 (MD) 分数的分布。线502代表5.4的阈值平方MD分数,以区分样品是落在轮廓图案的内部(正常)还是外部(异常)。

[0017] 图6A是示出当迭代一系列平方MD分数以确定阈值平方MD分数时的总体准确度的图。

[0018] 图6B是示出当迭代一系列平方MD分数以确定阈值平方MD分数时的灵敏度值的图。

[0019] 图7A是示出从傅里叶变换红外 (FTIR) 光谱检索到的原料奶的光谱数据的图。在该图中,光谱数据以吸光度相对于波数的关系绘制。在图7A中,带箭头的数字1到8指的是八个光谱区域,在所述区域中提取吸光度值用于层次群集学习以定义真实正常样品。

[0020] 图7B是树状图,其示出了层次群集学习中主分支内的样品群集。

[0021] 图7C是无根树,其示出了由层次群集学习观察到两个主要群集。在两个主要群集中,一个群集包括分支#1-3,另一群集包括分支#4-7。

[0022] 图8是可视化分类为国家标准内的113个异常样品(即异常)和198个真实正常样品的层次群集学习的热图。在图8中,图左侧的数字表示样品的树状图的分支编号,而图底部的数字表示从中提取吸光度值的区域。

[0023] 图9示出了在七个分支中不同异常成分的分布 (TN: 真实正常, AD: 掺杂物, N: 氮, Carb: 碳水化合物)。

[0024] 图10A是示出相应分支内真实正常样品的比率的图。图10B是示出相应分支内的富含氮的异常样品的比率的图。图10C是示出相应分支内基于碳水化合物的异常样品的比率的图。图10D是示出相应分支内作为异常样品的缓冲试剂的比率的图。在图10A至10D中,虚线和随附的百分比在图10A中表示真实正常的总百分比 ($198/311=63.7\%$)、在图10B中表

示富含氮的异常样品的总百分比 ($26/311=8.4\%$)、在图10C中表示基于碳水化合物的异常样品的总百分比 ($65/311=20.9\%$) 以及在图10D中表示缓冲试剂的总百分比 ($22/311=7.1\%$)。

[0025] 图11A示出了根据一个实施例的由当前预测模型学习的XGBoost模型的提升树。

[0026] 图11B是示出根据所学习的XGBoost模型的每个成分特征(即,物理化学特性)f1-f8的相对重要性的图。f1-f8分别代表脂肪、蛋白质、总固体、非脂肪固体、乳糖、相对密度、冰点和酸度。

[0027] 图12示出了计算机系统1200的框图,该计算机系统1200适合用作用于物质的异常检测的装置。

[0028] 图13是示出用于将化学指纹和生物指纹转换成标准化格式的数据流框架1300的实施例的图。标准化格式包括关于数值和分类检验结果(num_cat_testing)、光谱检验结果(spec_testing)、数值和分类检验模板(num_cat_testing_template)和光谱检验结果模板(spec_testing_template)的模式。本领域技术人员可以理解,标准化格式可以包括关于样品的其他物理化学特性的其他模式。

[0029] 图14是示出图13中描绘的模式的实施例的图。在该实施例中,示例性表格被示出为具有在相应模式num_cat_testing、spec_testing、num_cat_testing_template和spec_testing_template中捕获的数据列。

[0030] 本领域技术人员将理解,图中的元素是为了简单和清楚而示出的,并且不一定按比例描绘的。例如,图示、框图或流程图的一些元素的尺寸可能相对于其他元素被夸大以帮助改进对当前实施例的理解。

具体实施方式

[0031] 将参考附图仅以示例的方式描述实施例。附图中相同的附图标记和字符指代相同的元件或等同物。

[0032] 根据算法和对计算机存储器内的数据的运算的函数或符号表示,来明确或隐含地展现说明书中的以下一些部分。这些算法描述和函数或符号表示是数据处理领域的技术人员将他们的工作内容最有效地传达给本领域的其他技术人员的手段。在本文中算法通常被认为是导致期望结果的自洽步骤序列。这些步骤是需要对物理量进行物理操纵的步骤,诸如能够存储、传输、组合、比较和以其他方式操纵的电、磁或光信号。

[0033] 除非另有明确说明并且从以下内容显而易见,否则应当理解,在整个说明书中,使用诸如“获得”、“转换”、“构建”、“优化”、“计算”、“排列”、“确定”、“分级”、“迭代”、“训练”之类的术语的讨论是指计算机系统或类似电子装置的动作和过程,其将表示为计算机系统内的物理量的数据操纵和转换为类似地表示为计算机系统或其他信息存储、传输或显示装置内的物理量的其他数据。

[0034] 本说明书还公开了用于执行所述方法的操作的设备。这样的设备可以为了所需目的而专门构建,或者可以包括计算机或由存储在计算机中的计算机程序选择性地激活或重新配置的其他装置。本文呈现的算法和显示与任何特定的计算机或其他设备无内在关联。根据本文的教导,各种机器可以与程序一起使用。替代地,构建更专业的设备来执行所需的方法步骤可能是合适的。适合于执行本文描述的各种方法/过程的计算机的结构将从下面

的描述中显而易见。

[0035] 此外,本说明书还隐含地公开了一种计算机程序,因为对于本领域的技术人员显而易见的是,本文描述的方法的各个步骤可以通过计算机代码来实施。计算机程序不旨在限于任何特定的编程语言及其实现方式。应当理解,可以使用多种编程语言及其编码来实现本文包含的说明书的教导。而且,计算机程序不旨在限于任何特定的控制流。计算机程序具有许多其他变体,它们可以使用不同的控制流而不背离本发明的精神或范围。

[0036] 此外,计算机程序的一个或多个步骤可以并行而不是顺序地执行。这样的计算机程序可以被存储在任何计算机可读介质上。计算机可读介质可以包括存储装置,诸如磁盘或光盘、存储芯片或其他适合与计算机接口的存储装置。计算机可读介质还可以包括诸如在因特网系统中示例的硬连线介质,或者诸如在GSM移动电话系统中示例的无线介质。当在这种计算机上加载并执行该计算机程序时,该计算机程序有效地产生了实现优选方法的步骤的设备。

[0037] 本说明书结合系统、装置和计算机程序组件使用术语“配置为”。对于待被配置为执行特定操作或动作的一个或多个计算机的系统而言,意味着系统已在其上安装了软件、固件、硬件或它们的组合,这些软件、固件、硬件或它们的组合在操作中使得系统执行这些操作或动作。对于待被配置为执行特定操作或动作的一个或多个计算机程序而言,意味着该一个或多个程序包括指令,当由数据处理设备执行时,这些指令使得该设备执行这些操作或动作。对于待被配置为执行特定操作或动作的专用逻辑电路而言,意味着该电路具有执行这些操作或动作的电子逻辑。

[0038] 本申请的实施例提供用于非针对性检测物质中的异常的方法。这些方法利用机器学习方法来学习物质的化学指纹,从而构建和优化预测模型以识别以前遇到过的或未遇到过的异常。在本申请中,异常可互换地称为异常样品或掺杂物(AD)。

[0039] 本申请提供的这种预测模型提供了多种优点。首先,不需要使用昂贵的专业基础设施和机器进行额外的劳动密集型样品预处理或检验,从而减少了人为错误的机会和相关的时间成本。其次,在预测模型的训练阶段之后,不需要先验知道是否存在异常,因为纵向检验数据的化学指纹有助于发现牛奶成分的变化并随着时间的推移识别异常图案。第三,不需要手动建立已知真实参考产品(即真实正常样品)的公差水平,因为在本申请中提供了有效的数学公式,用于建立定义异常的阈值分数(可互换地称为截止分数),这继而节省了时间并提高了异常检测的准确度。最后也是最重要的是,预测模型能够识别每个样品存在一个以上的异常,而无需事先了解其组成。现实中多个未知异常可以共存于一个样品中,并且它们之间的相互作用可能导致信号干扰和完全不同的化学指纹。可以学习由预测模型识别的新遇到的化学指纹以进行进一步表征,这可以解决实际场景中食品掺假的实际问题。

[0040] 图1示出了用于物质的异常检测的装置100的示意图。

[0041] 装置100至少包括一个或多个处理器102和存储器104。该至少一个处理器102和存储器104互连。存储器104包括计算机程序代码(图1中未示出),其用于由该至少一个处理器102执行以进行如图2所示并在本申请中描述的异常检测的步骤。

[0042] 在步骤202,计算机程序代码指示该至少一个处理器102获得第一组化学指纹。第一组化学指纹的每个化学指纹均指示物质的一组正常样品中的每个样品的多个物理化学特性。

[0043] 在一个实施例中,该物质是牛奶。在一个示例中,该物质的一组正常样品是由蒙牛乳业(集团)有限公司(中国内蒙古和林)从2017年至2019年使用MilkoScan FT120(丹麦FOSS Analytical)采集的63473个正常原料奶样品。正常原料奶样品的化学指纹通过傅里叶变换红外(FTIR)光谱仪采集并存储在数据库中。这样的数据库可以建立在装置100的内部存储器或存储组件或装置100外部的存储空间上。

[0044] 每个化学指纹均指示每个样品的多变性。例如,化学指纹可以包括两种数据格式:(1)每个样品在1000-3550 cm^{-1} 范围内在吸光度相对于波数方面的光谱数据;以及(2)每个样品的具有多个成分特征(可互换地称为物理化学特性)的成分数据。在一个实施例中,多个物理化学特性包括八(8)个物理化学特性,诸如脂肪、蛋白质、固体非脂肪、总固体、乳糖、相对密度、冰点降低和牛奶的酸度。

[0045] 本领域技术人员可以理解,成分数据可以包括针对不同种类的牛奶的不同组的物理化学特性。因此化学指纹将反映这样的物理化学特性。

[0046] 此外,由于当前方法可以用于检测除牛奶以外的物质中的异常,因此本领域技术人员可以理解,成分数据可以包括针对各种类型物质(食品或者非食品)的各组的物理化学特性。因此化学指纹将反映这样的物理化学特性。

[0047] 在本实施例中,第一组化学指纹包括该组63473个正常原料奶样品的63473个化学指纹,每个化学指纹均指示在该组63473个正常原料奶样品中每个样品的八个物理化学特性和/或光谱数据。在步骤202,从数据库中获得第一组63473个化学指纹。

[0048] 在步骤204,计算机程序代码指示该至少一个处理器102将第一组化学指纹转换为多维主成分分析(PCA)图中的数据点群集。多维PCA图的每个维度均基于主成分(PC)。在本申请中,每个PC均对应于多个物理化学特性中的一个物理化学特性。在该意义上,PC在以下描述中可以互换地称为物理化学特性。在本申请中,为了便于参考,可以将该数据群集称为第一包络。

[0049] 转换的PCA图的一个实施例在图3A中示出。如图3A所示,PCA图300包括用于一组63473个正常原料奶样品的所有63473个化学指纹的数据点群集。然而,该群集是密集的,如果为构建预测模型而将群集中的所有数据点包含在内,则这将需要大量的计算。该技术挑战通过本申请的以下步骤有利地解决。

[0050] 在步骤206,计算机程序代码指示该至少一个处理器102将数据点群集的轮廓图案构建为预测模型。该预测模型可以将具有落在轮廓图案的外部的化学指纹的新样品识别为异常。

[0051] 例如,当物质为牛奶时,异常包括硫酸钾、重铬酸钾、柠檬酸、硫酸铵、三聚氰胺、尿素、乳糖、葡萄糖、蔗糖、麦芽糖糊精和果糖、水、柠檬酸钠和现实生活中的场景(诸如牛奶具有奶牛气味和牛奶被不当储存36小时)中的一种或多种。本领域技术人员可以理解,当异常变化时,变化是基于不同的物质。在一个实施例中,如果预测模型接收到新牛奶样品的化学指纹,该化学指纹在转换为PCA图中的新数据点时落在PCA图中的轮廓图案之外,则新牛奶样品将被预测模型识别为异常。同样,如果预测模型接收到新牛奶样品的化学指纹,该化学指纹在转换为PCA图中的新数据点时落在PCA图中的轮廓图案上或其内部,则新牛奶样品将被预测模型识别为正常。

[0052] 图3B中示出了构建轮廓图案的步骤206的实施例。如图所示,在轮廓图案的构建过

程中,计算机程序代码指示至少一个处理器102执行以下子步骤:

[0053] 首先,在子步骤206a,对于群集中每个数据点,计算所有平方PC总和的平方根。

[0054] 此后,在子步骤206b,基于数据点的所计算的总和的平方根的相应值,将数据点群集排列成预定数量的间隔。

[0055] 此后,在子步骤206c,通过从群集中去除落入具有超过预定数量的数据点的一个或多个排列中的数据点,来获得PCA图中的轮廓图案。

[0056] 在图3B所示的实施例中,间隔的预定数量为20,数据点的预定数量为1000。本领域技术人员可以理解,预定数量可以变化,并且可以根据实际需要来确定。在该实施例中,根据子步骤206a、206b和206c从PCA图中去除总共61,782个数据点,并且由PCA图中剩余的1691个数据点来构建轮廓图案350。这样的轮廓图案350示出了数据点群集的“核心”和“皮肤”,其保持了数据点群集的结构。在本申请中,为了便于参考,所构建的轮廓图案可以称为第二包络。

[0057] 在上述方式中,所构建的轮廓图案有利地使本申请能够将所需的计算从63473个数据点显著减少到仅1691个数据点,同时保留所有63473个化学指纹的数据点群集的结构。也就是说,其提高了计算效率,同时保持了预测模型的准确度。

[0058] 在步骤208,计算机程序代码指示该至少一个处理器102使用包括多个正常检验样品和多个异常检验样品的一组检验样品的第二组化学指纹来优化预测模型。

[0059] 在一个实施例中,该组检验样品包括1087个检验样品,该1087个检验样品包括976个正常原料奶样品和111个异常原料奶样品。在异常原料奶样品中,加标了各种浓度的以下化学品以模拟现实生活中的掺杂物(AD):硫酸钾、重铬酸钾、柠檬酸、硫酸铵、三聚氰胺、尿素、乳糖、葡萄糖、蔗糖、麦芽糖糊精和果糖、水、柠檬酸钠以及现实生活中的场景(诸如牛奶具有奶牛气味和牛奶被不当储存36小时)。

[0060] 表1示出了在该组检验样品中加标相应浓度的掺杂物异常样品的数量。

[0061]

添加的掺杂物(AD)	异常样品的数量
常见化学品	
重铬酸钾*	9
硫酸钾*	9
柠檬酸钠	4
柠檬酸*	9
基于氮的掺杂物	
硫酸铵*	5
尿素*	5
三聚氰胺*	10
基于碳水化合物的掺杂物	
蔗糖#	5
葡萄糖#	5
乳糖#	5
果糖#	5
麦芽糊精#	5

不当储存36小时	23
水	2
奶牛气味	10
总数	111

[0062] 表1添加了加标掺杂物的异常样品的数量。

[0063] *掺杂物的添加浓度为每100克原料奶0.01、0.02、0.05、0.1、0.2克；#掺杂物的添加浓度为每100克原料奶0.1、0.2、0.5、1、2克。

[0064] 与第一组化学指纹相似，第二组化学指纹中的每个化学指纹均指示该组检验样品中的每个正常检验样品或异常检验样品的多变性。例如，化学指纹可以包括两种数据格式：(1) 每个样品在1000-3550 cm^{-1} 范围内在吸光度相对于波数方面的光谱数据；以及(2) 每个正常或异常检验样品的具有多个物理化学特性的成分数据。在一个实施例中，多个物理化学特性包括每个正常原料奶样品或每个异常原料奶样品的八(8)个物理化学特性，诸如脂肪、蛋白质、固体非脂肪、总固体、乳糖、相对密度、冰点降低和酸度。

[0065] 优化预测模型的步骤208的实施例反映在图4A至图4C中。如图所示，在优化预测模型的过程期间，计算机程序代码指示该至少一个处理器102执行以下子步骤：

[0066] 首先，在子步骤208a，对于该组检验样品，将第二组化学指纹转换为PCA图中的新的数据点。从第二组化学指纹转换的新的数据点的一个实施例被描绘成被包络形状402包裹的点和在PCA图400、450中被发现位于包络形状402之外的点，如图4A和图4B中的替代视角所示。在PCA图400、450中，在步骤206获得的轮廓图案以形成包络形状402的点描绘。

[0067] 此后，在子步骤208b，对于该组检验样品中的每个样品，计算样品与PCA图中的轮廓图案中的质心之间的平方马氏距离(MD)分数，并确定阈值平方MD分数，以区分样品是落在轮廓图案的内部还是外部。

[0068] 在一个实施例中，平方MD分数可以通过R或RStudio上的MD分数计算脚本如下进行计算，其中考虑前X个主成分(PC)。应理解，斜体式名称可以根据实际需要替换为自己的名称：

```
[0069] name1<-prcomp(matrixdataframe,scale=TRUE)
```

```
[0070] name2<-as.data.frame(name1$x[,1:X])
```

```
[0071] MD<-mahalanobis(name2,center=colMeans(name2),cov=cov(name2))
```

```
[0072] name2$MD<-round(MD,3)
```

[0073] 如图4A的实施例中所示，包裹在轮廓图案(以形成包络形状402的点描绘)内的样品(以被包络形状402包裹的点描绘)是来自一组检验样品的正常检验样品，而图4B示出了在轮廓图案之外发现的样品(例如，以诸如点406之类的点描绘)是来自该组检验样品的异常检验样品。

[0074] 图4C示出了具有不同数量(在该示例中从1到8)的主成分(PC)的轮廓图案的累积百分比方差的陡坡图。在图4C中，其示出了四个PC负责轮廓图案的94.9%的方差。因此，本申请可以通过仅使用四个最重要的PC来进行子步骤208b中的平方MD分数计算而有利地提高了计算效率。关于图11A和11B描述了四个最重要的PC的细节。

[0075] 此外，如下表4所示，与用于第一组化学指纹(即第一包络)的数据点群集相比，轮廓图案(即第二包络)产生更好的准确度(95.22%相对于89.97%)、特异性(99.45%相对于

91.68%)、精密度(96.21%相对于64.98%)和CSI(70.95%相对于56.40%),具有相当的BA(86.22%相对于86.36%),但是具有较低的灵敏度(72.99%相对于81.03%)。其表明,使用子步骤208b中基于4个最重要的PC的平方MD分数计算的预测模型可以通过有效地区分工业乳品环境中的正常样品和异常样品来实现准确的预测结果,其性能优于国家标准。

[0076] 在一个实施例中,在子步骤208b中,计算机程序代码指示该至少一个处理器102执行以下子步骤以确定阈值平方MD分数。

[0077] 首先,在子步骤208b1中,对该组检验样品的平方MD分数进行排列。

[0078] 此后,在子步骤208b2中,迭代一系列平方MD分数作为灵敏度计算中的阈值。

[0079] 此后,在子步骤208b3中,将在所迭代的一系列平方MD分数内具有最高灵敏度值和超过95%的总体准确度的平方MD分数确定为阈值平方MD分数。

[0080] 在一些实施例中,灵敏度值基于以下等式计算:

$$[0081] \quad \text{灵敏度值} = \frac{\text{正确地识别为异常的样品数}}{\text{平方 MD 分数大于阈值的样品数}} \times 100\%。$$

[0082] 在一些实施例中,总体准确度基于以下等式计算:

$$[0083] \quad \text{总体准确度} = \frac{\text{正确地识别为真实正常样品和异常的样品数}}{\text{该组检验样品中所有样品数}} \times 100\%。$$

[0084] 图5示出了图表500,其描绘了根据一个实施例的一组检验样品中的每个样品的平方马氏距离(MD)分数的分布。在该实施例中,线502表示5.4的阈值平方MD分数,以区分样品是落在轮廓图案的内部(正常)还是外部(异常)。

[0085] 表2示出了当在子步骤208b中使用5.4的阈值MD分数以区分该组检验样品中的每个检验样品是落在轮廓图案的内部(正常)还是外部(异常)时的总体准确度、灵敏度和特异性、精密度(或阳性预测值)、阴性预测值(NPV)、关键成功指数(CSI)和平衡准确度(BA)。在该实施例中,阈值平方MD分数5.4分别提供了95.86%、72.99%和97.39%的总准确度、灵敏度和特异性。此外,精密度(或阳性预测值)为65.13%,阴性预测值(NPV)为98.18%,关键成功指数(CSI)为52.48%,并且平衡准确度(BA)为85.19%。

	正确预测	真实正常	2536
		真实异常 (AD)	127
	错误预测	虚假正常	47
		虚假异常 (AD)	68
[0086]	预测指标	准确度	95.86%
		灵敏度	72.99%
		特异性	97.39%
		精密度	65.13%
		阴性预测值 (NPV)	98.18%
		关键成功指数 (CSI)	52.48%
		平衡准确度 (BA)	85.19%

[0087] 表2

[0088] 图6A示出了描绘当迭代一系列MD平方分数以确定阈值平方MD分数时的总体准确度的图。图6B示出了描绘当迭代一系列平方MD分数以确定阈值平方MD分数时的灵敏度值的图。样品分组表明,阈值平方MD分数为5.4对识别正常的原料奶样品非常有效,其中正确预测率为99.5%。该阈值平方MD分数在检测异常样品方面也很有效,其中提供了94.6%的正确预测率,如图6A和6B所示。

[0089] 表3示出了当在子步骤208b中使用的阈值MD分数为5.4时,一组检验样品中的每个异常(掺杂物)的检测极限。使用Welch双样品t检验(2面)比较正常样品和异常样品之间的平方MD分数,产生 6.523×10^{-8} 的p值,并且该组检验样品中正常样品的所有平方MD分数位于轮廓图案内。此外,平方MD值甚至可以检测出已通过国家标准的异常样品:将“国家标准内的加标异常样品”(N=44)与正常样品进行比较的t检验得出0.01359的p值。在加标异常样品中,用水稀释的原料奶(N=2)和劣质原料奶(带有“奶牛气味”,N=10)很容易被检测出来,其中正确预测率达到100%。大部分因储存条件不当而变质的原料奶(N=23)和加标有各种缓冲试剂或无机化学品(硫酸钾、重铬酸钾、柠檬酸、柠檬酸钠)的原料奶也能被检测到,从而导致正确预测率分别为87.0%和71.0%。葡萄糖、麦芽糖糊精和果糖在2%时可检测到,而蔗糖仅在1%时就能检测到。虽然可以检测到低至0.02% (w/w)的三聚氰胺,但是硫酸铵和尿素分别在0.1%和0.2%时可检测到。富氮(硫酸铵、三聚氰胺、尿素)或基于碳水化合物(乳糖、葡萄糖、蔗糖、麦芽糖糊精、果糖)的掺杂物不易检测到。即使在2%的浓度下也无法检测到乳糖。

[0090]

添加了掺杂物	原料奶样品的数量	截至浓度 (g/100g 原料奶)
常见化学品:		
重铬酸钾*	9	0.2
硫酸钾*	9	0.1
柠檬酸钠	4	0.01
柠檬酸*	9	0.05
基于氮的掺杂物		
硫酸铵*	5	0.1
尿素*	5	0.2
三聚氰胺*	10	0.02
基于碳水化合物的掺杂物:		
蔗糖 #	5	1
葡萄糖 #	5	2
乳糖 #	5	未检测到
果糖 #	5	2

[0091]	麦芽糊精#	5	2
	存放不当 36 小时	23	正确预测率 87%
	水	2	正确预测率 100%
	奶牛气味	10	正确预测率 100%

[0092] 表3

[0093] 为了比较第二“包络”与第一“包络”的性能,通过将子步骤208b应用于数据点群集(第一包络)中的质心而不是图案轮廓(第二包络)中的质心,来获得8.6的截止平方MD分数。虽然第一包络包含更多样品而导致更高的总准确度和灵敏度,但是精密度和CSI显著降低,并且蔗糖变得无法检测到。因此,根据本申请的使用图案轮廓(第二包络)的预测模型在准确度、特异性、精密度和关键成功指数(CSI)方面优于使用第一组化学指纹的数据点群集(第一包络)的预测模型,如以下的表4所示。

预测模型方法		数据点群集的轮廓图案 (即第二包络)	第一组化学指纹的数据点群集 (即第一包络)
阈值平方 MD 分数 (即平方 MD 截止分数)		5.4	8.6
正确预测	真实正常	908	837
	真实异常 (AD)	127	141
错误预测	虚假正常	47	33
	虚假异常 (AD)	5	76
预测指标	准确度	95.22%	89.97%
	灵敏度	72.99%	81.03%
	特异性	99.45%	91.68%
	精密度	96.21%	64.98%
	阴性预测值 (NPV)	95.08%	96.21%
	关键成功指数 (CSI)	70.95%	56.40%

[0095]	平衡准确度 (BA)	86.22%	86.36%
--------	------------	--------	--------

[0096] 表4第一“包络”和第二“包络”之间的预测性能比较

[0097] 在一些实施例中,为了便于进行总体准确度的计算,计算机程序代码可以进一步指示至少一个处理器102训练预测模型,以基于对一组检验样品的光谱数据的层次群集学习来定义真实正常样品。

[0098] 图7A是示出从傅里叶变换红外 (FTIR) 光谱检索到的原料奶的光谱数据的图。如上

所述,光谱数据可以包含在每个样品的化学指纹中。在该图中,光谱数据以吸光度相对于波数的关系绘制。在图7A中,带箭头的数字1至8指的是八(8)个光谱区域,在这些区域中提取吸光度值用于层次群集学习以定义真正正常样品。

[0099] 在一个实施例中,为了检验对严格异常样品进行分类的有效性,对具有已通过国家标准的低浓度AD的40个异常样品和68个正常样品进行无监督层次群集。如图7A所示,从每个样品的光谱数据中提取出光谱区域1000-1100、1500-1600、1730-1800、2840-2940、3450-3550 cm^{-1} 内七个峰值的吸光度值和1250-1450 cm^{-1} 的平均吸光度值。使用欧几里得距离和沃德连锁法进行无监督层次聚集。聚集通过树状图和热图可视化,如图7B、图7C和图8所示。

[0100] 图7B是树状图,其示出了层次群集学习中主分支内的样品群集。图7B示出了基于碳水化合物的掺杂物聚集在分支#1和#6中,富含氮的掺杂物聚集在分支#7中,缓冲试剂或无机化学品聚集在分支#4中,而正常样品聚集在分支#2、#3、#5中。

[0101] 图7C是无根树,其示出了通过层次群集学习观察到两个主要群集。在两个主要群集中,一个群集包括分支#1至3,另一群集包括分支#4至7。

[0102] 图8示出了可视化分类为国家标准内的113个异常样品(即异常)和198个真正正常样品的层次群集学习的热图。在图8中,图左侧的数字表示样品的树状图的分支编号#,而图底部的数字表示从中提取吸光度值的区域。图8示出了在群集#2、#3和#5以及分支#1和#6中,正常原料奶样品和基于碳水化合物的异常样品中分别存在子类。另一方面,富含氮的掺杂物和缓冲试剂对FTIR光谱的影响更强,仅观察到一个主要分支。

[0103] 图9示出了7个分支中不同异常成分的分布(TN:真正正常,AD:掺杂物,N:氮,Carb:碳水化合物)。图9进一步示出了第一个群集包含大部分正常样品(真正正常%=81.2%;总数=101,真正正常=82,基于碳水化合物的AD=19),而第二个群集包含更多不同类型的异常样品(真正正常%=55.2%;总计=210,真正正常=116,富含氮的AD=26,基于碳水化合物的AD=46,缓冲试剂=22)。

[0104] 图10A是示出相应分支内真正正常样品的比率的图。图10B是示出相应分支内的富含氮的异常样品的比率的图。图10C是示出相应分支内基于碳水化合物的异常样品的比率的图。图10D是示出相应分支内作为异常样品的缓冲试剂的比率的图。在图10A至10D中,虚线和随附的百分比在图10A中表示真正正常的总百分比($198/311=63.7\%$)、在图10B中表示富含氮的异常样品的总百分比($26/311=8.4\%$)、在图10C中表示基于碳水化合物的异常样品的总百分比($65/311=20.9\%$)以及在图10D中表示缓冲试剂的总百分比($22/311=7.1\%$)。

[0105] 在一些实施例中,为了优化总体准确度,计算机程序代码可以进一步指示该至少一个处理器102基于对一组检验样品的多个物理化学特性的Extratree或XGBoost学习来训练预测模型。由预测模型学习到,某些物理化学特性比其他物理化学特性对于轮廓图案更有指示性。

[0106] 图11A示出了根据一个实施例的由当前预测模型学习的XGBoost模型的提升树。

[0107] 图11B是示出根据所学习的XGBoost模型的每个成分特征(即物理化学特性)f1-f8的相对重要性的图,f1-f8分别代表脂肪、蛋白质、总固体、非脂肪固体、乳糖、相对密度、冰点、酸度。

[0108] 图11A和11B都示出了非脂肪固体是确定原料奶样品是否经过处理的最具指示性的因素;而蛋白质和总固体乳糖是检测潜在掺假或污染样品为异常的最具指示性的因素。

[0109] 图12示出了计算机系统1200的框图,该计算机系统1200适合用作如本文所述的用于物质的异常检测的装置100。

[0110] 计算机系统/计算装置1200的以下描述仅作为示例提供并且不旨在进行限制。

[0111] 如图12所示,示例计算装置1200包括用于执行软件例程的处理器1204。尽管为了清楚起见而示出了单个处理器,但是计算装置1200也可以包括多处理器系统。处理器1204连接到通信基础设施1206以用于与计算装置1200的其他组件通信。通信基础设施1206可以包括例如通信总线、交叉开关或网络。

[0112] 计算装置1200还包括诸如随机存取存储器(RAM)的主存储器1208和辅助存储器1210。辅助存储器1210可以包括例如硬盘驱动器1212和/或可移动存储驱动器1214,该可移动存储驱动器可以包括磁带驱动器、光盘驱动器等。可移动存储驱动器1214以众所周知的方式从可移动存储单元1218读取和/或写入到可移动存储单元1218。可移动存储单元1218可以包括由可移动存储驱动器1214读取和写入到可移动存储驱动器1214的磁带、光盘等。如相关领域的技术人员将理解,可移动存储单元1218包括其中存储有计算机可执行程序代码指令和/或数据的计算机可读存储介质。

[0113] 在替代实施中,辅助存储器1210可以附加地或替代地包括用于允许将计算机程序或其他指令加载到计算装置1200中的其他类似装置。这样的装置可以包括例如可移动存储单元1222和接口1220。可移动存储单元1222和接口1220的示例包括可移动存储芯片(诸如EPROM或PROM)和相关联的插槽以及其他可移动存储单元1222和接口1220,该接口1220允许将软件和数据从可移动存储单元1222传输到计算机系统1200。

[0114] 计算装置1200还包括至少一个通信接口1224。通信接口1224允许通过通信路径1226将软件和数据在计算装置1200和外部装置之间传输。在各个实施例中,通信接口1224允许数据在计算装置1200和数据通信网络(诸如公共数据或私有数据通信网络)之间传输。通信接口1224可以用于在不同的计算装置1200之间交换数据,这些不同的计算装置1200形成互连计算机网络的一部分。通信接口1224的示例可以包括调制解调器、网络接口(诸如以太网卡)、通信端口、具有相关联电路的天线等。通信接口1224可以是有线的或者可以是无线的。通过通信接口1224传输的软件和数据呈信号的形式,这些信号可以是电子信号、电磁信号、光学信号或能够被通信接口1224接收的其他信号。这些信号通过通信路径1226提供到通信接口。

[0115] 可选地,计算装置1200还包括显示器接口1202和音频接口1232,该显示器接口1202执行用于将图像渲染到相关联显示器1230的操作,该音频接口1232用于执行通过(一个或多个)相关联扬声器1234播放音频内容的操作。

[0116] 如本文所使用的,术语“计算机程序产品”可以部分地指代可移动存储单元1218、可移动存储单元1222、安装在硬盘驱动器1212中的硬盘、或者通过通信路径1226(无线链路或电缆)连接到通信接口1224的承载软件的载波。计算机可读存储介质是指向计算装置1200提供所记录的指令和/或数据以供执行和/或处理的任何非暂时性有形存储介质。这种存储介质的示例包括软盘、磁带、CD-ROM、DVD、蓝光™光盘、硬盘驱动器、ROM或集成电路、USB存储器、磁光盘或计算机可读卡(诸如PCMCIA卡等),无论这些装置是在计算装置1200的内

部还是外部。也可以参与将软件、应用程序、指令和/或数据提供到计算装置1200的临时或无形计算机可读传输介质的示例包括无线电或红外线传输信道以及到另一计算机或联网装置的网络连接,以及包括电子邮件传输和记录在网站等上的信息的因特网或内联网。

[0117] 计算机程序(也称为计算机程序代码)被存储在主存储器1208和/或辅助存储器1210中。计算机程序也可以经由通信接口1224接收。这样的计算机程序在执行时能够启用计算装置1200来执行本文讨论的实施例的一个或多个特征。在各个实施例中,计算机程序在被执行时使处理器1204能够执行上述实施例的特征。因此,这样的计算机程序代表计算机系统1200的控制器。

[0118] 软件可以存储在计算机程序产品中并利用可移动存储驱动器1214、硬盘驱动器1212或接口1220加载到计算装置1200中。替代地,可以将计算机程序产品通过通信路径1226下载到计算机系统1200。该软件在由处理器1204执行时使计算装置1200执行本文描述的实施例的功能。

[0119] 应当理解,图12的实施例仅以示例的方式呈现。因此,在一些实施例中,可以省略计算装置1200的一个或多个特征。此外,在一些实施例中,计算装置1200的一个或多个特征可以被组合在一起。另外,在一些实施例中,计算装置1200的一个或多个特征可以被分成一个或多个组成部分。

[0120] 本说明书中描述的技术产生一种或多种技术效果。如上所述,本申请的实施例提供了用于对物质的异常进行非针对性检测的方法。这些方法利用机器学习方法来学习物质的化学指纹,从而构建和优化预测模型以识别以前遇到或未遇到的异常。在本申请中,异常可以互换地称为异常样品或掺杂物(AD)。

[0121] 此外,如本领域技术人员所理解的,可以利用新牛奶样品的化学指纹进行连续更新来训练预测模型。更新可以包括尽可能多的因素,其包括关键的化学计量学和牛的所有已知特征,诸如地理季节性物流变化、奶牛品种、饲料、年龄等以及有关错误添加特定成分(诸如非乳蛋白、非法防腐剂、高含量合法防腐剂、抗生素、杀虫剂)的信息。化学指纹应扩展到FTIR之外,以还覆盖质谱、MMR、红外光谱、液相色谱、气相色谱等。化学指纹可以扩大到还包括生物指纹,诸如来自下一代测序(NGS)的数据,以提供有价值的欺诈信息,诸如不同种类的牛奶、植物和动物脂肪;对品种和地域培养的欺诈性声明;和病原微生物的污染。化学指纹和生物指纹可以是各种异构的报告格式,包括数值、数据点、图形、图像、数字表示等。这些化学指纹和生物指纹可以被转换为呈模式(schemas)形式的标准化格式,以便于数据库中的数据存储。图13和14中示出了模式的一个实施例。在图13中,化学指纹和生物指纹在示例性数据流框架1300中被称为“数据”。在图14中,示例性数据表格被示出为具有在如图13所示的相应模式num_cat_testing、spec_testing、num_cat_testing_template和spec_testing_template中捕获的数据列。

[0122] 这些数据可以在数据处理模块1302中处理并转换成标准化格式以存储在数据库1304中。标准化格式包括关于数字和分类检验结果(num_cat_testing)、光谱检验结果(spec_testing)、数值和分类检验模板(num_cat_testing_template)和光谱检验结果模板(spec_testing_template)的模式。本领域技术人员可以理解,标准化格式可以包括关于样品的其他物理化学特性的其他模式。可以将标准化格式的存储数据提供到机器学习模块1306,以基于前述段落中描述的实施例来训练用于物质的异常检测的预测模型。

[0123] 在持续机器学习的情况下,预测模型的预测能力将进一步增强,使得本申请中提供的方法可以用作对广泛的食物商品的标准筛选。

[0124] 本领域技术人员将理解,在不背离宽泛描述的本发明的精神或范围的情况下,可以对如具体实施例中所示的本发明进行多种变化和/或修改。因此,当前实施例在所有方面都被认为是说明性的而不是限制性的。

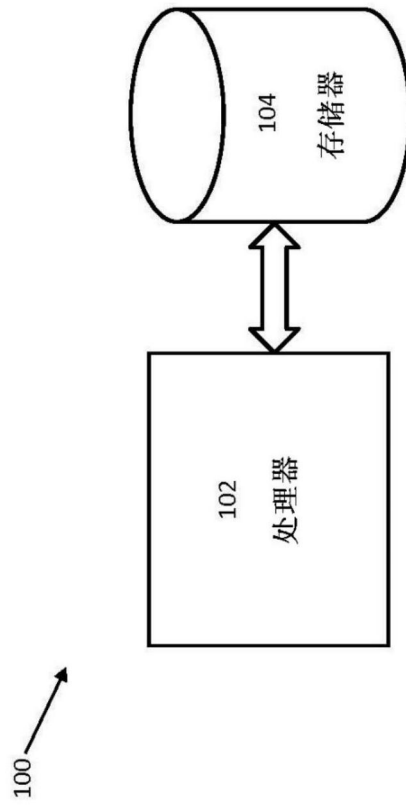


图1

200

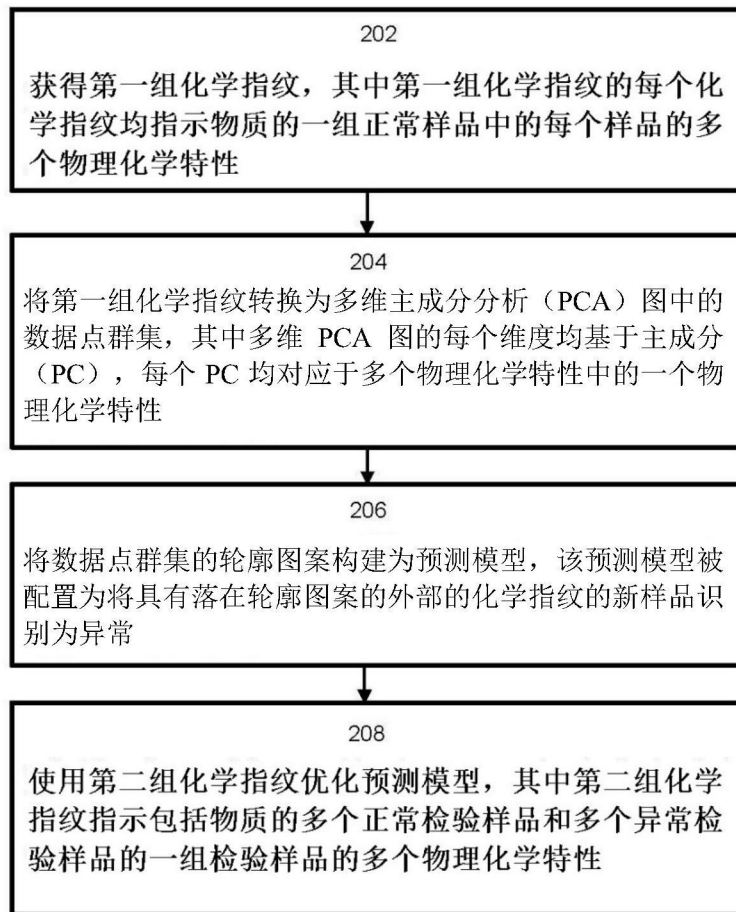


图2

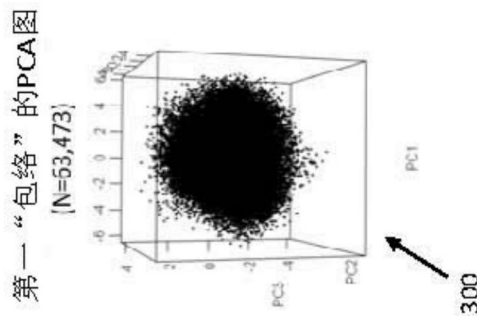


图3A

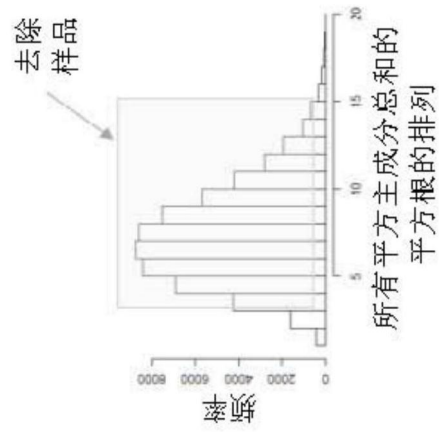


图3B

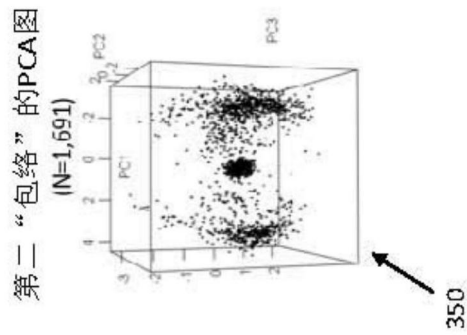


图3C

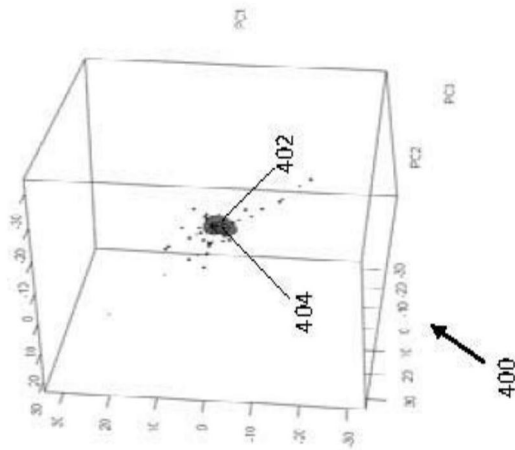


图4A

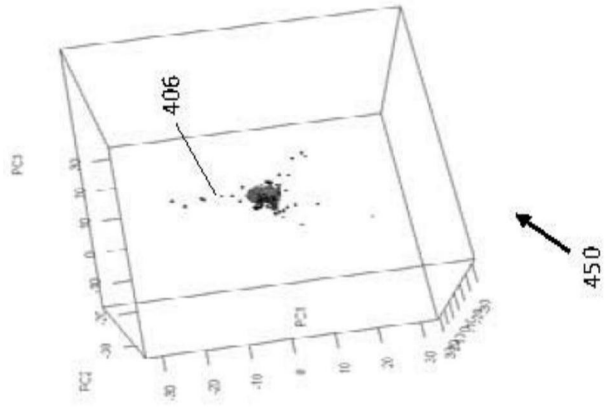


图4B

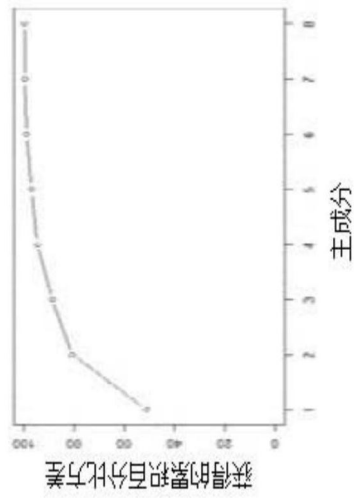


图4C

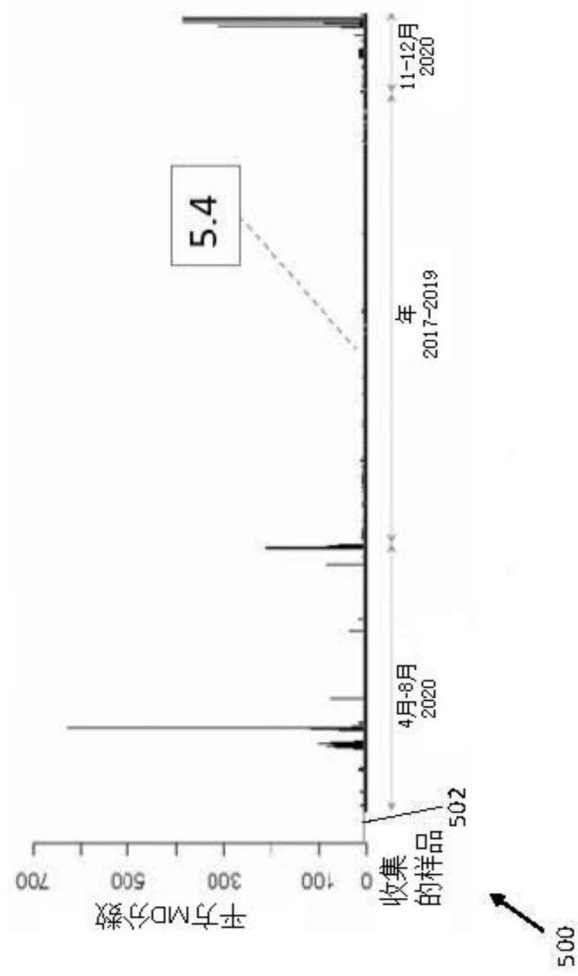


图5

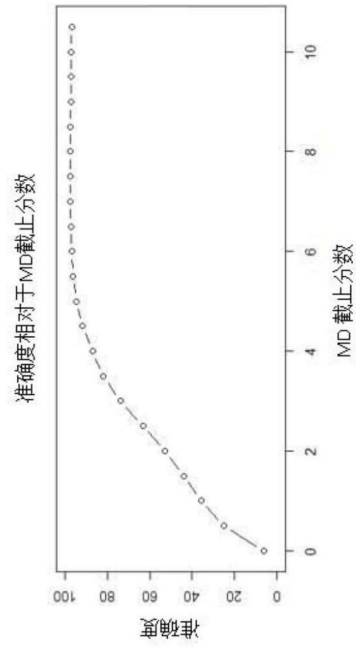


图6A

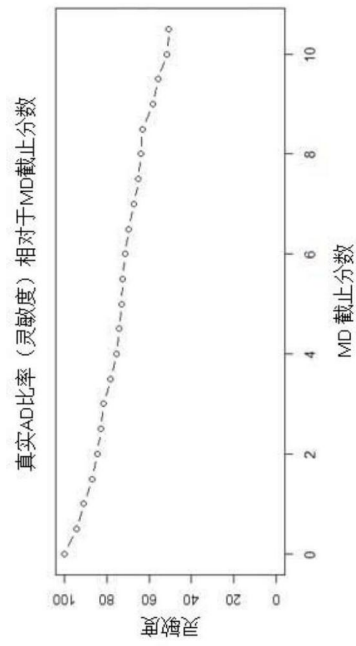


图6B

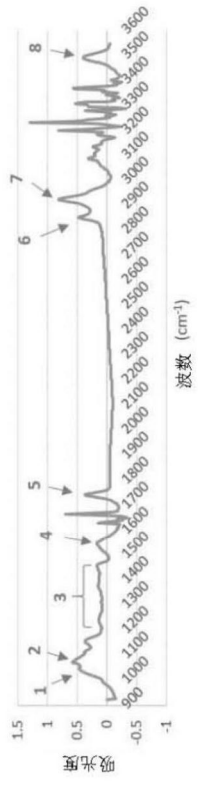


图7A

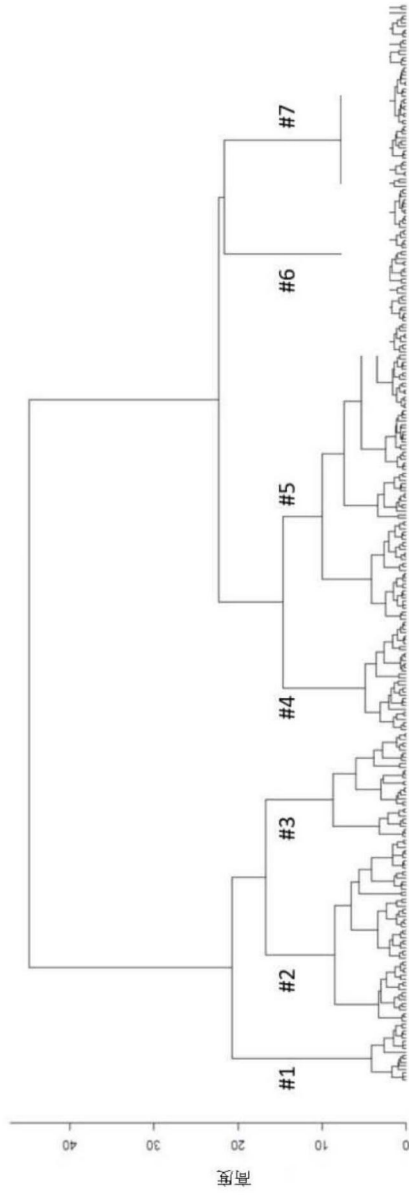


图7B

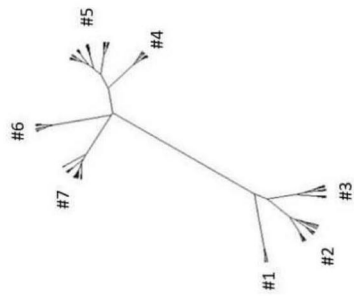


图7C

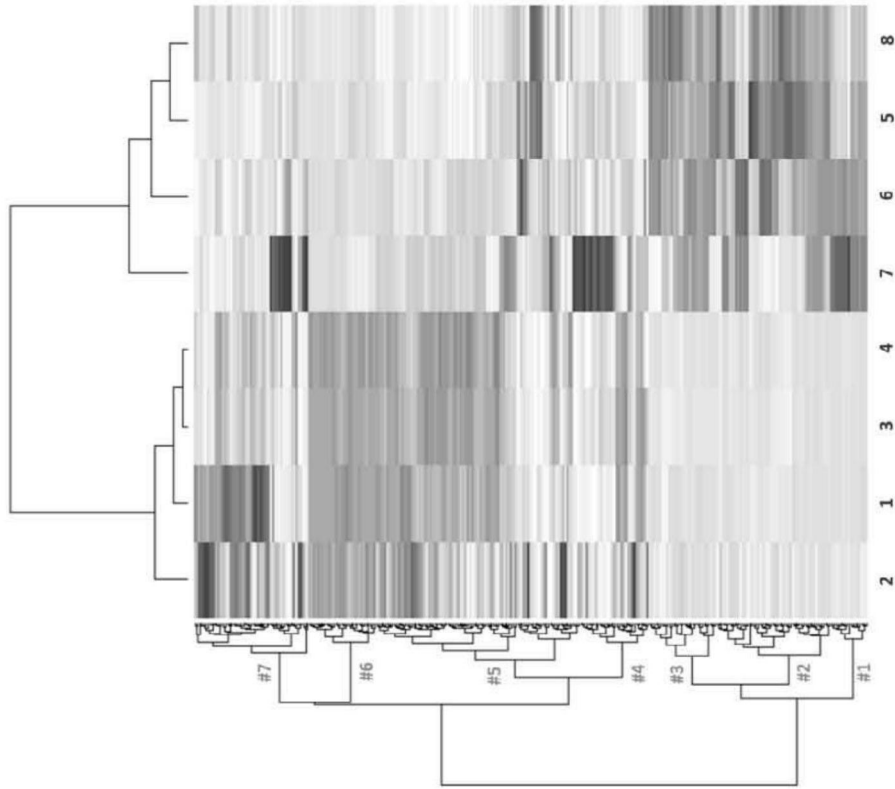


图8

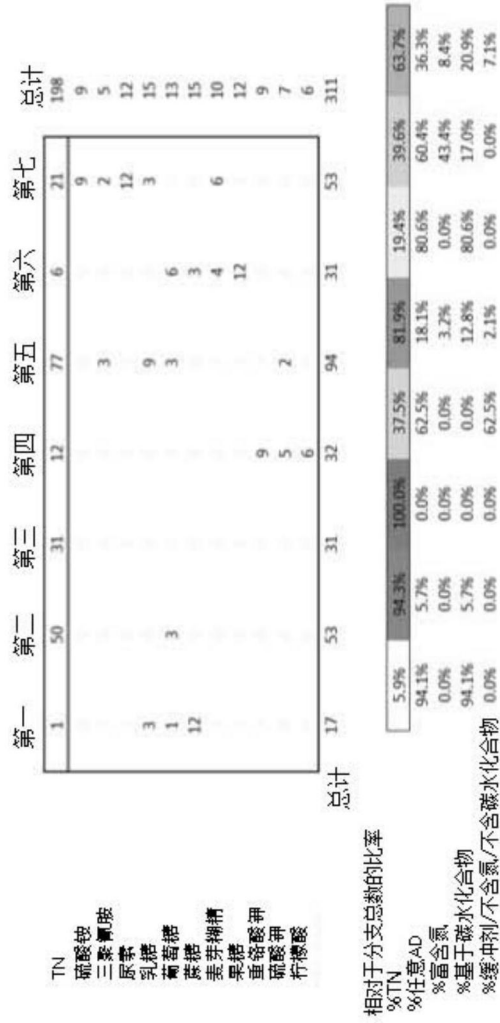


图9

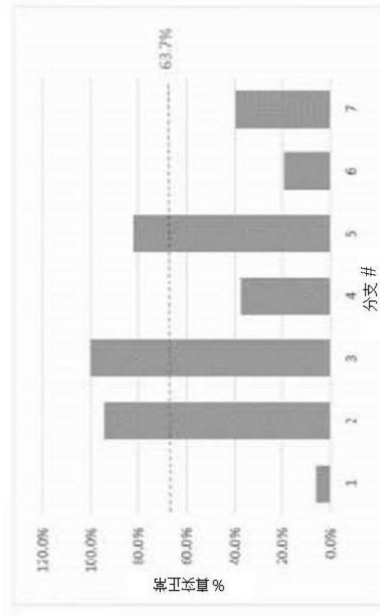


图10A

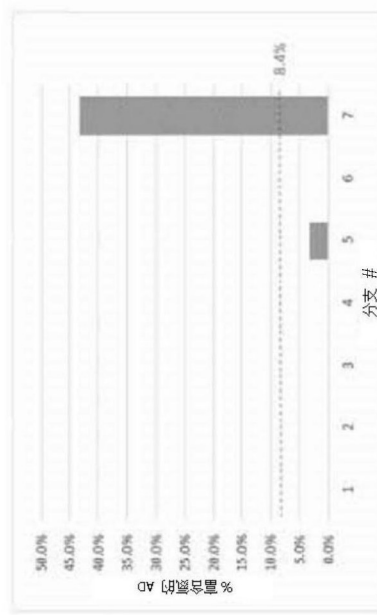


图10B

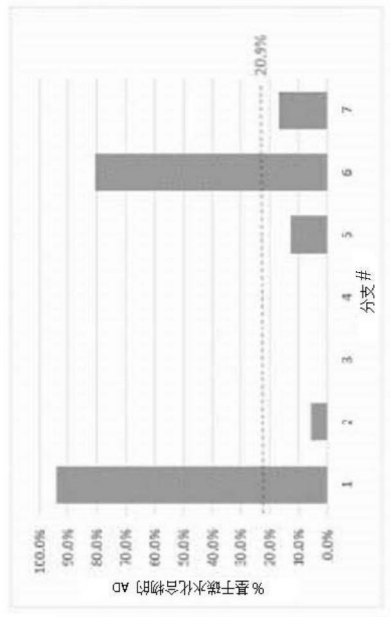


图10C

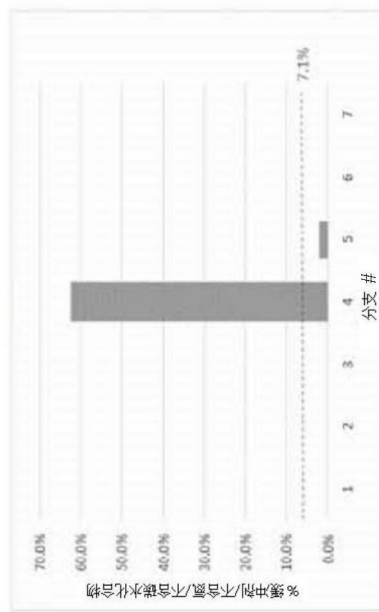


图10D

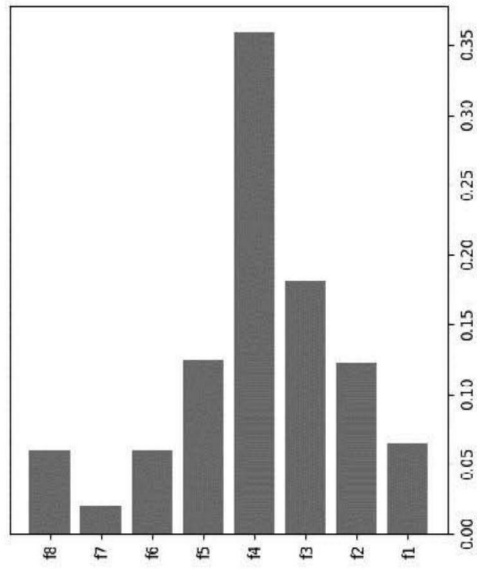


图11B

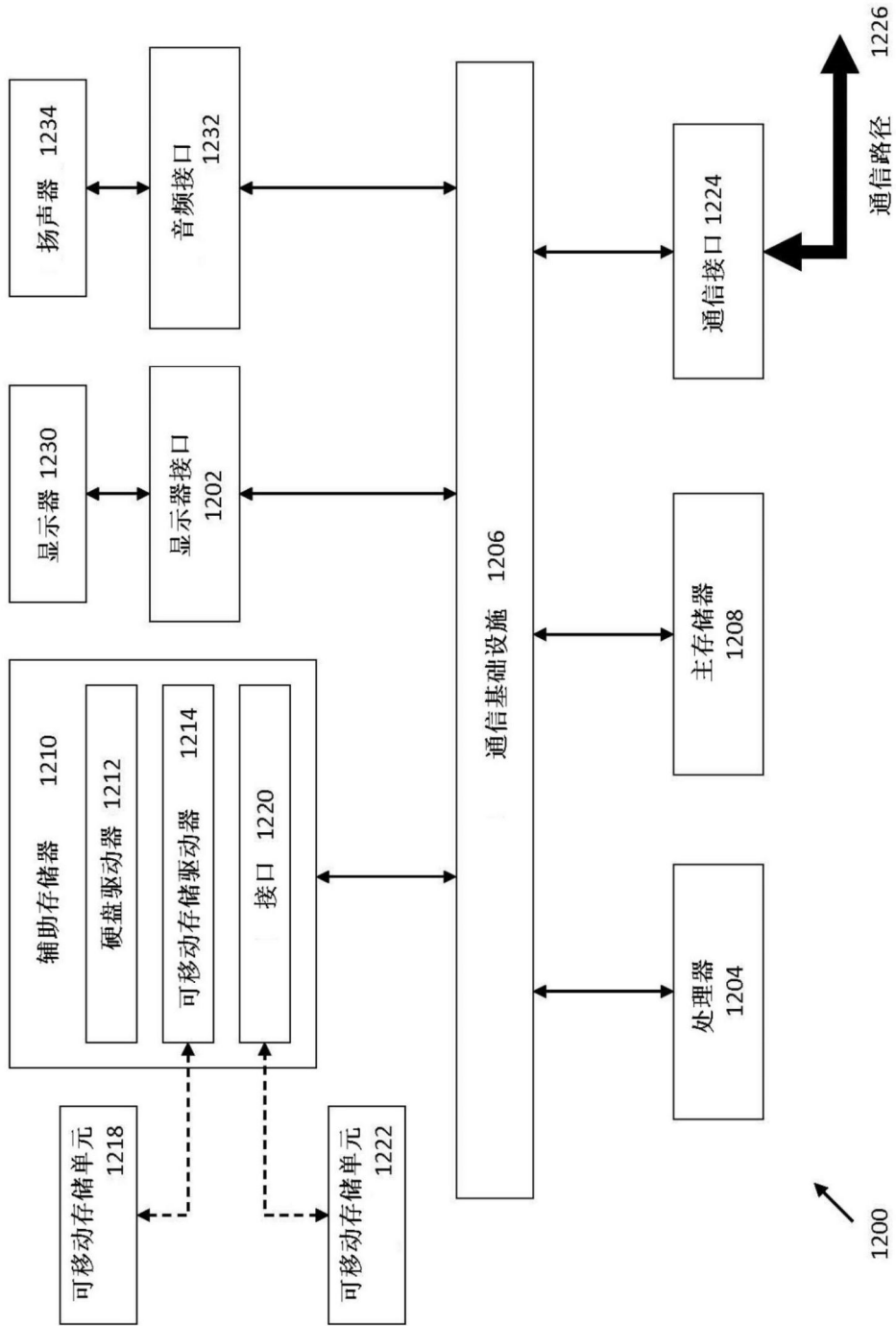
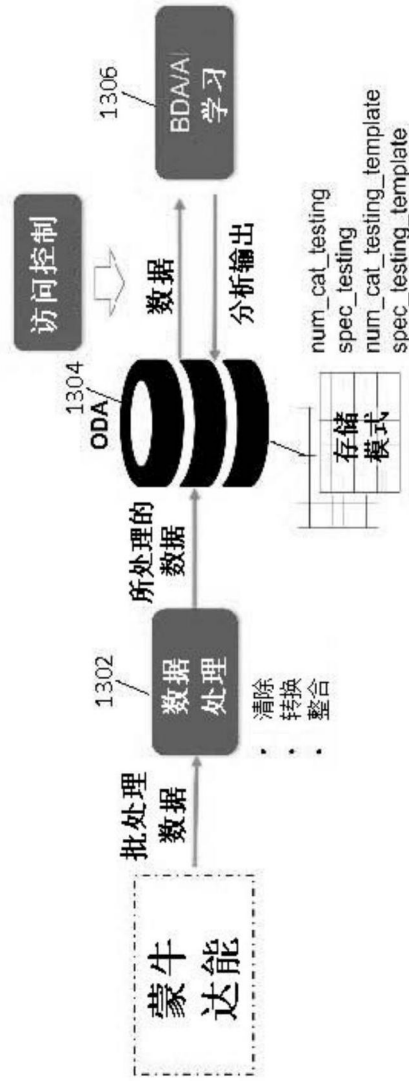


图12

数据流框架



ODBA: Oracle 数据库设备 BDA: 大数据分析, AI: 人工智能

1300

图13

数据表格

- **num_cat_testing**
 - ◆ 数字和分类检验结果
- **spec_testing**
 - ◆ 光谱检验结果
- **num_cat_testing_template**
 - ◆ 数值和分类检验模板
- **spec_testing_template**
 - ◆ 光谱检验结果模板

表格名称: num_cat_testing_template

列

-	产品类型
-	工厂
-	材料
-	过程
-	样品特性
-	批次编号
-	项目名称
-	输入值
-	输出值
-	项目状态
-	数据累积

表格名称: num_cat_testing

列

-	日期
-	产品类型
-	样品参考编号
-	工厂
-	材料
-	样品检验状态
-	数据累积
-	批次编号
-	过程
-	样品特性
-	批次编号
-	项目名称
-	输入值
-	输出值
-	项目状态
-	数据累积

表格名称: spec_testing_template

列

-	产品类型
-	频谱图序列

表格名称: spec_testing

列

-	参考编号
-	文件名称(pdf)
-	文件名称(png)
-	文件名称(jmeh)
-	比率(吸光度与微素)
-	样品名称
-	脂肪
-	蛋白质
-	AD2
-	时间
-	日期
-	评论
-	复制品
-	频谱图序列
-	其他分析结果
-	总固体
-	非脂肪
-	乳糖
-	密度
-	冰点
-	酸度
-	标签
-	AD1

1400

图14