

说明书

发明名称: 细胞培养液代谢物浓度确定方法、装置、设备和介质

[0001] 本申请是以CN申请号为202311548419.6, 申请日为2023年11月20日的申请为基础, 并主张其优先权, 该CN申请的公开内容在此作为整体引入本申请中。

技术领域

[0002] 本公开涉及生物检测技术领域, 尤其涉及一种细胞培养液代谢物浓度确定方法、装置、设备和介质。

背景技术

[0003] 在生物技术的研究与应用过程中, 很多场景都需要监测细胞培养过程中细胞培养液中的细胞代谢物浓度。拉曼光谱技术具有无损、高灵敏度和非标记等优点, 逐渐在细胞培养代谢物的检测中得到应用。

[0004] 发明内容

[0005] 第一方面, 本公开提供了一种细胞培养液代谢物浓度确定方法, 该方法包括:

[0006] 获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数;

[0007] 将所述光谱数据和所述预设细胞培养环境参数输入到目标浓度预测模型, 得到所述目标检测细胞培养液在所述目标时刻对应的代谢物浓度预测结果;

[0008] 其中, 所述目标浓度预测模型包括多个数据特征提取模块, 所述多个数据特征提取模块分别用于从不同特征分析维度对所述光谱数据和所述预设细胞培养环境参数进行特征提取。

[0009] 第二方面, 本公开提供了一种细胞培养液代谢物浓度确定装置, 该装置包括:

[0010] 数据获取单元, 用于获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数;

[0011] 浓度预测单元, 用于将所述光谱数据和所述预设细胞培养环境参数输入到目标浓度预测模型, 得到所述目标检测细胞培养液在所述目标时刻对应的代谢物浓度预测结果;

- [0012] 其中，所述目标浓度预测模型包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对所述光谱数据和所述预设细胞培养环境参数进行特征提取。
- [0013] 第三方面，本公开还提供了一种计算机设备，所述计算机设备包括：
- [0014] 一个或多个处理器；
- [0015] 存储器，用于存储一个或多个程序；
- [0016] 当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如本公开任意实施例所提供的细胞培养液代谢物浓度确定方法。
- [0017] 第四方面，本公开还提供了一种非瞬态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本公开任意实施例所提供的细胞培养液代谢物浓度确定方法。

附图说明

- [0018] 图1是根据本公开的一些实施例的细胞培养液代谢物浓度确定方法的流程图；
- [0019] 图2是根据本公开的一些实施例的细胞培养液代谢物浓度确定方法的流程图；
- [0020] 图3是根据本公开的一些实施例的目标浓度预测模型的训练过程的示意图；
- [0021] 图4是根据本公开的一些实施例的实验中罐子6的组分趋势变化图；
- [0022] 图5是根据本公开的一些实施例的实验中罐子9的组分趋势变化图；
- [0023] 图6是根据本公开的一些实施例的细胞培养液代谢物浓度确定方法的流程图；
- [0024] 图7是根据本公开的一些实施例的目标浓度预测模型的应用实例的示意图；
- [0025] 图8是根据本公开的一些实施例的自动控制的葡萄糖浓度与蠕动泵匹配的示意图；
- [0026] 图9是根据本公开的一些实施例的手动控制的葡萄糖浓度与蠕动泵匹配的示意图；
- [0027] 图10是根据本公开的一些实施例的细胞培养液代谢物浓度确定装置的结构示意图；
- [0028] 图11是根据本公开的一些实施例的计算机设备的结构示意图。

具体实施方式

- [0029] 下面结合附图和实施例对本公开作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释本公开，而非对本公开的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与本公开相关的部分而非全部结构。
- [0030] 在实现本公开的过程中，发现由于拉曼光谱数据的高维复杂度和数据之间的多样性，采用简单的单模态机器学习或者深度学习(例如文本卷积模型)进行代谢物浓度预测具有一定的局限性，预测误差较大。
- [0031] 本公开提供了一种细胞培养液代谢物浓度确定方法、装置、设备和介质，可以基于被测细胞培养液的多模态的数据进行数据特征分析，得到更加准确的细胞培养液浓度预测结果。
- [0032] 具体地，本公开通过获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数，以及将光谱数据和预设细胞培养环境参数输入到目标浓度预测模型，得到目标检测细胞培养液在目标时刻对应的代谢物浓度预测结果。目标浓度预测模型包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对光谱数据和预设细胞培养环境参数进行特征提取。本公开实施例的技术方案解决了现有细胞培养液浓度分析模型的预测结果不准确的问题，可以基于被测细胞培养液的多模态的数据进行数据特征分析，得到更加准确的细胞培养液浓度预测结果。
- [0033] 图1为根据本公开的一些实施例的细胞培养液代谢物浓度确定方法的流程图。本实施例可适用于生物实验的场景，特别是在生物反应器中进行细胞培养过程中对细胞代谢物浓度进行预测的情况。该方法可以由细胞培养液代谢物浓度确定装置执行，该装置可以由软件和/或硬件的方式来实现，集成于具有应用开发功能的计算机设备中。
- [0034] 如图1所示，本实施例的细胞培养液代谢物浓度确定方法包括以下步骤S110和S120。
- [0035] 在步骤S110处，获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数。

- [0036] 具体地，目标检测细胞培养液可以是任意一个生物反应过程中，需要进行细胞培养液中细胞代谢物浓度分析的细胞培养液。目标时刻可以是目标检测细胞培养液在生物反应过程中的任意一个时刻。
- [0037] 光谱数据可以是任意一种能够反映细胞培养液中不同物质浓度差异的光谱信息。例如，光谱数据可以是拉曼光谱数据。可以通过拉曼光谱仪获得目标检测细胞培养液在目标时刻的在线光谱数据。光谱数据的预设长度可以是基于目标浓度预测模型的训练过程中经过试验确定的一种较优的数据范式，也可以是根据相关技术人员的经验确定的。
- [0038] 在以往的细胞培养液中代谢物浓度检测过程中，通常仅依赖光谱数据进行浓度分析，得到一个代谢物浓度预测结果。但是，考虑到生物反应过程中，细胞培养过程中的环境参数也会对生物反应过程有一定的影响，在本实施例中进行代谢物浓度分析时，还会综合环境参数进行代谢物浓度预测分析预测，以得到一个更加准确的预测结果。可以基于目标检测细胞培养液的不同检测方式获取到的不同维度的多模态数据，对目标检测细胞培养液中代谢物浓度进行分析。
- [0039] 具体地，预设细胞培养环境参数可以包括但不限于细胞系、克隆、培养工艺等、细胞培养实验相关的离散元数据，如细胞株种类、培养基类型、细胞培养过程中的温度和PH等参数。预设细胞培养环境参数可以是上述各参数中的一种或几种的组合。在一些实施例中，预设细胞培养环境参数包括细胞系参数、细胞克隆参数、培养工艺参数及细胞培养过程中各项参数中的至少一种参数。
- [0040] 在步骤S120处，将光谱数据和预设细胞培养环境参数输入到目标浓度预测模型，得到目标检测细胞培养液在目标时刻对应的代谢物浓度预测结果。
- [0041] 具体地，目标浓度预测模型可以是经过预先训练的神经网络模型，其可以基于输入数据对细胞培养液中代谢物浓度进行预测，输出对应的代谢物浓度预测结果。之所以目标浓度预测模型能够更加准确地预测细胞培养液中代谢物浓度，是由于目标浓度预测模型可以包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对光谱数据和预设细胞培养环境参数进行特征提取。

- [0042] 此外，在一些实施例中，目标浓度预测模型还包括特征融合模块和浓度预测模块。特征融合模块用于将所述多个数据特征提取模块分别提取的数据特征进行特征融合，得到目标融合特征。浓度预测模块用于对所述目标融合特征进行分析，得到所述代谢物浓度预测结果。
- [0043] 在一些实施例中，目标浓度预测模型的训练是通过使用拉曼光谱仪从生物反应器的细胞培养液中获得不同组分的浓度按时间变化的在线光谱数据，收集对应的培养环境参数值，同时在实验过程中定时取样并使用离线检测设备获得各批次的离线目标值，对上述结构的模型进行训练确定的。这一过程中例如可以结合为不同数据特性设计的数据特征提取模块和特征融合模块，为整个模型的浓度预测模块提供丰富的上下文信息，从而提高目标浓度预测模型的预测和分析精度。
- [0044] 因此，将获取到的光谱数据和预设细胞培养环境参数输入到目标浓度预测模型，便可以得到目标检测细胞培养液在目标时刻对应的代谢物浓度预测结果。
- [0045] 在一些实施例中，目标浓度预测模型中的数据特征提取模块包括第一数据特征提取模块、第二数据特征提取模块和第三数据特征提取模块。第一数据特征提取模块用于对所述预设细胞培养环境参数进行特征编码分析，得到目标环境参数特征。第二数据特征提取模块用于提取所述光谱数据中不同频段的光谱数据之间的光谱特征。第三数据特征提取模块用于提取所述光谱数据的预设光谱统计量特征、预设谱特征和预设光谱峰值分析特征中的至少一种光谱特征。
- [0046] 这里需要说明的是，将光谱数据的预设光谱统计量(如均值，中位数等)、预设谱特征(如谱能量，谱熵等)，以及预设光谱峰值分析特征(如峰值数量，平均峰值高度，峰值间距离等)加入到算法中来，是因为不仅仅是拉曼设备间光信号的差异很大，同时相同设备内差异也是存在的(使用存在不同程度的漂移)。设备间差异常常需要根据光信号数据建立一套转换的光学物理和数学策略，这种做法耗时耗力且准确性不足。设备内不同通道的差异目前的做法是通过光学标准品对光纤、光学耦合传感器、激光光源进行校准，将不同的光通道校准到相同的量级，这种做法在实验过程中进行则会带来不必要的培养风险例如染菌等。而

本公开实施例通过加入这些统计量等则可以有效消除设备间和设备内的光谱差异，从而可以使预测得到的结果准确度更高。

- [0047] 在一些实施例中，第二数据特征提取模块可以用于通过具有自注意力机制的神经网络模块，提取所述光谱数据中不同频段的光谱数据之间的光谱特征。在对光谱数据进行特征处理的过程中，可以引入自注意力机制。在变换器(Transformer)模型中，自注意力机制作为核心组件，可以提高模型的并行处理能力和理解深度。
- [0048] 采用自注意力机制是基于拉曼光谱数据的特性与自然语言数据存在相似性进行设置的。首先，拉曼光谱中不同频段的光谱强度之间的相关性，可以与自然语言中单词或短语之间的语义和语法关系相类比。在拉曼光谱中，不同频段可能反映了相同或相似化学成分的特性，因此会存在内在的相互关联。同样地，在自然语言文本中，不同的词汇或短语可能存在语义依赖和相互影响。其次，对拉曼光谱和自然语言数据的分析都需要捕捉这些复杂的内部关系。基于自注意力机制的方法正是强调了这一点，通过对不同部分的动态权重分配，可以捕捉拉曼光谱中不同频段之间的相关性，这一特性可类比于自注意力机制善于处理自然语言中的长距离依赖关系的特点。再者，基于自注意力机制的神经网络模型可以有效地突破光谱频段距离限制，有效捕捉代谢物和各个波段光谱值的关联性，这一规律的总结是可以自动进行的，无需依托人为的特征选择等步骤，在一定程度上提升了建模和推理过程的效率。最后，无论是拉曼光谱还是自然语言数据，其结构都可以在一定程度上表示为序列。正因为这一相似性，自注意力机制能够跨越这两个领域，为拉曼光谱的定量分析以及自然语言的深层理解提供了一种统一而高效的方法。基于上述综合数据特征训练得到的目标浓度预测模型，经实际效果的验证，确实可以得到更加准确的预测结果。
- [0049] 本实施例的技术方案，通过获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数；将所述光谱数据和预设细胞培养环境参数输入到目标浓度预测模型，得到所述目标检测细胞培养液在所述目标时刻对应的代谢物浓度预测结果；所述目标浓度预测模型包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对所述光谱数据和预

设细胞培养环境参数进行特征提取。本公开实施例的技术方案解决了现有细胞培养液浓度分析模型的预测结果不准确的问题，可以基于被测细胞培养液的多模态的数据进行数据特征分析，得到更加准确的细胞培养液浓度预测结果，特别是基于自注意力机制对目标检测细胞培养液在目标时刻的预设长度的光谱数据进行分析，能够捕捉到更深层次的数据特征，以反应最终的预测结果。

- [0050] 图2为根据本公开的一些实施例的细胞培养液代谢物浓度确定方法的流程图，本实施例与上述实施例中细胞培养液代谢物浓度确定方法属于同一个发明构思，进一步地描述了目标浓度预测模型训练的过程。该方法可以由细胞培养液代谢物浓度确定装置执行，该装置可以由软件和/或硬件的方式来实现，集成于具有应用开发功能的计算机设备中。
- [0051] 如图2所示，本实施例的细胞培养液代谢物浓度确定方法包括以下步骤S210至S240。
- [0052] 在步骤S210处，获取细胞培养液的不同组分的浓度在不同时间的光谱样本数据、对应的预设细胞培养环境参数样本数据和离线浓度检测值。
- [0053] 例如，得到上述实施例中目标浓度预测模型的过程大致可分为数据处理、模型搭建、数据训练以及模型验证几个环节。
- [0054] 在本步骤中，首先进行数据处理，以构造用于进行模型训练的样本数据。
- [0055] 具体的，可以通过例如使用拉曼光谱仪从生物反应器的细胞培养液中获得细胞培养液的不同组分的浓度随着时间变化的在线光谱数据，并且收集培养环境参数值，同时在实验过程中定时取样并使用离线检测设备获得细胞培养液的不同组分的浓度的离线目标值。
- [0056] 示例性的，光谱数据是通过将预设拉曼光谱仪的探针安装在生物反应器中，直接浸入细胞培养液中获取的。在整个实验过程中记录不同生物反应器的拉曼光谱。对于单个记录的光谱，可以用10s暴露时间捕获30个后续光谱并取平均值，每次生物反应器的采集间隔约为5min。光谱仪工作过程中激光的激发波长为785nm，提供 $100\text{--}3425\text{cm}^{-1}$ 的光谱覆盖(拉曼位移)。
- [0057] 例如，预设细胞培养环境参数样本数据可以包括：细胞在种子阶段摇床中培养时的条件诸如温度、摇床转速、二氧化碳浓度水平、培养基参数，生产培养

阶段使用的反应器大小(如3L和250L)，以及细胞培养液的初始培养体积、培养温度、酸碱度设定值、溶氧饱和度、初始接种密度以及补料培养基参数等。

- [0058] 不同组分的离线浓度检测值可以是按时间每天5次取样后检测的目标值，例如但不限于通过细胞计数仪(诸如Vi-CELL XR)检测的活细胞密度(VCD)、细胞活性(Viability)和平均细胞直径(AvgDiam)；通过生化分析仪(诸如Cedex Bio HT)确定的葡萄糖(glucose)、乳酸(lactate)、铵根离子(NH_4^+)、谷氨酸(glutamate)、谷氨酰胺(glutamine)、铁离子(Fe)、磷酸根离子(Phosphate)、乳酸脱氢酶(LDH)和目标蛋白浓度(titer)；通过渗透压仪(诸如OsmoPRO)检测的渗透压(Osmolality)；以及通过血气分析仪检测的酸度值(pH)、二氧化碳分压(pCO_2)、钠离子(Na^+)和钾离子(K^+)。
- [0059] 在步骤S220处，以所述光谱样本数据和所述预设细胞培养环境参数样本数据作为模型训练输入数据，并以对应时间的所述离线浓度检测值为模型训练标签数据，对预设初始浓度预测模型进行训练，以得到目标浓度预测模型。
- [0060] 每个模型训练样本都是以光谱样本数据和所述预设细胞培养环境参数样本数据作为模型训练输入数据，并以对应时间的离线浓度检测值为模型训练标签数据。光谱的检测时间和离线浓度检测值的检测时间可以一一对应。
- [0061] 为了对获取到的多模态数据进行分析，在构造模型时，预设初始浓度预测模型包括多个数据特征提取模块，用于从不同维度对数据特征的提取，得到更高阶的数据特征。在一些实施例中，数据特征提取模块包括第一数据特征提取模块、第二数据特征提取模块和第三数据特征提取模块；其中，第一数据特征提取模块用于对所述预设细胞培养环境参数进行特征编码分析，得到目标环境参数特征；第二数据特征提取模块用于提取所述光谱数据中不同频段的光谱数据之间的光谱特征；第三数据特征提取模块用于提取所述光谱数据的预设光谱统计量特征、预设谱特征和预设光谱峰值分析特征中的至少一种光谱特征。在一些示例中，第二数据特征提取模块为具有自注意力机制的神经网络模块，可以提取所述光谱数据中不同频段的光谱数据之间的光谱特征。注意力机制的主要思想是，在对序列数据进行处理时，通过给不同位置的输入信号分配不同的权重，使得模型更加关注重要的输入。在深度学习模型中，注意力机制通常是通过

添加额外的网络层实现的，这些层可以学习到如何计算权重，并将这些权重应用于输入信号。

[0062] 示例性的，对预设初始浓度预测模型进行训练的过程可参考如图3所示的过程。在图3中，历史数据库中的数据即为模型训练的样本数据，将样本数据输入到模型中，分别通过三条路径(对应于三个数据特征提取模块)对样本数据进行特征提取。其中，对预设细胞培养环境参数样本数据进行独热编码后，经过双通道的归一化层和全连接网络得到环境参数高阶特征，这对应于第一数据特征提取模块。对光谱信号进行频道扩增，然后经过双通道的归一化层、多头注意力网络、全连接网络再经过一次池化，即可得到分频段光谱高阶特征，这对应于第二数据特征提取模块。此外，还会计算并选择光谱数据的全局特征，经过双通道的归一化层和全连接网络得到全局光谱的高阶特征，这对应于第三数据特征提取模块。

[0063] 其中，多头注意力(Multi-Head Attention)是注意力机制的一种扩展形式，可以在处理序列数据时更有效地提取信息。在多头注意力中，使用多组注意力权重，每组权重可以学习到不同的语义信息，并且每组权重都会产生一个上下文向量。最后，这些上下文向量会被拼接起来，再通过一个线性变换得到最终的输出。

[0064] 进一步的，经过图3中从多个维度对光谱数据和细胞培养环境参数的特征提取，还会对提取到的环境参数高阶特征、分频段光谱高阶特征以及全局光谱高阶特征进行多模态特征融合，并将融合结果经过浓度预测模块进行多检测项组分预测，即预测细胞培养液中各个组分的浓度，得到最终的预测结果。

[0065] 此外，这里需要说明的是，在进行模型训练阶段，除了图3中所示的模型结构(基于注意力的多模态拉曼光谱变换器模型，RAFORMER-多模态)之外，本实施例中还构建了多种机器学习模型，包括偏最小二阶乘回归方法(Partial Least Squares Regression, PLSR)、梯度提升决策树(Gradient Boosting Decision Tree, GBDT)、文本卷积(TextCNN)模型，以及只采用拉曼光谱数据的基于注意力的拉曼光谱变换器模型(RAFORMER-单模态)。进一步的，还基于模型训练样本对构造的各个模型结构进行训练，以及验证和测试比较。具体的，首先，

根据测试集的验证结果以确保训练的各个模型有稳定的推理能力。然后，在测试集表现稳定的基础上，为了确保模型的生产可用，通过湿实验来评估所有的模型效果。湿实验流程描述如下：与历史训练数据相同的培养条件，检测了多批次细胞培养实验，并扫描生物反应器收集拉曼光谱数据，每日收集四个点，同一时间采集样品收集离线数据，收集完整一个培养周期。进而，基于采集到的数据进行模型评估，从模型预测值和离线实测值的均方根误差值来看，基于注意力的多模态拉曼光谱变换器模型精确度更高、表现较优，且该模型未通过超参数选择等步骤去针对单一运用场景调试模型，大大节约了人工干预在建模过程中所需要付出的努力。这一点也提升了RAFORMER的通用性和真正大规模在工业生产中应用的可能性。经过多种模型结构的比较分析，也可以确定RAFORMER-多模态模型在考虑了与细胞培养实验相关的环境参数后性能上的优越性。此外，基于测试生物反应器在培养工艺等方面的多样性，多模态模型相较于单模态模型更有潜力通用于多种克隆和细胞系，可以有效克服单模态模型仅强调对于光谱数据解析的局限性，减少预测精度的损失。

[0066] 此外，还比较了RAFORMER-多模态模型预测的变化趋势与基线模型之间的对比差异。其中，用皮尔逊相关系数来评估这一趋势契合度。通过对比RAFORMER-多模态模型与GBDT和PLSR这两个传统模型在上述不同生物反应器上的表现，对于所有浓度检测项，RAFORMER-多模态模型相对于GBDT和PLSR在皮尔逊相关系数上均有所提升。

[0067] 综上，可以确定本实施例中训练得到的目标浓度预测模型可以在对多模态数据进行分析的基础上，得到更加准确的细胞液代谢物浓度的预测结果。

[0068] 具体的，针对上述模型进行评估的详细过程如下：

[0069] 利用经过训练的RAFORMER-多模态模型、RAFORMER-单模态模型、GBDT模型、TextCNN模型及PLSR模型，分别对选定的两个生物反应器中的细胞培养液中代谢物浓度进行预测分析。其中，两个生物反应器分别标记为罐子6和罐子9。选取这两个生物反应器的评估结果作为结果展示的原因是因为它们分别代表了3L，200L培养流程，以及强化型流加培养和传统的流加培养的

培养工艺，可以较为全面地反映模型在不同条件下的适用性。对于罐子6和罐子9的评估对比结果请分别参见表1和表2：

[0070] 表1：生物反应器罐子6的模型预测值和离线实测值的均方根误差

工艺参数	RAFORMER-多模态	GBDT	PLSR	单位	参数范围
VCD	3.713315	9.701639	5.935412	E6cell/mL	0.70 – 9.00
Glucose	0.843273	1.380617	1.553841	g/L	0.50 -9.00
Iron	37.43598	73.21201	73.89616	mg/L	85-180
Glutamine	0.546186	1.745281	0.940769	mM	0.45 -4.00
pH	0.072274	0.175187	0.12444		6.58 -7.27
pCO ₂	6.738848	11.43872	8.539133	mmHg	12.1 -71.9
Na ⁺	3.652582	6.236039	8.572321	mM	88 -133
K ⁺	3.462971	3.694135	3.80827	mM	3.85 – 8.10
Lactate	0.22798	0.269963	0.198929	g/L	0.15 -2. 45
NH ₄ ⁺	1.058867	1.811047	1.957382	mM	0.45 -9.00
Glutamate	0.610238	2.664974	2.014847	mM	2.70 -15.00
OSMO	21.04932	31.00926	31.21049	mOsm/kg	260 -450
Viability	5.289168	8.086848	6.899572	%	25 – 99.9
AvgDiam	0.737902	2.86519	0.783846	um	11.50-19.50
Titer	0.928424	3.276577	2.882383	g/L	0 – 6.0
LDH	441.2631	1770.916	1993.386	U/L	0 - 6000
Phosphate	71.60987	90.70235	83.46153	mg/L	350 -1200

[0071] 表2：生物反应器罐子9的模型预测值和离线实测值的均方根误差

工艺参数	RAFORMER-多模态	GBDT	PLSR	单位	参数范围
VCD	0.913781	1.880418	1.962796	E6cell/mL	0.70 – 49.00
Glucose	0.54204	1.012293	1.137349	g/L	0.50 -9.00
Iron	2.273163	7.044774	6.938764	mg/L	85 - 180
Glutamine	0.452604	0.886814	0.87953	mM	0.45 -4.00
pH	0.063159	0.080681	0.093612		6.58 -7.27
pCO ₂	12.78213	11.81791	12.53938	mmHg	12.1 -71.9
Na ⁺	6.129993	4.648543	5.108044	mM	88 -133
K ⁺	0.728452	1.113267	1.098925	mM	3.85 – 18.10
Lactate	0.255035	0.224591	0.173402	g/L	0.15 -2. 45
NH ₄ ⁺	0.586976	0.606469	0.726414	mM	0.45 -9.00
Glutamate	0.429346	0.761154	0.643775	mM	2.70 -15.00
OSMO	7.006372	28.54601	19.3163	mOsm/kg	260 -450
Viability	7.032988	2.806012	5.726478	%	25 – 99.9
AvgDiam	0.689337	1.12204	1.820591	um	11.50 -19.50
Titer	0.355118	0.591324	0.449733	g/L	0 – 6.0
LDH	657.2798	352.9957	253.5386	U/L	0 - 6000
Phosphate	34.93441	109.8995	89.16007	mg/L	350 -1200

- [0072] 其中，两个表格中VCD表示活细胞密度，Glucose表示葡萄糖，Iron表示铁离子，Glutamine表示谷氨酰胺，PH表示酸碱度， $p\text{CO}_2$ 表示二氧化碳分压， Na^+ 表示钠离子， K^+ 表示钾离子，Lactate表示乳酸， NH_4^+ 表示铵根离子，Glutamate表示谷氨酸，OSMO表示渗透压，Viability表示细胞活率，AvgDiam表示平均细胞直径，Titer表示目标蛋白浓度，LDH表示乳酸脱氢酶，Phosphate表示磷酸根离子。
- [0073] 通过对比RAFORMER-多模态模型与GBDT和PLSR这两个传统模型在上述两个生物反应器上的表现，对于所有17个检测项，可以确定在罐子6反应器上，RAFORMER-多模态模型相对于GBDT和PLSR在均方根误差上分别平均提升了47.57%和36.86%的精确性；在罐子9反应器上，RAFORMER-多模态模型相对于GBDT和PLSR在均方根误差上分别平均提升了14.67%和17.47%的精确性。具体到每一个检测项，可以发现在一共17个检测项中，罐子6反应器上RAFORMER在16个检测项上为最优表现，罐子9反应器上RAFORMER在12个检测项上有最优表现。最优表现指在RAFORMER-多模态模型和基线模型GBDT与PLSR的比较中取得最小的均方根误差。另外值得一提的是，相较于基线模型GBDT和PLSR，RAFORMER-多模态模型并未通过超参数选择等步骤去针对单一运用场景调试模型，大大节约了人工干预在建模过程中所需要付出的努力。这一点也提升了RAFORMER-多模态模型的通用性和真正大规模在工业生产中应用的可能性。
- [0074] 为了验证多模态模型的优越性，且保证模型表现比较的公平性，还进一步对比了RAFORMER-多模态模型与仅有光谱数据输入的RAFORMER-单模态模型在罐子6和罐子9这两个生物反应器上的预测表现。对于罐子6和罐子9的评估对比结果请分别参见表3和表4：
- [0075] 表3：生物反应器罐子6的模型预测值和离线实测值的均方根误差

工艺参数	RAFORMER-多模态	RAFORMER-单模态	单位	参数范围
VCD	3.713314701	3.172992268	E6cell/mL	0.70 – 49.00
Glucose	0.843273361	0.957784274	g/L	0.50 -9.00
Iron	37.43597591	60.44640728	mg/L	85 - 180
Glutamine	0.546186403	0.574262464	mM	0.45 -4.00
pH	0.072274445	1.282868082		6.58 -7.27
pCO ₂	6.738848485	9.015574411	mmHg	12.1 -71.9
Na ⁺	3.652581943	15.88855617	mM	88 -133
K ⁺	3.46297119	3.577128264	mM	3.85 – 18.10
Lactate	0.227980319	0.300729799	g/L	0.15 -2. 45
NH ₄ ⁺	1.058867455	1.137171702	mM	0.45 -9.00
Glutamate	0.61023846	0.657835621	mM	2.70 -15.00
OSMO	21.04932497	33.04471882	mOsm/kg	260 -450

Viability	5.28916793	9.309222936	%	25 – 99.9
AvgDiam	0.737902106	2.284397101	um	11.50 -19.50
Titer	0.928424497	1.263533082	g/L	0 – 6.0
LDH	441.2630814	522.993953	U/L	0 - 6000
Phosphate	71.60986569	147.5284062	mg/L	350 -1200

[0076] 表4生物反应器罐子9的模型预测值和离线实测值的均方根误差

工艺参数	RAFORMER-多模态	RAFORMER-单模态	单位	参数范围
VCD	0.913780577	1.712740586	E6cell/mL	0.70 – 49.00
Glucose	0.54203976	0.862373849	g/L	0.50 -9.00
Iron	2.273162991	22.83208488	mg/L	85 - 180
Glutamine	0.452603719	0.607982898	mM	0.45 -4.00
pH	0.063159219	1.257158596		6.58 -7.27
pCO ₂	12.78212886	15.37680782	mmHg	12.1 -71.9
Na ⁺	6.12999264	12.65110701	mM	88 -133
K ⁺	0.72845242	1.092446034	mM	3.85 – 18.10
Lactate	0.255034973	0.279460445	g/L	0.15 -2. 45
NH ₄ ⁺	0.586975846	0.528184514	mM	0.45 -9.00
Glutamate	0.429346213	1.320674032	mM	2.70 -15.00
OSMO	7.006372312	22.67719824	mOsm/kg	260 -450
Viability	7.032987838	8.659964682	%	25 – 99.9
AvgDiam	0.689336691	2.726716687	um	11.50 -19.50
Titer	0.355118119	0.526046097	g/L	0 – 6.0
LDH	657.2797856	556.4714381	U/L	0 - 6000
Phosphate	34.93440647	122.9367997	mg/L	350 -1200

[0077] 通过对比RAFORMER–单模态模型与RAFORMER–多模态模型在上述两个生物反应器上的表现，对于所有17个检测项，可以发现在罐子6反应

器上，RAFORMER–多模态模型相对于RAFORMER–单模态模型在均方根误

差上平均提升了30.40%的精确性；在罐子9反应器上，RAFORMER-多模态模型相对于RAFORMER-单模态模型在均方根误差上平均提升了41.75%的精确性。具体到每一个检测项，可以发现在一共17个检测项中，罐子6反应器上RAFORMER-多模态模型在16个检测项上为最优表现，罐子9反应器上RAFORMER-多模态模型在15个检测项上有最优表现。最优表现指在多模态RAFORMER和RAFORMER-单模态模型的比较中取得最小的均方根误差。

[0078] 这一结果可以有利的说明RAFORMER-多模态模型在考虑了与细胞培养实验相关的环境参数后性能上的优越性。此外，基于两个测试生物反应器在培养工艺等方面的多样性，RAFORMER-多模态模型相较于单模态模型更有潜力通用于多种克隆和细胞系，可以有效克服单模态模型仅强调对于光谱数据解析的局限性，减少预测精度的损失。

[0079] 除了评估模型的均方根误差之外，我们也评测了RAFORMER-多模态模型预测的变化趋势与基线模型之间的对比差异。其中用来评估这一趋势契合度的指标我们选择了皮尔逊相关系数。详细的评估结果请参看表5和表6。

[0080] 表5：生物反应器罐子6的模型预测值和离线实测值的皮尔逊相关系数

工艺参数	RAFORMER-多模态模型	GBDT	PLSR
VCD	0.905372297	0.502574985	0.553810369
Glucose	0.953500414	0.736731995	0.92980928
Iron	0.170242406	-0.700499926	-0.738249779
Glutamine	0.944098713	0.715915947	0.931215158
pH	0.846043273	0.604745873	0.599429936
pCO2	0.568536766	-0.100024754	0.243373543
Na+	0.308478187	0.079462401	-0.007145086
K+	0.593442311	0.552581153	0.592347291
Lactate	0.955488716	0.881222719	0.946135467
NH4+	0.558043378	0.481936269	0.480796009
Glutamate	0.83773165	0.396497076	0.660377411
OSMO	0.531828217	0.1289012	0.041187009
Viability	0.949010384	0.80936637	0.908831869
AvgDiam	0.938260499	-0.349261029	0.853397329
Titer	0.989585427	0.910202719	0.959285842
LDH	0.984040009	0.854746669	0.9211994
Phosphate	0.912219423	0.72356352	0.885837973

[0081] 表6：生物反应器罐子9的模型预测值和离线实测值的皮尔逊相关系数

工艺参数	RAFORMER	GBDT	PLSR
VCD	0.974184	0.884372	0.881397
Glucose	0.957969	0.458617	0.835644
Iron	0.983576	0.885184	0.953718
Glutamine	0.976015	0.939541	0.893836
pH	0.92558	0.799699	0.719813
pCO ₂	-0.45484	-0.26878	-0.17165
Na ⁺	0.820508	0.850363	0.80105
K ⁺	0.904431	0.777484	0.853984
Lactate	0.988182	0.92142	0.976415
NH ₄ ⁺	0.963026	0.90085	0.880874
Glutamate	0.987507	0.975352	0.973965
OSMO	0.983426	0.961691	0.973604
Viability	0.97151	0.97526	0.959102
AvgDiam	0.909183	0.783828	0.08637
Titer	0.971853	0.945887	0.983593
LDH	0.99595	0.954765	0.981075
Phosphate	0.993444	0.936647	0.972697

[0082] 通过对比RAFORMER–多模态模型与GBDT和PLSR这两个传统模型在上述两个生物反应器上的表现，对于所有17个检测项，可以发现在罐子6反应器上，RAFORMER–多模态模型相对于GBDT和PLSR在皮尔逊相关系数上分别平均提升了0.34和0.19；在罐子9反应器上，RAFORMER–多模态模型相对于GBDT和PLSR在均方根误差上分别平均提升了0.07和0.08。罐子6和罐子9的典型参数趋势变化图可参考图4和图5(趋势线为模型预测值，叉号标记为离线检测得到的真实值)。

[0083] 在步骤S230处，获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数。

[0084] 具体地，目标检测细胞培养液可以是任意一个生物反应过程中，需要进行细胞培养液中细胞代谢物浓度分析的细胞培养液。目标时刻可以是目标检测细胞培养液在生物反应过程中的任意一个时刻。

[0085] 预设长度的光谱数据和预设细胞培养环境参数，可以是经过数据预处理，与上述模型训练过程中，输入范式相同的模型输入数据。

- [0086] 在步骤S240处，将光谱数据和预设细胞培养环境参数输入到目标浓度预测模型，得到目标检测细胞培养液在目标时刻对应的代谢物浓度预测结果。
- [0087] 经过训练的目标浓度预测模型，可以直接用于相同的应用场景中，对任意生物反应器中需要进行浓度预测的细胞培养液中的细胞代谢物浓度进行预测。
- [0088] 本实施例的技术方案，通过获取细胞培养液的不同组分的浓度在不同时间的光谱样本数据、对应的预设细胞培养环境参数样本数据和离线浓度检测值；以所述光谱样本数据和所述预设细胞培养环境参数样本数据作为模型训练输入数据，并以对应时间的所述离线浓度检测值为模型训练标签数据，对预设初始浓度预测模型进行训练，以得到目标浓度预测模型，其中，目标浓度预测模型包括多个数据特征提取模块，用于从不同维度对数据特征的提取；进而应用训练得到的模型，获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数，将所述光谱数据和预设细胞培养环境参数输入到所述目标浓度预测模型，得到所述目标检测细胞培养液在所述目标时刻对应的代谢物浓度预测结果。本公开实施例的技术方案解决了现有细胞培养液浓度分析模型的预测结果不准确的问题，可以基于被测细胞培养液的多模态的数据进行数据特征分析，得到更加准确的细胞培养液浓度预测结果，特别是基于自注意力机制对目标检测细胞培养液在目标时刻的预设长度的光谱数据进行分析，能够捕捉到更深层次的数据特征，以反应最终的预测结果。
- [0089] 图6为根据本公开的一些实施例的细胞培养液代谢物浓度确定方法的流程图，本实施例与上述实施例中细胞培养液代谢物浓度确定方法属于同一个发明构思，进一步说明了细胞培养液浓度预测结果的应用过程，特别是基于预测结果操作生物反应器的场景。该方法可以由细胞培养液代谢物浓度确定装置执行，该装置可以由软件和/或硬件的方式来实现，集成于具有应用开发功能的计算机设备中。
- [0090] 如图6所示，本实施例的细胞培养液代谢物浓度确定方法包括以下步骤S310至S340。
- [0091] 在步骤S310处，获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数。

- [0092] 在步骤S320处，将光谱数据和预设细胞培养环境参数输入到目标浓度预测模型，得到目标检测细胞培养液在目标时刻对应的代谢物浓度预测结果。
- [0093] 具体地，所述目标浓度预测模型可以包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对光谱数据和预设细胞培养环境参数进行特征提取。在一些实施例中，所述多个数据特征提取模块包括用于通过自注意力机制提取光谱数据中不同频段的光谱数据之间的光谱特征的数据特征提取模块。在一些实施例中，然后可以将多模态的高阶特征进行融合，以预测目标检测细胞培养液在目标时刻对应的代谢物浓度预测结果。
- [0094] 在步骤S330处，根据代谢物浓度预测结果中的至少一种物质成分的浓度预测结果，确定针对目标检测细胞培养液所在目标生物反应器进行操作的控制策略。
- [0095] 例如，代谢物浓度预测结果可以包括二氧化碳浓度以及各种营养物质浓度，还有细胞浓度等信息。还可以根据这些浓度信息确定渗透压状态、温度状态以及通气搅拌状态等信息。当各项浓度值对应的调整参数指标不符合标准条件时，便可进行调整，从而生成控制策略。
- [0096] 在一些示例中，生物反应器的控制策略可以包括培养环境控制策略和补料策略。例如，培养环境控制策略可以包含二氧化碳分压策略和降温策略等。
- [0097] 示例性的，细胞培养过程中，二氧化碳既是细胞的代谢产物，也是细胞生长的必需成分，又与维持培养液的酸碱度有关。在细胞培养过程中随着二氧化碳代谢释放量增多，培养液会偏酸性，添加碱性溶液或增大通气维持培养液的酸碱度。反之培养液会偏碱性，此时可主动通入二氧化碳维持酸碱度。通过代谢物浓度预测结果中的二氧化碳浓度和酸碱度等参数，可以自动地根据相应的策略增加或减少二氧化碳浓度，这样调节更具有实时性，调节的准确度更高。
- [0098] 再如，细胞培养过程中当活细胞密度达到一定的量后，主动降低培养温度可以有效地维持细胞活率和提高目标蛋白的产量。可以根据代谢物浓度预测结果中的细胞密度等参数及时调整温度。
- [0099] 此外，细胞培养过程中需要根据细胞的营养消耗速率补加相应的营养物质。由于没有实时反馈的离线检测数据，补料策略只能根据历史数据和实验人员的经验设定细胞培养周期过程中的补料次数、补料比例、葡萄糖浓度。操作人员按

照经验公式计算获得补料参数后，人工依靠秤与蠕动泵手动将补料补加入生物反应器中。基于本实施例中的实时代谢物浓度预测结果，可以根据预设补料规则，及时进行补充，提高细胞培养的成功率。

- [0100] 在步骤S340处，基于控制策略操作目标生物反应器。
- [0101] 将上述步骤确定的控制策略，同步到目标生物反应器中进行执行便可以实现相应的控制目标，以确保生物反应的顺利进行。具体可参考从图3中所示的过程，经过多检测项组分预测后，可以基于预测结果生成相应的控制策略，将控制策略对应的指令下发到目标生物反应器中进行反馈控制。与此同时，还可以对生物反应器中细胞培养液进行抽样检测，以增加数据库中历史数据。
- [0102] 在一些实例中，可以通过物联网平台实时传输拉曼光谱数据至配置有自注意力编码器模型(目标浓度预测模型)的计算机设备(可以是边缘计算网关设备)，通过模型预测组分指标生成各项指标的状态值，根据状态值对应自动化控制生物反应器，如自动反馈控制生物反应器中的葡萄糖浓度。在一些实施例中，可以将目标浓度预测模型进行边缘设备的部署，还集成拉曼光谱探测仪、拉曼工控机、边缘计算网关、物联网平台、工控机和生物反应器组成人工智能物联网装置进行生物培养器的监测和反控，如图7所示。人工智能物联网装置可实时抓取拉曼光谱数据来预测各种组分的浓度，将浓度结合其他相关参数计算出状态值，对应自动化控制策略来实现生物反应器中细胞培养的自动化的控制。
- [0103] 进一步的，通过图7所述的系统进行生物反应过程控制以及通过人工进行生物反应过程控制中，葡萄糖浓度变化的对比可参考图8和图9。图8中展示了8天时间内，模型反馈控制的葡萄糖浓度与蠕动泵匹配的示意图。模型反馈控制的葡萄糖遵循设定值4.0g/L。图9中展示了11天时间内，手动控制的葡萄糖遵循实验设计的6-8g范围的示意图。11天时间内手动控制补料7次总量270g，自动多次少量补料总量216g，所以在实时控制的同时，减少了补料的量约20%。目标蛋白的产量并未减少，反而增加了约4.03%。可以确定基于本实施例中目标浓度预测模型的用于自动化控制生物反应器的控制策略更优。
- [0104] 本实施例的技术方案，通过获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数；将所述光谱数据和预设细胞培养环境参数

输入到目标浓度预测模型，得到所述目标检测细胞培养液在所述目标时刻对应的代谢物浓度预测结果；根据所述代谢物浓度预测结果中的至少一种物质成分的浓度预测结果，确定针对所述目标检测细胞培养液所在目标生物反应器进行操作的策略；基于所述策略操作所述目标生物反应器。本公开实施例的技术方案解决了现有细胞培养液浓度分析模型的预测结果不准确的问题，可以基于被测细胞培养液的多模态的数据进行数据特征分析，得到更加准确的细胞培养液浓度预测结果，特别是基于自注意力机制对目标检测细胞培养液在目标时刻的预设长度的光谱数据进行分析，能够捕捉到更深层次的数据特征，以反应最终的预测结果。进而可以实时地、更加准确地、自动化地调节生物反应器，是生物反应顺利进行。

- [0105] 图10为根据本公开的一些实施例的细胞培养液代谢物浓度确定装置的结构示意图，本实施例可适用于生物实验的场景，特别是在生物反应器中需要在细胞培养过程中对细胞代谢物浓度进行预测的情况，该细胞培养液代谢物浓度确定装置可以由软件和/或硬件的方式来实现，集成于具有应用开发功能的计算机终端设备中。
- [0106] 如图10所示，细胞培养液代谢物浓度确定装置包括：数据获取单元410和浓度预测单元420。
- [0107] 数据获取单元410用于获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数。浓度预测单元420用于将所述光谱数据和所述预设细胞培养环境参数输入到目标浓度预测模型，得到所述目标检测细胞培养液在所述目标时刻对应的代谢物浓度预测结果。所述目标浓度预测模型包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对所述光谱数据和所述预设细胞培养环境参数进行特征提取。
- [0108] 本实施例的技术方案，通过获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数；将光谱数据和预设细胞培养环境参数输入到目标浓度预测模型，得到目标检测细胞培养液在目标时刻对应的代谢物浓度预测结果；目标浓度预测模型包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对光谱数据和预设细胞培养环境参数进

行特征提取。本公开实施例的技术方案解决了现有细胞培养液浓度分析模型的预测结果不准确的问题，可以基于被测细胞培养液的多模态的数据进行数据特征分析，得到更加准确的细胞培养液浓度预测结果。

- [0109] 在一些实施例中，所述多个数据特征提取模块包括第一数据特征提取模块、第二数据特征提取模块和第三数据特征提取模块；
- [0110] 其中，所述第一数据特征提取模块用于对所述预设细胞培养环境参数进行特征编码分析，得到目标环境参数特征；
- [0111] 所述第二数据特征提取模块用于提取所述光谱数据中不同频段的光谱数据之间的光谱特征；
- [0112] 所述第三数据特征提取模块用于提取所述光谱数据的预设光谱统计量特征、预设谱特征和预设光谱峰值分析特征中的至少一种光谱特征。
- [0113] 在一些实施例中，所述第二数据特征提取模块用于：
- [0114] 通过具有自注意力机制的神经网络模块，提取所述光谱数据中不同频段的光谱数据之间的光谱特征。
- [0115] 在一些实施例中，所述预设细胞培养环境参数包括细胞系参数、细胞克隆参数、培养工艺参数及细胞培养过程中各项参数中的至少一种参数。
- [0116] 在一些实施例中，所述目标浓度预测模型还包括特征融合模块和浓度预测模块；
- [0117] 其中，所述特征融合模块，用于将所述多个数据特征提取模块分别提取的数据特征进行特征融合，得到目标融合特征；
- [0118] 所述浓度预测模块，用于对所述目标融合特征进行分析，得到所述代谢物浓度预测结果。
- [0119] 在一些实施例中，细胞培养液代谢物浓度确定装置还包括模型训练模块，用于训练得到目标浓度预测模型，所述目标浓度预测模型的训练过程包括：
- [0120] 获取细胞培养液的不同组分的浓度在不同时间的光谱样本数据、对应的预设细胞培养环境参数样本数据和离线浓度检测值；

- [0121] 以所述光谱样本数据和所述预设细胞培养环境参数样本数据作为模型训练输入数据，并以对应时间的所述离线浓度检测值为模型训练标签数据，对预设初始浓度预测模型进行训练，以得到所述目标浓度预测模型。
- [0122] 在一些实施例中，细胞培养液代谢物浓度确定装置还包括反应器操作策略确定模块，用于：
- [0123] 根据所述代谢物浓度预测结果中的至少一种物质成分的浓度预测结果，确定针对所述目标检测细胞培养液所在目标生物反应器进行操作的控制策略。
- [0124] 本公开实施例所提供的细胞培养液代谢物浓度确定装置可执行本公开任意实施例所提供的细胞培养液代谢物浓度确定方法，具备执行方法相应的功能模块和有益效果。
- [0125] 图11为根据本公开的一些实施例的计算机设备的结构示意图。图11示出了适于用来实现本公开实施例的示例性计算机设备12的框图。图11显示的计算机设备12仅仅是一个示例，不应对本公开实施例的功能和使用范围带来任何限制。计算机设备12可以是任意具有计算能力的终端设备，如智能控制器及服务器、手机等终端设备。
- [0126] 如图11所示，计算机设备12以通用计算设备的形式表现。计算机设备12的组件可以包括但不限于：一个或者多个处理器或者处理单元16，系统存储器28，连接不同系统组件(包括系统存储器28和处理单元16)的总线18。
- [0127] 总线18表示几类总线结构中的一种或多种，包括存储器总线或者存储器控制器，外围总线，图形加速端口，处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说，这些体系结构包括但不限于工业标准体系结构(Industrial Standard Architecture, ISA)总线，微通道体系结构(Micro-Channel Architecture, MAC)总线，增强型ISA总线、视频电子标准协会(Video Electronics Standards Association, VESA)局域总线以及外围组件互连(Peripheral Component Interconnect, PCI)总线。
- [0128] 计算机设备12典型地包括多种计算机系统可读介质。这些介质可以是任何能够被计算机设备12访问的可用介质，包括易失性和非易失性介质，可移动的和不可移动的介质。

- [0129] 系统存储器28可以包括易失性存储器形式的计算机系统可读介质，例如随机存取存储器(Random Access Memory, RAM)30和/或高速缓存存储器32。计算机设备12可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例，存储系统34可以用于读写不可移动的、非易失性磁介质(图11未显示，通常称为“硬盘驱动器”)。尽管图11中未示出，可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器，以及对可移动非易失性光盘(例如紧凑磁盘只读存储器(Compact Disc Read-Only Memory, CD-ROM)，数字多功能盘只读存储器(Digital Versatile Disc Read-Only Memory, DVD-ROM)或者其它光介质)读写的光盘驱动器。在这些情况下，每个驱动器可以通过一个或者多个数据介质接口与总线18相连。系统存储器28可以包括至少一个程序产品，该程序产品具有一组(例如至少一个)程序模块，这些程序模块被配置以执行本公开各实施例的功能。
- [0130] 具有一组(至少一个)程序模块42的程序/实用工具40，可以存储在例如系统存储器28中，这样的程序模块42包括但不限于操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块42通常执行本公开所描述的实施例中的功能和/或方法。
- [0131] 计算机设备12也可以与一个或多个外部设备14(例如键盘、指向设备、显示器24等)通信，还可与一个或者多个使得用户能与该计算机设备12交互的设备通信，和/或与使得该计算机设备12能与一个或多个其它计算设备进行通信的任何设备(例如网卡，调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口22进行。并且，计算机设备12还可以通过网络适配器20与一个或者多个网络(例如局域网(Local Area Network, LAN)，广域网(Wide Area Network, WAN)和/或公共网络，例如因特网)通信。如图所示，网络适配器20通过总线18与计算机设备12的其它模块通信。应当明白，尽管图11中未示出，可以结合计算机设备12使用其它硬件和/或软件模块，包括但不限于：微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、冗余独立磁盘阵列(Redundant Array of Independent Disks, RAID)系统、磁带驱动器以及数据备份存储系统等。

- [0132] 处理单元16通过运行存储在系统存储器28中的程序，从而执行各种功能应用以及数据处理，例如实现本公开实施例所提供的细胞培养液代谢物浓度确定方法，该方法包括：
- [0133] 获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数；
- [0134] 将所述光谱数据和预设细胞培养环境参数输入到目标浓度预测模型，得到所述目标检测细胞培养液在所述目标时刻对应的代谢物浓度预测结果；
- [0135] 其中，所述目标浓度预测模型包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对所述光谱数据和预设细胞培养环境参数进行特征提取。
- [0136] 本公开实施例还提供了一种非瞬态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本公开任意实施例所提供的细胞培养液代谢物浓度确定方法，该方法包括：
- [0137] 获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数；
- [0138] 将所述光谱数据和预设细胞培养环境参数输入到目标浓度预测模型，得到所述目标检测细胞培养液在所述目标时刻对应的代谢物浓度预测结果；
- [0139] 其中，所述目标浓度预测模型包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对所述光谱数据和预设细胞培养环境参数进行特征提取。
- [0140] 本公开实施例的计算机存储介质，可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是但不限于：电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器(RAM)、只读存储器(Read-Only Memory, ROM)、可擦式可编程只读存储器(Erasable Programmable Read-Only Memory, EPROM或闪存)、光纤、便携式紧凑磁盘只读存储器(CD-

ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

- [0141] 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。
- [0142] 计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：无线、电线、光缆、射频(Radio Frequency, RF)等等，或者上述的任意合适的组合。
- [0143] 可以以一种或多种程序设计语言或其组合来编写用于执行本公开操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言，诸如Java、Smalltalk、C++，还包括常规的过程式程序设计语言，诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络，包括局域网(LAN)或广域网(WAN)，连接到用户计算机，或者，可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。
- [0144] 本领域普通技术人员应该明白，上述的本公开的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个计算装置上，或者分布在多个计算装置所组成的网络上，可选地，他们可以用计算机装置可执行的程序代码来实现，从而可以将它们存储在存储装置中由计算装置来执行，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本公开不限制于任何特定的硬件和软件的结合。

[0145] 注意，上述仅为本公开的较佳实施例及所运用技术原理。本领域技术人员会理解，本公开不限于这里所述的特定实施例，对本领域技术人员来说能够进行各种明显的变化、重新调整和替代而不会脱离本公开的保护范围。因此，虽然通过以上实施例对本公开进行了较为详细的说明，但是本公开不仅仅限于以上实施例，在不脱离本公开构思的情况下，还可以包括更多其他等效实施例，而本公开的范围由所附的权利要求范围决定。

权利要求书

- [权利要求 1] 一种细胞培养液代谢物浓度确定方法，包括：
- 获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数；
- 将所述光谱数据和所述预设细胞培养环境参数输入到目标浓度预测模型，得到所述目标检测细胞培养液在所述目标时刻对应的代谢物浓度预测结果；
- 其中，所述目标浓度预测模型包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对所述光谱数据和所述预设细胞培养环境参数进行特征提取。
- [权利要求 2] 根据权利要求1所述的方法，其中，所述多个数据特征提取模块包括第一数据特征提取模块、第二数据特征提取模块和第三数据特征提取模块；
- 其中，所述第一数据特征提取模块用于对所述预设细胞培养环境参数进行特征编码分析，得到目标环境参数特征；
- 所述第二数据特征提取模块用于提取所述光谱数据中不同频段的光谱数据之间的光谱特征；
- 所述第三数据特征提取模块用于提取所述光谱数据的预设光谱统计量特征、预设谱特征和预设光谱峰值分析特征中的至少一种光谱特征。
- [权利要求 3] 根据权利要求2所述的方法，其中，所述第二数据特征提取模块用于：
- 通过具有自注意力机制的神经网络模块，提取所述光谱数据中不同频段的光谱数据之间的光谱特征。
- [权利要求 4] 根据权利要求1所述的方法，其中，所述预设细胞培养环境参数包括细胞系参数、细胞克隆参数、培养工艺参数及细胞培养过程中各项参数中的至少一种参数。

- [权利要求 5] 根据权利要求1-4中任一项所述的方法，其中，所述目标浓度预测模型还包括特征融合模块和浓度预测模块；
其中，所述特征融合模块用于将所述多个数据特征提取模块分别提取的数据特征进行特征融合，得到目标融合特征；
所述浓度预测模块用于对所述目标融合特征进行分析，得到所述代谢物浓度预测结果。
- [权利要求 6] 根据权利要求1-4中任一项所述的方法，其中，所述目标浓度预测模型的训练过程包括：
获取细胞培养液的不同组分的浓度在不同时间的光谱样本数据、对应的预设细胞培养环境参数样本数据和离线浓度检测值；
以所述光谱样本数据和所述预设细胞培养环境参数样本数据作为模型训练输入数据，并以对应时间的所述离线浓度检测值为模型训练标签数据，对预设初始浓度预测模型进行训练，以得到所述目标浓度预测模型。
- [权利要求 7] 根据权利要求1-4中任一项所述的方法，所述方法还包括：
根据所述代谢物浓度预测结果中的至少一种物质成分的浓度预测结果，确定针对所述目标检测细胞培养液所在目标生物反应器进行操作的策略。
- [权利要求 8] 一种细胞培养液代谢物浓度确定装置，包括：
数据获取单元，用于获取目标检测细胞培养液在目标时刻的预设长度的光谱数据和预设细胞培养环境参数；
浓度预测单元，用于将所述光谱数据和所述预设细胞培养环境参数输入到目标浓度预测模型，得到所述目标检测细胞培养液在所述目标时刻对应的代谢物浓度预测结果；
其中，所述目标浓度预测模型包括多个数据特征提取模块，所述多个数据特征提取模块分别用于从不同特征分析维度对所述光谱数据和所述预设细胞培养环境参数进行特征提取。
- [权利要求 9] 一种计算机设备，其中，所述计算机设备包括：

一个或多个处理器；

存储器，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求1-7中任一项所述的细胞培养液代谢物浓度确定方法。

[权利要求 10]

一种非瞬态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权利要求1-7中任一项所述的细胞培养液代谢物浓度确定方法。