



(12) 发明专利

(10) 授权公告号 CN 116631572 B

(45) 授权公告日 2023. 12. 19

(21) 申请号 202310908402.0

G06F 18/213 (2023.01)

(22) 申请日 2023.07.24

G06F 18/243 (2023.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 116631572 A

(56) 对比文件

CN 116245019 A, 2023.06.09

CN 116434960 A, 2023.07.14

(43) 申请公布日 2023.08.22

CN 110021397 A, 2019.07.16

(73) 专利权人 中国人民解放军总医院

AU 2020103309 A4, 2021.01.14

地址 100853 北京市海淀区复兴路28号

CN 113158555 A, 2021.07.23

(72) 发明人 何昆仑 孙宇慧

审查员 张鑫

(74) 专利代理机构 北京预立生科知识产权代理

有限公司 11736

专利代理师 李红伟

(51) Int. Cl.

G16H 20/10 (2018.01)

G16H 50/20 (2018.01)

G16H 70/40 (2018.01)

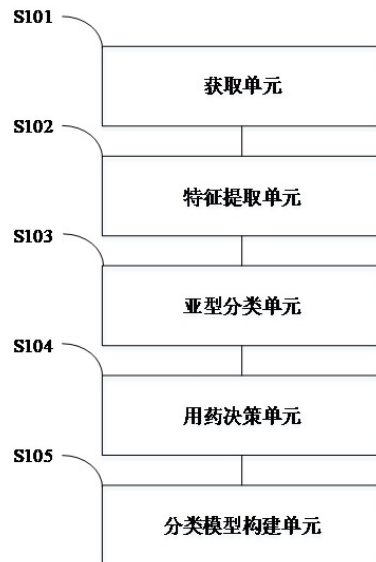
权利要求书2页 说明书11页 附图5页

(54) 发明名称

基于人工智能的急性心肌梗死临床决策支持系统及设备

(57) 摘要

本发明提供了基于人工智能的急性心肌梗死临床决策支持系统及设备,涉及智能医疗领域。所述系统包括:获取单元、特征提取单元、亚型分类单元、用药决策单元和分类模型构建单元。所述单元依次用于获取患者样本的遗传信息、特征提取、输入分类模型进行分类、基于分类结果给予用药建议和构建分类模型。本发明公开的急性心肌梗死临床决策支持系统,该系统基于患者的遗传信息给予用药建议,为急性心肌梗死临床诊断提供了分子指标;该系统使用自适应剪枝优化的随机森林分类算法对数据集进行分类,提高了急性心肌梗死临床决策的效率,减少患者等待时间以及医生的工作负担。



1. 一种基于人工智能的急性心肌梗死临床决策支持系统,其特征在于,所述系统包括:
 获取单元,用于获取患者样本的遗传信息;
 特征提取单元,用于对所述遗传信息进行特征提取,得到遗传特征;
 亚型分类单元,用于基于所述遗传特征输入分类模型对所述患者进行分类,判断患者是否为药物敏感型;

用药决策单元,用于基于患者的分类结果给予用药建议,当患者为药物敏感型时,用药建议为给予心肌梗死药物治疗;

分类模型构建单元,用于构建所述分类模型,所述分类模型的构建方法包括:

获取急性心肌梗死患者的遗传信息和分类标签;

对所述遗传信息进行特征提取,得到遗传特征;

将所述遗传特征输入深度森林中进行模型构建,得到初步分类结果,将所述初步分类结果与所述分类标签比对,生成损失函数,基于损失函数优化机器学习模型,得到训练好的分类模型;

所述深度森林仅在最后一层与原始输入进行级联,并对深度森林算法中的决策树进行自适应剪枝;

所述决策树进行自适应剪枝的过程包括:

步骤1:从叶节点开始逐层向上选择每个内部节点;

步骤2:计算节点 r_t 的损失函数;

步骤3:计算决策树的整体损失函数;

步骤4:对于每个节点,计算其在验证集上的损失函数值,基于损失函数值决定是否进行剪枝。

2. 根据权利要求1所述的基于人工智能的急性心肌梗死临床决策支持系统,其特征在于,所述节点 r_t 的损失函数为:

$$C_{\alpha}(r_t) = C(r_t) + \alpha|rT_{r_t}|$$

其中, $C(r_t)$ 表示节点 r_t 的训练误差, $|rT_{r_t}|$ 表示节点 r_t 的叶节点个数, α 是一个非负超参数,用于平衡训练误差和叶节点个数。

3. 根据权利要求1所述的基于人工智能的急性心肌梗死临床决策支持系统,其特征在于,所述决策树的整体损失函数为:

$$C_{\alpha}(rT) = C(rT) + \alpha|rT|$$

其中, $C(rT)$ 表示决策树 rT 的训练误差, $|rT|$ 表示决策树 rT 的叶节点个数,通过调整超参数 α ,控制剪枝程度。

4. 根据权利要求1所述的基于人工智能的急性心肌梗死临床决策支持系统,其特征在于,所述心肌梗死药物包括以下任意一种或几种:Epinephrine、Dobutamine、Norepinephrine、Aramine、Isoproterenol。

5. 根据权利要求1所述的基于人工智能的急性心肌梗死临床决策支持系统,其特征在于,采用Lasso回归对所述遗传信息进行特征提取。

6. 根据权利要求1所述的基于人工智能的急性心肌梗死临床决策支持系统,其特征在于,所述遗传特征包括以下基因中的一种或几种:NISCH、PLAGL2、TLR1、ARID5B、TAF9B、

SERINC1、CARKD、EPG5、ANKLE2、BC043227、FAM188A、ERVH.4、ZNF26、ERO1LB、TMEM208、EXOC2、SGMS1、BRPF3。

7. 根据权利要求1所述的基于人工智能的急性心肌梗死临床决策支持系统,其特征在于,所述遗传特征为NISCH和PLAGL2,基于所述NISCH和PLAGL2对患者进行分类,判断患者是否为药物敏感型。

8. 一种基于人工智能的急性心肌梗死临床决策支持设备,其特征在于,所述设备包括存储器和处理器;

所述存储器用于存储程序指令;

所述处理器用于调用程序指令,当程序指令被执行时,用于执行以下基于人工智能的急性心肌梗死临床决策支持方法,包括:

获取患者样本的遗传信息;

对所述遗传信息进行特征提取,得到遗传特征;

将所述遗传特征输入分类模型对所述患者进行分类,判断患者是否为药物敏感型;

基于患者的分类结果给予用药建议,当患者为药物敏感型时,用药建议为给予心肌梗死药物治疗;

构建所述分类模型,所述分类模型的构建方法包括:

获取急性心肌梗死患者的遗传信息和分类标签;

对所述遗传信息进行特征提取,得到遗传特征;

将所述遗传特征输入深度森林中进行模型构建,得到初步分类结果,将所述初步分类结果与所述分类标签比对,生成损失函数,基于损失函数优化机器学习模型,得到训练好的分类模型;

所述深度森林仅在最后一层与原始输入进行级联,并对深度森林算法中的决策树进行自适应剪枝的随机森林;

所述决策树进行自适应剪枝的过程包括:

步骤1:从叶节点开始逐层向上选择每个内部节点;

步骤2:计算节点rt的损失函数;

步骤3:计算决策树的整体损失函数;

步骤4:对于每个节点,计算其在验证集上的损失函数值,基于损失函数值决定是否进行剪枝。

9. 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,所述计算机程序被处理器执行时实现权利要求8中基于人工智能的急性心肌梗死临床决策支持设备中的基于人工智能的急性心肌梗死临床决策支持方法。

基于人工智能的急性心肌梗死临床决策支持系统及设备

技术领域

[0001] 本发明属于智能医疗领域,更具体地,涉及一种基于人工智能的急性心肌梗死临床决策支持系统、设备、计算机可读存储介质及其应用。

背景技术

[0002] 急性心肌梗死(Acute Myocardial Infarction,简称AMI)是心肌梗死的一种临床表现,是一种心血管急症,属于冠心病的一种表现形式。与稳定性心绞痛相比,急性心肌梗死的特点是冠状动脉痉挛或阻塞导致心肌缺血坏死的过程突然发生,并伴有明显的心肌损伤和心功能障碍。目前急性心肌梗死的临床决策主要基于患者的年龄、性别、合并症、新功能、症状、临床表现、时间窗口、冠脉病变情况等方面做出的。

[0003] 随着人工智能和生物信息学的发展,收集并分析患者分子层面的临床信息,并基于这些信息为患者提供个性化的临床决策成为了可能。其中,分子层面的临床信息是指分析患者的基因组、转录组、蛋白质组、代谢组等分子水平的数据获得的个性化信息。通过对这些数据的收集和分析,我们可以更好地了解患者的疾病风险、病理机制和治疗反应,从而为患者制定更加精确和有效的治疗方案。

[0004] 人工智能在这个过程中发挥着关键作用。通过机器学习技术,人工智能可以从大量的分子数据中提取有价值的信息,并构建预测模型和分类算法,以帮助医生做出临床决策。例如,基于患者的基因组数据,人工智能可以预测患者对某种药物的反应,从而指导药物选择和剂量调整。另外,人工智能还可以通过整合多种类型的分子数据,帮助医生发现新的生物标志物,以用于疾病的早期诊断和预测。

发明内容

[0005] 本申请依托人工智能和生物信息学技术的进步,收集并分析急性心肌梗死患者分子层面的临床信息,预测患者对不同药物的临床反应,为患者提供个性化的用药决策,以改善疾病的诊断和治疗效果,提高患者的生活质量。

[0006] 本发明公开一种基于人工智能的急性心肌梗死临床决策支持系统,所述系统包括:

[0007] 获取单元,用于获取患者样本的遗传信息;

[0008] 特征提取单元,用于对所述遗传信息进行特征提取,得到遗传特征;

[0009] 亚型分类单元,用于基于所述遗传特征输入分类模型对所述患者进行分类,判断患者是否为药物敏感型;

[0010] 用药决策单元,用于基于患者的分类结果给予用药建议,当患者为药物敏感型时,用药建议为给予心肌梗死药物治疗;

[0011] 分类模型构建单元,用于构建所述分类模型,所述分类模型的构建方法包括:

[0012] 获取急性心肌梗死患者的遗传信息和分类标签;

[0013] 对所述遗传信息进行特征提取,得到遗传特征;

[0014] 将所述遗传特征输入随机森林中进行模型构建,得到初步分类结果,将所述初步分类结果与所述分类标签比对,生成损失函数,基于损失函数优化机器学习模型,得到训练好的分类模型;

[0015] 所述随机森林为对随机森林算法中的决策树进行自适应剪枝的随机森林。

[0016] 进一步,所述决策树进行自适应剪枝的过程包括:

[0017] 步骤1:从叶节点开始逐层向上选择每个内部节点;

[0018] 步骤2:计算节点 r_t 的损失函数;

[0019] 步骤3:计算决策树的整体损失函数;

[0020] 步骤4:对于每个节点,计算其在验证集上的损失函数值,基于损失函数值决定是否进行剪枝。

[0021] 进一步,所述节点 r_t 的损失函数为:

$$[0022] \quad C_{\alpha}(r_t) = C(r_t) + \alpha|rT_{rt}|$$

[0023] 其中, $C(r_t)$ 表示节点 r_t 的训练误差, $|rT_{rt}|$ 表示节点 r_t 的叶节点个数, α 是一个非负超参数,用于平衡训练误差和叶节点个数。

[0024] 进一步,所述决策树的整体损失函数为:

$$[0025] \quad C_{\alpha}(rT) = C(rT) + \alpha|rT|$$

[0026] 其中, $C(rT)$ 表示决策树 rT 的训练误差, $|rT|$ 表示决策树 rT 的叶节点个数,通过调整超参数 α ,控制剪枝程度。

[0027] 进一步,所述心肌梗死药物包括以下任意一种或几种:Epinephrine、Dobutamine、Norepinephrine、Aramine、Isoproterenol。

[0028] 进一步,采用Lasso回归对所述遗传信息进行特征提取。

[0029] 进一步,所述遗传特征包括以下基因中的一种或几种:NISCH、PLAGL2、TLR1、ARID5B、TAF9B、SERINC1、CAR KD、EPG5、ANKLE2、BC043227、FAM188A、ERVH.4、ZNF26、ER01LB、TMEM208、EXOC2、SGMS1、BRPF3。

[0030] 进一步,所述遗传特征为NISCH和PLAGL2,基于所述NISCH和PLAGL2对患者进行分类,判断患者是否为药物敏感型。

[0031] 一种基于人工智能的急性心肌梗死临床决策支持设备,所述设备包括存储器和处理器;

[0032] 所述存储器用于存储程序指令;

[0033] 所述处理器用于调用程序指令,当程序指令被执行时,用于执行以下基于人工智能的急性心肌梗死临床决策支持方法,包括:

[0034] 获取患者样本的遗传信息;

[0035] 对所述遗传信息进行特征提取,得到遗传特征;

[0036] 将所述遗传特征输入分类模型对所述患者进行分类,判断患者是否为药物敏感型;

[0037] 基于患者的分类结果给予用药建议,当患者为药物敏感型时,用药建议为给予心肌梗死药物治疗;

[0038] 构建所述分类模型,所述分类模型的构建方法包括:

- [0039] 获取急性心肌梗死患者的遗传信息和分类标签；
- [0040] 对所述遗传信息进行特征提取,得到遗传特征；
- [0041] 将所述遗传特征输入随机森林中进行模型构建,得到初步分类结果,将所述初步分类结果与所述分类标签比对,生成损失函数,基于损失函数优化机器学习模型,得到训练好的分类模型；
- [0042] 所述随机森林为对随机森林算法中的决策树进行自适应剪枝的随机森林。
- [0043] 一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现基于人工智能的急性心肌梗死临床决策支持设备中的基于人工智能的急性心肌梗死临床决策支持方法。
- [0044] 本发明的优点：
- [0045] 1. 本申请公开了一种基于人工智能的急性心肌梗死临床决策支持系统,该系统通过获取患者遗传信息并进行特征提取,基于提取的遗传特征对患者进行分类,判断患者是否为药物敏感型,相比传统的临床判断方法,该系统能够基于患者的遗传信息判断患者是否适合心肌梗死药物治疗。
- [0046] 2. 本申请公布的系统能够根据患者的分类结果,给出个性化的用药建议。当患者为药物敏感型时,系统建议给予心肌梗死药物治疗,提高治疗效果,这种个性化的用药建议可以更好地满足患者的特定需求,避免不必要的药物治疗。
- [0047] 3. 本申请公布的系统所使用的随机森林分类算法中的决策树进行自适应剪枝优化,具备快速决策的能力,可以在短时间内对患者进行分类和给出用药建议,该算法的改进帮助决策支持系统提高了临床决策的效率,减少患者等待时间以及医生的工作负担。
- [0048] 4. 本申请公布的系统通过机器学习模型对患者进行分类,能够自动学习和优化模型,具有很高的可扩展性和灵活性,可以根据需要添加更多的遗传特征和基因,以及调整机器学习模型的算法,不断优化和改进系统,适应不同的临床需求和研究进展。

附图说明

- [0049] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获取其他的附图。
- [0050] 图1是一种本发明实施例提供的基于人工智能的急性心肌梗死临床决策支持系统示意图；
- [0051] 图2是一种本发明实施例提供的18个重要基因相关性分析示意图；
- [0052] 图3是一种本发明实施例提供的基因NISCH和PLAGL2在对照组、药物敏感组和药物耐受组三组间基因表达量差异的箱线图,其中A是基因NISCH,B是基因PLAGL2；
- [0053] 图4是一种本发明实施例提供的改进前的深度森林分类模型图解的示意图；
- [0054] 图5是一种本发明实施例提供的改进后的深度森林分类模型图解的示意图。
- [0055] 图6是一种本发明实施例提供的基于人工智能的急性心肌梗死临床决策支持设备示意图；
- [0056] 图7是一种本发明实施例提供的基于人工智能的急性心肌梗死临床决策支持方法

流程示意图；

具体实施方式

[0057] 为了使本技术领域的人员更好地理解本发明方案，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述。

[0058] 在本发明的说明书和权利要求书及上述附图中的描述的一些流程中，包含了按照特定顺序出现的多个操作，但是应该清楚了解，这些操作可以不按照其在本文中出现的顺序来执行或并行执行，操作的序号如S101、S102等，仅仅是用于区分各个不同的操作，序号本身不代表任何的执行顺序。另外，这些流程可以包括更多或更少的操作，并且这些操作可以按顺序执行或并行执行。需要说明的是，本文中的“第一”、“第二”等描述，是用于区分不同的消息、设备、模块等，不代表先后顺序，也不限定“第一”和“第二”是不同的类型。

[0059] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域技术人员在没有做出创造性劳动前提下所获取的所有其他实施例，都属于本发明保护的范围。

[0060] 图1是本发明实施例提供的一种基于人工智能的急性心肌梗死临床决策支持系统，包括：

[0061] S101：获取单元，用于获取患者样本的遗传信息。

[0062] 在一个实施例中，样本包括以下任意一种或几种：血液、组织、唾液；遗传信息包括以下任意一种或几种：基因型、基因变异信息、基因表达信息、表观遗传信息。

[0063] 基因变异信息是指个体基因组中发生的与正常基因序列不一致的变化，可能与疾病易感性、药物敏感性相关，基因变异主要包括以下几类：

[0064] 1.单核苷酸多态性(Single Nucleotide Polymorphism, SNP)：SNP是最常见的基因变异类型，指在基因组中的一个核苷酸位置上，至少有两个不同的核苷酸存在。

[0065] 2.插入缺失变异(Insertion/Deletion, Indel)：Indel是指在基因组中发生插入或缺失碱基的变异，导致基因序列长度的变化。

[0066] 3.拷贝数变异(Copy Number Variation, CNV)：CNV是指在基因组中某一段DNA序列的拷贝数发生变化。它可以导致基因拷贝数的增加或减少，从而对基因表达产生重要影响。

[0067] 4.基因重排(Gene Rearrangement)：基因重排是指基因组中的某些基因片段发生断裂和重组，导致基因的位置、顺序和方向发生改变。

[0068] 5.基因扩增(Gene Amplification)：基因扩增是指基因组中某个基因的拷贝数目增加，导致该基因的表达增强。

[0069] 表观遗传信息是指在基因组中发生的不涉及DNA序列改变的遗传变化，通过影响基因的表达和调控来影响细胞和个体的功能。以下是表观遗传信息的几种主要类型：

[0070] 1.基因扩增(Gene Amplification)：基因扩增是指基因组中某个基因的拷贝数目增加，导致该基因的表达增强。基因扩增与肿瘤的发生和发展密切相关。

[0071] 2.组蛋白修饰：组蛋白是染色质的主要组成部分，它可以通过多种化学修饰来调控基因的表达。这些修饰包括乙酰化、甲基化、磷酸化、泛素化等。不同的修饰方式可以影响

染色质的结构和紧密度,从而调节基因的可及性和转录活性。

[0072] 3.非编码RNA:非编码RNA(non-coding RNA, ncRNA)是指在转录过程中产生的不参与编码蛋白质的RNA分子。这些ncRNA分子可以通过多种机制调节基因表达,包括miRNA、siRNA、lncRNA等。它们可以与DNA、RNA和蛋白质相互作用,参与转录调控、RNA剪接、转运、翻译和稳定性等过程。

[0073] 4.染色质构象:染色质构象是指染色质在三维空间中的结构和组织方式。染色质的三维结构可以通过染色质相互作用、染色质回路和染色质域的形成来调控基因的表达。染色质构象的改变可以影响基因的可及性和转录效率。

[0074] 在一个具体实施例中,从NCBI下载GSE66360数据集,该数据集包括来自49个心梗患者的血液样本和50个健康对照组样本。利用平台注释文件GPL570和R语言AnnoProbe包(<https://github.com/ableno/AnnoProbe>)对其注释,有多个探针对应的基因取其平均值作为基因表达量,经此处理之后,得到99个样本21655个基因的表达矩阵。

[0075] S102:特征提取单元,用于对所述遗传信息进行特征提取,得到遗传特征。

[0076] 在一个实施例中,采用Lasso回归对所述遗传信息进行特征提取。

[0077] Lasso回归是一种用于特征选择和正则化的线性回归方法。Lasso回归通过在损失函数中添加L1正则化项来实现特征选择,即通过惩罚系数使得一些特征的系数变为0,从而实现特征的稀疏性约束。Lasso回归的优点是可以自动选择重要的特征,减少模型的复杂度和泛化误差。同时,Lasso回归还可以解决多重共线性问题,即当特征之间存在高度相关性时,可以通过L1正则化将其中一个特征的系数置为0,从而减少冗余。

[0078] 在一个实施例中,遗传特征包括以下基因中的一种或几种:NISCH、PLAGL2、TLR1、ARID5B、TAF9B、SERINC1、CARKD、EPG5、ANKLE2、BC043227、FAM188A、ERVH.4、ZNF26、ERO1LB、TMEM208、EXOC2、SGMS1、BRPF3。

[0079] 其中,NISCH编码一种与神经元相干的蛋白质,参与神经元的发育和功能调节。PLAGL2编码一种转录因子蛋白质,参与胚胎发育、细胞增殖和分化等生物过程。TLR1编码Toll样受体1,参与免疫系统的免疫应答,识别和应答来自细菌和其他微生物的信号。ARID5B编码一种转录调节因子,参与胚胎发育、细胞增殖和分化等生物过程。TAF9B编码一种转录因子蛋白质,参与基因的转录调控。SERINC1编码一种膜蛋白,参与细胞膜的代谢和功能调节。CARKD编码一种酮糖酸激酶,参与细胞的能量代谢。EPG5编码一种蛋白质,参与细胞的自噬过程。ANKLE2编码一种蛋白质,参与细胞的分裂和染色体稳定性。BC043227编码一种长链非编码RNA。ERVH.4编码一种内源性逆转录病毒。ZNF26编码一种锌指转录因子,参与基因的转录调控。ERO1LB编码一种蛋白质,在细胞内负责调节蛋白质的氧化折叠过程。TMEM208编码一种膜蛋白。EXOC2编码一种蛋白质,参与细胞分泌途径的调节。SGMS1编码鞘脂合酶A,参与细胞膜的代谢和功能调节。BRPF3编码一种蛋白质,参与基因的转录调控。

[0080] 在一个具体实施例中,利用NMF算法的共识聚类分析对49例心梗患者进行亚组划分。NMF的目的是通过将原始矩阵分解为两个非负矩阵来识别基因表达谱中的潜在特征。综合相关系数用于确定最佳k值。将最佳总群集数设置为 $k = 2$ (两个亚组分别称为药物敏感组和药物耐受组)。例如,当 $k = 2$ 时,共识矩阵热图保持清晰且清晰的边界,表明样本具有稳定且健壮的簇。其中,药物敏感组和cultureB亚组分别包含15及34例心梗患者。

[0081] 使用R 4.0.5语言的limma对得到的两个疾病亚群进行差异表达分析,差异表达

mRNA筛选标准：校正后 $pvalue < 0.05$, $|\logFC| > 1$ ，共获得8482个差异表达mRNA。利用Lasso回归分析对8482个差异表达基因进行高维变量筛选和特征选择，筛选出18个重要特征。利用R语言的corrgram包对18个重要基因进行相关性分析，结果如图2所示。利用R语言的randomForest包对上面得到的18个重要基因进行机器学习，以筛选最优生物标志物，通过随机森林算法构建分类模型，按Mean Decrease Accuary值从大到小对18个mRNA进行重要性排序，依次为：NISCH、PLAGL2、TLR1、ARID5B、TAF9B、SERINC1、CARKD、EPG5、ANKLE2、BC043227、FAM188A、ERVH.4、ZNF26、ER01LB、TMEM208、EXOC2、SGMS1、BRPF3。

[0082] S103:亚型分类单元,用于基于所述遗传特征对所述患者进行分类,判断患者是否为药物敏感型。

[0083] 在一个实施例中,所述遗传特征为NISCH和PLAGL2,基于所述NISCH和PLAGL2对患者进行分类,判断患者是否为药物敏感型。

[0084] 在一个具体实施例中,按随机森林排序结果的顺序,从上而下依次增加1个mRNA,分别利用随机森林(R语言randomForest包)算法进行分类,并用十折交叉验证(10-fold cross-validation)过程求准确率和AUC。从图中可以看出,当mRNA的数量达到2的时候,AUC值和准确性达到最大值,所以选取前2个mRNA作为最优的生物标志物。在一个实施例中,最优的生物标志物为NISCH和PLAGL2。

[0085] 对上面筛选的2个mRNA应用支持向量机(R语言e1071包)来构建诊断模型,并用随机森林算法构建分类模型与之比较,发现2个模型AUC都很高。分类器的ROC曲线图和模型内2个基因单独诊断两个亚型的ROC曲线,这两个基因在三组间的基因表达量差异如图3所示,图3中的A是基因NISCH在三组间的基因表达量差异的箱线图,图3中的B是基因PLAGL2在三组间的基因表达量差异的箱线图。

[0086] S104:用药决策单元,用于基于患者的分类结果给予用药建议,当患者为药物敏感型时,用药建议为给予心肌梗死药物治疗。

[0087] 在一个实施例中,所述心肌梗死药物包括以下任意一种或几种:Epinephrine、Dobutamine、Norepinephrine、Aramine、Isoproterenol。

[0088] 所述5种药物中,Epinephrine是指肾上腺素,在心肌梗死的急救过程中,肾上腺素通过激活 β -肾上腺素能受体来增强心肌收缩力和心率,以恢复心脏功能。Dobutamine是指多巴酚丁胺,一种正性肌力药物,常用于心脏功能不全的治疗,包括心肌梗死患者,它通过激活 β_1 -肾上腺素能受体来增加心肌收缩力,帮助改善心脏泵血功能。Norepinephrine是指去甲肾上腺素,在心肌梗死患者中,去甲肾上腺素通常用于治疗低血压和休克等情况,它通过激活 α -肾上腺素能受体来收缩血管,增加血压和组织灌注。Aramine是指阿拉明,一种 α -肾上腺素能受体激动剂,可用于治疗心肌梗死引起的低血压,它通过收缩血管,提高血压和组织灌注。Isoproterenol是指异丙肾上腺素,一种 β -肾上腺素能受体激动剂,可用于治疗心脏传导阻滞和心肌梗死引起的心律失常,它通过增加心肌收缩力和心率,帮助改善心脏功能。

[0089] 在一个具体实施例中,在iLINC数据库中对Dopamine、Epinephrine、Dobutamine、Norepinephrine、Aramine、Levosimendan、Isoproterenol、Milrinone这8种药物进行药物靶点检索,检索到药物对应的作用靶点参考表1。

[0090] 表1 药物靶点对应表

药物	基因
Dopamine	ADRB1 DRD1 DRD2 DRD3 DRD4 DRD5
Epinephrine	ADRA1A ADRA1B ADRA1D ADRA2A ADRA2B ADRA2C ADRB1 ADRB2 ADRB3
[0091] Dobutamine	ADRB1 ADRB2
Norepinephrine	ADRA1A ADRA1B ADRA1D ADRA2A ADRA2B ADRA2C ADRB1 ADRB3
Aramine	ADRA1A ADRA1B ADRA1D
Levosimendan	
Isoproterenol	ADRB1 ADRB2 ADRB3
Milrinone	PDE3A

[0092] 所述8种药物中,1. Dopamine是指多巴胺,一种神经递质,也可用作药物。它主要通过多巴胺受体相互作用来产生效应。多巴胺受体包括D1、D2、D3、D4和D5亚型。多巴胺在中枢神经系统中起着调节运动、情绪和奖赏等功能,同时也参与调节心血管系统的功能。2. Epinephrine是指肾上腺素,一种重要的激素和神经递质,在机体应激反应中起着关键的作用。肾上腺素主要通过 α -肾上腺素受体和 β -肾上腺素受体相互作用来产生效应。肾上腺素的作用包括增加心率、促进心肌收缩、扩张支气管、增加血糖浓度等。3. Dobutamine是指多巴酚丁胺,一种 β_1 -肾上腺素能受体激动剂。它通过与 β_1 -肾上腺素受体结合来增加心肌收缩力和心输出量,从而用于治疗心力衰竭和心肌缺血等心血管疾病。4. Norepinephrine是指去甲肾上腺素,一种重要的激素和神经递质,与肾上腺素类似。它主要通过 α -肾上腺素受体和 β -肾上腺素受体相互作用来产生效应。去甲肾上腺素的作用包括收缩血管、增加心率和收缩力等。5. Aramine是指阿拉明,一种强力的 α -肾上腺素能受体激动剂。它主要通过 α -肾上腺素受体结合来收缩血管,从而用于治疗低血压和休克等状况。6. Levosimendan是指左西孟旦,一种钙敏感蛋白调节剂,用于增强心肌收缩力。它通过增加心肌细胞对钙离子的敏感性,从而增强心肌收缩力,改善心力衰竭和心肌缺血等疾病。7. Isoproterenol是指异丙肾上腺素,一种非选择性 β -肾上腺素能受体激动剂。它通过与 β -肾上腺素受体结合来扩张支气管、增加心率和心肌收缩力等。8. Milrinone是指米力农,一种磷二酯酶III抑制剂,用于增强心肌收缩力。它通过抑制磷二酯酶III的活性,从而增加细胞内环磷酸腺苷(cAMP)水平,增强心肌收缩力,改善心力衰竭等疾病。

[0093] 使用R 4.0.5语言的limma对得到的两个疾病亚群分别与健康对照组进行差异表达分析,差异表达mRNA筛选标准:校正后pvalue < 0.05,|logFC|>1,药物敏感组共获得6065个差异表达mRNA,药物敏感组共获得618个差异表达mRNA。

[0094] 将得到的差异表达基因与药物靶标进行韦恩图分解结果显示Epinephrine、Dobutamine、Norepinephrine、Aramine、Isoproterenol五个药物的靶标基因与药物敏感组的差异表达基因均存在交集。其中ADRB2与ADRA1A为Epinephrine的作用靶点,并且这两个基因在药物敏感组与正常对照组中存在差异,用药建议为给予药物敏感组患者心肌梗死药物Epinephrine治疗。同理,ADRB2为Dobutamine和Isoproterenol的作用靶点,并且该基因在药物敏感组与正常对照组中存在显著差异,用药建议为给予药物敏感组患者心肌梗死药物Dobutamine和Isoproterenol治疗。ADRA1A为Norepinephrine和Aramine的作用靶点,并

且该基因在药物敏感组与正常对照组中存在差异,用药建议为给予药物敏感组患者心肌梗死药物Norepinephrine和Aramine治疗。综上所述,药物敏感组患者,可以采用以下任意一种或几种心梗药物进行治疗:Epinephrine、Dobutamine、Norepinephrine、Aramine、Isoproterenol。

[0095] S105:分类模型构建单元,用于构建所述分类模型,所述分类模型的构建方法包括:

[0096] 获取急性心肌梗死患者的遗传信息和分类标签;

[0097] 对所述遗传信息进行特征提取,得到遗传特征;

[0098] 将所述遗传特征输入随机森林中进行模型构建,得到初步分类结果,将所述初步分类结果与所述分类标签比对,生成损失函数,基于损失函数优化机器学习模型,得到训练好的分类模型;

[0099] 所述随机森林为对随机森林算法中的决策树进行自适应剪枝的随机森林。

[0100] 在一个实施例中,所述决策树进行自适应剪枝的过程包括:

[0101] 步骤1:从叶节点开始逐层向上选择每个内部节点;

[0102] 步骤2:计算节点rt的损失函数;

[0103] 步骤3:计算决策树的整体损失函数;

[0104] 步骤4:对于每个节点,计算其在验证集上的损失函数值,基于损失函数值决定是否进行剪枝。

[0105] 在一个具体实施例中,所述节点rt的损失函数为:

$$[0106] \quad C_{\alpha}(rt) = C(rt) + \alpha|rT_{rt}|$$

[0107] 其中,C(rt)表示节点rt的训练误差, $|rT_{rt}|$ 表示节点rt的叶节点个数, α 是一个非负超参数,用于平衡训练误差和叶节点个数。

[0108] 所述决策树的整体损失函数为:

$$[0109] \quad C_{\alpha}(rT) = C(rT) + \alpha|rT|$$

[0110] 其中,C(rT)表示决策树rT的训练误差, $|rT|$ 表示决策树rT的叶节点个数,通过调整超参数 α ,控制剪枝程度。

[0111] 在一个实施例中,分类标签是对某一种或几种药物是否有效的药敏分类结果,具体的,所述药物包括以下任意一种或几种:Dopamine、Epinephrine、Dobutamine、Norepinephrine、Aramine、Levosimendan、Isoproterenol、Milrinone。

[0112] 随机森林(Random Forest)是一种集成学习算法,通过建立多个决策树来进行分类和回归。它通过随机选择特征和样本子集来构建决策树,并通过投票或平均的方式进行预测。随机森林的特点是具有较高的准确性和鲁棒性,能够处理高维数据和大型数据集。

[0113] 在一个实施例中,特征选择完成后,利用得到的样本特征训练改进的深度森林分类模型。

[0114] 深度森林分类模型采用级联结构进行连接,每一层森林则为决策树的集成。深度森林分类模型自动确定深度森林级联的层数,其自动确定深度森林级联的层数的方法为:每个森林通过k折交叉验证生成类向量,即每个样本数据作为k-1次的训练样本,产生k-1个类向量,并根据待分类的图像得到验证数据,当拓展产生新的一层森林时,整个深度森林框

架的性能会根据验证数据进行评估,如果整个深度森林框架的性能没有明显改善,则深度森林的层数不再增加。在深度森林分类模型的训练过程中,每层使用 Q 个随机森林和 P 个完全随机森林,第一层将产生 $(Q + P) \times C$ 个特征,产生的特征向量将与原始特征向量连接并输入到下一级,其中, C 为类别个数;之后的每一层与此类似,直至倒数第二层。

[0115] 决策树是一种树形结构,其中每个内部节点表示一个属性上的测试,每个分支代表一个测试输出,每个叶节点代表一种类别,决策树是一个预测模型,它代表的是对象属性与对象值之间的一种映射关系。

[0116] 随机森林是利用多棵决策树对样本数据进行训练并预测的一种分类器,每棵树的产生方法为,在整个特征空间中随机选择 \sqrt{d} 个特征作为候选特征,其中, d 为输入特征的个数,然后选择具有最佳Gini值的特征作为节点的分裂特征。

[0117] 完全随机森林是利用多棵决策树对样本进行训练并预测的一种分类器,完全随机森林中每棵树的产生方法为,在整个特征空间中随机选择1个特征作为节点的分裂特征。

[0118] Gini值是指基尼系数,在决策树之CART算法中用Gini指数来衡量数据的不纯度或者不确定性,同时用Gini指数来决定类别变量的最优二分值的切分问题。

[0119] 当深度森林分类模型训练完成后,利用训练后的深度森林分类模型对实际需要处理的图像进行测试,并得到目标域图像的分类结果;其中,取深度森林分类模型中最后一层森林的结果的平均值中的最大值作为深度森林分类模型输出的分类结果所对应的类别:

$$[0120] \quad \text{Fin}(c) = \text{Max}_y \{ \text{Ave}_m [c_{11}, c_{12}, \dots, c_{1y}, c_{21}, c_{22}, \dots, c_{2y}, \dots, c_{m1}, c_{m2}, \dots, c_{my}] \}$$

[0121] 其中, m 为深度森林每层包含森林的个数, y 为该数据集的类别的个数, c 为数据集的分类的类别, $\text{Fin}(c)$ 为深度森林分类模型输出的分类结果, Max_y 为深度森林分类模型中最后一层森林的结果的平均值中的最大值, Ave_m 为深度森林分类模型中最后一层森林的结果的平均值。

[0122] 改进前的深度森林分类模型图解的示意图如图4所示

[0123] 传统的随机森林算法由随机森林和完全随机森林组成,但两种森林都无法自适应剪枝技术,模型的泛化能力较低,且容易产生过拟合现象。

[0124] 基于此,本发明提出的改进深度森林中,利用自适应剪枝随机森林替代完全随机森林。

[0125] 具体的,自适应剪枝随机森林分类器通过对每个决策树进行剪枝,以提高模型的泛化能力和减少过拟合现象。

[0126] 假设对于一个包含 rN 个样本的训练集,每个样本由 rM 个特征表示,共有 rC 个类别。将训练集表示为:

$$[0127] \quad rD = \{(rx_1, ry_1), (rx_2, ry_2), \dots, (rx_{rN}, ry_{rN})\}$$

[0128] 其中, $rx_i = (rx_{i1}, rx_{i2}, \dots, rx_{iM})$ 表示第 i 个样本的特征向量, ry_i 表示对应的类别。

[0129] 随机森林构建时,由 rT 个决策树构成,每个决策树的构建过程如下:

[0130] 1、从训练集 rD 中使用有放回抽样的方式,构造一个大小为 rN 的新数据集 $D_{\text{bootstrap}}$ 。

[0131] 2、从 rM 个特征中随机选择 rm 个特征,用于构建决策树。

[0132] 3、使用 $D_{\text{bootstrap}}$ 和选定的 rm 个特征构建一棵决策树。构建过程可以使用CART算法

或其他决策树构建算法。

[0133] 4、对构建好的决策树进行自适应剪枝,以提高模型的泛化能力。具体的:

[0134] 1) 剪枝过程:

[0135] 对每个内部节点进行剪枝,从叶节点开始逐层向上。对于每个节点,计算其在验证集上的损失函数值,如果剪枝后损失函数值较小,则进行剪枝操作。

[0136] 2) 损失函数:

[0137] 定义节点 r_t 的损失函数为:

$$[0138] \quad C_{\alpha}(r_t) = C(r_t) + \alpha|rT_{r_t}|$$

[0139] 其中, $C(r_t)$ 表示节点 r_t 的训练误差, $|rT_{r_t}|$ 表示节点 r_t 的叶节点个数, α 是一个非负超参数,用于平衡训练误差和叶节点个数。

[0140] 3) 剪枝策略:

[0141] 定义决策树的整体损失函数为:

$$[0142] \quad C_{\alpha}(rT) = C(rT) + \alpha|rT|$$

[0143] 其中, $C(rT)$ 表示决策树 rT 的训练误差, $|rT|$ 表示决策树 rT 的叶节点个数。通过调整超参数 α ,可以控制剪枝程度。

[0144] 此外,由于传统的随机森林逐层级联,容易造成特征的冗余,影响训练速度。本发明提出的改进深度森林仅在最后一层与原始输入进行级联,极大地降低了数据的冗余程度,改进后的深度森林分类模型图解的示意图如图5所示。

[0145] 图6是本发明实施例提供的基于人工智能的急性心肌梗死临床决策支持设备,包括:存储器和处理器;

[0146] 所述存储器用于存储程序指令;

[0147] 所述处理器用于调用程序指令,当程序指令被执行时,用于执行以下基于人工智能的急性心肌梗死临床决策支持方法,所述方法流程图见图7,包括:

[0148] S701:获取患者样本的遗传信息;

[0149] S702:对所述遗传信息进行特征提取,得到遗传特征;

[0150] S703:基于所述遗传特征对所述患者进行分类,判断患者是否为药物敏感型;

[0151] S704:基于患者的分类结果给予用药建议,当患者为药物敏感型时,用药建议为给予心肌梗死药物治疗;

[0152] S705:构建所述分类模型,所述分类模型的构建方法包括:

[0153] 获取急性心肌梗死患者的遗传信息和分类标签;

[0154] 对所述遗传信息进行特征提取,得到遗传特征;

[0155] 将所述遗传特征输入随机森林中进行模型构建,得到初步分类结果,将所述初步分类结果与所述分类标签比对,生成损失函数,基于损失函数优化机器学习模型,得到训练好的分类模型;

[0156] 所述随机森林为对随机森林算法中的决策树进行自适应剪枝的随机森林。

[0157] 一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现上述基于心源性休克分期的治疗辅助决策方法。

[0158] 本验证实施例的验证结果表明,为适应症分配固有权重相对于默认设置来说可以

适度改善本方法的性能。

[0159] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统,装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0160] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统,装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0161] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0162] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0163] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成,该程序可以存储于一计算机可读存储介质中,存储介质可以包括:只读存储器(ROM,Read Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁盘或光盘等。

[0164] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分步骤是可以通程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0165] 以上对本发明所提供的一种计算机设备进行了详细介绍,对于本领域的一般技术人员,依据本发明实施例的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本发明的限制。

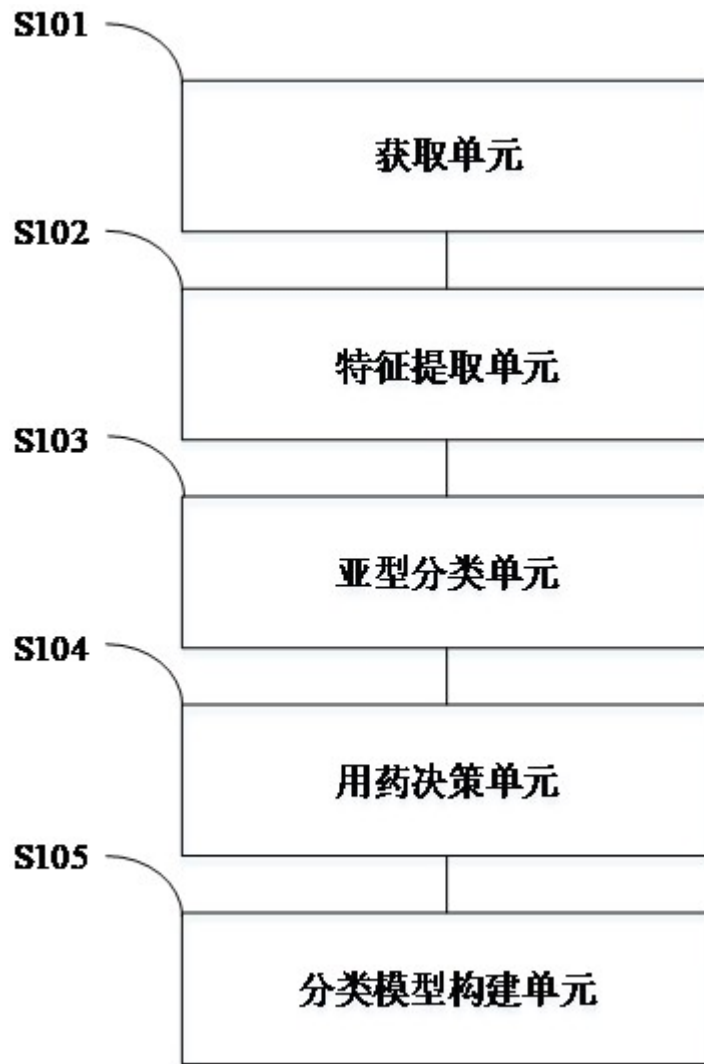


图1

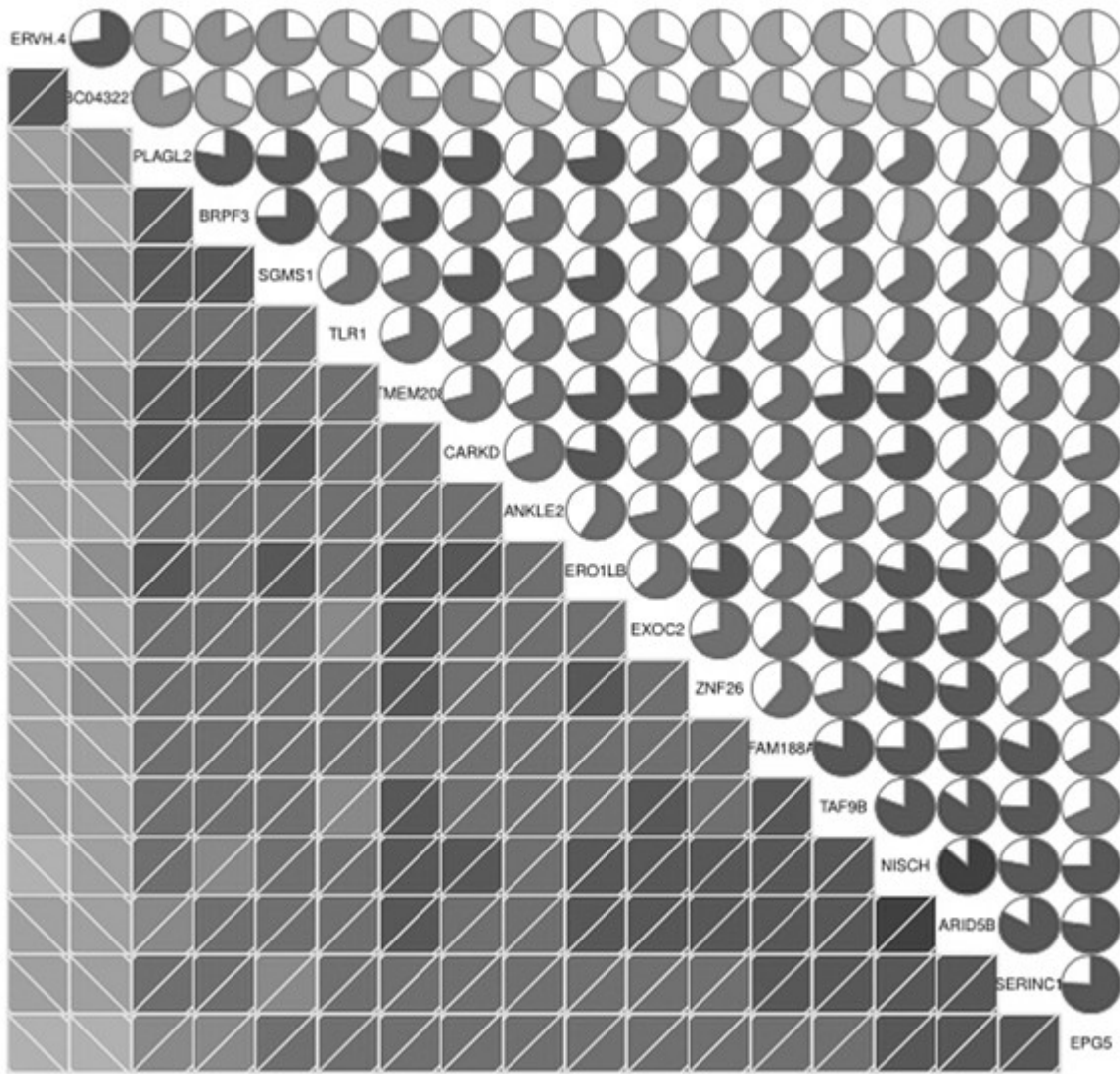


图2

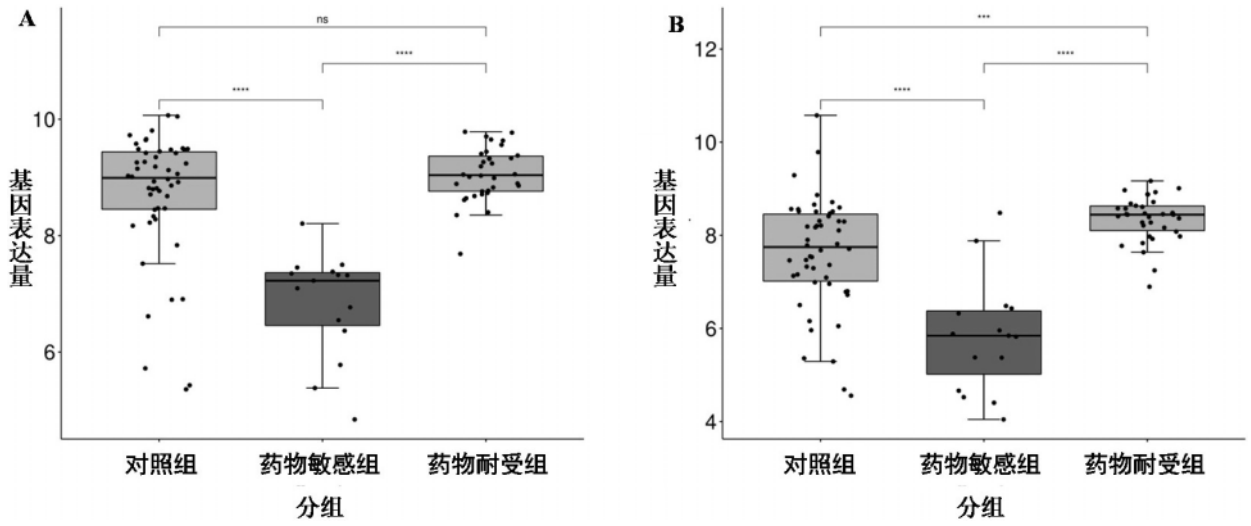


图3

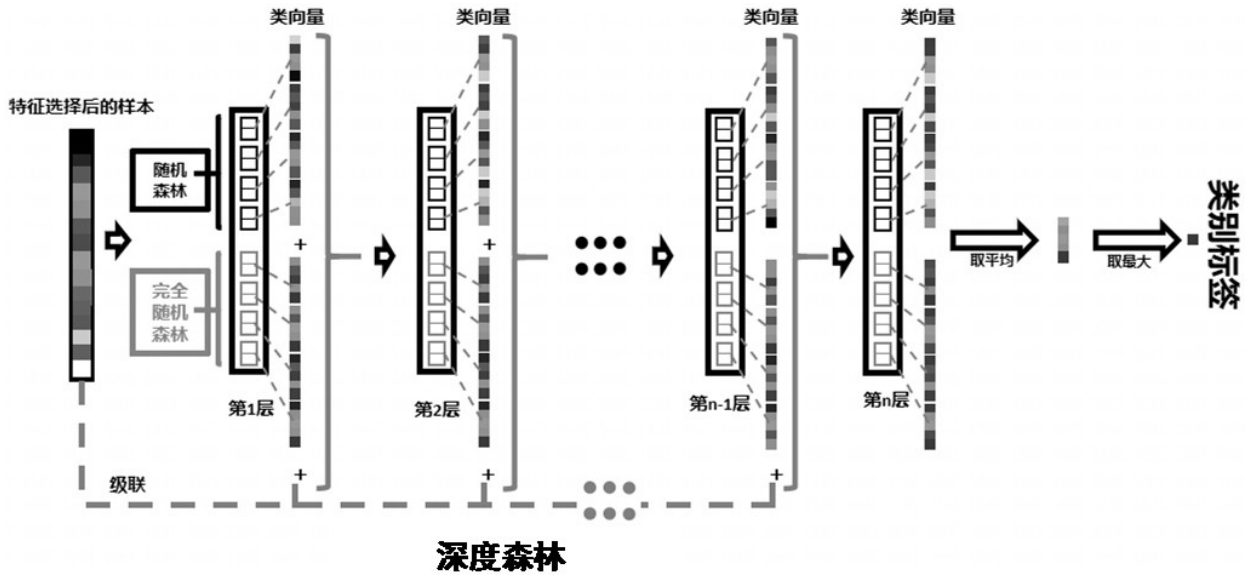


图4

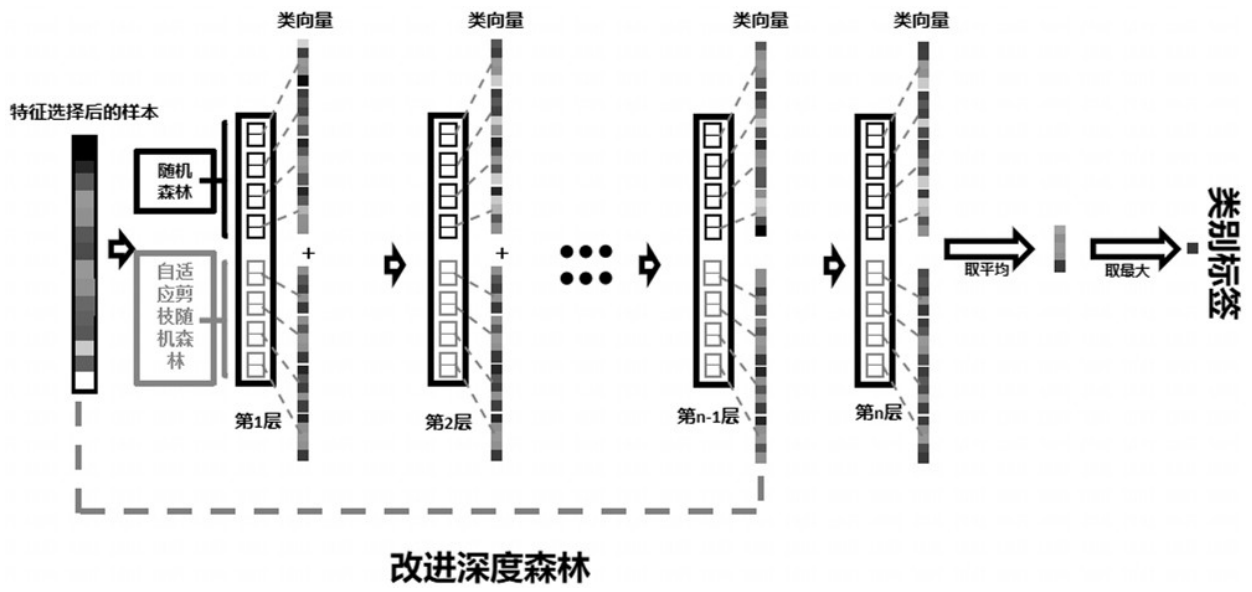


图5

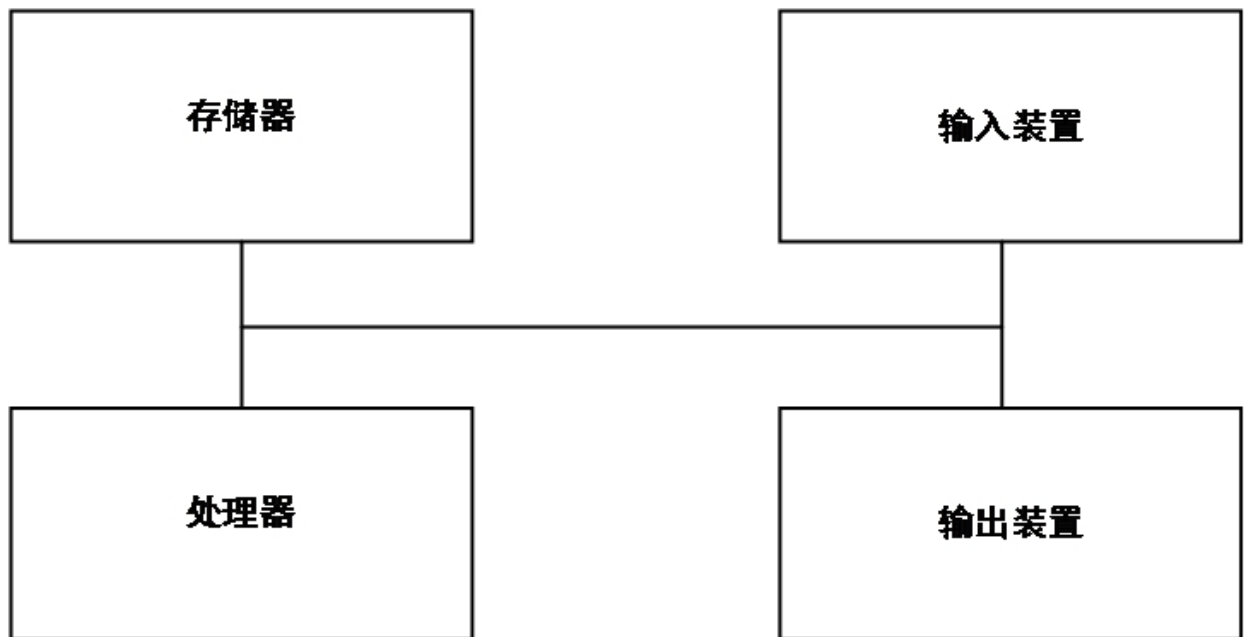


图6

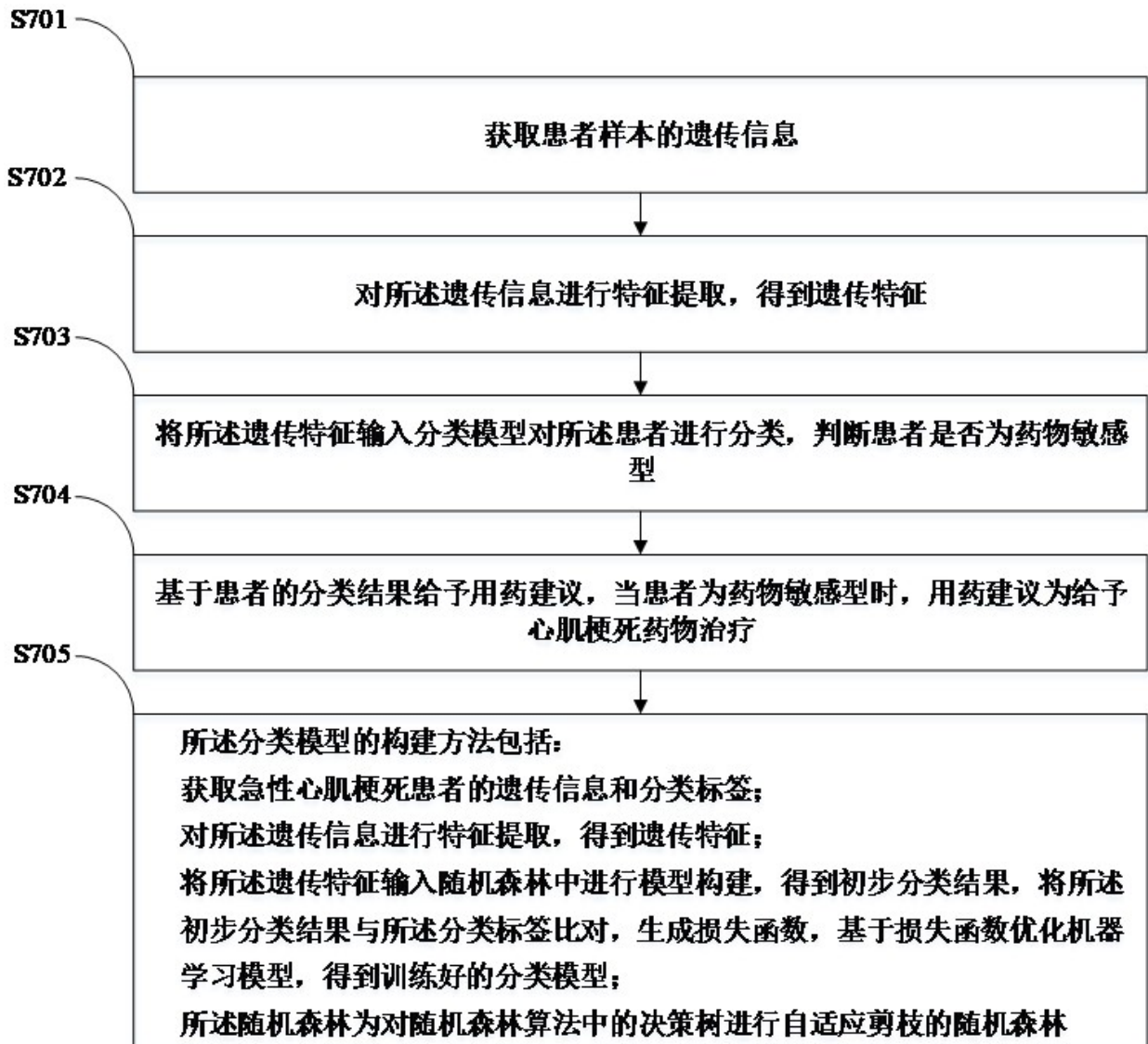


图7