

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022 年 10 月 6 日 (06.10.2022)



(10) 国际公布号
WO 2022/206604 A1

(51) 国际专利分类号:
G16B 40/00 (2019.01)

(21) 国际申请号: PCT/CN2022/083074

(22) 国际申请日: 2022 年 3 月 25 日 (25.03.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110355164.6 2021年4月1日 (01.04.2021) CN

(71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(72) 发明人: 杨子翊(YANG, Ziyi); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。 叶兆丰(YE, Zhaofeng); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。 廖奔奔(LIAO, Benben); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。 张胜誉(ZHANG, Shengyu); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(74) 代理人: 华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路 6 号 4501 房 (部位: 自编 01-03 和 08-12 单元) (仅限办公用途), Guangdong 510623 (CN)。

(54) Title: CLASSIFICATION MODEL TRAINING METHOD AND APPARATUS, CLASSIFICATION METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 分类模型训练和分类方法、装置、计算机设备和存储介质

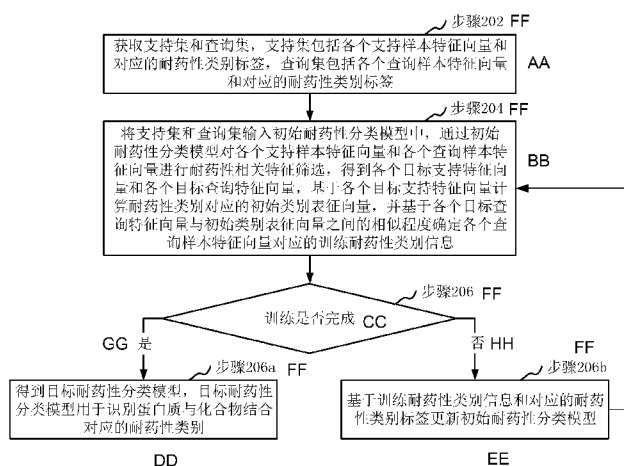


图 2

- AA Acquire a support set and a query set, wherein the support set comprises each support sample feature vector and a corresponding drug resistance class label, and the query set comprises each query sample feature vector and a corresponding drug resistance class label
- BB Input the support set and the query set into an initial drug resistance classification model; carry out the screening of drug resistance related features on each support sample feature vector and each query sample feature vector by means of the initial drug resistance classification model, so as to obtain each target support feature vector and each target query feature vector; and calculate, on the basis of each target support feature vector, an initial class representation vector corresponding to a drug resistance class, and determine, on the basis of the similarity between each target query feature vector and the initial class representation vector, training drug resistance class information corresponding to each query sample feature vector
- CC Is training completed?
- DD Obtain a target drug resistance classification model, wherein the target drug resistance classification model is used for recognizing a drug resistance class corresponding to a combination of proteins and a compound
- EE Update the initial drug resistance classification model on the basis of the training drug resistance class information and a corresponding drug resistance class label
- FF Step
- GG Yes
- HH No

(57) Abstract: The present application relates to a classification model training method and apparatus, and a computer device and a storage medium. The method comprises: acquiring a support set and a query set; inputting the support set and the query set into an initial drug resistance classification model; the initial drug resistance classification model carrying out drug resistance related feature screening on the basis of each support sample feature vector and each query sample feature vector, so as to obtain each target support feature vector and each target query feature vector; the initial drug resistance classification model calculating, on the basis of each

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

target support feature vector, an initial class representation vector corresponding to a drug resistance class, and determining, on the basis of the similarity between each target query feature vector and the initial class representation vector, training drug resistance class information corresponding to each query sample feature vector; and updating the initial drug resistance classification model on the basis of the training drug resistance class information and a corresponding drug resistance class label, and carrying out iterative execution, so as to obtain a target drug resistance classification model. By means of the method, the drug resistance classification accuracy of a target drug resistance classification model can be improved.

(57) 摘要: 本申请涉及一种分类模型训练方法、装置、计算机设备和存储介质。该方法包括: 获取支持集和查询集; 将支持集和查询集输入初始耐药性分类模型中, 初始耐药性分类模型基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选, 得到各个目标支持特征向量和各个目标查询特征向量, 基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量, 并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息; 基于训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型并迭代执行, 得到目标耐药性分类模型。采用本方法能够提高目标耐药性分类模型的耐药性分类准确性。

发明名称：分类模型训练和分类方法、装置、计算机设备和存储介质

本申请要求于 2021 年 04 月 01 日提交中国专利局，申请号为 2021103551646，申请名称为“分类模型训练和分类方法、装置、计算机设备和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机技术领域，特别是涉及一种分类模型训练和分类方法、装置、计算机设备和存储介质。

背景技术

随着人工智能技术的发展，出现了使用人工智能技术来预测靶向蛋白质突变所引发的耐药性分类问题。通过获取到大量的有标签的靶向蛋白质突变所引发的耐药性分类数据，然后训练得到人工智能模型，使用人工智能模型进行耐药性分类。然后，由于耐药性分类数据收集困难，导致训练人工智能模型时使用的样本量小，并且数据集之间的特征分布差异大，从而使得训练得到的人工智能模型在进行耐药性分类时的准确性低。

发明内容

基于此，有必要针对上述技术问题，提供一种能够提升耐药性分类准确性的分类模型训练和分类方法、装置、计算机设备和存储介质。

一种分类模型训练方法，所述方法包括：

获取支持集和查询集，支持集包括各个支持样本特征向量和对应的耐药性类别标签，查询集包括各个查询样本特征向量和对应的耐药性类别标签；

将支持集和查询集输入初始耐药性分类模型中，初始耐药性分类模型基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息；

基于训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，目标耐药性分类模型用于识别蛋白质与化合物结合对应的耐药性类别。

一种分类模型训练装置，所述装置包括：

数据获取模块，用于获取支持集和查询集，支持集包括各个支持样本特征向量和对应的耐药性类别标签，查询集包括各个查询样本特征向量和对应的耐药性类别标签；

初始分类模块，用于将支持集和查询集输入初始耐药性分类模型中，初始耐药性分类模型基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息；

迭代训练模块，用于基于训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，目标耐药性分类模型用于识别蛋白质与化合物结合对应的耐药性类别。

一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，所述处理器执行所述计算机可读指令时实现以下步骤：

获取支持集和查询集，支持集包括各个支持样本特征向量和对应的耐药性类别标签，查询集包括各个查询样本特征向量和对应的耐药性类别标签；

将支持集和查询集输入初始耐药性分类模型中，初始耐药性分类模型基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息；

基于训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，目标耐药性分类模型用于识别蛋白质与化合物结合对应的耐药性类别。

一种计算机可读存储介质，其上存储有计算机可读指令，所述计算机可读指令被处理器执行时实现以下步骤：

获取支持集和查询集，支持集包括各个支持样本特征向量和对应的耐药性类别标签，查询集包括各个查询样本特征向量和对应的耐药性类别标签；

将支持集和查询集输入初始耐药性分类模型中，初始耐药性分类模型基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息；

基于训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，目标耐药性分类模型用于识别蛋白质与化合物结合对应的耐药性类别。

上述分类模型训练方法、装置、计算机设备和存储介质，通过获取支持集和查询集，将支持集和查询集输入初始耐药性分类模型中，初始耐药性分类模型基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，从而使得训练时使用的特征更加的准确，然后使用各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，能够使计算得到的初始类别表征向量更加准确。此时，计算各个目标查询特征向量与初始类别表征向量之间的相似程度从而确定各个查询样本特征向量对应的训练耐药性类别信息，能够使得到的训练耐药性类别信息更加的准确。然后使用训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，从而能够使训练得到的目标耐药性分类模型能够提高耐药性分类的准确性。

一种分类方法，所述方法包括：

获取待分类数据和样本数据，待分类数据中包括待分类特征向量，样本数据中包括各个样本特征向量和对应的样本类别标签；

将待分类数据和样本数据输入耐药性分类模型中，耐药性分类模型基于待分类特征向量和各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量，基于各个目标样本特征向量计算样本类别对应的目标类别表征向量，并基于目标待

分类特征向量与目标类别表征向量之间的相似程度确定待分类特征向量对应的耐药性类别信息；

耐药性分类模型输出待分类数据对应的耐药性类别信息。

一种分类装置，所述装置包括：

待分类数据获取模块，用于获取待分类数据和样本数据，待分类数据中包括待分类特征向量，样本数据中包括各个样本特征向量和对应的样本类别标签；

分类模块，用于将待分类数据和样本数据输入耐药性分类模型中，耐药性分类模型基于待分类特征向量和各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量，基于各个目标样本特征向量计算样本类别对应的目标类别表征向量，并基于目标待分类特征向量与目标类别表征向量之间的相似程度确定待分类特征向量对应的耐药性类别信息；

类别输出模块，用于耐药性分类模型输出待分类数据对应的耐药性类别信息。

一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，所述处理器执行所述计算机可读指令时实现以下步骤：

获取待分类数据和样本数据，待分类数据中包括待分类特征向量，样本数据中包括各个样本特征向量和对应的样本类别标签；

将待分类数据和样本数据输入耐药性分类模型中，耐药性分类模型基于待分类特征向量和各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量，基于各个目标样本特征向量计算样本类别对应的目标类别表征向量，并基于目标待分类特征向量与目标类别表征向量之间的相似程度确定待分类特征向量对应的耐药性类别信息；

耐药性分类模型输出待分类数据对应的耐药性类别信息。

一种计算机可读存储介质，其上存储有计算机可读指令，所述计算机可读指令被处理器执行时实现以下步骤：

获取待分类数据和样本数据，待分类数据中包括待分类特征向量，样本数据中包括各个样本特征向量和对应的样本类别标签；

将待分类数据和样本数据输入耐药性分类模型中，耐药性分类模型基于待分类特征向量和各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量，基于各个目标样本特征向量计算样本类别对应的目标类别表征向量，并基于目标待分类特征向量与目标类别表征向量之间的相似程度确定待分类特征向量对应的耐药性类别信息；

耐药性分类模型输出待分类数据对应的耐药性类别信息。

上述分类方法、装置、计算机设备和存储介质，通过获取待分类数据和样本数据，将待分类数据和样本数据输入耐药性分类模型中，耐药性分类模型基于待分类特征向量和各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量，从而能够减少与耐药性无法的特征，使得到的目标待分类特征向量更加的准确。然后基于各个目标样本特征向量计算样本类别对应的目标类别表征向量，并计算目标待分类特征向量与目标类别表征向量之间的相似程度从而确定待分类特征向量对应的耐药性类别信息，由于耐药性分类模型是使用与耐药性相关的特征进行训练得到，然后使用耐药性分类模型进行耐药性分类识别，得到待分类特征向量对应的耐药性类别信息，从而能够使得到的耐药性类别信

息更加的准确。

附图说明

为了更清楚地说明本申请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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

- 图 1 为一个实施例中分类模型训练方法的应用环境图；
- 图 2 为一个实施例中分类模型训练方法的流程示意图；
- 图 3 为一个实施例中得到样本特征向量的流程示意图；
- 图 4 为一个实施例中抽取查询集和支持集的流程示意图；
- 图 5 为一个实施例中得到各个目标查询特征向量的流程示意图；
- 图 6 为一个实施例中计算初始类别表征向量的流程示意图；
- 图 7 为另一个实施例中计算初始类别表征向量的流程示意图；
- 图 8 为一个实施例中得到耐药性类别信息的流程示意图；
- 图 9 为一个实施例中确定训练耐药性类别信息的流程示意图；
- 图 10 为另一个实施例中确定训练耐药性类别信息的流程示意图；
- 图 11 为一个具体实施例中原型网络的示意图；
- 图 12 为一个实施例中得到目标耐药性分类模型的流程示意图；
- 图 13 为一个实施例中分类方法的流程示意图；
- 图 14 为一个具体实施例中分类方法的流程示意图；
- 图 15 为一个具体实施例中分类模型训练方法的流程示意图；
- 图 16 为一个具体实施例中分类模型的架构示意图；
- 图 17 为一个具体实施例中测试评价指标的示意图；
- 图 18 为一个实施例中分类模型训练装置的结构框图；
- 图 19 为一个实施例中分类装置的结构框图；
- 图 20 为一个实施例中计算机设备的内部结构图；
- 图 21 为另一个实施例中计算机设备的内部结构图。

具体实施方式

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，对本申请进行进一步详细说明。应当理解，此处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。

本申请提供的分类模型训练方法，可以应用于如图 1 所示的应用环境中。其中，终端 102 通过网络与服务器 104 进行通信。服务器 104 从终端 102 中接收到训练指令，服务器 104 根据训练指令从数据库 102 中获取支持集和查询集，支持集包括各个支持样本特征向量和对应的耐药性类别标签，查询集包括各个查询样本特征向量和对应的耐药性类别标签；服务器 104 将支持集和查询集输入初始耐药性分类模型中，初始耐药性分类模型基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特

征向量对应的训练耐药性类别信息；服务器 104 基于训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，目标耐药性分类模型用于识别蛋白质与化合物结合对应的耐药性类别，然后可以将目标耐药性分类模型返回终端 102 进行展示。其中，终端 102 可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备，服务器 104 可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

在一个实施例中，如图 2 所示，提供了一种分类模型训练方法，以该方法应用于图 1 中的服务器为例进行说明，可以理解的是，该方法也可以应用于服务器，还可以应用于包括终端和服务器的系统，并通过终端和服务器的交互实现。本实施例中，包括以下步骤：

步骤 202，获取支持集和查询集，支持集包括各个支持样本特征向量和对应的耐药性类别标签，查询集包括各个查询样本特征向量和对应的耐药性类别标签。

其中，支持集和查询集是从样本数据集中抽取得到的样本数据集合，样本数据集合通常指样本容量非常小的样本数据集合，比如，样本容量小于或等于 30 的样本数据集合。样本数据集中包括各个样本特征向量和每个样本特征向量对应的耐药性类别标签。耐药性类别标签用于表征耐药性类别，包括已耐药类别和未耐药类别。已耐药类别是指靶向蛋白质突变后与化合物之间产生了耐药性。未耐药类别是指靶向蛋白质突变后与化合物之间未产生耐药性。支持集是用于确定每个耐药性类别对应原型表示的数据集合。查询集是用于预测耐药性类别的数据集合。支持样本特征向量是指支持集中的数据样本对应的特征向量。查询样本特征向量是指查询集中的数据样本对应的特征向量。

具体地，服务器可以直接从数据库中获取到支持集和查询集。服务器也可以从提供数据服务的服务器中获取到支持集和查询集。服务器也可以从互联网中采集到支持集和查询集。

在一个实施例中，服务器也可以获取到样本数据集合，然后从样本数据集中随机抽取支持集和查询集。在一个实施例中，服务器在获取到样本数据集中，先随机抽取得到查询集，然后从样本数据集中确定与查询集相似的各个样本数据，然后从相似的各个样本数据中抽取支持集。

在一个实施例中，服务器可以从互联网采集到靶向蛋白质突变前后的数据以及化合物数据，得到各个样本数据，然后从各个样本数据中提取得到各个样本特征向量，同时采集到耐药性类别信息，得到耐药性类别标签，从而得到样本数据集合，然后从样本数据集中随机抽取支持集和查询集。

步骤 204，将支持集和查询集输入初始耐药性分类模型中，通过初始耐药性分类模型对各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息。

其中，初始耐药性分类模型是指模型参数初始化的耐药性分类模型，其中，模型参数初始化可以是随机初始化，也可以是为零初始化等等。耐药性分类模型用于识别输入数据对应的耐药性类别，即预测蛋白质在突变后是否与化合物之间产生耐药性，从而为医生用药提供帮助。目标支持特征向量是指滤除掉支持特征向量中与耐药性分类识别无关的特征后得到的特征向量。目标查询特征向量是指滤除掉查询特征向量中与耐药性分类识别无关的特征后得到的特征向量。初始类别表征向量是指初始的耐药性类别对应的原型表示，即类别的中心。

训练耐药性类别信息是指训练时识别得到的耐药性类别的信息，每个查询样本特征向量都会识别得到对应的耐药性类别。

具体地，服务器将支持集和查询集输入初始耐药性分类模型中，通过使用初始耐药性分类模型分别对每个支持样本特征向量和每个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，从而能够使得到的目标支持特征向量和目标查询特征向量中的特征都是对耐药性类别识别有关的特征，有利于提升模型的识别准确性。然后使用各个目标支持特征向量计算每个耐药性类别的中心，得到每个耐药性类别对应的初始类别表征向量，最后，计算每个目标查询特征向量与初始类别表征向量之间的相似程度根据该相似程度确定各个查询样本特征向量对应的训练耐药性类别信息，其中，目标查询特征向量与初始类别表征向量的相似程度越高，该目标查询特征向量所属于的类别就为该初始类别表征向量所对应的耐药性类别。

步骤 206，判断训练是否完成，当训练完成时，执行步骤 206a，当训练未完成时，执行步骤 206b 并返回步骤 204 继续迭代执行。

步骤 206a，得到目标耐药性分类模型，目标耐药性分类模型用于识别蛋白质与化合物结合对应的耐药性类别。

步骤 206b，基于训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行。

其中，判断训练是否完成是指判断训练是否达到训练完成条件，该训练完成条件包括但不限于训练迭代次数达到最大次数、模型参数不再发生变化和模型损失信息达到预设阈值。模型损失信息是指训练结果与真实结果之间的误差。

具体地，服务器判断训练是否完成，当训练未完成时，基于训练耐药性类别信息和对应的耐药性类别标签计算得到模型的损失信息，使用模型的损失信息反向更新初始耐药性分类模型中的参数，得到更新后的耐药性分类模型，然后将更新后的耐药性分类模型作为初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤迭代执行。直到训练完成时，将训练完成时对应的初始耐药性分类模型作为目标耐药性分类模型，该目标耐药性分类模型用于识别蛋白质突变后与化合物进行结合时对应的耐药性类别。

上述分类模型训练方法，通过获取支持集和查询集，将支持集和查询集输入初始耐药性分类模型中，初始耐药性分类模型基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，从而使得训练时使用的特征更加的准确，然后使用各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，能够使计算得到的初始类别表征向量更加准确。此时，计算各个目标查询特征向量与初始类别表征向量之间的相似程度从而确定各个查询样本特征向量对应的训练耐药性类别信息，能够使得到的训练耐药性类别信息更加的准确。然后使用训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，从而能够使训练得到的目标耐药性分类模型能够提高耐药性分类的准确性。

在一个实施例中，获取支持集和查询集，包括：

获取样本数据集，样本数据集中包括各个训练样本对应的样本特征向量和耐药性类别标签，样本特征向量是基于训练样本进行特征提取得到的，训练样本包括野生型蛋白质信息、突变型蛋白质信息和化合物信息；从样本数据集中随机抽取支持集和查询集。

其中，样本数据集是训练样本数据的集合。野生型蛋白质信息是指野生型蛋白质的具体信息，包括但不限于野生型蛋白质的结构，野生型蛋白质的理化性质等等。突变型蛋白质信息是指突变型蛋白质的具体信息，包括但不限于突变型蛋白质的结构、突变型蛋白质的理化性质等等。化合物信息是指与野生型蛋白质以及突变型蛋白质能够产生相互作用的小分子化合物的具体信息，包括化合物的结构、化合物的理化性质等等。每个训练样本都包括野生型蛋白质信息、突变型蛋白质信息和化合物信息。

具体地，服务器获取到各个训练样本，即获取到每个训练样本中的野生型蛋白质信息、突变型蛋白质信息和化合物信息，然后对训练样本进行特征提取，得到样本特征向量，其中，提取的样本特征向量中的特征包括但不限于野生型蛋白质结构特征、突变型蛋白质结构特征、野生型蛋白质理化性质特征、突变型蛋白质理化性质特征、晶体蛋白与化合物相互作用的结构特征、化合物与残基相互作用的理化性质特征以及通过打分函数提取到的能量特征等等。此时，服务器得到样本数据集，然后从样本数据集中进行随机抽取支持集中的训练样本和查询集中的训练样本，其中，可以从样本数据集中有放回抽取，得到支持集和查询集。也可以是从样本数据集中无放回抽取，得到支持集和查询集。

在上述实施例中，通过从样本数据集中抽取得到支持集和查询集，然后使用支持集合查询集进行模型训练，有助于提升训练得到的耐药性分类模型进行分类的准确性。

在一个实施例中，在得到目标耐药性分类模型之后，还包括：

将目标耐药性分类模型作为初始耐药性分类模型，并返回从样本数据集中随机抽取支持集和查询集的步骤执行，直到达到最终训练完成条件时，将达到最终训练完成条件时的初始耐药性分类模型作为最终耐药性分类模型。

其中，最终训练完成条件是指训练得到最终耐药性分类模型的条件，包括训练次数达到最终训练的最大上限或者模型的参数不再发生变化或者模型的训练误差达到预设阈值。

具体地，服务器在得到目标耐药性分类模型时，还可以继续进行训练，即将目标耐药性分类模型作为初始耐药性分类模型，并返回从样本数据集中随机抽取支持集和查询集的步骤执行，即每次训练得到目标耐药性分类模型时，再从样本数据集中随机抽取支持集和查询集，再重新进行训练。直到达到最终训练完成条件时，将达到最终训练完成条件时的初始耐药性分类模型作为最终耐药性分类模型。

在一个具体的实施例中，可以使用 episodic（元学习策略）训练策略训练得到最终耐药性分类模型，即从样本数据集中随机采样的 2-way k-shot（2 个类别，每个类别中有 k 个样本）任务，每个任务中都包含随机抽取到的支持集和查询集，当所有任务都训练完成时，得到最终耐药性分类模型。

在上述实施例中，通过不断抽取支持集和查询集，使用支持集和查询集进行训练，并得到最终耐药性分类模型，从而提高了训练得到的最终耐药性分类模型泛化能力。

在一个实施例中，如图 3 所示，在获取样本数据集之前，还包括：

步骤 302，获取训练样本，训练样本包括野生型蛋白质信息、突变型蛋白质信息和化合物信息。

步骤 304，基于野生型蛋白质信息和化合物信息进行野生特征提取，得到野生特征向量。

其中，野生特征向量是指使用野生型蛋白质信息和化合物信息提取得到的野生特征的向量。野生特征是指野生型蛋白质信息和化合物信息对应的特征，包括但不限于结构特征、理化性质特征以及能量特征。理化性质（physicochemical property）衡量化学物质特性的指

标。是指物理性质和化学性质，物理性质包括熔沸点，常温下的状态，颜色，化学性质包括酸碱度。理化性质特征包括物理性质特征和化学性质特征。

具体地，服务器可以从数据库中获取训练样本，训练样本包括野生型蛋白质信息、野生型蛋白质信息对应的突变型蛋白质信息以及能够与野生型蛋白质和突变型化合物产生相互作用的化合物信息。然后服务器使用野生型蛋白质信息和化合物信息进行野生特征提取，得到野生特征向量。即可以通过野生型蛋白质信息和化合物信息中的结构信息提取到结构特征，比如，野生型蛋白质结构特征、化合物结构特征以及野生型蛋白质和化合物产生相互作用后的结构特征。可以通过野生型蛋白质信息和化合物信息中的理化性质信息提取到理化性质特征，比如，野生型蛋白质理化性质特征，化合物理化性质特征以及野生型蛋白质和化合物产生相互作用后的理化性质特征。还可以通过打分函数提取到野生型蛋白质和化合物产生相互作用时的能量特征，其中，可以通过基于经验的打分函数提取到非物理的能量特征，也可以通过基于物理和经验势能的能量函数提取到能量特征，还可以通过基于知识的打分函数提取到能量特征。

步骤 306，基于突变型蛋白质信息和化合物信息进行突变特征提取，得到突变特征向量。

其中，突变特征向量是指使用突变型蛋白质信息和化合物信息提取得到的突变特征的向量。突变特征是指突变型蛋白质信息和化合物信息对应的特征，包括但不限于结构特征、理化性质特征以及能量特征。

具体地，服务器使用突变型蛋白质信息和化合物信息进行突变特征提取，得到突变特征向量。即可以通过突变型蛋白质信息和化合物信息中的结构信息提取到结构特征，比如，突变型蛋白质结构特征、化合物结构特征以及突变型蛋白质和化合物产生相互作用后的结构特征。可以通过突变型蛋白质信息和化合物信息中的理化性质信息提取到理化性质特征，比如，突变型蛋白质理化性质特征，化合物理化性质特征以及突变型蛋白质和化合物产生相互作用后的理化性质特征。还可以通过打分函数提取到突变型蛋白质和化合物产生相互作用时的能量特征，其中，可以通过基于经验的打分函数提取到非物理的能量特征，也可以通过基于物理和经验势能的能量函数提取到能量特征，还可以通过基于知识的打分函数提取到能量特征。

步骤 308，基于野生特征向量和突变特征向量得到训练样本对应的样本特征向量。

其中，样本特征向量是指训练样本对应的样本特征的向量。

具体地，服务器将提取得到的野生特征向量和突变特征向量作为训练样本对应的样本特征向量。

在上述实施例中，通过提取得到野生特征向量和突变特征向量，然后基于野生特征向量和突变特征向量得到训练样本对应的样本特征向量，能够使得到的样本特征向量更加的准确。

在一个实施例中，如图 4 所示，从样本数据集中随机抽取支持集和查询集，包括：

步骤 402，从样本数据集中进行随机抽样，得到查询集；

步骤 404，计算查询集中各个查询样本特征向量分别与样本数据集中各个样本特征向量的相似程度；

其中，相似程度用于表征查询样本特征向量与样本特征向量的相似性，

具体地，服务器先从样本数据集中进行随机抽取训练样本，得到查询集。然后使用相似度算法计算查询集中每个查询样本特征向量分别与样本数据集中每个样本特征向量的相似程度。其中，相似度算法可以使用距离相似度算法，也可以使用余弦相似度算法等等。服务器就得到了每个查询样本特征向量分别与每个样本特征向量的相似程度。

步骤 406, 基于相似程度将样本数据集中的各个样本特征向量进行排序, 得到样本特征向量序列;

其中, 样本特征向量序列是指根据相似程度进行排序得到的样本特征向量的序列。

具体地, 服务器按照相似程度由高到低的顺序将各个样本特征向量进行排序, 得到样本特征向量序列, 也可以由低到高的顺序将各个样本特征向量进行排序, 得到样本特征向量序列。

步骤 408, 从样本特征向量序列中依次选取预设数量的样本特征向量, 得到待抽取样本数据集;

其中, 待抽取样本数据集是指抽取支持集时使用的部分样本数据集。预设数量是指预先设置好的要选取待抽取训练样本的数量。

具体地, 服务器按照相似程度从高到低的顺序依次从样本特征向量序列中选取预设数量的样本特征向量, 得到待抽取样本数据集。其中, 也可以是预先设置好的待抽取样本数据集的训练样本数量是样本数据集中样本总数量的一定比例的训练样本。比如, 预先设置好抽取 5% 的训练样本作为待抽取样本数据集。

步骤 410, 从待抽取样本数据集中进行随机抽取, 得到支持集。

具体地, 服务器从待抽取样本数据集中随机抽取训练样本, 得到支持集, 其中, 随机抽取可以有放回抽取, 也可以是无放回抽取。

在一个实施例中, 服务器使用非线性降维算法 t-SNE ((t-distributed stochastic neighbor embedding, t 分布随机邻域嵌入) 对查询集中的查询特征向量和样本数据集中的样本特征向量进行非线性降维, 得到降维后的查询特征向量和降维后的样本特征向量, 使用降维后的查询特征向量和降维后的样本特征向量进行相似程度的计算, 能够提高计算相似程度的效率。然后从降维后的样本特征向量按照相似程度得到待抽取样本数据集, 从待抽取样本数据集抽取得到降维后的支持特征向量, 使用降维后的支持特征向量和降维后的查询特征向量进行耐药性分类模型的训练, 进一步避免了数据集之间特征分布差异大的问题, 能够提高训练得到的耐药性分类模型的准确性,

在上述实施例中, 通过先抽取查询集, 然后计算查询集中查询特征向量与样本数据集中样本特征向量的相似程度, 然后根据相似程度选取次选取预设数量的样本特征向量, 得到待抽取样本数据集, 再从待抽取样本数据集抽取训练样本得到支持集, 从而使得抽取得到的支持集与查询集中的特征分布差异小, 然后使用支持集与查询集训练得到耐药性分类模型, 能够提高训练得到的耐药性分类模型在分类时的准确性。

在一个实施例中, 如图 5 所示, 步骤 204, 即基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选, 得到各个目标支持特征向量和各个目标查询特征向量, 包括:

步骤 502, 获取初始特征筛选参数。

步骤 504, 基于初始特征筛选参数分别对各个支持样本特征向量进行耐药性相关特征筛选, 得到各个目标支持特征向量。

其中, 特征筛选参数是指用于进行特征筛选参数的参数, 特征筛选参数是需要通过训练更新初始特征筛选参数后得到的。初始特征筛选参数是初始化的特征筛选参数。不同的样本特征有不同的特征筛选参数, 即每个样本特征都有对应的特征筛选参数。

具体地, 服务器获取到初始特征筛选参数, 该初始特征筛选参数可以是随机初始化得到

的，也可以是为零初始化得到的，还可以是从数据库中直接获取到的。然后服务器使用初始特征筛选参数与每个支持样本特征向量进行相乘，即进行耐药性相关特征筛选，得到各个目标支持特征向量。

步骤 506，基于初始特征筛选参数对各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标查询特征向量。

具体地，服务器使用初始特征筛选参数与每个查询样本特征向量进行相乘滤除与耐药性分类识别无法的特征，得到各个目标查询特征向量。

在上述实施例中，通过使用特征筛选参数对支持样本特征向量和查询样本特征向量进行相乘，从而滤除掉与耐药性分类识别无法的特征，得到目标支持特征向量和目标查询特征向量，然后使用目标支持特征向量和目标查询特征向量进行耐药性分类模型的训练，能够提高训练得到的耐药性分类模型的分类准确性。

在一个实施例中，如图 6 所示，步骤 204，即基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，包括：

步骤 602，将各个目标支持特征向量进行映射，得到各个映射特征向量。

其中，映射特征向量是指使用嵌入函数将目标支持特征向量映射到嵌入空间得到的向量。

具体地，服务器将每个目标支持特征向量通过嵌入函数映射到嵌入空间中，得到各个映射特征向量，其中，嵌入函数是通过训练得到的，可以是 embedding（（嵌入））函数等。

步骤 604，获取初始置信度计算参数，基于各个映射特征向量使用初始置信度计算参数计算得到各个映射特征向量对应的置信度。

其中，置信度计算参数是指计算映射特征向量对应的训练样本的置信度的参数，不同的映射特征向量有不同的置信度，即不同的训练样本有不同的置信度。置信度是用于表征训练样本的可信程度，置信度越高，使用对应训练样本进行训练的效率就越好。初始置信度计算参数是指初始化的置信度计算参数，可以是随机初始化得到的。

具体地，服务器可以直接从数据库中获取到初始置信度计算参数，也可以是通过随机初始化得到初始置信度计算参数，还可以获取到第三方服务器提供的初始置信度计算参数。然后服务器将每个映射特征向量与初始置信度计算参数进行相乘，得到每个映射特征向量对应的置信度。比如，服务器可以使用自适应样本加权策略 Meta-Weight-Net (MW-Net) 来计算映射特征向量对应的置信度。即将各个映射特征向量作为 MW-Net 的输入，输出训练样本对应的置信度，也即每个映射特征向量对应的置信度。

步骤 606，基于置信度对各个映射特征向量进行加权，得到各个加权特征向量。

具体地，服务器使用置信度对每个映射特征向量进行加权，即通过对训练样本进行重赋权，可以使用置信度对训练样本进行筛选，得到各个加权特征向量，使得到的加权特征向量更能表征对应的训练样本。

步骤 608，基于各个加权特征向量计算耐药性类别对应的初始类别表征向量。

具体地，服务器按照耐药性类别计算每个耐药性类别对应的各个加权特征向量的平均向量，就得到每个耐药性类别对应的初始类别表征向量。在一个实施例中，也可以计算各个加权特征向量的中位向量，将中位向量作为耐药性类别对应的初始类别表征向量。

在上述实施例中，通过计算每个映射特征向量对应的置信度，然后使用置信度对映射特征向量进行加权，得到各个加权特征向量，从而能够对训练样本按照置信度进行筛选，避免了训练样本中存在噪音样本的问题，然后使用加权特征向量得到初始类别表征向量，能够提

高得到初始类别表征向量的准确性。

在一个实施例中，耐药性类别包括已耐药类别和未耐药类别；

如图 7 所示，步骤 608，基于各个加权特征向量计算耐药性类别对应的初始类别表征向量，包括：

步骤 702，按照各个支持样本特征向量对应的耐药性类别标签将各个加权特征向量进行划分，得到已耐药类别对应的加权特征向量和未耐药类别对应的加权特征向量。

具体地，由于加权特征向量是基于支持样本特征向量得到的，则每个加权特征向量都有对应的耐药性类别标签，该耐药性类别标签包括已耐药类别对应的标签和未耐药类别对应的标签。服务器按照各个支持样本特征向量对应的耐药性类别标签将各个加权特征向量进行划分，就得到已耐药类别对应的各个加权特征向量和未耐药类别对应的各个加权特征向量。

步骤 704，基于已耐药类别对应的加权特征向量进行向量平均，得到已耐药类别对应的第一初始类别表征向量。

其中，第一初始类别表征向量是用于表征已耐药类别的向量。

具体地，服务器计算已耐药类别对应的各个加权特征向量的平均向量，将该平均向量作为已耐药类别对应的第一初始类别表征向量。

步骤 706，基于未耐药类别对应的加权特征向量进行向量平均，得到未耐药类别对应的第二初始类别表征向量。

其中，第二初始类别表征向量是用于表征未耐药类别的向量。

具体地，服务器计算未耐药类别对应的各个加权特征向量的平均向量，将该平均向量作为未耐药类别对应的第二初始类别表征向量。在一个具体的实施例中，已耐药是指化合物（配体）与野生型以及突变型的蛋白靶点（受体）之间的相对结合自由能差大于 1.36kcal/mol。未耐药是指化合物（配体）与野生型以及突变型的蛋白靶点（受体）之间的相对结合自由能差小于 1.36kcal/mol。

在上述实施例中，通过已耐药类别标签对应的各个加权特征向量进行平均，得到第一初始类别表征向量，并通过未耐药类别标签对应的各个加权特征向量进行平均，得到第二初始类别表征向量，能够提高得到初始类别表征向量的准确性，方便后续使用。

在一个实施例中，如图 8 所示，步骤 204，基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息，包括：

步骤 802，计算当前目标查询特征向量分别与第一初始类别表征向量和第二初始类别表征向量之间的距离，得到当前第一初始距离和当前第二初始距离。

其中，当前目标查询特征向量是指当前需要计算相似程度的目标查询特征向量，可以是将各个目标查询特征向量依次作为当前目标查询特征向量。当前第一初始距离是指当前目标查询特征向量与第一初始类别表征向量之间的相似度距离。当前第二初始距离是指当前目标查询特征向量与第二初始类别表征向量之间的相似度距离。

具体地，服务器使用距离相似度算法计算当前目标查询特征向量与第一初始类别表征向量之间的距离，得到当前第一初始距离，并计算当前目标查询特征向量与第二初始类别表征向量之间的距离，得到当前第二初始距离。其中，距离相似度算法可以是欧式距离算法等等。

步骤 804，比较当前第一初始距离和当前第二初始距离，当当前第一目标距离超过当前第二目标距离时，当前目标查询特征向量对应的训练耐药性类别信息为未耐药类别，当当前第一目标距离未超过当前第二目标距离时，当前目标查询特征向量对应的耐药性类别信息为

已耐药类别。

具体地，服务器将当前第一初始距离和当前第二初始距离的大小进行比较，当当前第一目标距离超过当前第二目标距离时，说明当前目标查询特征向量与第二初始类别表征向量的距离较近，则说明当前目标查询特征向量对应的训练耐药性类别信息为未耐药类别。当当前第一目标距离未超过当前第二目标距离时，说明当前目标查询特征向量与第一初始类别表征向量的距离较近，则说明当前目标查询特征向量对应的训练耐药性类别信息为已耐药类别。

在上述实施例中，通过计算目标查询特征向量与初始类别表征向量之间的距离，然后根据距离确定目标查询特征向量所对应的耐药性类别，提高了得到的耐药性类别的准确性。

在一个实施例中，初始耐药性分类模型包括初始特征筛选网络和初始分类网络。如图 9 所示，步骤 204，将支持集和查询集输入初始耐药性分类模型中，包括：

步骤 902，将支持集和查询集输入初始耐药性分类模型中，通过初始耐药性分类模型将各个支持样本特征向量和各个查询样本特征向量输入初始特征筛选网络中。

其中，初始特征筛选网络是指网络参数初始化的特征筛选网络，特征筛选网络是用于滤除与耐药性分类识别无关特征的网络。初始分类网络是初始化的分类网络，分类网络是用于进行耐药性分类识别的网络。

具体地，服务器将支持集和查询集输入初始耐药性分类模型中，即在初始耐药性分类模型中将各个支持样本特征向量和各个查询样本特征向量输入初始特征筛选网络中进行特征筛选。

步骤 904，通过初始特征筛选网络基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，并将各个目标支持特征向量和各个目标查询特征向量输入分类网络中。

具体地，初始特征筛选网络进行耐药性相关特征筛选，即通过对各个支持样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量，并通过对各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标查询特征向量。然后将各个目标支持特征向量和各个目标查询特征向量输入分类网络中。在一个具体的实施例中，该特征筛选网络可以是 Softmax（逻辑回归）网络，使用该初始的 softmax 网络进行耐药性相关特征筛选。即可以使用如下所示的公式（1）进行特征筛选。

$$\begin{aligned}
 x_{new} &= \beta(\theta) \odot x \triangleq f_{\theta}(x) \\
 \beta_i(\theta) &= \frac{\exp(\theta)_i}{\sum_j \exp(\theta)_j} \\
 \sum_i \beta_i(\theta) &= 1
 \end{aligned}
 \tag{1}$$

其中，f 表示 Softmax 网络， θ 表示特征筛选网络参数，x 是指输入的特征向量， x_{new} 是指输出的特征向量。 $x_{new} = \beta(\theta) \odot x$ 表示输入的特征向量与特征筛选网络参数对应位置元素相乘。exp 是指按照元素计算的指数运算符，i 是指第 i 个输入的特征向量，j 是指输

入特征向量的总数量。 $\sum_i \beta_i(\theta) = 1$ 表示所有网络参数向量归一化后的总和为 1。

$$\beta_i(\theta) = \frac{\exp(\theta)_i}{\sum_j \exp(\theta)_j}$$

表示将网络参数向量归一化。

步骤 906, 通过分类网络基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量, 并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息。

具体地, 分类网络获取到输入的各个目标支持特征向量和各个目标查询特征向量, 使用不同耐药性类别标签对应的目标支持特征向量进行平均计算, 得到不同耐药性类别对应的初始类别表征向量, 然后计算各个目标查询特征向量与初始类别表征向量之间的相似程度, 根据该相似程度确定各个查询样本特征向量对应的训练耐药性类别信息。

在一个具体的实施例中, 可以使用如下所示的公式 (2) 计算类别表征向量。

$$C_n = \frac{1}{|S_n|} \sum_{(x_i, y_i) \in S_n} g_\phi(f_\theta(x_i))$$

公式 (2)

其中, C_n 表示类别表征向量, n 表示类别, 本申请中 $n \in \{0, 1\}$, 是离散的。 S_n 表示耐药性类别 n 对应的各个支持特征向量。 y_i 表示第 i 个支持特征向量对应的耐药性类别标签。 x_i 表示第 i 个支持特征向量。 $f_\theta(x_i)$ 表示第 i 个支持特征向量 x 经过 softmax 网络层输出的目标支持特征向量。 g 表示嵌入函数, ϕ 是指映射参数。

然后, 可以使用如下所示的公式 (3) 计算各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息。

$$p_{\phi, \theta}(y|x, S) = \frac{\exp(-d(g_\phi(f_\theta(x_i)), C_n))}{\sum_n \exp(-d(g_\phi(f_\theta(x_i)), C_n))}$$

公式 (3)

其中, $p_{\phi, \theta}(y|x, S)$ 是指分类网络输出的查询集中查询样本特征向量 x 属于 y 类的概率。 S 是指支持集。 $d(g_\phi(f_\theta(x_i)), C_n)$ 表示查询样本特征向量 x 与类别表征向量之间的相似程度。

在上述实施例中, 通过初始特征筛选网络进行特征滤除与耐药性分类识别无关的特征, 得到目标支持特征向量, 然后通过使用初始分类网络计算耐药性类别对应的初始类别表征向量, 并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息, 从而使得到的训练耐药性类别信息更加的准确。

在一个实施例中，分类网络包括样本筛选网络和原型网络；如图 10 所示，步骤 904，将各个目标支持特征向量和各个目标查询特征向量输入分类网络中，包括：

步骤 1002，将各个目标支持特征向量输入样本筛选网络中，通过样本筛选网络将各个目标支持特征向量进行映射，得到各个映射特征向量；获取初始置信度计算参数，基于各个映射特征向量使用初始置信度计算参数计算得到各个映射特征向量对应的置信度；基于置信度对各个映射特征向量进行加权，得到各个加权特征向量，并将各个加权特征向量输入原型网络中。

其中，样本筛选网络是对输入的训练样本进行筛选的网络。

具体地，服务器将各个目标支持特征向量输入样本筛选网络中，样本筛选网络将各个目标支持特征向量进行嵌入空间的映射，得到各个映射特征向量。然后服务器获取到样本筛选网络中的初始置信度计算参数，计算各个映射特征向量分别与初始置信度计算参数的乘积，得到各个映射特征向量对应的置信度；然后使用置信度对各个映射特征向量进行加权，得到各个加权特征向量，最后将各个加权特征向量输入原型网络中。在一个具体的实施例中，服务器可以使用如下所示的公式（4）来加权特征向量。

$$v_i \bullet g_\phi \left(f_\theta \left(x_i \right) \right) \quad \text{公式 (4)}$$

其中， v_i 表示第 i 个支持集中支持特征向量对应的置信度， $v_i \in [0, 1]$ 表示置信度的范围在 0 到 1 之间。 $g_\phi \left(f_\theta \left(x_i \right) \right) v_i$ 表示第 i 个映射特征向量。同计算置信度与映射特征向量之间的乘积，得到加权特征向量。其中，置信度可以将映射特征向量输入到 MW-Net 中，得到输出的置信度，即可以使用如下所示的公式（5）来计算。

$$v_i = V \left(g_\phi \left(f_\theta \left(x_i \right) \right); \Theta \right) \quad \text{公式 (5)}$$

其中， V 表示样本筛选网络中的置信度计算网络， Θ 表示样本筛选网络中的置信度计算参数。

步骤 1004，通过原型网络基于各个加权特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息。

具体地，原型网络使用各个加权特征向量计算耐药性类别对应的初始类别表征向量，并计算各个目标查询特征向量与初始类别表征向量之间的相似程度，根据该相似程度得到各个查询样本特征向量对应的训练耐药性类别信息。

在一个具体的实施例中，服务器可以使用如下所示的公式（6）计算得到类别表征向量。

$$C_n = \frac{1}{|S_n|} \sum_{(x_i, v_i) \in S_n} V \left(g_\phi \left(f_\theta \left(x_i \right) \right); \Theta \right) \bullet g_\phi \left(f_\theta \left(x_i \right) \right) \quad \text{公式 (6)}$$

其中, $V\left(g_{\phi}\left(f_{\theta}\left(x_i\right)\right); \Theta\right) \bullet g_{\phi}\left(f_{\theta}\left(x_i\right)\right)$ 表示第 i 个加权特征向量。然后, 可以使用如下所示的公式 (7) 计算得到查询样本特征向量对应的训练耐药性类别信息。

$$p_{\phi, \theta, \Theta}(y|x, S) = \frac{\exp\left(-d\left(g_{\phi}\left(f_{\theta}\left(x_i\right)\right), C_n\right)\right)}{\sum_n \exp\left(-d\left(g_{\phi}\left(f_{\theta}\left(x_i\right)\right), C_n\right)\right)} \quad \text{公式 (7)}$$

其中, ϕ 是指将特征向量映射到嵌入空间中的映射参数, θ 表示特征筛选网络参数。 Θ 表示样本筛选网络中的置信度计算参数。如图 11 所示, 为类别表征向量的示意图, 其中, 通过小样本的支持集计算得到的已耐药类别对应的类别表征向量 C1 和未耐药类别对应的类别表征向量 C2, 然后计算查询集中查询特征向量 a 对应的训练耐药性类别信息, 该查询特征向量对应的目标查询特征向量与类别表征向量 C1 的相似程度较高, 则该查询特征向量 a 对应的训练耐药性类别信息为已耐药类别。

在上述实施例中, 通过样本筛选网络按照置信度进行样本筛选, 得到各个加权特征向量, 最后将各个加权特征向量输入原型网络中, 然后在通过原型网络计算类别表征向量, 通过类别表征向量与各个查询样本特征向量的相似程度确定对应的训练耐药性类别信息, 能够减少噪音数据, 提高得到的训练耐药性类别的准确性。

在一个实施例中, 如图 12 所示, 步骤 206b, 基于训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型, 并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行, 当训练完成时, 得到目标耐药性分类模型, 包括:

步骤 1202, 基于训练耐药性类别信息和对应的耐药性类别标签进行对数损失计算, 得到初始训练损失信息。

其中, 初始训练损失信息是指初始训练时计算得到的训练耐药性类别信息和对应的耐药性类别标签之间的误差。

具体地, 服务器使用对数损失函数计算训练耐药性类别信息和对应的耐药性类别标签之间的误差, 得到初始训练损失信息。

步骤 1204, 计算初始训练损失信息的梯度, 基于梯度反向初始耐药性分类模型, 得到更新耐药性分类模型。

步骤 1206, 将更新耐药性分类模型作为初始耐药性分类模型, 并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行, 直到达到训练完成条件时, 将达到训练完成条件时的初始耐药性分类模型作为目标耐药性分类模型。

具体地, 服务器使用梯度下降算法来反向更新初始耐药性分类模型。其中, 可以先判断是否达到训练完成条件, 比如, 可以比较初始训练损失信息是否达到预先设置好的损失阈值, 当未达到时, 说明训练未完成, 此时, 使用初始训练损失信息计算梯度, 基于梯度反向更新初始耐药性分类模型中的参数, 当更新完成时, 得到更新耐药性分类模型, 再将更新耐药性分类模型作为初始耐药性分类模型, 并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行, 直到达到训练完成条件时将达到训练完成条件时的初始耐药性分类模型作为目标耐药性分类模型。

在一个具体的实施例中, 使用如下所示的公式 (8) 来作为损失函数来更新初始耐药性分

类模型。

$$\phi, \theta, \Theta = \underset{\phi, \theta, \Theta}{\operatorname{argmin}} E_{d \subset D_s} \sum_{(x, y) \in Q} -\log p_{\phi, \theta, \Theta}(y|x, S) \quad \text{公式 (8)}$$

其中, $D_s = \{x_i, y_i\}_{i=1}^N$ 表示训练样本集, $x_i \in X$ 表示训练样本, X 表示训练样本空间。 $y_i \in Y$ 表示训练样本对应的耐药性类别标签, Y 表示标记空间。 N 是指所有训练样本的数目。每个训练样本 x_i 都是 D 维的样本特征向量, $y_i \in \{0, 1\}$ 是离散的。 S 是指支持集, Q 是指查询集。 $E_{d \subset D_s}$ 是指每次从训练样本集中抽取支持集和查询集进行训练的任务, 即每个 n -way k -shot 任务被定义为一个 episodic $d = (S; Q)$ 。即通过上述损失函数更新耐药性分类模型中的参数 ϕ, θ, Θ 。直到参数 ϕ, θ, Θ 最小化时, 将得到的参数 ϕ, θ, Θ 作为最终耐药性分类模型中的参数。

在上述实施例中, 通过将训练耐药性类别信息和对应的耐药性类别标签进行对数损失计算, 得到初始训练损失信息, 然后使用初始训练损失信息反向更新初始耐药性分类模型, 并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行, 直到达到训练完成条件时, 将达到训练完成条件时的初始耐药性分类模型作为目标耐药性分类模型, 即保证了训练得到的目标耐药性分类模型的准确性。

在一个实施例中, 如图 13 所示, 提供了一种分类方法, 以该方法应用于图 1 中的服务器为例进行说明, 可以理解的是, 该方法也可以应用于服务器, 还可以应用于包括终端和服务器的系统, 并通过终端和服务器的交互实现。本实施例中, 包括以下步骤:

步骤 1302, 获取待分类数据和样本数据, 待分类数据中包括待分类特征向量, 样本数据中包括各个样本特征向量和对应的样本类别标签。

其中, 待分类数据是指需要进行分类的数据, 待分类特征向量是指需要进行耐药性类别识别的特征向量。样本类别标签是指耐药性类别对应的标签。样本特征向量是指训练样本对应的特征向量。

具体地, 服务器可以直接从数据库获取待分类数据和样本数据。服务器也可以获取到需要进行耐药性分类的野生型蛋白质信息、突变型蛋白质信息以及化合物信息, 然后提取到野生型蛋白质信息、突变型蛋白质信息和化合物信息对应的待分类特征, 得到待分类特征向量。在一个具体的实施例中, 从野生型蛋白质信息、突变型蛋白质信息和化合物信息中提取到野生型蛋白质结构特征、突变型蛋白质结构特征、野生型蛋白质理化性质特征、突变型蛋白质理化性质特征、晶体蛋白与化合物相互作用的结构特征、化合物与残基相互作用的理化性质特征以及通过打分函数提取到的能量特征, 得到野生型特征向量和突变型特征向量, 然后计算野生型特征向量与突变型特征向量之间的差异, 得到待分类特征向量。然后从数据库中获取到各个样本特征向量和对应的样本类别标签。服务器也可以从终端中获取到待分类数据, 然后从数据库中查找到各个样本特征向量和对应的样本类别标签。

步骤 1304, 将待分类数据和样本数据输入耐药性分类模型中, 耐药性分类模型基于待分类特征向量和各个样本特征向量进行耐药性相关特征筛选, 得到目标待分类特征向量和各个

目标样本特征向量，基于各个目标样本特征向量计算样本类别对应的目标类别表征向量，并基于目标待分类特征向量与目标类别表征向量之间的相似程度确定待分类特征向量对应的耐药性类别信息。

其中，耐药性分类模型可以是上述耐药性分类模型训练方法中任意一实施例中训练得到的模型。

具体地，服务器将已训练的耐药性分类模型部署到服务器中，服务器在接收到待分类数据和样本数据时，将待分类数据和样本数据输入耐药性分类模型中进行耐药性分类识别。即耐药性分类模型基于待分类特征向量和各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量。基于各个目标样本特征向量计算样本类别对应的目标类别表征向量，并基于目标待分类特征向量与目标类别表征向量之间的相似程度确定待分类特征向量对应的耐药性类别信息。

在一个实施例中，将待分类数据和样本数据输入耐药性分类模型中，耐药性分类模型通过特征筛选网络将待分类特征向量和各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量，然后将各个目标样本特征向量通过样本筛选网络进行映射，得到各个映射特征向量；获取置信度计算参数，基于各个映射特征向量使用置信度计算参数计算得到各个映射特征向量对应的置信度；基于置信度对各个映射特征向量进行加权，得到各个加权特征向量，并将各个加权特征向量输入原型网络中，原型网络基于各个加权特征向量计算耐药性类别对应的类别表征向量，并计算目标待分类特征向量与类别表征向量之间的相似程度，根据该相似程度得到各个查询样本特征向量对应的训练耐药性类别信息。

步骤 1306，通过耐药性分类模型输出待分类数据对应的耐药性类别信息。

具体地，服务器中的耐药性分类模型将得到的待分类数据对应的耐药性类别信息进行输出，从而得到待分类数据对应的耐药性类别信息，然后将耐药性类别信息返回到终端进行展示。

在一个实施例中，服务器耐药性分类模型按照各个样本特征向量对应的样本类别标签将各个目标加权特征向量进行划分，得到已耐药类别对应的目标加权特征向量和未耐药类别对应的目标加权特征向量；基于已耐药类别对应的目标加权特征向量进行向量平均，得到已耐药类别对应的第一目标类别表征向量。基于未耐药类别对应的目标加权特征向量进行向量平均，得到未耐药类别对应的第二目标类别表征向量。然后计算目标待分类特征向量分别与第一目标类别表征向量和第二目标类别表征向量之间的距离，得到第一目标距离和第二目标距离，比较第一目标距离和第二目标距离，当第一目标距离超过第二目标距离时，待分类特征向量对应的耐药性类别信息为未耐药类别，当第一目标距离未超过第二目标距离时，待分类特征向量对应的耐药性类别信息为已耐药类别。

上述分类方法，通过获取待分类数据和样本数据，将待分类数据和样本数据输入耐药性分类模型中，耐药性分类模型基于待分类特征向量和各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量，从而能够减少与耐药性无法的特征，使得到的目标待分类特征向量更加的准确。然后基于各个目标样本特征向量计算样本类别对应的目标类别表征向量，并计算目标待分类特征向量与目标类别表征向量之间的相似程度从而确定待分类特征向量对应的耐药性类别信息，由于耐药性分类模型是使用与耐药性相关的特征进行训练得到，然后使用耐药性分类模型进行耐药性分类识别，得到待分类特征向

量对应的耐药性类别信息，从而能够使得到的耐药性类别信息更加的准确。

在一个具体的实施例中，如图 14 所示，为耐药性分类方法的具体流程示意图，具体来说：服务器获取到终端上传的待分类数据，该待分类数据中包括野生型蛋白质信息、突变型蛋白质信息和化合物信息对应的待分类特征，得到待分类特征向量，然后对待分类数据进行非物理模型的特征提取，即提取到结构特征、理化性质特征以及基于经验的打分函数提取到能量特征。在对待分类数据进行物理和经验势能的特征提取，即通过基混合的物理和经验势能的建模程序 Rosetta 计算能量特征，从而得到野生型特征向量和突变型特征向量，计算野生型特征向量和突变型特征向量之间的差值，得到待分类特征向量，然后服务器从数据库中获取到样本数据，将待分类特征向量、各个样本特征向量和对应的样本类别标签输入到已训练的耐药性分类模型中进行耐药性预测，得到输出的耐药性类别信息，根据耐药性类别信息确定该野生型蛋白质信息进行蛋白质突变后得到的突变型蛋白质与化合物结合时是否产生耐药性。

在一个具体的实施例中，如图 15 所示，提供一种分类模型训练方法，该方法执行于服务器中，具体包括以下步骤：

步骤 1502，获取样本数据集，样本数据集中包括各个训练样本对应的样本特征向量和耐药性类别标签，样本特征向量是基于训练样本进行特征提取得到的，训练样本包括野生型蛋白质信息、突变型蛋白质信息和化合物信息；

步骤 1504，从样本数据集中进行随机抽样，得到查询集。计算查询集中各个查询样本特征向量分别与样本数据集中各个样本特征向量的相似程度。基于相似程度将样本数据集中的各个样本特征向量进行排序，得到样本特征向量序列。从样本特征向量序列中依次选取预设数量的样本特征向量，得到待抽取样本数据集。从待抽取样本数据集中进行随机抽取，得到支持集。

步骤 1506，将支持集和查询集输入初始耐药性分类模型中，通过初始耐药性分类模型将各个支持样本特征向量和各个查询样本特征向量输入初始特征筛选网络中。

步骤 1507，通过初始特征筛选网络基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，将各个目标支持特征向量输入样本筛选网络中。

步骤 1508，通过样本筛选网络将各个目标支持特征向量进行映射，得到各个映射特征向量；获取初始置信度计算参数，基于各个映射特征向量使用初始置信度计算参数计算得到各个映射特征向量对应的置信度。基于置信度对各个映射特征向量进行加权，得到各个加权特征向量，并将各个加权特征向量输入原型网络中。

步骤 1509，通过原型网络基于各个加权特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息。

步骤 1510，基于训练耐药性类别信息和对应的耐药性类别标签进行对数损失计算，得到初始训练损失信息。计算初始训练损失信息的梯度，基于梯度反向初始耐药性分类模型，得到更新耐药性分类模型。

步骤 1511，将更新耐药性分类模型作为初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行，直到达到训练完成条件时，将达到训练完成条件时的初始耐药性分类模型作为目标耐药性分类模型。

步骤 1512, 将目标耐药性分类模型作为初始耐药性分类模型, 并返回从样本数据集中进行随机抽样, 得到查询集的步骤执行, 直到达到最终训练完成条件时, 将达到最终训练完成条件时的初始耐药性分类模型作为最终耐药性分类模型。

本申请还提供一种应用场景, 该应用场景应用上述的分类模型训练方法。具体地, 如图 16 所示, 为耐药性分类模型训练的架构示意图, 其中, 服务器从样本数据集中抽取得到查询集和支持集, 支持集中包括已耐药类别对应的训练样本 X^a 和未耐药类别对应的训练样本

X^b , 以及每个类别对应 K 个训练样本 $(X_1^a, \dots, X_k^a)$ 和 $(X_1^b, \dots, X_k^b)$ 。查询集中包括了每个耐药性类别对应的训练样本 x 。F 是指特征筛选网络, I 是指样本筛选网络, 网络参数都是初始化的, 需要进行训练的。即服务器将支持集中的训练样本的支持特征向量输入到初始耐药性分类模型的特征筛选网络中, 得到输出的目标支持特征向量, 将目标支持特征向量输入到样本筛选网络中进行样本筛选, 即使用置信度计算参数 V 进行样本筛选, 得到输出的加权特征向量, 基于各个加权特征向量计算耐药性类别对应的初始类别表征向量 e , 包

括已耐药类别对应的初始类别表征向量 e_1 和未耐药类别对应的初始类别表征向量 e_2 。然后通过欧式距离算法计算初始类别表征向量与查询集中经过特征筛选网络得到的目标查询特征向量之间的距离 d , 根据距离确定训练得到的耐药性类别, 然后通过损失函数计算耐药性类别和对应的耐药性类别标签之间的误差, 根据该误差反向更新初始耐药性分类模型, 直达训练完成时, 得到目标耐药性分类模型。此时, 完成了一次 episodic 任务, 然后进行下一轮的 episodic 任务, 即将目标耐药性分类模型作为初始耐药性分类模型, 在从训练样本集中抽取支持集和查询集, 并进行循环迭代, 直到所有的 episodic 任务完成时, 得到最终的耐药性分类模型。

然后可以基于 python (计算机编程语言) 语言和 pytorch (一个开源的 Python 机器学习库, 基于 Torch, 用于自然语言处理等应用程序) 库将耐药性分类模型部署到搭载 Linux 操作系统或 Windows 操作系统和 CPU (中央处理器, central processing unit) 计算资源的服务器上。

进一步, 可以对训练得到的最终耐药性分类模型进行对比测试, 具体来说:

服务器使用耐药性标准数据集 Platinum 和 TKI (一类能抑制酪氨酸激酶活性的化合物) 进行测试。服务器对数据集 Platinum 和 TKI 中的耐药性数据进行特征提取, 得到样本数据集。提取得到的样本数据集如下表 1 所示, 其中, 使用 RDKit (开源化学信息学与机器学习工具包), Biopython (生物信息学的资源库), FoldX (分子模拟工具), PLIP (蛋白配体非共价相互作用的分析工具), AutoDock (分子模拟软件) 等非物理模型工具生成对预测蛋白质突变后的结合自由能变化具有参考价值的特征。还使用了基于混合的物理和经验势能的建模程序 Rosetta 计算能量特征。

表 1 样本数据集

数据集	样本数	未耐药类别	已耐药类别	非物理模型特征	物理和经验势能特征	特征总数
PPlatinum	4484	3362	1109	1129	119	1148
TTKI	1144	1125	119	1129	119	1148

其中, 样本数据集中的样本特征总共有 148 种, 其中, 包括非物理模型特征 129 种以及基于物理和经验势能特征 19 种。然后从样本数据集中抽取支持集和查询集, 其中, 训练过程

和测试过程抽取的支持集和查询集如下表 2 所示。

表 2 样本抽取表

训练阶段	支持集	查询集
模型训练	2-way 5-shot (Platinum 数据集)	2-way 5-shot (Platinum 数据集)
模型验证	2-way 5-shot (Platinum 数据集) (top 5%, 多次样本相似性抽 样)	2-way 5-shot TKI 数据集 10 个样本
模型测试	2-way 5-shot (Platinum 数据集) (top 5%, 样本相似性抽样)	TKI 数据集

其中，在耐药性分类模型训练 (Meta-training) 过程中，按照 2-way 5-shot 的方法从 Platinum 数据集中抽取支持集和查询集，即已耐药类别抽取 5 个样本和未耐药类别抽取 5 个样本。在耐药性分类模型验证 (Meta-validation) 过程中按照 2-way 5-shot 的抽取方法从 Platinum 数据集中抽取支持集和查询集，同时从 TKI 数据集抽取 10 个样本也作为查询集对耐药性分类模型进行验证。在进行耐药性分类模型测试 (Meta-testing) 时，将 TKI 数据集中的每个样本作为待测试的样本数据，即作为测试过程中的查询集，从 Platinum 数据集中按照 2-way 5-shot 的抽取方法抽取得到测试过程中的支持集。然后使用抽取得到的测试数据对传统方法和本申请进行测试，其中，传统方法可以是基于分子动力学的方法、也可以是传统机器学习方法等等，得到的测试评价指标表如下表 3 所示。

表 3 测试评价指标表

方法	力场或打分函数	AUPRC		
		均值	最小值	最大值
传统方法 1	OPLS3	0.56	0.32	0.76
传统方法 2	Charmm22* and CGenFF v 3.0.1	0.25	0.12	0.48
传统方法 3	Amber99sb*-ILDN and GAFF v 2.1	0.56	0.32	0.77
传统方法 4	Amber99sb*-ILDN and GAFF v 2.1	0.51	0.27	0.75
传统方法 5	REF15	0.53	0.29	0.74
传统方法 6	β NOV16	0.39	0.18	0.60
传统方法 7	n/a	0.20	0.1	0.39
本申请	n/a	0.61	0.51	0.71

其中，使用 AUPRC (精确率-召回率曲线下面积) 作为测试的评价指标，其中，明显可以看出，本申请测试评价指标 AUPRC 的均值和最小值均优于其他传统方法，且方差下，耐药性

分类识别的性能更加的稳定，即本申请能够进一步提升耐药性分类识别的准确性。如图 17 所示，为测试评价指标 AUPRC 的具体示意图，其中，使用 random classifier（随机分类）的方法进行分类时测试评价指标 AUPRC 的值为 0.13。从该图 17 中明显可以看出，本申请中的分类识别的性能更加的稳定，能够进一步提升耐药性分类识别的准确性。

的应该理解的是，虽然图 2-15 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图 2-15 中的至少一部分步骤可以包括多个步骤或者多个阶段，这些步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

在一个实施例中，如图 18 所示，提供了一种分类模型训练装置 1800，该装置可以采用软件模块或硬件模块，或者是二者的结合成为计算机设备的一部分，该装置具体包括：数据获取模块 1802、初始分类模块 1804 和迭代训练模块 1806，其中：

数据获取模块 1802，用于获取支持集和查询集，支持集包括各个支持样本特征向量和对应的耐药性类别标签，查询集包括各个查询样本特征向量和对应的耐药性类别标签；

初始分类模块 1804，用于将支持集和查询集输入初始耐药性分类模型中，通过初始耐药性分类模型对各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息；

迭代训练模块 1806，用于基于训练耐药性类别信息和对应的耐药性类别标签更新初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，目标耐药性分类模型用于识别蛋白质与化合物结合对应的耐药性类别。

在一个实施例中，数据获取模块 1802，包括：

样本获取模块，用于获取样本数据集，样本数据集中包括各个训练样本对应的样本特征向量和耐药性类别标签，样本特征向量是基于训练样本进行特征提取得到的，训练样本包括野生型蛋白质信息、突变型蛋白质信息和化合物信息；

抽取模块，用于从样本数据集中随机抽取支持集和查询集。

在一个实施例中，分类模型训练装置 1800，还包括：

最终模型训练模块，用于将目标耐药性分类模型作为初始耐药性分类模型，并返回从样本数据集中随机抽取支持集和查询集的步骤执行，直到达到最终训练完成条件时，将达到最终训练完成条件时的初始耐药性分类模型作为最终耐药性分类模型。

在一个实施例中，分类模型训练装置 1800，还包括：

特征提取模块，用于获取训练样本，训练样本包括野生型蛋白质信息、突变型蛋白质信息和化合物信息。基于野生型蛋白质信息和化合物信息进行野生特征提取，得到野生特征向量；基于突变型蛋白质信息和化合物信息进行突变特征提取，得到突变特征向量；基于野生特征向量和突变特征向量得到训练样本对应的样本特征向量。

在一个实施例中，抽取模块还用于从样本数据集中进行随机抽样，得到查询集；计算查询集中各个查询样本特征向量分别与样本数据集中各个样本特征向量的相似程度；基于相似

程度将样本数据集中的各个样本特征向量进行排序，得到样本特征向量序列；从样本特征向量序列中依次选取预设数量的样本特征向量，得到待抽取样本数据集；从待抽取样本数据集上进行随机抽取，得到支持集。

在一个实施例中，初始分类模块 1804 还用于获取初始特征筛选参数；基于初始特征筛选参数分别对各个支持样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量；基于初始特征筛选参数对各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标查询特征向量。

在一个实施例中，初始分类模块 1804 还用于将各个目标支持特征向量进行映射，得到各个映射特征向量；获取初始置信度计算参数，基于各个映射特征向量使用初始置信度计算参数计算得到各个映射特征向量对应的置信度；基于置信度对各个映射特征向量进行加权，得到各个加权特征向量；基于各个加权特征向量计算耐药性类别对应的初始类别表征向量。

在一个实施例中，耐药性类别包括已耐药类别和未耐药类别；

初始分类模块 1804 还用于按照各个支持样本特征向量对应的耐药性类别标签将各个加权特征向量进行划分，得到已耐药类别对应的加权特征向量和未耐药类别对应的加权特征向量；基于已耐药类别对应的加权特征向量进行向量平均，得到已耐药类别对应的第一初始类别表征向量；基于未耐药类别对应的加权特征向量进行向量平均，得到未耐药类别对应的第二初始类别表征向量。

在一个实施例中，初始分类模块 1804 还用于计算当前目标查询特征向量分别与第一初始类别表征向量和第二初始类别表征向量之间的距离，得到当前第一初始距离和当前第二初始距离；比较当前第一初始距离和当前第二初始距离，当当前第一目标距离超过当前第二目标距离时，当前目标查询特征向量对应的训练耐药性类别信息为未耐药类别，当当前第一目标距离未超过当前第二目标距离时，当前目标查询特征向量对应的耐药性类别信息为已耐药类别。

在一个实施例中，初始耐药性分类模型包括初始特征筛选网络和初始分类网络；初始分类模块 1804，还用于将支持集和查询集输入初始耐药性分类模型中，通过初始耐药性分类模型将各个支持样本特征向量和各个查询样本特征向量输入初始特征筛选网络中；通过初始特征筛选网络基于各个支持样本特征向量和各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，并将各个目标支持特征向量和各个目标查询特征向量输入分类网络中；通过分类网络基于各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息。

在一个实施例中，分类网络包括样本筛选网络和原型网络；初始分类模块 1804 还用于将各个目标支持特征向量输入样本筛选网络中，通过样本筛选网络将各个目标支持特征向量进行映射，得到各个映射特征向量；获取初始置信度计算参数，基于各个映射特征向量使用初始置信度计算参数计算得到各个映射特征向量对应的置信度；基于置信度对各个映射特征向量进行加权，得到各个加权特征向量，并将各个加权特征向量输入原型网络中；通过原型网络基于各个加权特征向量计算耐药性类别对应的初始类别表征向量，并基于各个目标查询特征向量与初始类别表征向量之间的相似程度确定各个查询样本特征向量对应的训练耐药性类别信息。

在一个实施例中，迭代训练模块 1806，还用于基于训练耐药性类别信息和对应的耐药性

类别标签进行对数损失计算，得到初始训练损失信息；计算初始训练损失信息的梯度，基于梯度反向初始耐药性分类模型，得到更新耐药性分类模型；将更新耐药性分类模型作为初始耐药性分类模型，并返回将支持集和查询集输入初始耐药性分类模型中的步骤执行，直到达到训练完成条件时，将达到训练完成条件时的初始耐药性分类模型作为目标耐药性分类模型。

在一个实施例中，如图 19 所示，提供了一种分类装置 1900，该装置可以采用软件模块或硬件模块，或者是二者的结合成为计算机设备的一部分，该装置具体包括：待分类数据获取模块 1902、分类模块 1904 和类别输出模块 1906，其中：

待分类数据获取模块 1902，用于获取待分类数据和样本数据，待分类数据中包括待分类特征向量，样本数据中包括各个样本特征向量和对应的样本类别标签；

分类模块 1904，用于将待分类数据和样本数据输入耐药性分类模型中，通过耐药性分类模型基于待分类特征向量和各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量，基于各个目标样本特征向量计算样本类别对应的目标类别表征向量，并基于目标待分类特征向量与目标类别表征向量之间的相似程度确定待分类特征向量对应的耐药性类别信息；

类别输出模块 1906，用于通过耐药性分类模型输出待分类数据对应的耐药性类别信息。

关于分类模型训练装置以及分类装置的具体限定可以参见上文中对于分类模型训练方法以及分类方法的限定，在此不再赘述。上述分类模型训练装置以及分类装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 20 所示。该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储训练样本数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种分类模型训练方法以及分类方法。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是终端，其内部结构图可以如图 21 所示。该计算机设备包括通过系统总线连接的处理器、存储器、通信接口、显示屏和输入装置。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统和计算机可读指令。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的通信接口用于与外部的终端进行有线或无线方式的通信，无线方式可通过 WIFI、运营商网络、NFC（近场通信）或其他技术实现。该计算机可读指令被处理器执行时以实现一种分类模型训练方法以及分类方法。该计算机设备的显示屏可以是液晶显示屏或者电子墨水显示屏，该计算机设备的输入装置可以是显示屏上覆盖的触摸层，也可以是计算机设备外壳上设置的按键、轨迹球或触控板，还可以是外接的键盘、触控板或鼠标等。

本领域技术人员可以理解，图 19 以及图 20 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

在一个实施例中，还提供了一种计算机设备，包括存储器和处理器，存储器中存储有计算机可读指令，该处理器执行计算机可读指令时实现上述各方法实施例中的步骤。

在一个实施例中，提供了一种计算机可读存储介质，存储有计算机可读指令，该计算机可读指令被处理器执行时实现上述各方法实施例中的步骤。

在一个实施例中，提供了一种计算机程序产品或计算机程序，该计算机程序产品或计算机程序包括计算机指令，该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令，处理器执行该计算机指令，使得该计算机设备执行上述各方法实施例中的步骤。

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流程，是可以通过计算机程序来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器（Read-Only Memory, ROM）、磁带、软盘、闪存或光存储器等。易失性存储器可包括随机存取存储器（Random Access Memory, RAM）或外部高速缓冲存储器。作为说明而非局限，RAM 可以是多种形式，比如静态随机存取存储器（Static Random Access Memory, SRAM）或动态随机存取存储器（Dynamic Random Access Memory, DRAM）等。

以上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。

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。

权利要求书

1、一种分类模型训练方法，由计算机设备执行，其特征在于，所述方法包括：

获取支持集和查询集，所述支持集包括各个支持样本特征向量和对应的耐药性类别标签，所述查询集包括各个查询样本特征向量和对应的耐药性类别标签；

将所述支持集和所述查询集输入初始耐药性分类模型中，通过所述初始耐药性分类模型对所述各个支持样本特征向量和所述各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，基于所述各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于所述各个目标查询特征向量与所述初始类别表征向量之间的相似程度确定所述各个查询样本特征向量对应的训练耐药性类别信息；

基于所述训练耐药性类别信息和对应的所述耐药性类别标签更新所述初始耐药性分类模型，并返回将所述支持集和所述查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，所述目标耐药性分类模型用于识别蛋白质与化合物结合对应的耐药性类别。

2、根据权利要求1所述的方法，其特征在于，所述获取支持集和查询集，包括：

获取样本数据集，所述样本数据集中包括各个训练样本对应的样本特征向量和耐药性类别标签，所述样本特征向量是基于训练样本进行特征提取得到的；

从所述样本数据集中随机抽取所述支持集和所述查询集。

3、根据权利要求2所述的方法，其特征在于，在所述得到目标耐药性分类模型之后，还包括：

将所述目标耐药性分类模型作为初始耐药性分类模型，并返回从所述样本数据集中随机抽取所述支持集和所述查询集的步骤执行，直到达到最终训练完成条件时，将达到最终训练完成条件时的初始耐药性分类模型作为最终耐药性分类模型。

4、根据权利要求2所述的方法，其特征在于，在获取样本数据集之前，还包括：

获取训练样本，所述训练样本包括野生型蛋白质信息、突变型蛋白质信息和化合物信息；基于所述野生型蛋白质信息和所述化合物信息进行野生特征提取，得到野生特征向量；基于所述突变型蛋白质信息和所述化合物信息进行突变特征提取，得到突变特征向量；基于所述野生特征向量和所述突变特征向量得到所述训练样本对应的样本特征向量。

5、根据权利要求2所述的方法，其特征在于，所述从所述样本数据集中随机抽取所述支持集和所述查询集，包括：

从所述样本数据集中进行随机抽样，得到所述查询集；

计算所述查询集中各个查询样本特征向量分别与所述样本数据集中各个样本特征向量的相似程度；

基于所述相似程度将所述样本数据集中的各个样本特征向量进行排序，得到样本特征向量序列；

从所述样本特征向量序列中依次选取预设数量的样本特征向量，得到待抽取样本数据集；从所述待抽取样本数据集中进行随机抽取，得到所述支持集。

6、根据权利要求1所述的方法，其特征在于，所述基于所述各个支持样本特征向量和所述各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，包括：

获取初始特征筛选参数；

基于所述初始特征筛选参数分别对所述各个支持样本特征向量进行耐药性相关特征筛选，得到所述各个目标支持特征向量；

基于所述初始特征筛选参数分别对所述各个查询样本特征向量进行耐药性相关特征筛选，得到所述各个目标查询特征向量。

7、根据权利要求1所述的方法，其特征在于，所述基于所述各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，包括：

将所述各个目标支持特征向量进行映射，得到各个映射特征向量；

获取初始置信度计算参数，基于所述各个映射特征向量使用所述初始置信度计算参数计算得到所述各个映射特征向量对应的置信度；

基于所述置信度对所述各个映射特征向量进行加权，得到各个加权特征向量；

基于所述各个加权特征向量计算所述耐药性类别对应的初始类别表征向量。

8、根据权利要求7所述的方法，其特征在于，所述耐药性类别包括已耐药类别和未耐药类别；

所述基于所述各个加权特征向量计算所述耐药性类别对应的初始类别表征向量，包括：

按照所述各个支持样本特征向量对应的耐药性类别标签将所述各个加权特征向量进行划分，得到所述已耐药类别对应的加权特征向量和所述未耐药类别对应的加权特征向量；

基于所述已耐药类别对应的加权特征向量进行向量平均，得到所述已耐药类别对应的第一初始类别表征向量；

基于所述未耐药类别对应的加权特征向量进行向量平均，得到所述未耐药类别对应的第二初始类别表征向量。

9、根据权利要求8所述的方法，其特征在于，基于所述各个目标查询特征向量与所述初始类别表征向量之间的相似程度确定所述各个查询样本特征向量对应的训练耐药性类别信息，包括：

计算所述各个目标查询特征向量中当前目标查询特征向量分别与第一初始类别表征向量和所述第二初始类别表征向量之间的距离，得到当前第一初始距离和当前第二初始距离；

比较所述当前第一初始距离和所述当前第二初始距离，当所述当前第一目标距离超过所述当前第二目标距离时，所述当前目标查询特征向量对应的训练耐药性类别信息为未耐药类别，当所述当前第一目标距离未超过所述当前第二目标距离时，所述当前目标查询特征向量对应的耐药性类别信息为已耐药类别。

10、根据权利要求1所述的方法，其特征在于，所述初始耐药性分类模型包括初始特征筛选网络和初始分类网络；

所述将所述支持集和所述查询集输入初始耐药性分类模型中，包括：

将所述支持集和所述查询集输入初始耐药性分类模型中，通过所述初始耐药性分类模型将所述各个支持样本特征向量和所述各个查询样本特征向量输入所述初始特征筛选网络中；

通过所述初始特征筛选网络基于所述各个支持样本特征向量和所述各个查询样本特征向量进行耐药性相关特征筛选，得到所述各个目标支持特征向量和所述各个目标查询特征向量，并将所述各个目标支持特征向量和所述各个目标查询特征向量输入所述分类网络中；

通过所述分类网络基于所述各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于所述各个目标查询特征向量与所述初始类别表征向量之间的相似程度确定所述各个查询样本特征向量对应的训练耐药性类别信息。

11、根据权利要求 10 所述的方法，其特征在于，所述分类网络包括样本筛选网络 and 原型网络；

将所述各个目标支持特征向量和所述各个目标查询特征向量输入所述分类网络中，包括：

将所述各个目标支持特征向量输入所述样本筛选网络中，通过所述样本筛选网络将所述各个目标支持特征向量进行映射，得到各个映射特征向量；获取初始置信度计算参数，基于所述各个映射特征向量使用所述初始置信度计算参数计算得到所述各个映射特征向量对应的置信度；基于所述置信度对所述各个映射特征向量进行加权，得到各个加权特征向量，并将所述各个加权特征向量输入所述原型网络中；

通过所述原型网络基于所述各个加权特征向量计算所述耐药性类别对应的初始类别表征向量，并基于所述各个目标查询特征向量与所述初始类别表征向量之间的相似程度确定所述各个查询样本特征向量对应的训练耐药性类别信息。

12、根据权利要求 1 所述的方法，其特征在于，所述基于所述训练耐药性类别信息和对应的所述耐药性类别标签更新所述初始耐药性分类模型，并返回将所述支持集和所述查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，包括：

基于所述训练耐药性类别信息和对应的所述耐药性类别标签进行对数损失计算，得到初始训练损失信息；

计算所述初始训练损失信息的梯度，基于所述梯度反向初始耐药性分类模型，得到更新耐药性分类模型；

将所述更新耐药性分类模型作为初始耐药性分类模型，并返回将所述支持集和所述查询集输入初始耐药性分类模型中的步骤执行，直到达到训练完成条件时，将达到训练完成条件时的初始耐药性分类模型作为目标耐药性分类模型。

13、一种分类方法，由计算机设备执行，其特征在于，所述方法包括：

获取待分类数据和样本数据，所述待分类数据中包括待分类特征向量，所述样本数据中包括各个样本特征向量和对应的样本类别标签；

将所述待分类数据和样本数据输入耐药性分类模型中，通过所述耐药性分类模型基于所述待分类特征向量和所述各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量，基于所述各个目标样本特征向量计算样本类别对应的目标类别表征向量，并基于所述目标待分类特征向量与所述目标类别表征向量之间的相似程度确定所述待分类特征向量对应的耐药性类别信息；

通过所述耐药性分类模型输出所述待分类数据对应的耐药性类别信息。

14、一种分类模型训练装置，其特征在于，所述装置包括：

数据获取模块，用于获取支持集和查询集，所述支持集包括各个支持样本特征向量和对应的耐药性类别标签，所述查询集包括各个查询样本特征向量和对应的耐药性类别标签；

初始分类模块，用于将所述支持集和所述查询集输入初始耐药性分类模型中，通过所述初始耐药性分类模型对所述各个支持样本特征向量和所述各个查询样本特征向量进行耐药性相关特征筛选，得到各个目标支持特征向量和各个目标查询特征向量，基于所述各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于所述各个目标查询特征向量与所述初始类别表征向量之间的相似程度确定所述各个查询样本特征向量对应的训练耐药性类别信息；

迭代训练模块，用于基于所述训练耐药性类别信息和对应的所述耐药性类别标签更新所

述初始耐药性分类模型，并返回将所述支持集和所述查询集输入初始耐药性分类模型中的步骤执行，当训练完成时，得到目标耐药性分类模型，所述目标耐药性分类模型用于识别蛋白质与化合物结合对应的耐药性类别。

15、根据权利要求 14 所述的装置，其特征在于，所述数据获取模块，包括：

样本获取模块，用于获取样本数据集，所述样本数据集中包括各个训练样本对应的样本特征向量和耐药性类别标签，所述样本特征向量是基于训练样本进行特征提取得到的；

抽取模块，用于从所述样本数据集中随机抽取所述支持集和所述查询集。

16、根据权利要求 15 所述的装置，其特征在于，所述装置，还包括：

最终模型训练模块，用于将所述目标耐药性分类模型作为初始耐药性分类模型，并返回从所述样本数据集中随机抽取所述支持集和所述查询集的步骤执行，直到达到最终训练完成条件时，将达到最终训练完成条件时的初始耐药性分类模型作为最终耐药性分类模型。

17、根据权利要求 15 所述的装置，其特征在于，所述装置，还包括：

特征提取模块，用于获取训练样本，所述训练样本包括野生型蛋白质信息、突变型蛋白质信息和化合物信息；基于所述野生型蛋白质信息和所述化合物信息进行野生特征提取，得到野生特征向量；基于所述突变型蛋白质信息和所述化合物信息进行突变特征提取，得到突变特征向量；基于所述野生特征向量和所述突变特征向量得到所述训练样本对应的样本特征向量。

18、根据权利要求 15 所述的装置，其特征在于，所述抽取模块还用于从所述样本数据集中进行随机抽样，得到所述查询集；计算所述查询集中各个查询样本特征向量分别与所述样本数据集中各个样本特征向量的相似程度；基于所述相似程度将所述样本数据集中的各个样本特征向量进行排序，得到样本特征向量序列；从所述样本特征向量序列中依次选取预设数量的样本特征向量，得到待抽取样本数据集；从所述待抽取样本数据集中进行随机抽取，得到所述支持集。

19、根据权利要求 14 所述的装置，其特征在于，所述初始分类模块还用于获取初始特征筛选参数；基于所述初始特征筛选参数分别对所述各个支持样本特征向量进行耐药性相关特征筛选，得到所述各个目标支持特征向量；基于所述初始特征筛选参数分别对所述各个查询样本特征向量进行耐药性相关特征筛选，得到所述各个目标查询特征向量。

20、根据权利要求 14 所述的装置，其特征在于，所述初始分类模块还用于将所述各个目标支持特征向量进行映射，得到各个映射特征向量；获取初始置信度计算参数，基于所述各个映射特征向量使用所述初始置信度计算参数计算得到所述各个映射特征向量对应的置信度；基于所述置信度对所述各个映射特征向量进行加权，得到各个加权特征向量；基于所述各个加权特征向量计算所述耐药性类别对应的初始类别表征向量。

21、根据权利要求 20 所述的装置，其特征在于，所述耐药性类别包括已耐药类别和未耐药类别；

所述初始分类模块还用于按照所述各个支持样本特征向量对应的耐药性类别标签将所述各个加权特征向量进行划分，得到所述已耐药类别对应的加权特征向量和所述未耐药类别对应的加权特征向量；基于所述已耐药类别对应的加权特征向量进行向量平均，得到所述已耐药类别对应的第一初始类别表征向量；基于所述未耐药类别对应的加权特征向量进行向量平均，得到所述未耐药类别对应的第二初始类别表征向量。

22、根据权利要求 21 所述的装置，其特征在于，所述初始分类模块还用于计算所述各个

目标查询特征向量中当前目标查询特征向量分别与第一初始类别表征向量和所述第二初始类别表征向量之间的距离，得到当前第一初始距离和当前第二初始距离；比较所述当前第一初始距离和所述当前第二初始距离，当所述当前第一目标距离超过所述当前第二目标距离时，所述当前目标查询特征向量对应的训练耐药性类别信息为未耐药类别，当所述当前第一目标距离未超过所述当前第二目标距离时，所述当前目标查询特征向量对应的耐药性类别信息为已耐药类别。

23、根据权利要求 14 所述的装置，其特征在于，所述初始耐药性分类模型包括初始特征筛选网络和初始分类网络；

所述初始分类模块还用于将所述支持集和所述查询集输入初始耐药性分类模型中，通过所述初始耐药性分类模型将所述各个支持样本特征向量和所述各个查询样本特征向量输入所述初始特征筛选网络中；通过所述初始特征筛选网络基于所述各个支持样本特征向量和所述各个查询样本特征向量进行耐药性相关特征筛选，得到所述各个目标支持特征向量和所述各个目标查询特征向量，并将所述各个目标支持特征向量和所述各个目标查询特征向量输入所述分类网络中；通过所述分类网络基于所述各个目标支持特征向量计算耐药性类别对应的初始类别表征向量，并基于所述各个目标查询特征向量与所述初始类别表征向量之间的相似程度确定所述各个查询样本特征向量对应的训练耐药性类别信息。

24、根据权利要求 23 所述的装置，其特征在于，所述分类网络包括样本筛选网络和原型网络；

所述初始分类模块还用于将所述各个目标支持特征向量输入所述样本筛选网络中，通过所述样本筛选网络将所述各个目标支持特征向量进行映射，得到各个映射特征向量；获取初始置信度计算参数，基于所述各个映射特征向量使用所述初始置信度计算参数计算得到所述各个映射特征向量对应的置信度；基于所述置信度对所述各个映射特征向量进行加权，得到各个加权特征向量，并将所述各个加权特征向量输入所述原型网络中；通过所述原型网络基于所述各个加权特征向量计算所述耐药性类别对应的初始类别表征向量，并基于所述各个目标查询特征向量与所述初始类别表征向量之间的相似程度确定所述各个查询样本特征向量对应的训练耐药性类别信息。

25、根据权利要求 14 所述的装置，其特征在于，所述迭代训练模块还用于基于所述训练耐药性类别信息和对应的所述耐药性类别标签进行对数损失计算，得到初始训练损失信息；计算所述初始训练损失信息的梯度，基于所述梯度反向初始耐药性分类模型，得到更新耐药性分类模型；将所述更新耐药性分类模型作为初始耐药性分类模型，并返回将所述支持集和所述查询集输入初始耐药性分类模型中的步骤执行，直到达到训练完成条件时，将达到训练完成条件时的初始耐药性分类模型作为目标耐药性分类模型。

26、一种分类装置，其特征在于，所述装置包括：

待分类数据获取模块，用于获取待分类数据和样本数据，所述待分类数据中包括待分类特征向量，所述样本数据中包括各个样本特征向量和对应的样本类别标签；

分类模块，用于将所述待分类数据和样本数据输入耐药性分类模型中，所述耐药性分类模型基于所述待分类特征向量和所述各个样本特征向量进行耐药性相关特征筛选，得到目标待分类特征向量和各个目标样本特征向量，基于所述各个目标样本特征向量计算样本类别对应的目标类别表征向量，并基于所述目标待分类特征向量与所述目标类别表征向量之间的相似程度确定所述待分类特征向量对应的耐药性类别信息；

类别输出模块，用于所述耐药性分类模型输出所述待分类数据对应的耐药性类别信息。

27、一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现权利要求 1 至 13 中任一项所述的方法的步骤。

28、一种计算机可读存储介质，存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现权利要求 1 至 13 中任一项所述的方法的步骤。

29、一种计算机程序产品，包括计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现权利要求 1 至 13 中任一项所述的方法的步骤。

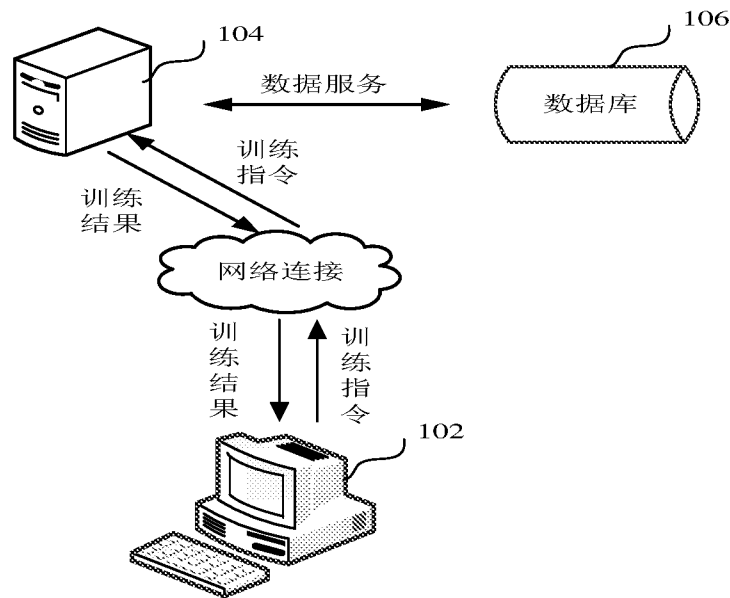


图 1

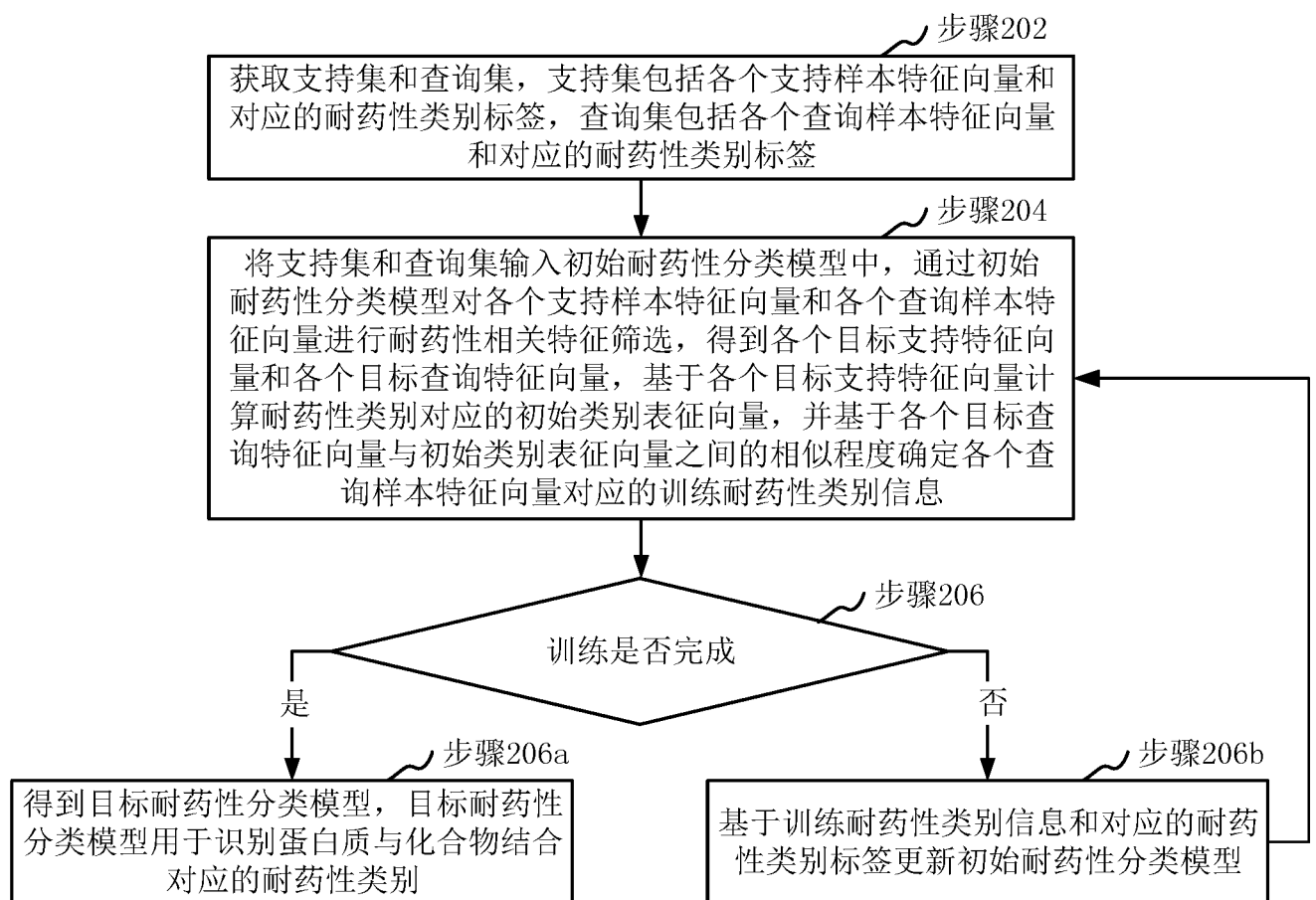


图 2

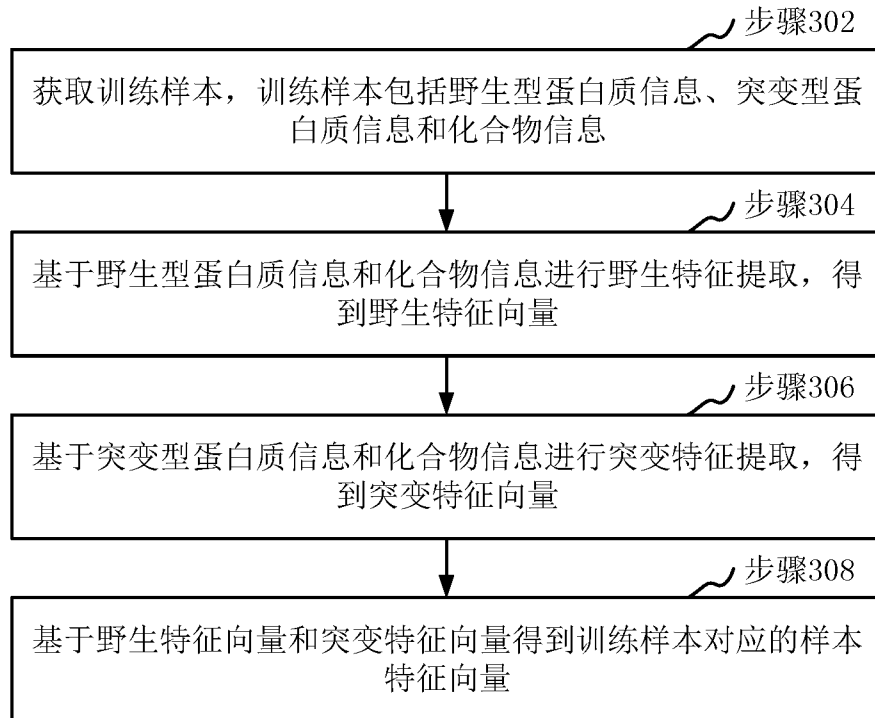


图 3

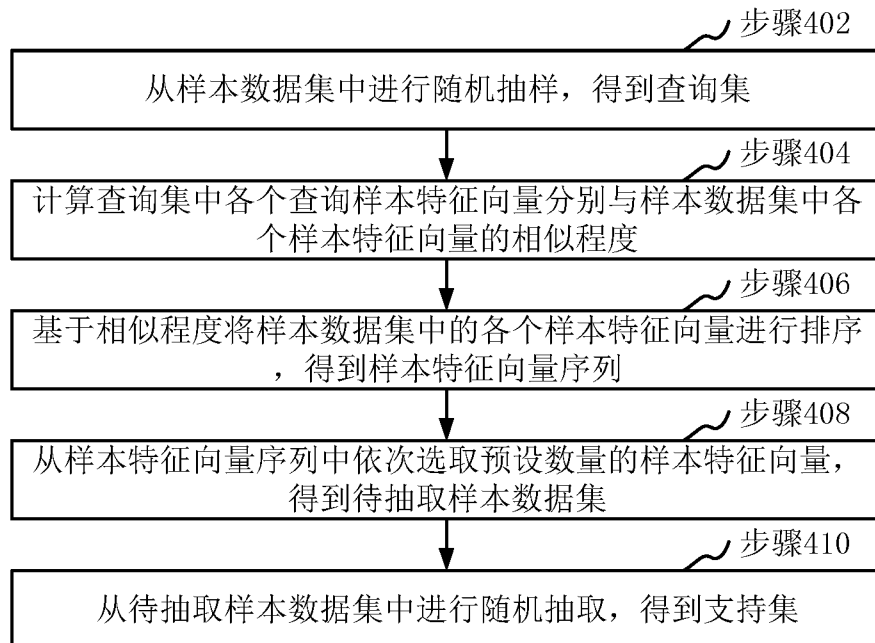


图 4

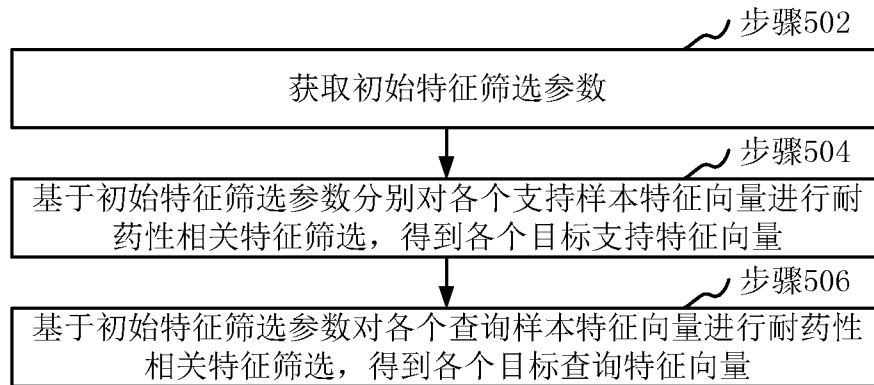


图 5

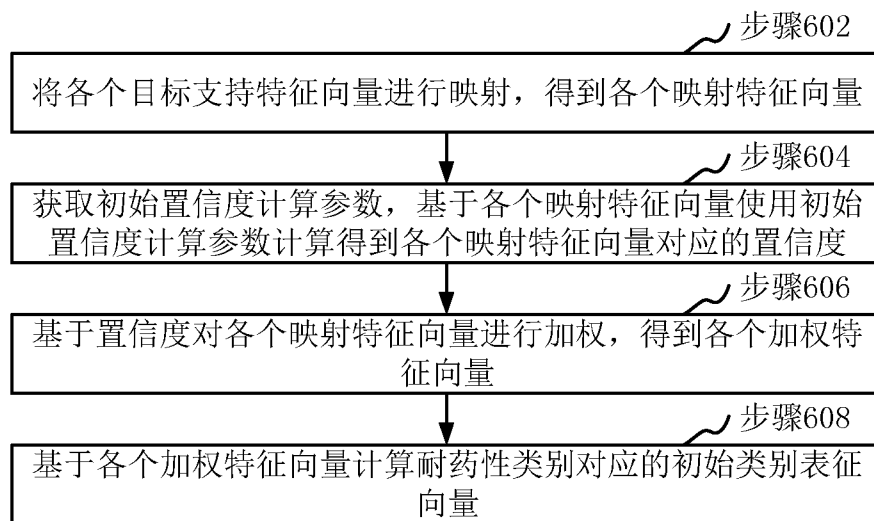


图 6

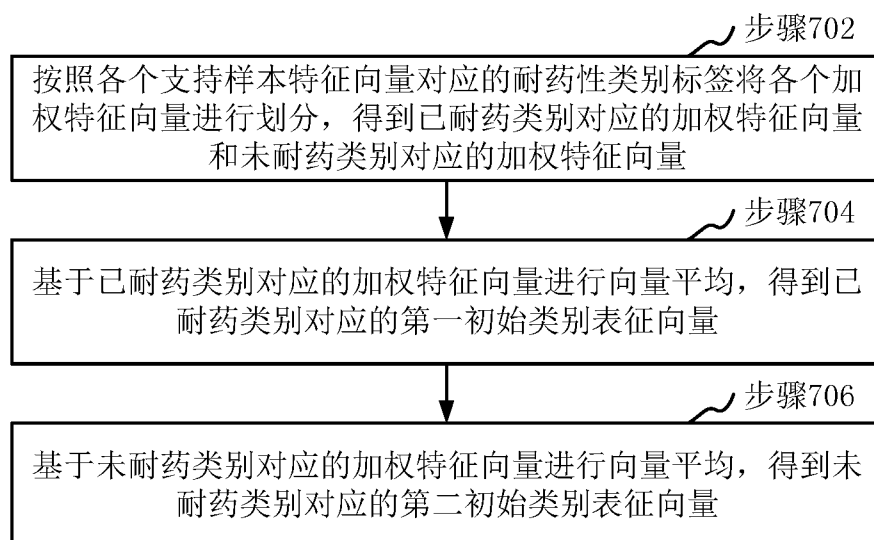


图 7

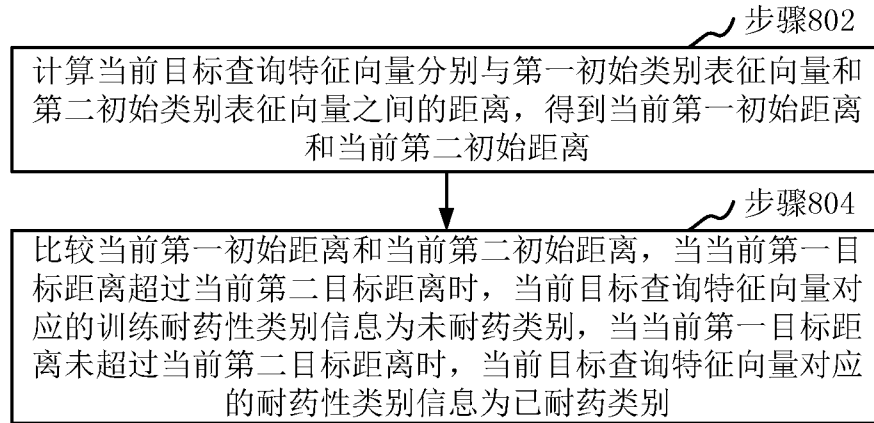


图 8

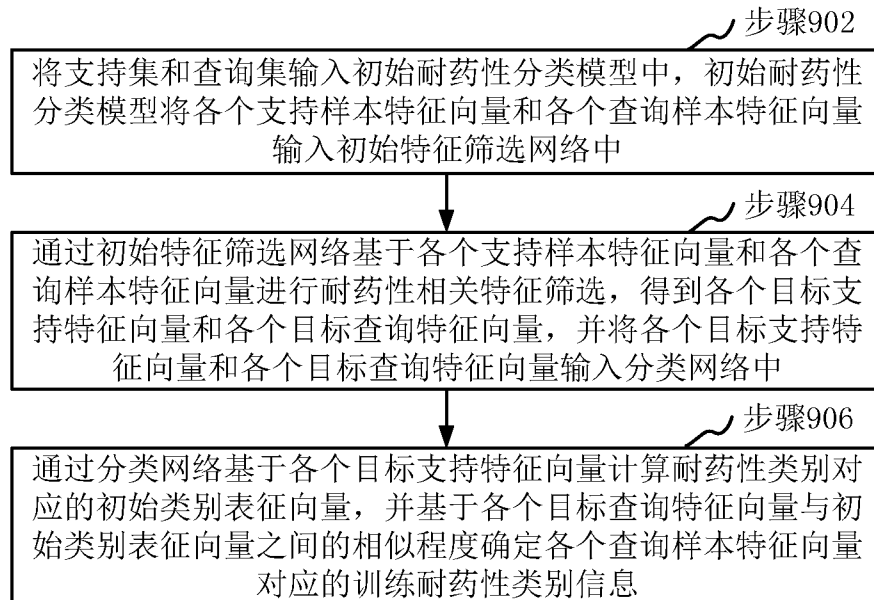


图 9

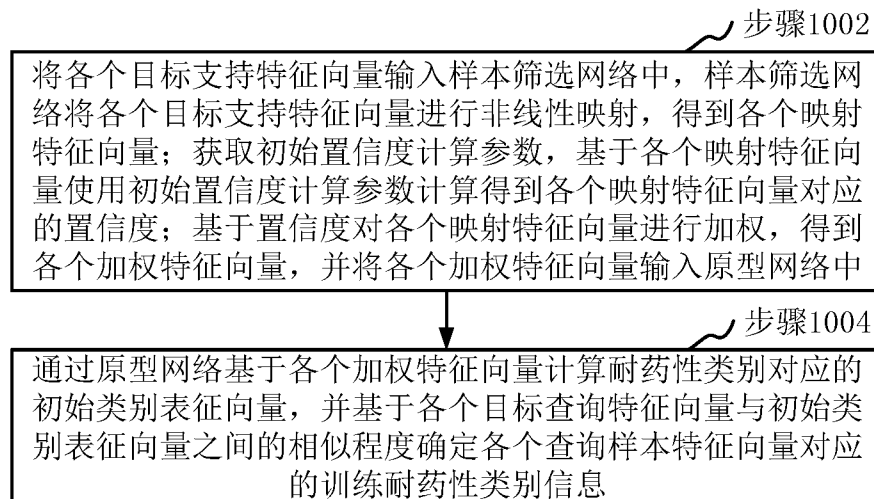


图 10

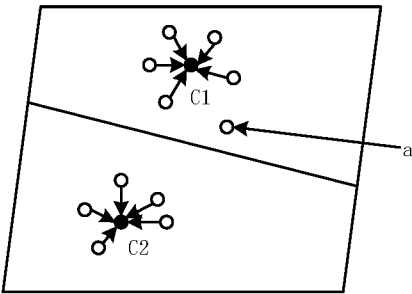


图 11

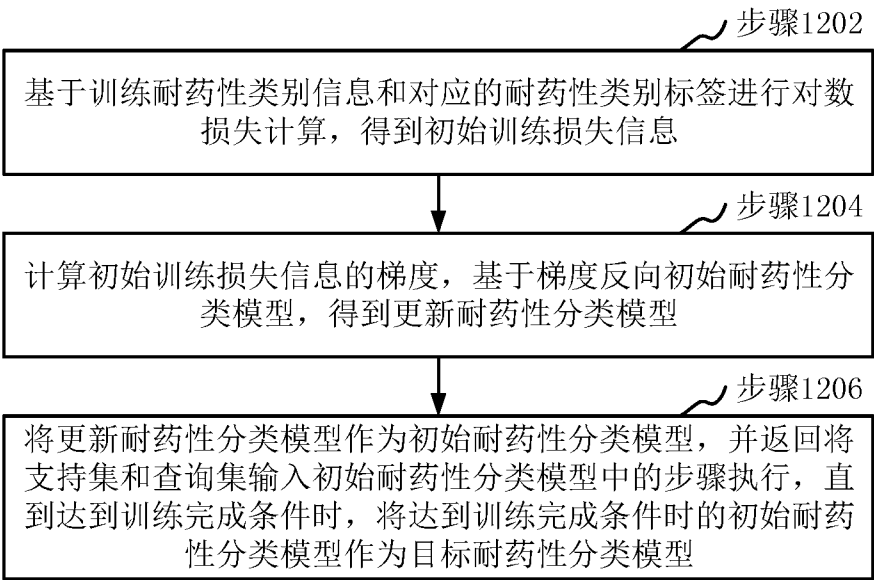


图 12

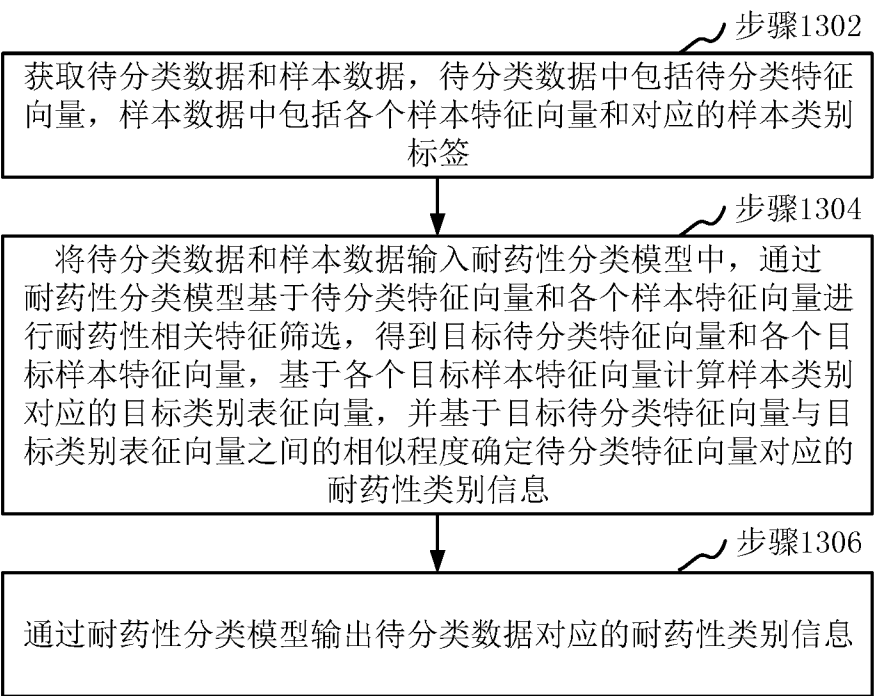


图 13

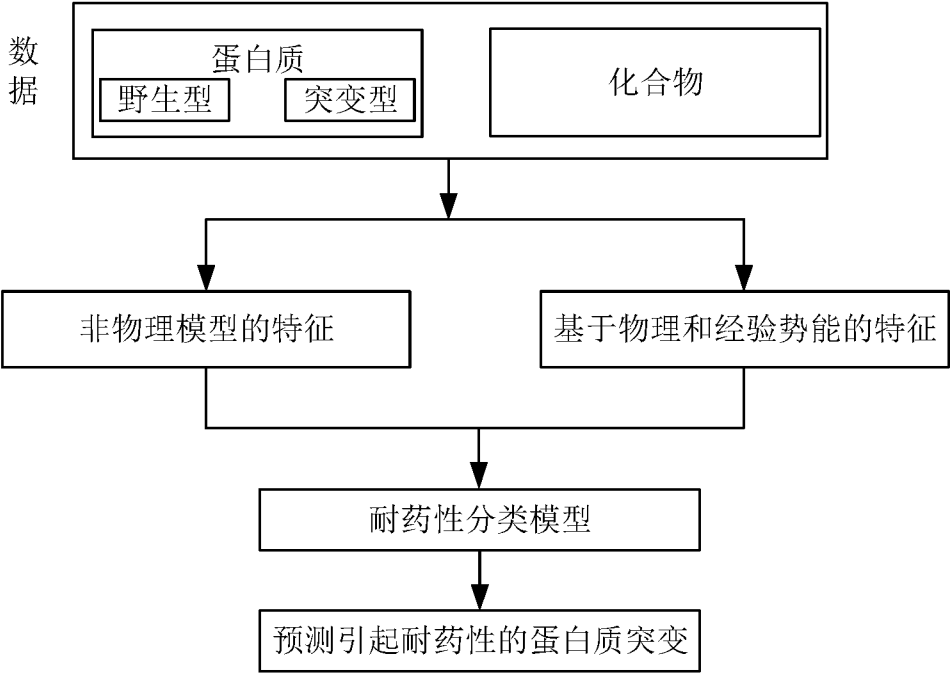


图 14

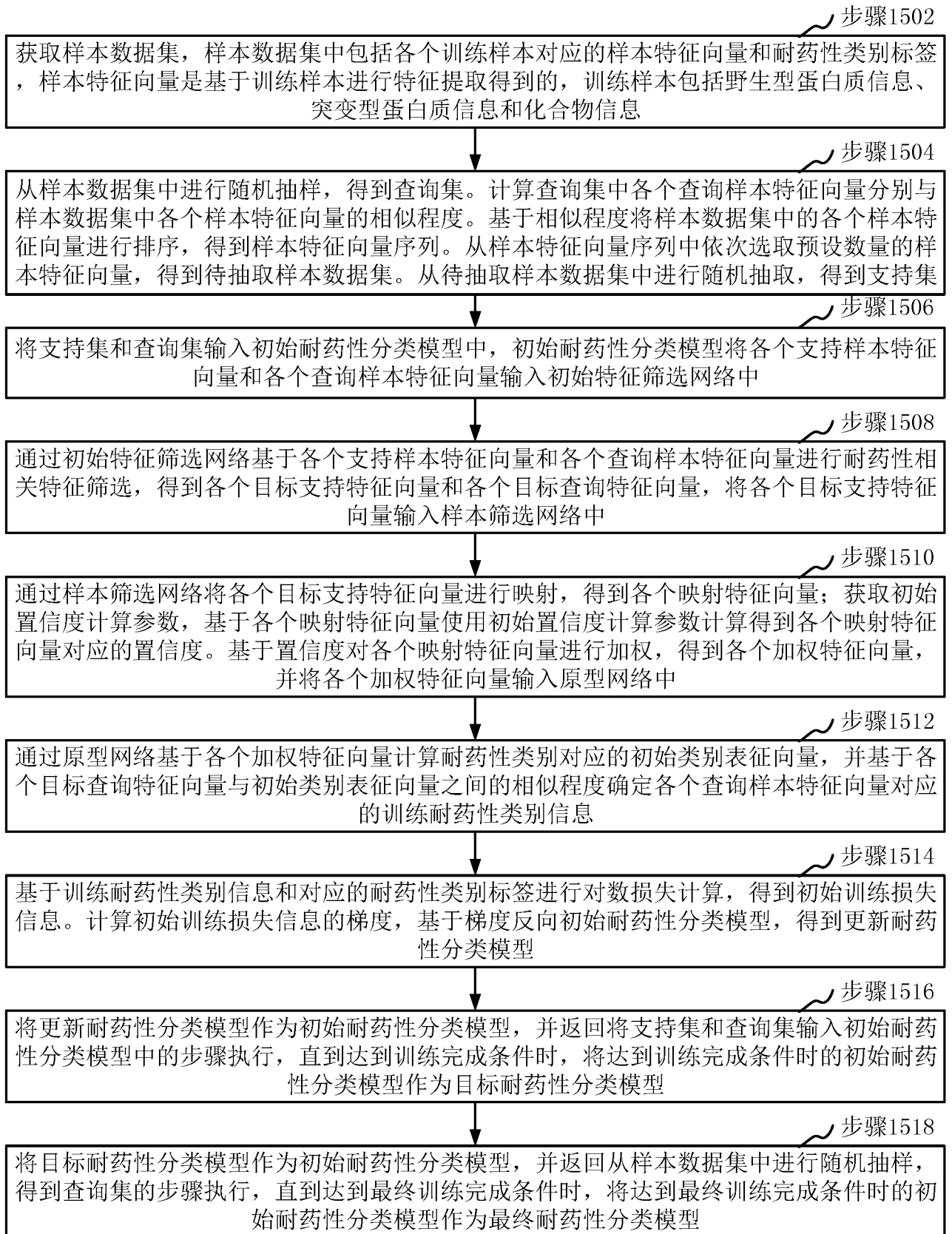


图 15

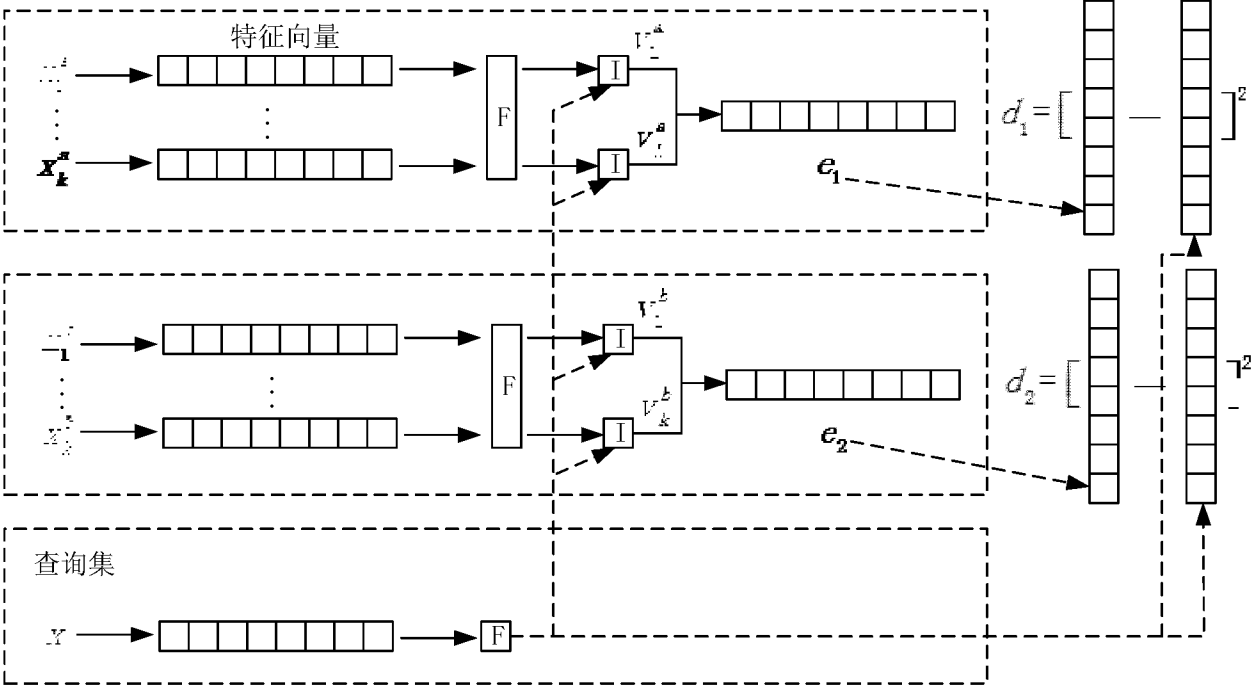


图 16

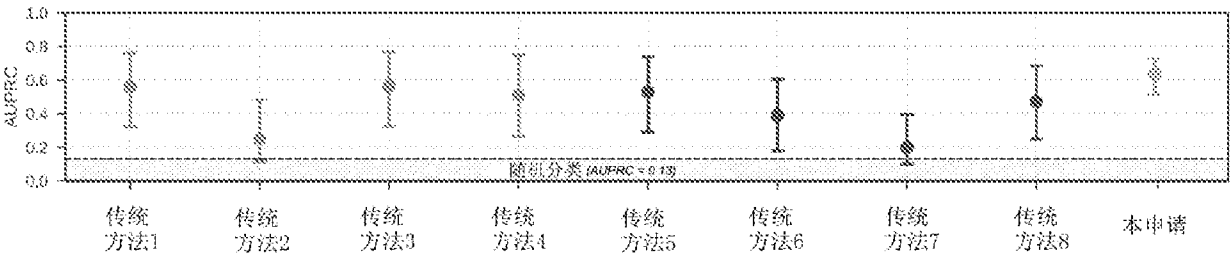


图 17

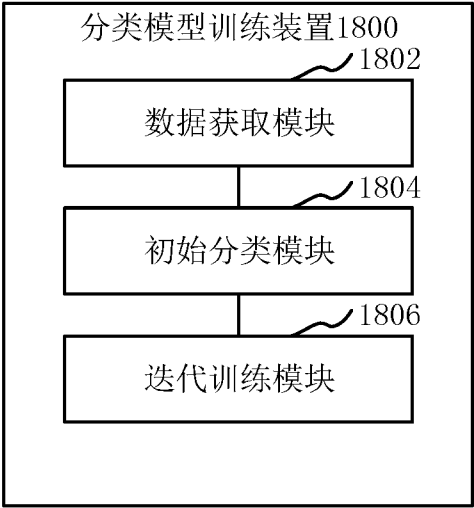


图 18

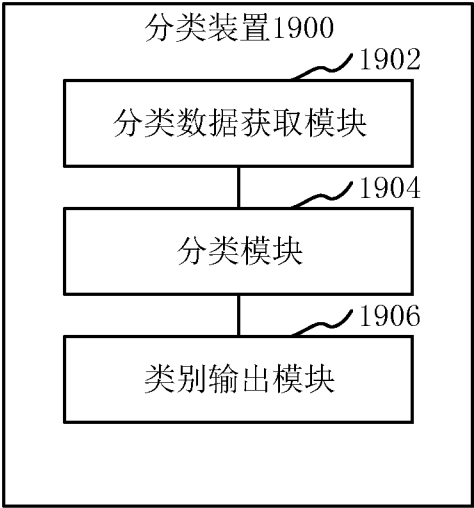


图 19

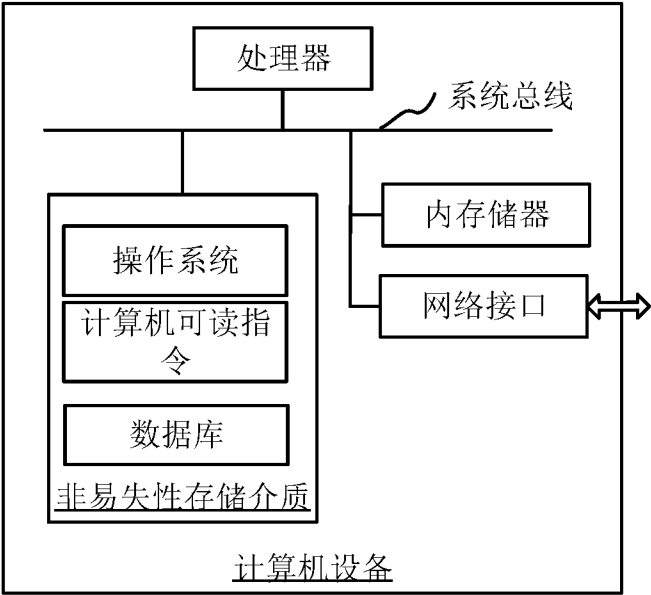


图 20

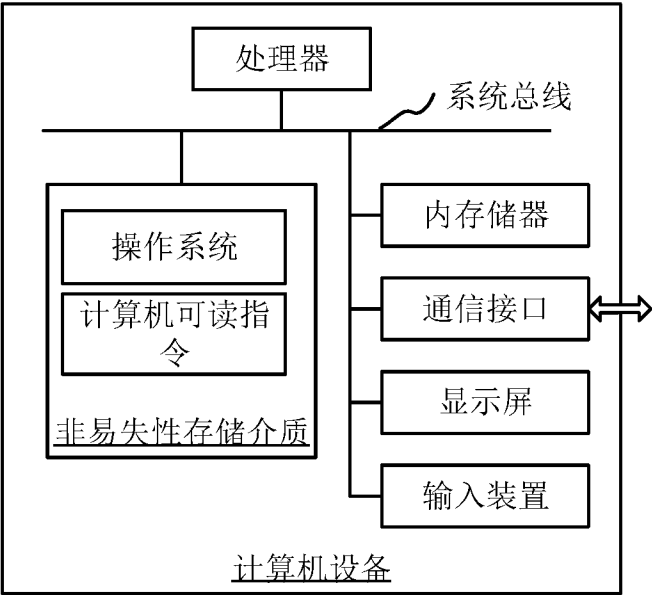


图 21

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/083074

A. CLASSIFICATION OF SUBJECT MATTER

G16B 40/00(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G16B

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

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

CNPAT, WPI, EPODOC, CNKI, IEEE: 分类, 模型, 支持集, 查询集, 特征向量, 相似, 训练, classification, model, support set, query set, feature, vector, similarity, train

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 113299346 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 24 August 2021 (2021-08-24) claims 1-15, and description, paragraphs [0123]-[0128], [0154]-[0159] and [0274]	1-29
Y	YANG, Ziyi et al. "Select-ProtoNet: Learning to Select for Few-Shot Disease Subtype Prediction" <i>arXiv</i> , 03 September 2020 (2020-09-03), pp. 1-11	1-29
Y	王纪防 等 (WANG, Jifang et al.). "靶蛋白耐药性变异的计算机预测 (Non-official translation: Computerized Prediction of Target Protein Resistance Variants)" <i>国外医学药学分册 (Foreign Medical Sciences (Section of Pharmacy))</i> , Vol. 32, No. 5, 31 October 2005 (2005-10-31), pp. 297-299	1-29
A	CN 111767400 A (PING AN INTERNATIONAL SMART CITY TECHNOLOGY CO., LTD.) 13 October 2020 (2020-10-13) entire document	1-29



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

02 June 2022

Date of mailing of the international search report

20 June 2022

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088, China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2022/083074

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	113299346	A	24 August 2021	None	
CN	111767400	A	13 October 2020	None	

国际检索报告

国际申请号

PCT/CN2022/083074

A. 主题的分类

G16B 40/00 (2019.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G16B

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

CNPAT, WPI, EPDOC, CNKI, IEEE: 分类, 模型, 支持集, 查询集, 特征向量, 相似, 训练, classification, model, support set, query set, feature, vector, similarity, train

C. 相关文件

类 型*

引用文件, 必要时, 指明相关段落

相关的权利要求

PX

CN 113299346 A (腾讯科技深圳有限公司) 2021年8月24日 (2021 - 08 - 24)
权利要求1-15, 说明书第[0123]-[0128], [0154]-[0159], [0274]段

1-29

Y

YANG, Ziyi 等. "Select-ProtoNet: Learning to Select for Few-Shot Disease Subtype Prediction"
arXiv, 2020年9月3日 (2020 - 09 - 03),
第1-11页

1-29

Y

王纪防 等. "靶蛋白耐药性变异的计算机预测"
国外医学药学分册, 第32卷, 第5期, 2005年10月31日 (2005 - 10 - 31),
第297-299页

1-29

A

CN 111767400 A (平安国际智慧城市科技股份有限公司) 2020年10月13日 (2020 - 10 - 13)
全文

1-29

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2022年6月2日

国际检索报告邮寄日期

2022年6月20日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

穆滢

电话号码 86-(10)-53961531

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/083074

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	113299346	A	2021年8月24日	无	
CN	111767400	A	2020年10月13日	无	