

# (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  
2019 年 7 月 18 日 (18.07.2019)



(10) 国际公布号  
**WO 2019/136806 A1**

- (51) 国际专利分类号:  
**G16H 50/70** (2018.01) **G06K 9/62** (2006.01)  
**G06F 17/30** (2006.01)
- (21) 国际申请号: PCT/CN2018/077475
- (22) 国际申请日: 2018 年 2 月 28 日 (28.02.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
201810030552.5 2018 年 1 月 12 日 (12.01.2018) CN
- (71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];  
中国广东省深圳市福田区八卦岭八卦三路平安大厦6楼, Guangdong 518000 (CN)。
- (72) 发明人: 鲁宁(LU, Ning); 中国广东省深圳市福田区八卦岭八卦三路平安大厦6楼, Guangdong 518000 (CN)。 周凡(ZHOU, Fan); 中国广东省深圳市福田区八卦岭八卦三路平安大厦6楼, Guangdong 518000 (CN)。 程诚(CHENG, Cheng);

中国广东省深圳市福田区八卦岭八卦三路平安大厦6楼, Guangdong 518000 (CN)。

- (74) 代理人: 深圳众鼎专利商标代理事务所(普通合伙)(ZHONGDING INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市龙岗区龙城街道中心城清林路546号城投商务中心4层/B, Guangdong 518172 (CN)。
- (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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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,

(54) Title: MEDICAL MODEL TRAINING METHOD AND APPARATUS, MEDICAL IDENTIFICATION METHOD AND APPARATUS, DEVICE, AND MEDIUM

(54) 发明名称: 医疗模型训练方法、医疗识别方法、装置、设备及介质

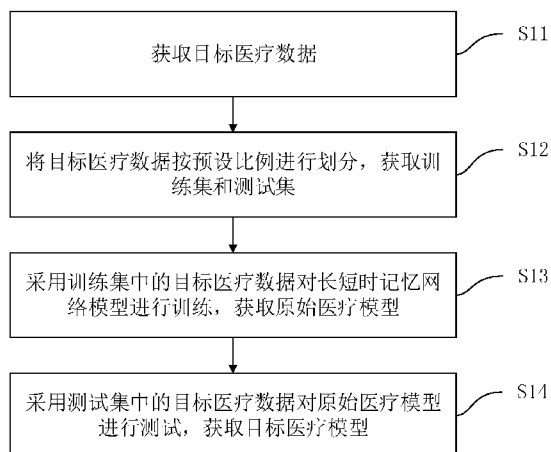


图 1

(57) Abstract: A medical model training method, comprising: obtaining target medical data (S11); dividing the target medical data according to a preset ratio to obtain a training set and a test set (S12); training a long short-term memory network model by using the target medical data in the training set to obtain an original medical model (S13); and testing the original medical model by using the target medical data in the test set to obtain a target medical model (S14). The medical model training method improves the accuracy of disease recognition. Also disclosed are a medical identification method and apparatus, a device, and a medium.

(57) 摘要: 一种医疗模型训练方法包括: 获取目标医疗数据(S11); 将目标医疗数据按预设比例进行划分, 获取训练集和测试集(S12); 采用训练集中的目标医疗数据对长短时记忆网络模型进行训练, 获取原始医疗模型(S13); 采用测试集中的目标医疗数据对原始医疗模型进行测试, 获取目标医疗模型(S14)。该医疗模型训练方法提高了疾病识别的准确率。还公开了一种医疗识别方法、装置、设备及介质。

- S11 Obtain target medical data  
S12 Divide the target medical data according to a preset ratio to obtain a training set and a test set  
S13 Train a long short-term memory network model by using the target medical data in the training set to obtain an original medical model  
S14 Test the original medical model by using the target medical data in the test set to obtain a target medical model

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))。

## 医疗模型训练方法、医疗识别方法、装置、设备及介质

本专利申请以 2018 年 1 月 12 日提交的申请号为 201810030552.5，名称为“医疗模型训练方法、医疗识别方法、装置、设备及介质”的中国发明专利申请为基础，并要求其优先权。

### 技术领域

本申请涉及数据识别领域，尤其涉及一种医疗模型训练方法、医疗识别方法、装置、设备及介质。

### 背景技术

当前疾病的识别通常由医生根据病人的症状确定，需依赖于医生的专业和经验，无法使用户根据自身的症状情况及时了解自身的健康情况。由于当前疾病的识别主要由医生进行诊断，一般用户都是在病情比较严重时才会意识到，进而到医疗机构就诊，可能延误病情。当前医疗网站中提供的症状、疾病和治疗等医疗数据没有形成完整的系统，用户查询时没有足够的专业知识，无法根据自己的症状情况及时了解自身的健康状态。

### 发明内容

本申请实施例提供一种医疗模型训练方法，以解决当前疾病的识别通常由医生根据病人的症状确定，需依赖于医生的专业和经验，无法使用户根据自身的症状情况及时了解自身的健康情况的问题。

本申请实施例提供一种医疗识别方法，以解决当前医疗网站中提供的症状、疾病和治疗等医疗数据没有形成完整的系统，用户查询时没有足够的专业知识，无法根据自己的症状情况及时了解自身的健康状态的问题。

第一方面，本申请实施例提供一种医疗模型训练方法，包括：

获取目标医疗数据；

将所述目标医疗数据按预设比例进行划分，获取训练集和测试集；

采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模

型；

采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试，获取目标医疗模型。

第二方面，本申请实施例提供一种医疗模型训练装置，包括：

目标医疗数据获取模块，用于获取目标医疗数据；

目标医疗数据划分模块，用于将所述目标医疗数据按预设比例进行划分，获取训练集和测试集；

原始医疗模型获取模块，用于采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模型；

目标医疗模型获取模块，用于采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试，获取目标医疗模型。

第三方面，本申请实施例提供一种医疗识别方法，包括：

获取用户的待测医疗数据，所述待测医疗数据包括至少一个症状特征；

将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，所述目标医疗模型是采用第一方面所述医疗模型训练方法获取的模型。

第四方面，本申请实施例提供一种医疗识别装置，包括：

待测医疗数据获取模块，用于获取用户的待测医疗数据，所述待测医疗数据包括至少一个症状特征；

目标医疗特征获取模块，用于将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，所述目标医疗模型是采用第一方面所述医疗模型训练方法获取的模型。

第五方面，本申请实施例提供一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现如下步骤：

获取目标医疗数据；

将所述目标医疗数据按预设比例进行划分，获取训练集和测试集；

采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模型；

采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试，获取目标医疗模型。

第六方面，本申请实施例提供一种终端设备，包括存储器、处理器以及存储在所述存

存储器中并可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现如下步骤：

获取用户的待测医疗数据，所述待测医疗数据包括至少一个症状特征；

将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，所述目标医疗模型是采用第一方面所述医疗模型训练方法获取的模型。

第七方面，本申请实施例提供一种计算机可读介质，所述计算机可读介质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如下步骤：

获取目标医疗数据；

将所述目标医疗数据按预设比例进行划分，获取训练集和测试集；

采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模型；

采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试，获取目标医疗模型。

第八方面，本申请实施例提供一种计算机可读介质，所述计算机可读介质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如下步骤：

获取用户的待测医疗数据，所述待测医疗数据包括至少一个症状特征；

将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，所述目标医疗模型是采用第一方面所述医疗模型训练方法获取的模型。

本申请实施例提供的医疗模型训练方法、装置、设备及介质中，先获取目标医疗数据，并对目标医疗数据进行划分，获取训练集和测试集，以便基于训练集中的目标医疗数据对长短时记忆网络模型进行训练，能够有效更新长短时记忆网络模型中各层的权值，以保证训练获取的原始医疗模型能够识别目标医疗特征。最后，再基于测试集中的目标医疗数据对原始医疗模型进行测试，以保证获取到的目标医疗特征识别目标医疗特征的准确率更高。

本申请实施例提供的医疗识别方法、装置、设备及介质中，通过获取用户的待测医疗数据，以便基于目标医疗模型对待测医疗数据进行识别，以获取目标医疗特征。通过目标医疗模型对待测医疗数据进行识别，以保证识别的准确率更高，较精准的识别目标医疗特征，以辅助用户判断患病风险，以便及时预防。

附图说明

为了更清楚地说明本申请实施例的技术方案，下面将对本申请实施例的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性的前提下，还可以根据这些附图获得其他的附图。

图 1 是实施例 1 中提供的医疗模型训练方法的一流程图。

图 2 是图 1 中步骤 S11 的一具体示意图。

图 3 是图 2 中步骤 S112 的一具体示意图。

图 4 是图 1 中步骤 S13 的一具体示意图。

图 5 是图 1 中步骤 S14 的一具体示意图。

图 6 是实施例 2 中提供的医疗模型训练装置的一原理框图。

图 7 是实施例 3 中提供的医疗识别方法的一流程图。

图 8 是图 7 中步骤 S22 的一具体示意图。

图 9 实施例 4 中提供的医疗识别装置的一原理框图。

图 10 是实施例 6 中提供的终端设备的一示意图。

## 具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

### 实施例 1

图 1 示出本实施例中医疗模型训练方法的流程图。该医疗模型训练方法可快速从网络中采集到大量的医疗数据，以便基于采集到的医疗数据进行医疗模型训练。该医疗模型训练方法可具体应用在医疗知识库管理系统这一数据管理系统中，用于对用户所输入的症状特征进行识别，为用户推荐疑似疾病特征列表，能够有效增加识别的准确率。如图 1 所示，该医疗模型训练方法包括如下步骤：

S11：获取目标医疗数据。

其中，目标医疗数据是用于进行模型训练的数据。该目标医疗数据包括但不限于本实施例中的症状特征、疾病特征等医疗数据。本实施例中，通过获取目标医疗数据，以便后续进行模型训练。

S12：将目标医疗数据按预设比例进行划分，获取训练集和测试集。

其中，预设比例是预先设定好的，用于对目标医疗数据进行分类的比例。该预设比例可以是根据历史经验获取的比例。其中，训练集 (training set) 是学习样本数据集，是通过匹配一些参数来建立分类器，即采用训练集中的目标医疗数据训练机器学习模型，以确定机器学习模型的参数。测试集 (test set) 是用于测试训练好的机器学习模型的分辨能力，如识别率。本实施例中，可按照 9:1 的比例对目标医疗数据进行分类，即可将 90% 的目标医疗数据作为训练集，剩余 10% 的数据作为测试集。

S13: 采用训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模型。

其中，长短时记忆网络 (Long Short-Term Memory, 简称 LSTM) 模型，是一种时间递归神经网络模型，用于训练具有时序性特点的数据，将该具有时序性特点的数据在长短时记忆网络模型训练，能够获取与该数据相对应的识别模型。本实施例中，该具有时序性特点的数据为目标医疗数据，通过训练集中的目标医疗数据获取的模型即为原始医疗模型。长短时记忆网络模型包括一输入层、一输出层和至少一隐藏层，长短时记忆网络模型中各层的权值是指神经网络模型中各层连接的权值，权值决定了各层最终输出的信息，并使得网络具有时序上的记忆功能。通过采用目标医疗数据对长短时记忆网络模型进行训练，能够有效更新长短时记忆网络模型中各层的权值。并且，长短时记忆网络模型通过对具有时序性特点的训练集中的目标医疗数据进行识别，可使目标医疗模型的识别结果更为准确。

本实施例中，LSTM 模型的输出层采用 Softmax (回归模型) 进行回归处理，用于分类输出权重矩阵。Softmax (回归模型) 是一种常用于神经网络的分类函数，它将多个神经元的输出，映射到 [0, 1] 区间内，可以理解成概率，计算起来简单方便，从而来进行多分类输出，使其输出结果更准确。

S14: 采用测试集中的目标医疗数据对原始医疗模型进行测试，获取目标医疗模型。

其中，目标医疗模型是经过不断对原始医疗模型进行测试、改进所得到的模型。本实施例中，将测试集中每一目标医疗数据输入到原始医疗模型进行测试，获取得到预测医疗特征；再基于预测医疗特征和测试集中的目标医疗数据，确定模型的准确率，以便基于模型的准确率，获取目标医疗模型。

本实施例中，先获取目标医疗数据，并对目标医疗数据进行划分，获取训练集和测试集，以便基于训练集中的目标医疗数据对长短时记忆网络模型进行训练，能够有效更新长短时记忆网络模型中各层的权值，以保证训练获取的原始医疗模型能够识别目标医疗特征。最后，再基于测试集中的目标医疗数据对原始医疗模型进行测试，以保证获取到的目

标医疗特征识别目标医疗特征的准确率更高。

在一具体实施方式中，如图 2 所示，步骤 S11 中，即获取目标医疗数据，具体包括如下步骤：

S111：获取目标网页地址。

其中，目标网页地址是预先定义好所要获取目标医疗数据对应的网页地址。例如：39 健康网站。通过获取到目标网页地址，以使后续采用网络爬虫技术爬取目标医疗数据时，为基于目标网页地址获取目标医疗数据提供支持。

S112：采用爬虫工具爬取目标网页地址对应的网页，获取原始医疗数据。

其中，爬虫工具是按照一定的规则自动爬取网页地址所对应的网页内容的工具，例如 Python 爬虫工具。具体地，采用爬虫工具爬取目标网页地址所对应的网页内容，以获取原始医疗数据，每一原始医疗数据都包含实际疾病特征和对应的至少一个症状特征，例如感冒（实际疾病特征）对应流鼻涕、头疼和发烧等症状特征。本实施例中，采用爬虫工具爬取目标网页地址对应的网页，无需人工搜索，有利于提高数据采集的效率。

本实施例中，爬虫工具会采取周期性爬取的方式对目标网页地址进行爬取数据，以使原始医疗数据具有时序性，以使后续模型训练获取到的目标医疗特征也具有时序性。

S113：对原始医疗数据进行数据清洗，获取目标医疗数据。

其中，数据清洗是指对原始医疗数据按照一定规则进行处理，获取纯净的目标医疗数据的方法。目标医疗数据是指按照数据清洗规则进行处理得到的纯净的数据。该数据清洗规则包括但不限于去除重复的数据。由于每一原始医疗数据均包含实际疾病特征和对应的至少一个症状特征，依据数据清洗规则进行清洗时，可将实际疾病特征和症状特征均相同的两个或两个以上原始医疗数据合并为一原始医疗数据。本实施例中，通过对原始医疗数据进行数据清洗，以获取目标医疗数据，能够有效提升目标医疗数据的质量，并且，去除重复的原始医疗数据，以使后续采用目标医疗数据进行训练时，无需对重复的原始医疗数据进行再次训练，能够有效减少训练时长，节省时间，提高训练效率。

本实施例中，先获取目标网页地址，然后采用爬虫工具爬取目标网页地址对应的网页，以获取原始医疗数据，无需人工搜索，有利于提高数据采集的效率。最后对原始医疗数据进行数据清洗，获取目标医疗数据，以使后续采用目标医疗数据进行训练时，无需对重复的数据进行再次训练，能够有效减少训练时长，节省时间，提高训练效率。

在一具体实施方式中，如图 3 所示，步骤 S112 中，即采用爬虫工具爬取目标网页地址对应的网页，获取原始医疗数据，具体包括如下步骤：



S1121: 采用爬虫工具, 依据深度优先算法或广度优先算法爬取目标网页地址所链接的至少一个访问地址, 每一访问地址对应一网页。

其中, 深度优先算法是指网络爬虫会从起始页开始, 一个链接一个链接跟踪下去, 处理完这条线路之后再转入下一个起始页, 继续追踪链接。目标网页地址对应一目标网页, 该目标网页包括起始页和至少一个起始页所链接的至少一个访问地址, 每一访问地址对应一访问网页。爬虫工具包括网页提取工具和网页下载工具。网页提取工具是用于提取访问地址的工具, 步骤 S1121 具体是采用网页提取工具爬取目标网页地址所链接的至少一个访问地址。网页下载工具是用于下载访问地址对应的网页的工具。

本实施例中, 还可采用广度优先算法不断从当前页面上爬取新的网页地址放入待下载消息队列中, 直到预设停止条件满足时停止执行爬虫任务。其中, 广度优先算法是指将新下载网页发现的链接直接插入到待抓取消息队列的末尾, 也就是指网络爬虫会先抓取起始页中的所有网页, 然后在选择其中的一个链接的网页, 继续抓取在此网页中链接的所有网页。

S1122: 将至少一个访问地址存储在待下载消息队列中。

具体地, 将步骤 S121 中爬取到的每一访问地址依据爬取到的时间先后顺序存储在待下载消息队列中, 以便在执行步骤 S123 时, 可基于待下载消息队列中的网页地址进行爬取数据。待下载消息队列依据先进先入的方式对访问地址进行处理, 可使爬取访问地址和基于访问地址爬取原始医疗数据异步处理, 有利于提高获取原始医疗数据效率。

具体地, 爬虫工具中的网页提取工具会先抓取起始页中的所有网页内容, 然后再选择起始页所链接的至少一个访问地址, 继续抓取此访问地址链接的网页。本实施例中, 起始页为目标网页地址。

S1123: 采用爬虫工具对待下载消息队列中的每一访问地址对应的网页进行数据提取, 获取原始医疗数据。

具体地, 采用网页下载工具根据待下载消息队列的每一访问地址自动下载该访问地址对应的网页中所有医疗数据。本实施例中, 待下载消息队列中存储有多个包含原始医疗数据的网页地址, 爬虫工具的网页下载工具依序从待下载消息队列中逐一获取访问地址并下载该访问地址对应的网页中的原始医疗数据。具体地, 爬虫工具从待下载消息队列的队头获取到一访问地址并对该访问地址对应的网页进行下载, 将下载的原始医疗数据存储在数据库后, 注销待下载消息队列中相应的网页地址, 重复上述步骤直至待下载消息队列中不存在访问地址, 以获取爬虫工具爬取的所有网页中的原始医疗数据。本实施例中, 爬虫工

具中的网页下载工具根据用户设置的爬虫任务自动从互联网上爬取包含原始医疗数据的网页地址，无需人工搜索，有利于提高数据采集效率。

本实施例中，采用爬虫工具，依据深度优先算法或广度优先算法爬取目标网页地址所链接的至少一个访问地址，并将获取到的访问地址存储在待下载消息队列中，再采用爬虫工具基于待下载消息队列中获取的访问地址下载原始医疗数据，使得访问地址下载的原始医疗数据异步处理，有利于提高原始医疗数据的获取效率。本实施例中，通过采用网页提取工具和网页下载工具异步处理获取原始医疗数据，有利于提高获取原始医疗数据的效率。

在一具体实施方式中，如图4所示，步骤S13中，即采用训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模型，具体包括如下步骤：

S131：初始化长短时记忆网络模型。

本实施例中，对长短时记忆网络模型进行初始化操作，其中，长短时记忆神经网络是在时间上相互连接的网络，其基本单元称为神经元。长短时记忆网络模型包括一输入层、一输出层和至少一隐藏层，其隐藏层包括输入门、遗忘门、输出门、神经元状态和神经元输出，长短时记忆网络模型中的每一层可以包括多个神经元。遗忘门决定了在神经元状态中所要丢弃的信息。输入门决定了在神经元中所要增加的信息。输出门决定了在神经元中所要输出的信息。神经元状态决定了各个门丢弃、增加和输出的信息，具体表示为与各个门之间连接的权值。神经元输出决定了与下一层的连接权值。可以理解地，初始化长短时记忆网络模型，即为设置长短时记忆网络模型各层之间连接的权值以及隐藏层中输入门、遗忘门、输出门、神经元状态和神经元输出之间的初始权值，本实施例中初始权值可设为1。

S132：在长短时记忆网络模型中输入训练集中的目标医疗数据，计算长短时记忆网络模型各层的输出值。

本实施例中，采用按单元时间间隔在一预设时间段内获取的训练集中的目标医疗数据输入到长短时记忆网络模型中，分别计算各层的输出值，包括计算训练集中的目标医疗数据在输入门、遗忘门、输出门、神经元状态和神经元输出的输出。其中，一个神经元包括有三种激活函数  $f(\text{sigmoid})$ 、 $g(\text{tanh})$  和  $h(\text{softmax})$ 。激活函数能够将权值结果转化成分类结果，其作用是能够给神经网络加入一些非线性因素，使得神经网络可以更好地解决较为复杂的问题。

一个神经元所接收和处理的数据包括：输入的训练集中的目标医疗数据： $x$ ，状态数

据:  $s$ 。此外, 以下提及的参数还包括: 神经元的输入用  $a$  表示, 输出用  $b$  表示。下标  $i$ ,  $\phi$  和  $\omega$  分别表示输入门、遗忘门和输出门。下标  $c$  表示神经元,  $t$  代表时刻。神经元跟输入门、遗忘门和输出门连接的权值分别记做  $w_{cl}$ 、 $w_{c\phi}$  和  $w_{c\omega}$ 。  $S_c$  表示神经元状态。  $I$  表示输入层的神经元的个数,  $H$  是隐层神经元的个数,  $C$  是神经元状态的神经元个数, 这里取  $C=H$ 。

输入门接收当前时刻的样本  $X_t$ 、上一时刻的输出值  $b^{t-1}_h$  以及上一时刻神经元的状态数据  $S^{t-1}_c$ , 通过连接输入的训练集中的目标医疗数据与输入门的权值  $w_{il}$ 、连接上一时刻的输出值与输入门的权值  $w_{hl}$  和连接神经元与输入门的权值  $w_{cl}$ , 根据公式  $a^t_l = \sum_{i=1}^I w_{il} x^t_i + \sum_{h=1}^H w_{hl} b^{t-1}_h + \sum_{c=1}^C w_{cl} s^{t-1}_c$  计算得到输入门的输出  $a^t_l$ , 将激活函数  $f$  作用于  $a^t_l$ , 由公式  $b^t_l = f(a^t_l)$  得到一个 0-1 区间的标量。此标量控制了神经元根据当前状态和过去状态的综合判断所接收当前信息的比例。

遗忘门接收当前时刻的样本  $X_t$ 、上一时刻的输出值  $b^{t-1}_h$  以及上一时刻的状态数据  $S^{t-1}_c$ , 通过连接输入的训练集中的目标医疗数据与遗忘门的权值  $w_{i\phi}$ 、连接上一时刻的输出值与遗忘门的权值  $w_{h\phi}$  和连接神经元与遗忘门的权值  $w_{c\phi}$ , 根据公式  $a^t_\phi = \sum_{i=1}^I w_{i\phi} x^t_i + \sum_{h=1}^H w_{h\phi} b^{t-1}_h + \sum_{c=1}^C w_{c\phi} s^{t-1}_c$  计算得到遗忘门的输出  $a^t_\phi$ , 将激活函数  $f$  作用于  $a^t_\phi$  由公式  $b^t_\phi = f(a^t_\phi)$  得到一个 0-1 区间的标量, 此标量控制了神经元根据当前状态和过去状态的综合判断所接收过去信息的比例。

神经元接收当前时刻的样本  $X_t$ 、上一时刻的输出值  $b^{t-1}_h$  以及上一时刻的状态数据  $S^{t-1}_c$ 、连接神经元与输入的训练集中的目标医疗数据的权值  $w_{ic}$ 、连接神经元与上一时刻的输出值的权值  $w_{hc}$  以及输入门、遗忘门的输出标量, 根据式  $a^t_c = \sum_{i=1}^I w_{ic} x^t_i + \sum_{h=1}^H w_{hc} b^{t-1}_h$ ,  $s^t_c = b^t_\phi s^{t-1}_c + b^t_l g(a^t_c)$  计算当前时刻的神经元状态  $S^t_c$ 。

输出门接收当前时刻的样本以及当前时刻的状态数据  $X_t$ , 上一时刻的输出值  $b^{t-1}_h$  以及当前时刻的状态数据  $S^t_c$ , 通过连接输入的训练集中的目标医疗数据与输出门的权值  $w_{i\omega}$ 、

连接上一时刻的输出值与输出门的权值  $w_{hw}$  以及连接神经元与输出门的权值  $w_{cw}$ ，根据公式

$$a_w^t = \sum_{i=1}^I w_{iw} x_{it} + \sum_{h=1}^H w_{hw} b_h^{t-1} + \sum_{c=1}^C w_{cw} s_c^t \text{ 计算输出门的输出 } a_w^t, \text{ 将激活函数 } f \text{ 作用于 } a_w^t \text{ 上由公式}$$

$b_w^t = f(a_w^t)$  得到一个 0-1 区间的标量。

神经元输出  $b_c^t$  根据输出门输出的标量计算。具体地，神经元输出的输出根据公式  $b_c^t = b_w^t h(s_c^t)$  计算得出。由上述对训练集中的目标医疗数据在各层间的计算可获取长短时记忆网络模型各层的输出值。

S133: 根据输出值对长短时记忆网络模型各层进行误差反传更新，获取更新后的各层的权值，其中，误差反传更新的表达式为  $\frac{\partial E}{\partial W} = \sum_{t=1}^T \frac{\partial E}{\partial S_c^t} \frac{\partial S_c^t}{\partial B} \frac{\partial B}{\partial W} = \sum_{t=1}^T \delta S_c^{t-1} B(1-B)(b^{t-1}_h)^T$ ，其中， $T$  表示时刻， $W$  表示权值， $B$  表示输出值， $\delta$  表示误差项， $S_c^{t-1}$  为上一时刻神经元的状态数据， $b^{t-1}_h$  为上一时刻的输出值。

本实施例中，根据获取长短时记忆网络模型各层的输出值对长短时记忆网络模型各层进行误差反传更新。具体地，首先根据误差项的表达式  $\varepsilon_c^t = \frac{\partial E}{\partial b_c^t}$ ， $\varepsilon_s^t = \frac{\partial E}{\partial s_c^t}$ ，可求出各层的误差项。其中， $\varepsilon$  和  $\delta$  均表示误差项，特别地， $\varepsilon_c^t$  表示神经元输出反传的误差项， $\delta_c^t$  表示神经元状态反传的误差项，两者均表示误差项，但具体含义不同。在以下表达式中，神经元的输入用  $a$  表示，输出用  $b$  表示。下标  $\iota$ ， $\phi$  和  $\omega$  分别表示输入门、遗忘门和输出门。下标  $c$  表示神经元， $t$  代表时刻。神经元跟输入门、遗忘门和输出门连接的权值分别记做  $w_{c\iota}$ 、 $w_{c\phi}$  和  $w_{c\omega}$ 。  $S_c$  表示神经元状态，控制门的激活函数用  $f$  (sigmoid) 表示， $g$  (tanh) 和  $h$  (softmax) 分别表示神经元的输入激活函数和输出激活函数。  $K$  是输出层神经元的个数， $H$  是隐层神经元的个数， $C$  是神经元状态的神经元个数，这里取  $C=H$ 。则输入门反传的误差

项为  $\delta_\iota^t = f'(a_\iota^t) \sum_{c=1}^C g(a_c^t) \varepsilon_s^t$ ，遗忘门反传的误差项为  $\delta_\phi^t = f'(a_\phi^t) \sum_{c=1}^C s_c^{t-1} \varepsilon_s^t$ ，神经元状态反传的

误差项为  $\delta_c^t = b_\iota^t g'(a_c^t) \varepsilon_s^t$ ，其中， $\varepsilon_s^t = b_w^t h'(s_c^t) \varepsilon_c^t + b_\phi^{t+1} \varepsilon_s^{t+1} + w_{c\iota} \delta_\iota^{t+1} + w_{c\phi} \delta_\phi^{t+1} + w_{cw} \delta_w^t$ ，输出门

反传的误差项为  $\delta_w^t = f'(a_w^t) \sum_{c=1}^C h(s_c^t) \varepsilon_c^t$ ，神经元输出反传的误差项为

$\varepsilon_c^t = \sum_{k=1}^K w_{ck} \delta_k^t + \sum_{h=1}^H w_{ch} \delta_h^{t+1}$ 。根据获得的各层误差项，再进行权值梯度的计算即可更新各层的

权值，其中，权值更新的表达式为  $\frac{\partial E}{\partial W} = \sum_{t=1}^T \frac{\partial E}{\partial S_c^t} \frac{\partial S_c^t}{\partial B} \frac{\partial B}{\partial W} = \sum_{t=1}^T \delta S_c^{t-1} B(1-B)(b^{t-1}_h)^T$ ，式中  $T$  表示时刻， $W$  表示权值，如  $w_{cl}$ 、 $w_{c\phi}$  和  $w_{cw}$  等连接权值。 $B$  表示输出值，如  $b'_l$ 、 $b'_\phi$  和  $b'_w$  等输出。 $\delta$  表示误差项，如  $\delta'_l$ 、 $\delta'_\phi$  和  $\delta'_w$  等误差项。 $S_c^{t-1}$  为上一时刻神经元的状态数据， $b^{t-1}_h$  为上一时刻的输出值。上述表达式各参数需相对应，如更新的具体权值为  $w_{cl}$  时，则输出  $B$  为相对应的  $b'_l$ ，误差项  $\delta$  为相对应的  $\delta'_l$ 。根据步骤 S132 和步骤 S133 的表达式可获得该权值更新表达式的所需参数值。则根据该权值更新的表达式进行运算即可获取更新后各层的权值。

S134：基于更新后的各层的权值，获取原始医疗模型。

本实施例中，将获取的更新后的各层的权值，应用到长短时记忆网络模型中即可获取原始医疗模型。进一步地，该原始医疗模型中各层之间的权值实现了原始医疗模型决定丢弃哪些旧信息、增加哪些新信息以及输出哪些信息的功能。在原始医疗模型的输出层最终会输出概率值，该概率值表示信息在通过原始医疗模型处理后与该原始医疗模型的贴近程度，可广泛应用于医疗特征识别，以达到准确识别医疗特征的目的。

在一具体实施方式中，如图 5 所示，步骤 S14 中，即采用测试集中的目标医疗数据对原始医疗模型进行测试，获取目标医疗模型，具体包括如下步骤：

S141：将每一目标医疗数据中的至少一个症状特征输入到原始医疗模型进行测试，获取对应的预测医疗特征。

其中，每一目标医疗数据包含至少一个症状特征和对应的实际疾病特征，如流鼻涕和/或发烧（症状特征）对应感冒（实际疾病特征）。具体地，将测试集中的每一目标医疗数据输入到原始医疗模型进行测试，即执行步骤 S31-S34，以获取对应的预测医疗特征。

S142：获取同一目标医疗数据中实际疾病特征与预测医疗特征相匹配的识别数量。

具体地，将步骤 S41 获取到的预测疾病特征与同一目标医疗数据中的实际疾病特征相匹配，若匹配成功，则将识别结果准确的数量加 1，若匹配不成功，则识别结果准确的数量不变，最后统计识别结果准确的数量即为识别数量。

S143：基于识别数量和测试集中的目标医疗数据的数据总量，获取识别准确率。

具体地，将步骤 S42 获取到的识别数量即识别结果准确的数量除以测试集中所有目标

医疗数据的数据总量，以获取识别准确率。可以理解地，识别准确率可以采用公式  $P = \frac{T}{N}$  进行计算获取，其中，T 为识别结果准确的数量，N 为目标医疗数据的数据总量。

S144：若识别准确率大于预设准确率，则获取目标医疗模型。

其中，预设准确率是预先设置的用于评价模型好坏的概率值。具体地，若识别准确率大于预设准确率，则认定该原始医疗模型较准确，以将该原始医疗模型作为目标医疗模型；反之，若识别准确率不大于预设准确率，则认定该原始医疗模型不够准确，需采用更多的目标医疗数据重新训练原始医疗模型，在训练获取的原始医疗模型的识别准确率大于预设准确率时，认为模型训练完成，以获取目标医疗模型。

本实施例中，将每一目标医疗数据中的至少一个症状特征输入到原始医疗模型进行测试，获取对应的预测医疗特征，然后基于同一目标医疗数据中实际疾病特征与预测医疗特征相匹配的识别数量和测试集中的目标医疗数据的数据总量，以获取识别准确率。最后基于识别准确率，获取目标医疗模型，以提高医疗特征识别的准确率。

本实施例中，先获取目标医疗数据，并对目标医疗数据进行划分，获取训练集和测试集，以便基于训练集中的目标医疗数据对长短时记忆网络模型进行训练，能够有效更新长短时记忆网络模型中各层的权值，使得通过目标医疗数据训练得到的原始医疗特征识别效果更精准。最后，再基于测试集中的目标医疗数据对原始医疗模型进行测试，以保证获取到的目标医疗特征识别目标医疗特征的准确率更高。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不对本申请实施例的实施过程构成任何限定。

## 实施例 2

图 6 示出与实施例 1 中医疗模型训练方法一一对应的医疗模型训练装置的原理框图。如图 6 所示，该医疗模型训练装置包括目标医疗数据获取模块 11、目标医疗数据划分模块 12、原始医疗模型获取模块 13 和目标医疗模型获取模块 14。其中，目标医疗数据获取模块 11、目标医疗数据划分模块 12、原始医疗模型获取模块 13 和目标医疗模型获取模块 14 的实现功能与实施例中医疗模型训练方法对应的步骤一一对应，为避免赘述，本实施例不一一详述。

目标医疗数据获取模块 11，用于获取目标医疗数据。

目标医疗数据划分模块 12，用于将目标医疗数据按预设比例进行划分，获取训练集和测试集。

原始医疗模型获取模块 13，用于采用训练集中的目标医疗数据对长短时记忆网络模

型进行训练，获取原始医疗模型。

目标医疗模型获取模块 14，用于采用测试集中的目标医疗数据对原始医疗模型进行测试，获取目标医疗模型。

优选地，目标医疗数据获取模块 11 包括目标网页地址获取单元 111、原始医疗数据获取单元 112 和目标医疗数据获取单元 113。

目标网页地址获取单元 111，用于获取目标网页地址。

原始医疗数据获取单元 112，用于采用爬虫工具爬取目标网页地址对应的网页，获取原始医疗数据。

目标医疗数据获取单元 113，用于对原始医疗数据进行数据清洗，获取目标医疗数据。

优选地，原始医疗数据获取单元 112 包括访问地址获取子单元 1121、访问地址存储子单元 1122 和原始医疗数据获取子单元 1123。

访问地址获取子单元 1121，用于采用爬虫工具，依据深度优先算法或广度优先算法爬取目标网页地址所链接的至少一个访问地址，每一访问地址对应一网页。

访问地址存储子单元 1122，将至少一个访问地址存储在待下载消息队列中。

原始医疗数据获取子单元 1123，采用爬虫工具对待下载消息队列中的每一访问地址对应的网页进行数据提取，获取原始医疗数据。

优选地，原始医疗模型获取模块 13 包括网络模型初始化单元 131、输出值计算单元 132、权值更新单元 133 和原始医疗模型获取单元 134。

网络模型初始化单元 131，用于初始化长短时记忆网络模型。

输出值计算单元 132，用于在长短时记忆网络模型中输入正负样本，计算长短时记忆网络模型各层的输出值。

权值更新单元 133，用于根据输出值对长短时记忆网络模型各层进行误差反传更新，获取更新后的各层的权值。

原始医疗模型获取单元 134，用于基于更新后的各层的权值，获取原始医疗模型。

优选地，目标医疗模型获取模块 14 包括预测医疗特征获取单元 141、识别数量获取单元 142、识别准确率获取单元 143 和目标医疗模型获取单元 144。

预测医疗特征获取单元 141，用于将每一目标医疗数据中的至少一个症状特征输入到原始医疗模型进行测试，获取对应的预测医疗特征。

识别数量获取单元 142，用于获取同一目标医疗数据中实际疾病特征与预测医疗特征相匹配的识别数量。

识别准确率获取单元 143，用于基于识别数量和测试集中的目标医疗数据的数据总量，获取识别准确率。

目标医疗模型获取单元 144，用于识别准确率大于预设准确率时，则获取目标医疗模型。

### 实施例 3

图 7 示出本实施例中医疗识别方法的一流程图。该医疗识别方法可应用在医疗机构或者其他机构的终端设备上，以便对症状特征进行识别，达到智能识别疾病的效果。如图 7 所示，该医疗识别方法包括如下步骤：

S21：获取用户的待测医疗数据，待测医疗数据包括至少一个症状特征。

其中，待测医疗数据是指用户选择的用于识别目标医疗特征概率的医疗数据。本实施例中，待测医疗数据包括至少一个症状特征，如流鼻涕、体温和血压等症状特征。

S22：将至少一个症状特征输入到目标医疗模型进行识别，获取目标医疗特征，目标医疗模型是采用实施例 1 中医疗模型训练方法获取的模型。

其中，目标医疗特征是采用目标医疗模型对症状特征进行识别所得到的最大概率的患病特征。具体地，将至少一个症状特征输入到目标医疗模型进行识别，即执行实施例 1 中步骤 S131-S134，以获取目标医疗特征。

本实施例中，先获取用户的待测医疗数据，待测医疗数据包括至少一个症状特征，以便至少一个症状特征输入到目标医疗模型进行识别，获取目标医疗特征，有利保障获取识别概率值的准确性，以辅助用户判断患病风险。

在一具体实施方式中，如图 8 所示，步骤 S22，即将至少一个症状特征输入到目标医疗模型进行识别，获取目标医疗特征，具体包括如下步骤：

S221：基于目标医疗模型对至少一个症状特征进行识别，获取至少一个识别医疗特征和对应的识别概率值。

具体地，将至少一个症状特征输入到目标医疗模型中进行识别，在目标医疗模型中对输入的至少一个症状特征进行基于各层间权值的转换处理，在输出层输出至少一个识别医疗特征和对应识别概率值。该识别医疗特征是采用目标医疗模型对至少一个症状特征进行识别时，识别出其可能患有的疾病特征。识别概率值是采用目标医疗模型对至少一个症状特征进行识别时，识别其可能为识别医疗特征的可能性。其中，该识别概率值可以为 0-1 之间的实数。

S222：选取最大的识别概率值对应的识别医疗特征作为目标医疗特征。



具体地，选取最大的识别概率值对应的识别医疗特征作为目标医疗特征。其中，数据库还会关联存储有目标医疗特征所对应的科室和至少一个药品名称。本实施例中，当用户输入至少一个症状特征进行识别时，除了输出目标医疗特征，还会关联输出对应的科室和至少一个药品名称，以增强医疗识别模型的实用性。

本实施例中，基于目标医疗模型对至少一个症状特征进行识别，获取至少一个识别医疗特征和对应的识别概率值，以便基于识别概率值确定目标医疗特征，即确定疑似患病的疾病特征，以保证医疗特征识别结果更精确可靠。

本实施例中，通过获取用户的待测医疗数据，以便基于目标医疗模型对待测医疗数据进行识别，以获取目标医疗特征。通过目标医疗模型对待测医疗数据进行识别，以保证识别的准确率更高，较精准的识别目标医疗特征，以辅助用户判断患病风险，以便及时预防。

#### 实施例 4

图 9 示出与实施例 1 中医疗识别方法一一对应的医疗识别装置的原理框图。如图 9 所示，该医疗识别装置包括待测医疗数据获取模块 21 和目标医疗特征获取模块 22。其中，待测医疗数据获取模块 21 和目标医疗特征获取模块 22 的实现功能与实施例中医疗识别方法对应的步骤一一对应，为避免赘述，本实施例不一一详述。

待测医疗数据获取模块 21，用于获取用户的待测医疗数据，待测医疗数据包括至少一个症状特征。

目标医疗特征获取模块 22，用于将至少一个症状特征输入到目标医疗模型进行识别，获取目标医疗特征。

优选地，目标医疗特征获取模块 22 包括识别概率值获取单元 221 和目标医疗特征获取单元 222。

识别概率值获取单元 221，用于基于目标医疗模型对至少一个症状特征进行识别，获取至少一个识别医疗特征和对应的识别概率值。

目标医疗特征获取单元 222，用于选取最大的识别概率值对应的识别医疗特征作为目标医疗特征。

#### 实施例 5

本实施例提供一计算机可读存储介质，该计算机可读存储介质上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实施例 1 中医疗模型训练方法，为避免重复，这里不再赘述。或者，该计算机可读指令被处理器执行时实现实施例 2 中医疗模型训练装置中各模块/单元的功能，为避免重复，这里不再赘述。或者，该计算机可读指令被处理

器执行时实现实施例 3 中医疗识别方法中各步骤的功能，为避免重复，此处不一一赘述。或者，该计算机可读指令被处理器执行时实现实施例 4 中医疗识别装置中各模块/单元的功能，为避免重复，此处不一一赘述。

该计算机可读存储介质可以包括：能够携带所述计算机可读指令代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、电载波信号、电信信号以及软件分发介质等。

### 实施例 6

图 10 是本申请一实施例提供的终端设备的示意图。如图 10 所示，该实施例的终端设备 100 包括：处理器 101、存储器 102 以及存储在存储器 102 中并可在处理器 101 上运行的计算机可读指令 103。处理器 101 执行计算机可读指令 103 时实现实施例 1 中的医疗模型训练方法各步骤的功能，例如图 1 所示的步骤 S11 至 S14。或者，处理器 101 执行计算机可读指令 103 时实现实施例 2 中的医疗模型训练装置各模块/单元的功能，例如图 6 所示模块 11 至 14 的功能。或者，处理器 101 执行计算机可读指令 103 时实现实施例 3 中医疗识别方法各步骤的功能，为避免重复，此处不一一赘述。或者，该计算机可读指令被处理器 101 执行计算机可读指令 103 时实现实施例 4 中医疗识别装置中各模块/单元的功能。为避免重复，此处不一一赘述。

示例性的，计算机可读指令 103 可以被分割成一个或多个模块/单元，一个或者多个模块/单元被存储在存储器 102 中，并由处理器 101 执行，以完成本申请。一个或多个模块/单元可以是能够完成特定功能的一系列计算机可读指令 103 的指令段，该指令段用于描述计算机可读指令 103 在终端设备 100 中的执行过程。例如，计算机可读指令 103 可以被分割成实施例 2 中的目标医疗数据获取模块 11、目标医疗数据划分模块 12、原始医疗模型获取模块 13 和目标医疗模型获取模块 14，或者实施例 4 中的待测医疗数据获取模块 21 和目标医疗特征获取模块 22，各模块的具体功能如实施例 2 或实施例 4 所述，在此不一一赘述。

另外，在本申请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。

以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施

例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

## 权 利 要 求 书

1. 一种医疗模型训练方法，其特征在于，包括：

获取目标医疗数据；

将所述目标医疗数据按预设比例进行划分，获取训练集和测试集；

采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模型；

采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试，获取目标医疗模型。

2. 如权利要求 1 所述的医疗模型训练方法，其特征在于，所述获取目标医疗数据包括：

获取目标网页地址；

采用爬虫工具爬取所述目标网页地址对应的网页，获取原始医疗数据；

对所述原始医疗数据进行数据清洗，获取目标医疗数据。

3. 如权利要求 1 所述的医疗模型训练方法，其特征在于，所述采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模型，包括：

初始化长短时记忆网络模型；

在所述长短时记忆网络模型中输入所述训练集中的目标医疗数据，计算所述长短时记忆网络模型各层的输出值；

根据所述输出值对所述长短时记忆网络模型各层进行误差反传更新，获取更新后的所述各层的权值，其中，所述误差反传更新的表达式为

$$\frac{\partial E}{\partial W} = \sum_{t=1}^T \frac{\partial E}{\partial S_c^t} \frac{\partial S_c^t}{\partial B} \frac{\partial B}{\partial W} = \sum_{t=1}^T \delta S_c^{t-1} B(1-B)(b^{t-1}_h)^T, \text{ 其中, } T \text{ 表示时刻, } W \text{ 表示权值, } B \text{ 表示输出值, } \delta \text{ 表示误差项, } S_c^{t-1} \text{ 为上一时刻神经元的状态数据, } b^{t-1}_h \text{ 为上一时刻的输出值;}$$

基于更新后的所述各层的权值，获取原始医疗模型。

4. 如权利要求 1 所述的医疗模型训练方法，其特征在于，所述目标医疗数据包括一个实际疾病特征和对应的至少一个症状特征；

所述采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试，获取目标医疗模型，包括：

将每一所述目标医疗数据中的至少一个所述症状特征输入到所述原始医疗模型进行测试，获取对应的预测医疗特征；

获取同一所述目标医疗数据中所述实际疾病特征与所述预测医疗特征相匹配的识别数量；

基于所述识别数量和所述测试集中所述的目标医疗数据的数据总量，获取识别准确率；

若所述识别准确率大于预设准确率，则获取所述目标医疗模型。

5. 一种医疗识别方法，其特征在于，包括：

获取用户的待测医疗数据，所述待测医疗数据包括至少一个症状特征；

将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，所述目标医疗模型是采用权利要求 1-4 任一项所述医疗模型训练方法获取的模型。

6. 如权利要求 5 所述的医疗识别方法，其特征在于，将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，包括：

基于所述目标医疗模型对至少一个所述症状特征进行识别，获取至少一个识别医疗特征和对应的识别概率值；

选取最大的所述识别概率值对应的识别医疗特征作为所述目标医疗特征。

7. 一种医疗模型训练装置，其特征在于，包括：

目标医疗数据获取模块，用于获取目标医疗数据；

目标医疗数据划分模块，用于将所述目标医疗数据按预设比例进行划分，获取训练集和测试集；

原始医疗模型获取模块，用于采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模型；

目标医疗模型获取模块，用于采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试，获取目标医疗模型。

8. 一种医疗识别装置，其特征在于，包括：

待测医疗数据获取模块，用于获取用户的待测医疗数据，所述待测医疗数据包括至少一个症状特征；

目标医疗特征获取模块，用于将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，所述目标医疗模型是采用权利要求 1-4 任一项所述医疗模型训练方法获取的模型。

9. 一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现如下

步骤:

获取目标医疗数据;

将所述目标医疗数据按预设比例进行划分, 获取训练集和测试集;

采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练, 获取原始医疗模型;

采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试, 获取目标医疗模型。

10. 如权利要求 9 所述的终端设备, 其特征在于, 所述获取目标医疗数据包括:

获取目标网页地址;

采用爬虫工具爬取所述目标网页地址对应的网页, 获取原始医疗数据;

对所述原始医疗数据进行数据清洗, 获取目标医疗数据。

11. 如权利要求 9 所述的终端设备, 其特征在于, 所述采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练, 获取原始医疗模型, 包括:

初始化长短时记忆网络模型;

在所述长短时记忆网络模型中输入所述训练集中的目标医疗数据, 计算所述长短时记忆网络模型各层的输出值;

根据所述输出值对所述长短时记忆网络模型各层进行误差反传更新, 获取更新后的所述各层的权值, 其中, 所述误差反传更新的表达式为

$$\frac{\partial E}{\partial W} = \sum_{t=1}^T \frac{\partial E}{\partial S_c^t} \frac{\partial S_c^t}{\partial B} \frac{\partial B}{\partial W} = \sum_{t=1}^T \delta S_c^{t-1} B(1-B)(b^{t-1}_h)^T, \text{ 其中, } T \text{ 表示时刻, } W \text{ 表示权值, } B \text{ 表示输出值, } \delta \text{ 表示误差项, } S_c^{t-1} \text{ 为上一时刻神经元的状态数据, } b^{t-1}_h \text{ 为上一时刻的输出值;}$$

基于更新后的所述各层的权值, 获取原始医疗模型。

12. 如权利要求 9 所述的终端设备, 其特征在于, 所述目标医疗数据包括一个实际疾病特征和对应的至少一个症状特征;

所述采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试, 获取目标医疗模型, 包括:

将每一所述目标医疗数据中的至少一个所述症状特征输入到所述原始医疗模型进行测试, 获取对应的预测医疗特征;

获取同一所述目标医疗数据中所述实际疾病特征与所述预测医疗特征相匹配的识别数量;

基于所述识别数量和所述测试集中所述的目标医疗数据的数据总量，获取识别准确率；

若所述识别准确率大于预设准确率，则获取所述目标医疗模型。

13. 一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现如下步骤：

获取用户的待测医疗数据，所述待测医疗数据包括至少一个症状特征；

将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，所述目标医疗模型是采用权利要求 1-4 任一项所述医疗模型训练方法获取的模型。

14. 如权利要求 13 所述的终端设备，其特征在于，将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，包括：

基于所述目标医疗模型对至少一个所述症状特征进行识别，获取至少一个识别医疗特征和对应的识别概率值；

选取最大的所述识别概率值对应的识别医疗特征作为所述目标医疗特征。

15. 一种计算机可读存储介质，所述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如下步骤：

获取目标医疗数据；

将所述目标医疗数据按预设比例进行划分，获取训练集和测试集；

采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模型；

采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试，获取目标医疗模型。

16. 如权利要求 15 所述的计算机可读存储介质，其特征在于，所述获取目标医疗数据包括：

获取目标网页地址；

采用爬虫工具爬取所述目标网页地址对应的网页，获取原始医疗数据；

对所述原始医疗数据进行数据清洗，获取目标医疗数据。

17. 如权利要求 15 所述的计算机可读存储介质，其特征在于，所述采用所述训练集中的目标医疗数据对长短时记忆网络模型进行训练，获取原始医疗模型，包括：

初始化长短时记忆网络模型；

在所述长短时记忆网络模型中输入所述训练集中的目标医疗数据，计算所述长短时记忆网络模型各层的输出值；

根据所述输出值对所述长短时记忆网络模型各层进行误差反传更新，获取更新后的所述各层的权值，其中，所述误差反传更新的表达式为

$$\frac{\partial E}{\partial W} = \sum_{t=1}^T \frac{\partial E}{\partial S_c^t} \frac{\partial S_c^t}{\partial B} \frac{\partial B}{\partial W} = \sum_{t=1}^T \delta S_c^{t-1} B(1-B)(b^{t-1}_h)^T, \text{ 其中, } T \text{ 表示时刻, } W \text{ 表示权值, } B \text{ 表示输出值, } \delta \text{ 表示误差项, } S_c^{t-1} \text{ 为上一时刻神经元的状态数据, } b^{t-1}_h \text{ 为上一时刻的输出值;}$$

基于更新后的所述各层的权值，获取原始医疗模型。

18. 如权利要求 15 所述的计算机可读存储介质，其特征在于，所述目标医疗数据包括一个实际疾病特征和对应的至少一个症状特征；

所述采用所述测试集中的目标医疗数据对所述原始医疗模型进行测试，获取目标医疗模型，包括：

将每一所述目标医疗数据中的至少一个所述症状特征输入到所述原始医疗模型进行测试，获取对应的预测医疗特征；

获取同一所述目标医疗数据中所述实际疾病特征与所述预测医疗特征相匹配的识别数量；

基于所述识别数量和所述测试集中所述的目标医疗数据的数据总量，获取识别准确率；

若所述识别准确率大于预设准确率，则获取所述目标医疗模型。

19. 一种计算机可读存储介质，所述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如下步骤：

获取用户的待测医疗数据，所述待测医疗数据包括至少一个症状特征；

将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，所述目标医疗模型是采用权利要求 1-4 任一项所述医疗模型训练方法获取的模型。

20. 如权利要求 19 所述的计算机可读存储介质，其特征在于，将至少一个所述症状特征输入到所述目标医疗模型进行识别，获取目标医疗特征，包括：

基于所述目标医疗模型对至少一个所述症状特征进行识别，获取至少一个识别医疗特征和对应的识别概率值；

选取最大的所述识别概率值对应的识别医疗特征作为所述目标医疗特征。



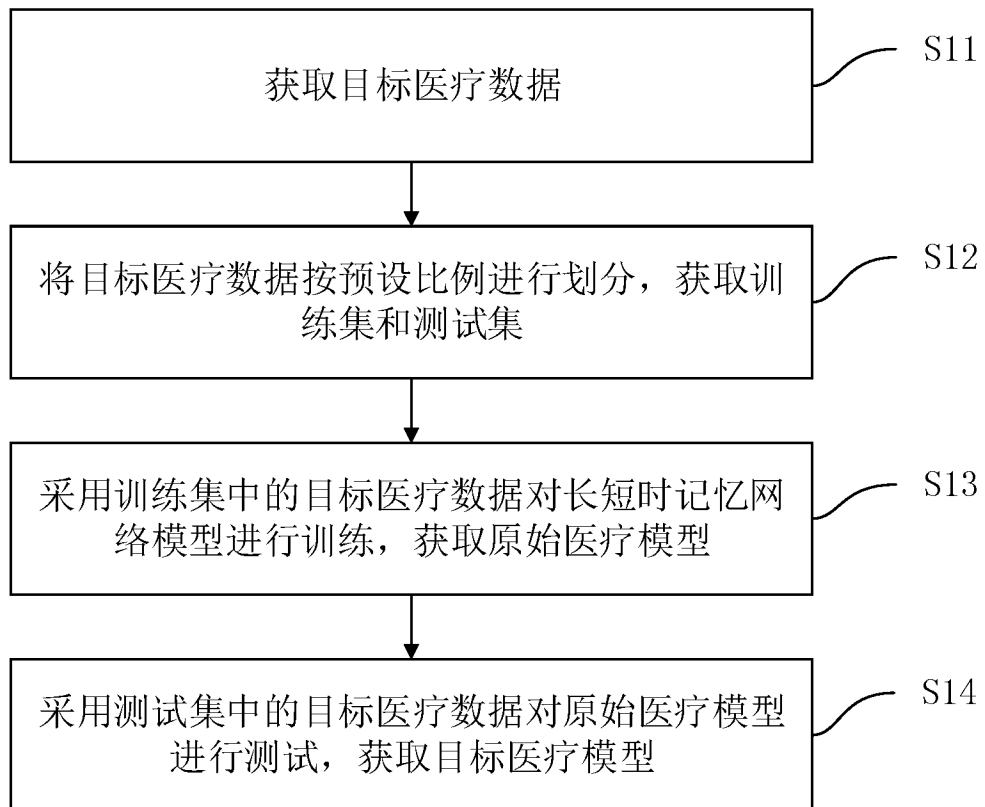


图 1

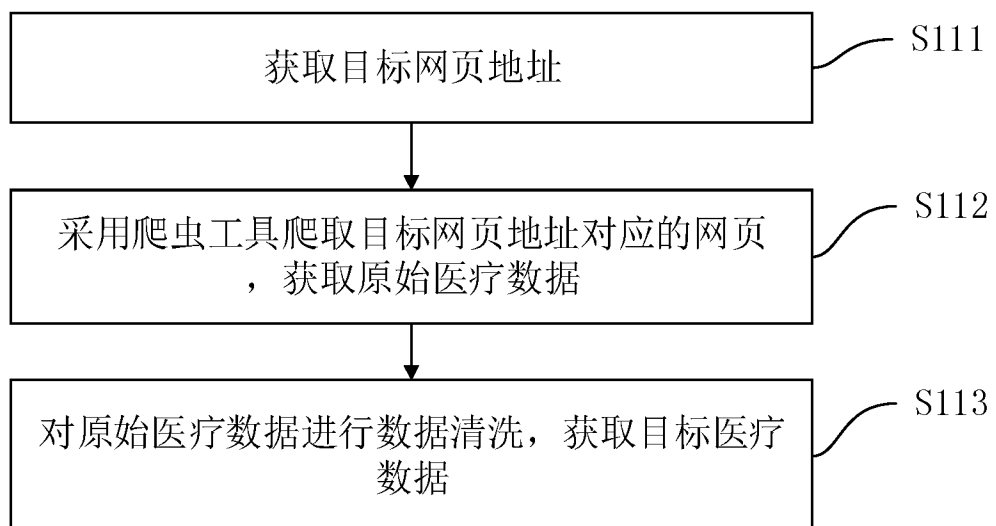


图 2

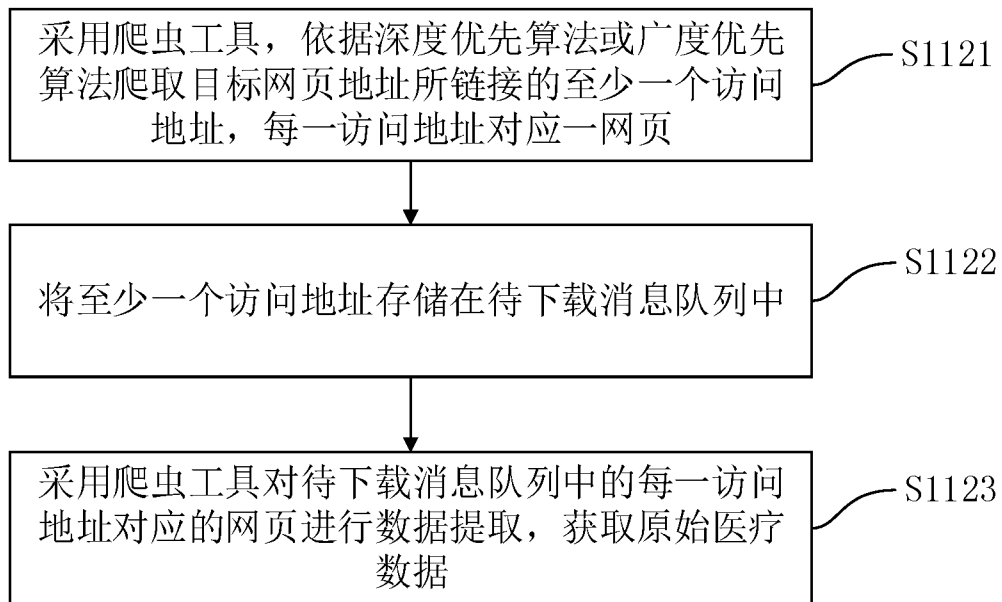


图 3

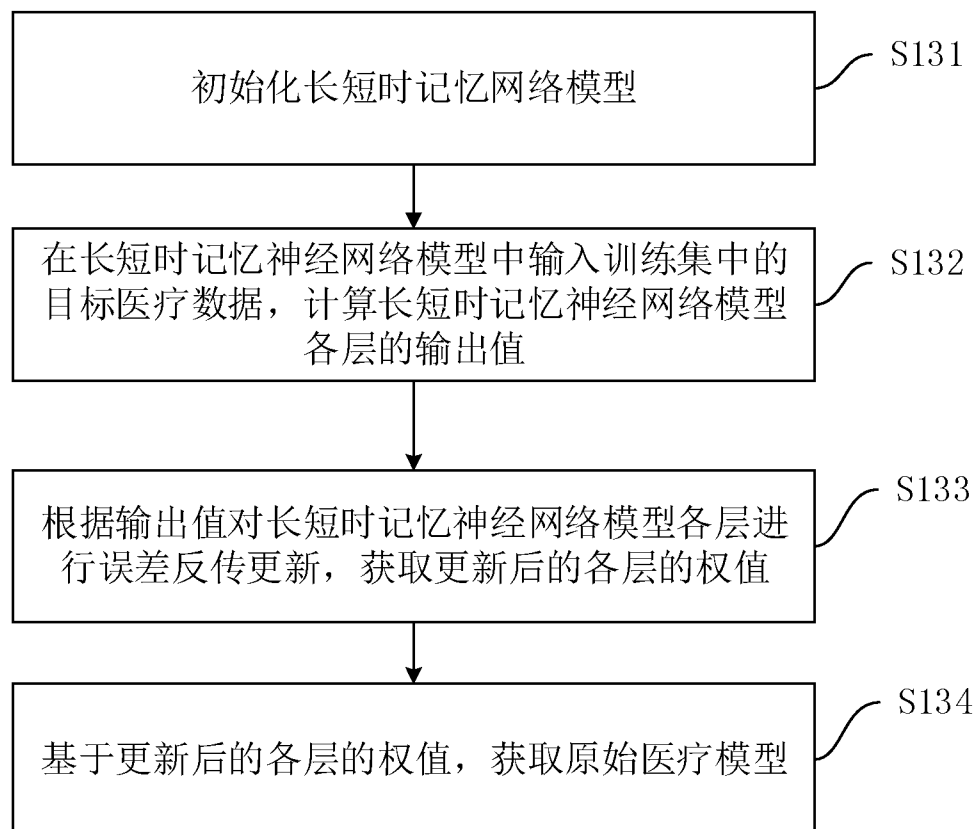


图 4

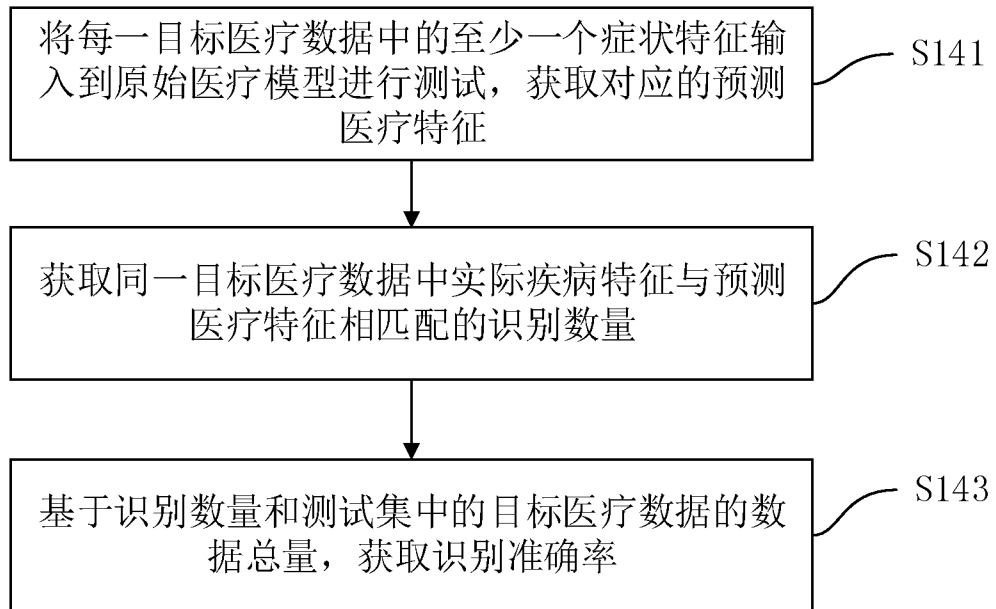


图 5

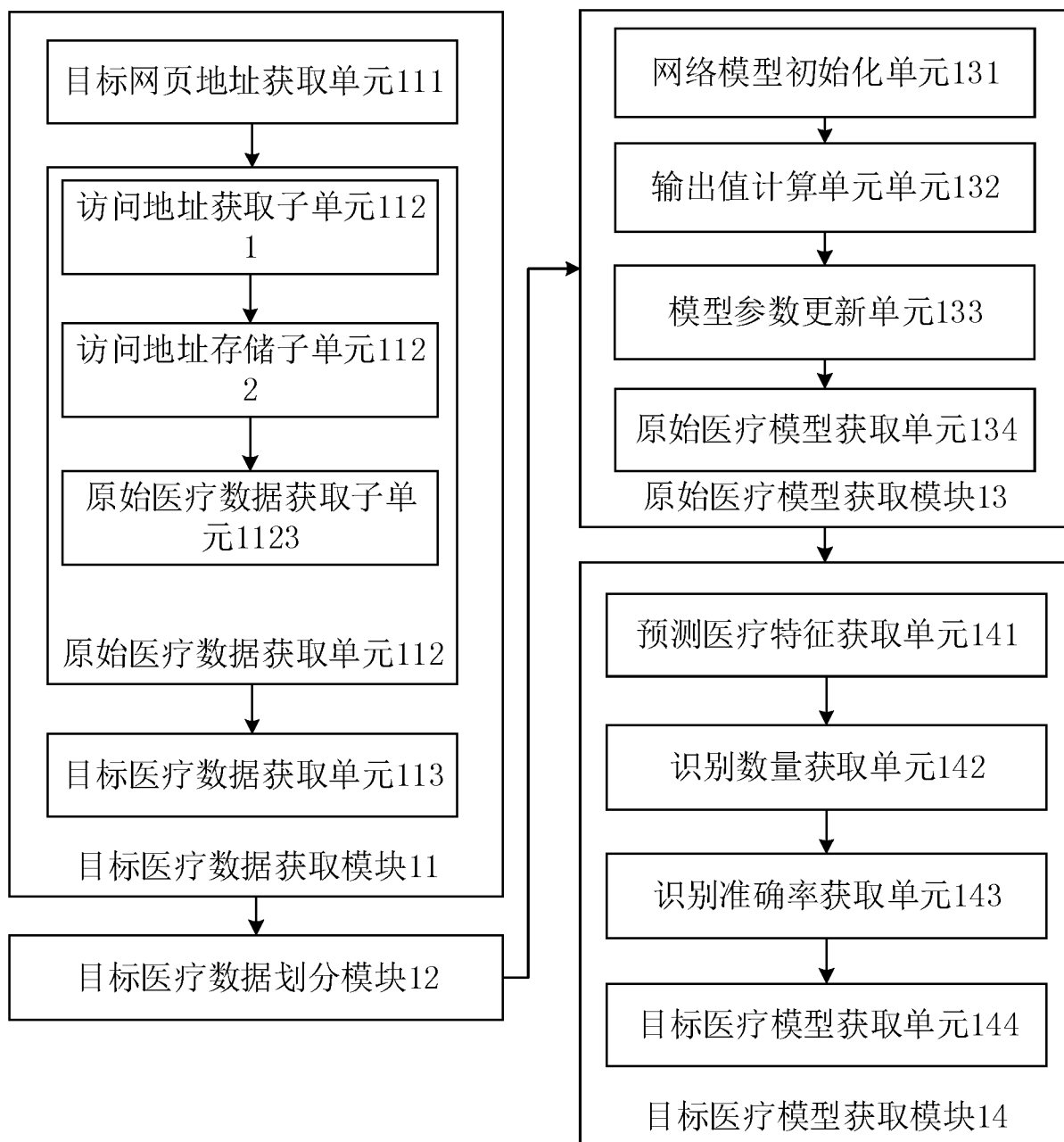


图 6

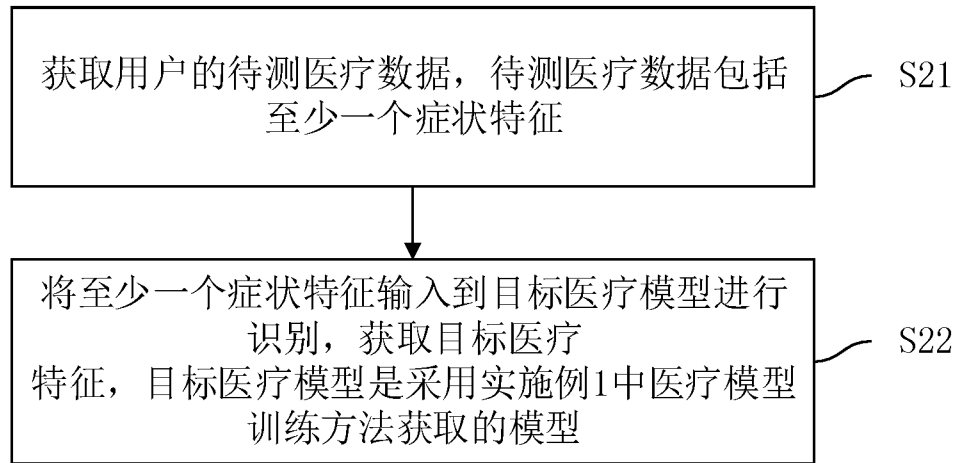


图 7

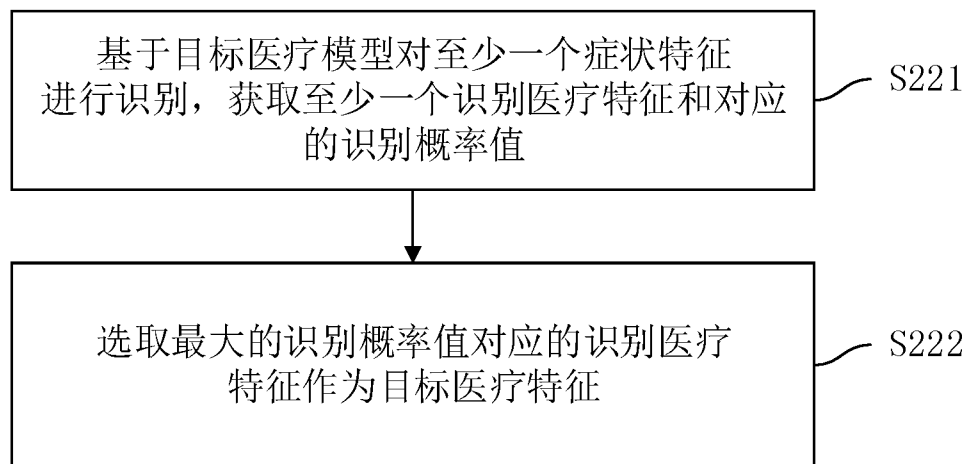


图 8

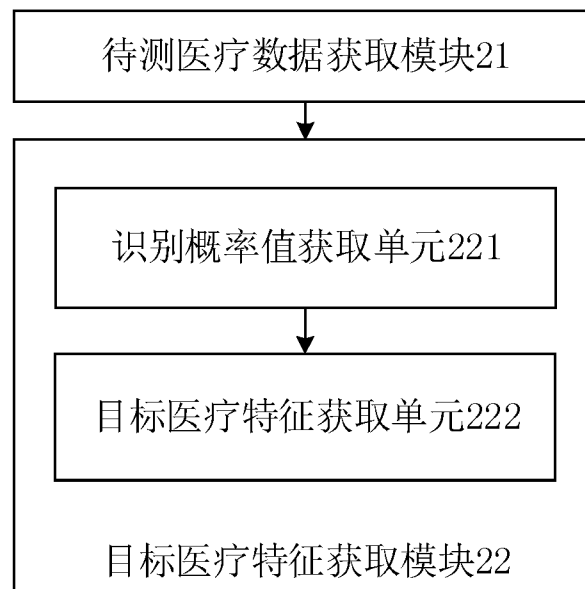


图 9

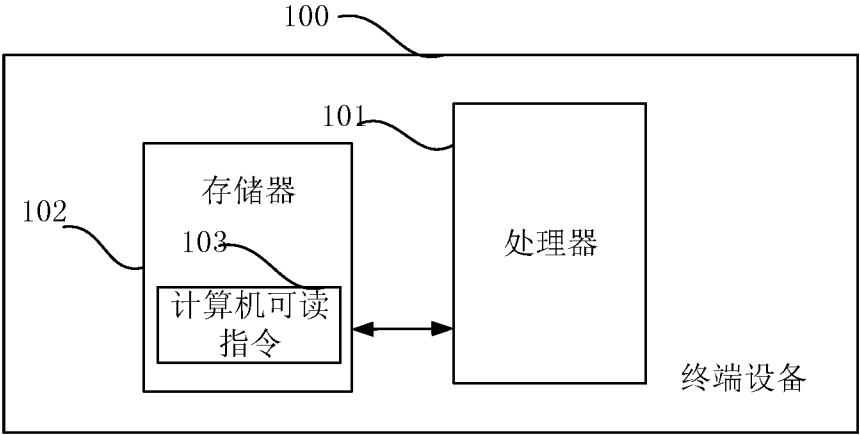


图 10

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/077475

**A. CLASSIFICATION OF SUBJECT MATTER**

G16H 50/70(2018.01)i; G06F 17/30(2006.01)i; G06K 9/62(2006.01)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)

G06, G16

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)

CPRSABS; CNTXT; VEN: 诊断, 模型, 长短时, 训练, 学习, diagnosis, short w term, model, train+, learn+

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                  | Relevant to claim No. |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | CN 106250707 A (WANG, SHUANGKUN ET AL.) 21 December 2016 (2016-12-21)<br>description, paragraphs [0025]-[0048], and figures 1 and 2 | 1-4, 7-20             |
| A         | CN 106980899 A (BEIJING CURACLOUD TECHNOLOGY CO., LTD.) 25 July 2017<br>(2017-07-25)<br>entire document                             | 1-4, 7-20             |
| A         | CN 105786830 A (ALIBABA GROUP HOLDING LIMITED) 20 July 2016 (2016-07-20)<br>entire document                                         | 1-4, 7-20             |
| A         | CN 106407387 A (SOUTH CHINA NORMAL UNIVERSITY) 15 February 2017<br>(2017-02-15)<br>entire document                                  | 1-4, 7-20             |
| A         | CN 106980811 A (SENSETIME GROUP LTD.) 25 July 2017 (2017-07-25)<br>entire document                                                  | 1-4, 7-20             |
| A         | US 2004181512 A1 (LOCKHEED MARTIN CORP) 16 September 2004 (2004-09-16)<br>entire document                                           | 1-4, 7-20             |
| A         | WO 2004088464 A2 (DUN & BRADSTREET INC) 14 October 2004 (2004-10-14)<br>entire document                                             | 1-4, 7-20             |

☐ 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

“&amp;” document member of the same patent family

Date of the actual completion of the international search

29 September 2018

Date of mailing of the international search report

18 October 2018

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China (ISA/  
CN)  
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing  
100088  
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/CN2018/077475**

## Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: **5,6**  
because they relate to subject matter not required to be searched by this Authority, namely:  
[1] A diagnostic method implemented on a human body or an animal body (PCT Rule 39.1(iv)).
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).



**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2018/077475**

| Patent document<br>cited in search report |            |    | Publication date<br>(day/month/year) | Patent family member(s) |            |    | Publication date<br>(day/month/year) |
|-------------------------------------------|------------|----|--------------------------------------|-------------------------|------------|----|--------------------------------------|
| CN                                        | 106250707  | A  | 21 December 2016                     | None                    |            |    |                                      |
| CN                                        | 106980899  | A  | 25 July 2017                         | CN                      | 107977709  | A  | 01 May 2018                          |
| CN                                        | 105786830  | A  | 20 July 2016                         | None                    |            |    |                                      |
| CN                                        | 106407387  | A  | 15 February 2017                     | None                    |            |    |                                      |
| CN                                        | 106980811  | A  | 25 July 2017                         | None                    |            |    |                                      |
| US                                        | 2004181512 | A1 | 16 September 2004                    | None                    |            |    |                                      |
| WO                                        | 2004088464 | A2 | 14 October 2004                      | US                      | 2004225543 | A1 | 11 November 2004                     |
|                                           |            |    |                                      | US                      | 9268803    | B2 | 23 February 2016                     |
|                                           |            |    |                                      | AU                      | 2004225099 | A1 | 14 October 2004                      |
|                                           |            |    |                                      | EP                      | 1609044    | A2 | 28 December 2005                     |
|                                           |            |    |                                      | US                      | 2011055252 | A1 | 03 March 2011                        |
|                                           |            |    |                                      | CA                      | 2555220    | A1 | 14 October 2004                      |
|                                           |            |    |                                      | US                      | 7707164    | B2 | 27 April 2010                        |
|                                           |            |    |                                      | AU                      | 2004225099 | B2 | 29 October 2009                      |

## 国际检索报告

国际申请号

PCT/CN2018/077475

## A. 主题的分类

G16H 50/70(2018.01)i; G06F 17/30(2006.01)i; G06K 9/62(2006.01)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)

G06, G16

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))

CPRSABS;CNTXT;VEN:诊断, 模型, 长短时, 训练, 学习, diagnosis, short w term, model, train+, learn+

## 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   |
|------|-----------------------------------------------------------------------------------|-----------|
| X    | CN 106250707 A (王双坤 等) 2016年 12月 21日 (2016 - 12 - 21)<br>说明书第[0025]-[0048]段, 图1-2 | 1-4, 7-20 |
| A    | CN 106980899 A (北京昆仑医云科技有限公司) 2017年 7月 25日 (2017 - 07 - 25)<br>全文                 | 1-4, 7-20 |
| A    | CN 105786830 A (阿里巴巴集团控股有限公司) 2016年 7月 20日 (2016 - 07 - 20)<br>全文                 | 1-4, 7-20 |
| A    | CN 106407387 A (华南师范大学) 2017年 2月 15日 (2017 - 02 - 15)<br>全文                       | 1-4, 7-20 |
| A    | CN 106980811 A (商汤集团有限公司) 2017年 7月 25日 (2017 - 07 - 25)<br>全文                     | 1-4, 7-20 |
| A    | US 2004181512 A1 (LOCKHEED MARTIN CORP) 2004年 9月 16日 (2004 - 09 - 16)<br>全文       | 1-4, 7-20 |
| A    | WO 2004088464 A2 (DUN & BRADSTREET INC) 2004年 10月 14日 (2004 - 10 - 14)<br>全文      | 1-4, 7-20 |

☐ 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

“&amp;” 同族专利的文件

国际检索实际完成的日期

2018年 9月 29日

国际检索报告邮寄日期

2018年 10月 18日

ISA/CN的名称和邮寄地址

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

受权官员

张清楠

传真号 (86-10)62019451

电话号码 (86-10)62085610

## 第II栏 某些权利要求被认为是不能检索的意见(续第1页第2项)

根据条约第17条(2)(a)，对某些权利要求未做国际检索报告的理由如下：

1. ☒ 权利要求： 5, 6  
因为它们涉及不要求本单位进行检索的主题，即：  
[1] 在人体或动物体上实施的诊断方法（PCT Rule 39.1(iv)）。
2. ☐ 权利要求：  
因为它们涉及国际申请中不符合规定的要求的部分，以致不能进行任何有意义的国际检索， 具体地说：
3. ☐ 权利要求：  
因为它们是从属权利要求，并且没有按照细则6.4(a)第2句和第3句的要求撰写。

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2018/077475

| 检索报告引用的专利文件 |            |    | 公布日<br>(年/月/日) | 同族专利 |            |    | 公布日<br>(年/月/日) |
|-------------|------------|----|----------------|------|------------|----|----------------|
| CN          | 106250707  | A  | 2016年 12月 21日  | 无    |            |    |                |
| CN          | 106980899  | A  | 2017年 7月 25日   | CN   | 107977709  | A  | 2018年 5月 1日    |
| CN          | 105786830  | A  | 2016年 7月 20日   | 无    |            |    |                |
| CN          | 106407387  | A  | 2017年 2月 15日   | 无    |            |    |                |
| CN          | 106980811  | A  | 2017年 7月 25日   | 无    |            |    |                |
| US          | 2004181512 | A1 | 2004年 9月 16日   | 无    |            |    |                |
| WO          | 2004088464 | A2 | 2004年 10月 14日  | US   | 2004225543 | A1 | 2004年 11月 11日  |
|             |            |    |                | US   | 9268803    | B2 | 2016年 2月 23日   |
|             |            |    |                | AU   | 2004225099 | A1 | 2004年 10月 14日  |
|             |            |    |                | EP   | 1609044    | A2 | 2005年 12月 28日  |
|             |            |    |                | US   | 2011055252 | A1 | 2011年 3月 3日    |
|             |            |    |                | CA   | 2555220    | A1 | 2004年 10月 14日  |
|             |            |    |                | US   | 7707164    | B2 | 2010年 4月 27日   |
|             |            |    |                | AU   | 2004225099 | B2 | 2009年 10月 29日  |

表 PCT/ISA/210 (同族专利附件) (2015年1月)