(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115398486 A
(43)申请公布日 2022.11.25

(21)申请号 202180025012.8

(22)申请日 2021.01.29

(30)优先权数据
20155469.8 2020.02.04 EP

(85)PCT国际申请进入国家阶段日
2022.09.27

(86)PCT国际申请的申请数据
PCT/EP2021/052170 2021.01.29

(87)PCT国际申请的公布数据
WO2021/156152 EN 2021.08.12

(71)申请人 豪夫迈·罗氏有限公司
地址 瑞士巴塞尔
申请人 基因泰克公司

(72)发明人 F·阿尔卡杜
B·古铁雷斯-贝克尔 A·塔哈默
M·普鲁诺托 Y·S·吴

(74)专利代理机构 北京坤瑞律师事务所 11494
专利代理师 封新琴

(51)Int.Cl.
G06V 10/764(2022.01)
G06V 10/82(2022.01)

权利要求书3页 说明书29页 附图6页

(54)发明名称

内镜下疾病的自动评估

(57)摘要

本申请涉及用于分析结肠镜检查视频或其部分、以及用于通过分析从受试者体内获取的结肠镜检查视频来评估所述受试者体内的溃疡性结肠炎严重程度的设备和方法。分析结肠镜检查视频包括使用第一深度神经网络分类器将来自所述受试者的结肠镜检查视频或其部分的图像数据至少分类为第一严重程度类别(严重程度较高的内镜下病变)和第二严重程度类别(严重程度较低的内镜下病变),其中所述第一深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练,所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧,其中一组中的帧具有相同的严重程度类别标签。还描述了用于提供用于分析结肠镜检查视频的工具的设备和方法。

CN 115398486 A

1.一种分析结肠镜检查视频或其部分的方法,所述方法包括使用第一深度神经网络分类器将来自所述结肠镜检查视频或其部分的图像数据至少分类为第一严重程度类别和第二严重程度类别,所述第一严重程度类别所关联的内镜下病变比所述第二严重程度类别的更严重,其中所述第一深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练,其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧,其中一组中的帧具有相同的严重程度类别标签。

2.根据权利要求1所述的方法,其中所述方法进一步包括使用第二深度神经网络分类器至少以第一质量类别和第二质量类别对来自所述结肠镜检查视频或其部分的图像数据进行分类,其中所述第一质量类别所关联的图像的质量比所述第二质量类别的更好,其中所述第一质量类别中的图像数据被提供给所述第一深度神经网络分类器,可选地,其中所述第二深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练,其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧,其中一组中的帧具有相同的质量类别标签,优选地,其中所述训练图像数据中的每组连续帧已经通过对包括相应组连续帧的视频片段的目视检查而被分配类别标签。

3.根据前述权利要求中任一项所述的方法,其中所述训练图像数据中的每组帧中的所述帧对应于结肠的单个解剖学区段。

4.根据前述权利要求中任一项所述的方法,其中对于所述训练图像数据中的每组帧,如果目视检查将包括所述组帧的训练结肠镜检查视频片段与第一范围内的内镜下严重程度得分相关联,则所述组帧被分配第一严重程度类别标签,可选地,其中对于所述训练图像数据中的每组帧,如果两次独立的目视检查将包括所述组帧的训练结肠镜检查视频片段与第一范围内的内镜下严重程度得分相关联,则所述组帧被分配第一严重程度类别标签。

5.根据权利要求4所述的方法,其中所述内镜下严重程度得分是梅奥诊所内镜下子评分(MCES),可选地,其中所述第一范围是MCES>1或MCES>2。

6.根据权利要求1至5中任一项所述的方法,其中所述第一深度神经网络分类器以三个或更多个严重程度类别对图像数据进行分类,其中对于所述训练图像数据中的每组帧,如果目视检查将包括所述组帧的训练结肠镜检查视频片段与针对所述三个或更多个严重程度类别中的各严重程度类别的预定不同范围内的内镜下严重程度得分相关联,则所述组帧被分配三个或更多个严重程度类别标签中的一个,可选地,其中对于所述训练图像数据中的每组帧,如果两次独立的目视检查将包括所述组帧的训练结肠镜检查视频片段与在与所述三个或更多个严重程度类别标签中的同一个严重程度类别标签相关联的范围内的内镜下严重程度得分相关联,则所述组帧被分配所述严重程度类别标签,优选地,其中所述内镜下严重程度得分是MCES,并且所述第一深度神经网络分类器将图像数据分类为四个严重程度类别,每个严重程度类别与不同的MCES相关联。

7.根据前述权利要求中任一项所述的方法,其中来自所述结肠镜检查视频或其部分的所述图像数据包括多个连续帧,并且其中分析所述结肠镜检查视频或其部分包括使用所述第一深度神经网络分类器和可选的所述第二深度神经网络分类器对来自所述结肠镜检查视频或其部分的所述图像数据中的所述多个帧各自进行分类。

8.根据权利要求7所述的方法,其中对各个帧进行分类包括为所述多个帧中的每一帧提供所述帧属于所述第一类别的概率和/或所述帧属于所述第二类别的概率。

9.根据权利要求7或8中任一项所述的方法,其中分析所述结肠镜检查视频或其部分进一步包括基于来自所述第一深度神经网络分类器的针对所述多个帧进行的各自分类为所述结肠镜检查视频或其部分分配总体严重程度类别,可选地,其中对各个帧进行分类包括为所述多个帧中的每一帧提供所述帧属于所述第一严重程度类别的概率,并且基于针对所述多个帧进行的各自分类为所述结肠镜检查视频或其部分分配总体严重程度类别包括:如果所述帧属于所述第一严重程度类别的概率的平均值高于阈值,则分配所述第一严重程度类别,或者如果分配给所述第一严重程度类别的帧的比例高于阈值,则分配所述第一严重程度类别。

10.根据前述权利要求中任一项所述的方法,其中所述第一、第二和/或第三深度神经网络分类器是卷积神经网络(CNN),任选地,其中所述CNN已经在不相关的图像数据上被预训练,和/或其中所述CNN是50层CNN,和/或其中所述CNN是已经使用深度残差学习框架预先训练的CNN。

11.根据前述权利要求中任一项所述的方法,其中所述内镜下病变指示溃疡性结肠炎,和/或其中所述第一严重程度类别所关联的溃疡性结肠炎比所述第二严重程度类别的更严重。

12.一种评估受试者体内的溃疡性结肠炎(UC)严重程度的方法,所述方法包括使用根据前述权利要求中任一项所述的方法分析来自所述受试者的结肠镜检查视频或其部分。

13.根据权利要求12所述的方法,其中分析来自所述受试者的结肠镜检查视频或其部分包括:

(i)使用第一深度神经网络分类器分析所述结肠镜检查视频或其部分,所述第一深度神经网络分类器在对应于MCES>1的第一严重程度类别和对应于MCES≤1的第二严重程度类别之间对图像数据进行分类;

(ii)使用第一深度神经网络分类器分析所述结肠镜检查视频或其部分,所述第一深度神经网络分类器在对应于MCES>2的第一严重程度类别和对应于MCES≤2的第二严重程度类别之间对图像数据进行分类;以及

(iii)选择所述受试者进行针对UC的治疗,如果在步骤(ii)中所述视频的至少一个片段被分配了第一严重程度类别,则选用第一治疗,并且如果在步骤(ii)中所述视频没有片段被分配第一严重程度类别,但在步骤(i)中所述视频的至少一个片段被分配了第一严重程度类别,则选用第二治疗;可选地,其中所述方法进一步包括治疗所述受试者。

14.一种提供用于分析结肠镜检查视频或其部分的工具的方法,所述方法包括:

获取包括来自多个训练结肠镜检查视频的多组连续帧的训练图像数据,其中一组中的帧具有相同的严重程度类别标签,所述严重程度类别标签至少选自第一严重程度类别和第二严重程度类别,其中所述第一严重程度类别所关联的内镜下病变比所述第二严重程度类别的更严重,以及

训练第一深度神经网络至少以第一严重程度类别和第二严重程度类别对图像数据进行分类,其中使用所述训练图像数据和所述严重程度类别标签以弱监督的方式执行所述训练,

其中所述方法可选地进一步包括使用第二深度神经网络分类器至少以第一质量类别和第二质量类别对包括多个帧的训练图像数据进行分类,其中所述第一质量类别所关联的

图像的质量比所述第二类别的更好，并且其中训练所述第一深度神经网络是使用被所述第二深度神经网络分类器分类在所述第一质量类别中的所述训练图像数据来执行的。

15.一种用于根据从受试者体内获取的结肠镜检查视频来评估所述受试者体内的溃疡性结肠炎严重程度的系统，所述系统包括：

至少一个处理器；以及

至少一个非暂时性计算机可读介质，所述至少一个非暂时性计算机可读介质包含指令，当由所述至少一个处理器执行时，所述指令使所述至少一个处理器执行包括以下各项的操作：

将来自所述结肠镜检查视频或其部分的图像数据至少分类为第一严重程度类别和第二严重程度类别，其中所述第一严重程度类别所关联的溃疡性结肠炎比所述第二严重程度类别的更严重，其中所述第一深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练，其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧，其中一组中的帧具有相同的严重程度类别标签，可选地，其中当由所述至少一个处理器执行时，所述指令进一步使所述处理器执行根据权利要求2至14中任一项所述的操作。

内镜下疾病的自动评估

技术领域

[0001] 本发明涉及用于分析结肠镜检查视频的计算机实现的方法,以及实现所述方法的计算设备。本发明的方法和设备可应用于炎症性肠病(如溃疡性结肠炎)的临床评估。因此,本发明还涉及评估患者炎症性肠病(特别是溃疡性结肠炎)的严重程度的方法。

背景技术

[0002] 内镜下评估内镜下病变的存在和严重程度是炎症性肠病如溃疡性结肠炎(UC)临床评估的一个既定部分。这种评估具有很高的可变性和偏差(关于述评,参见Panes等人,Journal of Crohn's and Colitis,2016,S542-S547)。中心内镜识读已经成为缓解这些问题的一种可能的方法,通过不与病人接触的受过专门训练的识读者对内镜数据进行独立评估。然而,这一过程甚至比"传统"临床评估更加人力资源密集,限制了其实际可行性。

[0003] 有人建议使用机器学习方法来自动化结肠镜检查视频的评估。特别地,Stidham等人(JAMA Network Open.2019年5月;2(5):e193963)报告了一项确定深度学习模型是否可以像有经验的人类审阅员一样对UC的内镜下严重程度进行分级的诊断研究。这项研究使用机器学习算法,这些算法使用专家为训练算法的特定目的选择并逐个评分(这为训练提供了高质量的"地面真值"数据)的静态图像进行了训练。虽然这种方法已经取得了一些成功,但它的实际适用性受到了胃肠病专家对静态图像的仔细手工选择的要求以及与此相关的潜在偏差的限制(这些偏差类似于中心内镜识读旨在减少的偏差)。

[0004] 因此,仍然需要改进的方法根据结肠镜检查视频自动评估内镜下病变的严重程度。

发明内容

[0005] 本发明的发明人开发了一种用于使用深度神经网络分类器分析结肠镜检查视频的新设备和新方法,特别是用于将严重程度类别与来自此类视频的图像数据相关联的设备和方法。所述方法和设备源于这样的发现,即可以通过使用已经使用原始结肠镜检查视频数据训练过的深度神经网络分类器分析原始结肠镜检查视频或其部分来获取对严重程度的有临床意义的评估,其中训练数据中的整个视频或其片段与相同的类别标签相关联。以前自动化内镜下评估的方法使用了机器学习算法,这些算法使用专家为训练算法的特定目的选择并分别评分的静态图像来训练,为训练提供"地面真值"数据。相比之下,本发明的发明人已经发现,可以使用深度神经网络分类器将结肠镜检查视频准确地分类到不同的严重程度类别,所述深度神经网络分类器使用原始的(即,不是在逐帧基础上选择的)专家注释的结肠镜检查视频数据(因为它通常作为例如临床试验的输出可用)在没有用于训练的"地面真值"数据的情况下以弱监督的方式训练。

[0006] 专家注释的结肠镜检查视频,如那些作为炎症性肠病临床评估的一部分制作的视频,即使是以更严格的形式,如中心内镜识读,也依赖于专家对代表结肠解剖学区段的每个视频或视频片段的全局严重程度得分的分配。因此,不是这样的视频中的所有帧都将实际

显示使专家分配得分的病变，并且视频的严重程度得分不会被期望准确地捕捉组成视频的每个单独帧的状态。此外，质量和信息内容可以在视频中高度可变。结果，这个数据是有噪声且不精确的。本发明的发明人惊讶地发现，可以使用已使用这种原始的(即，不是在逐帧基础上选择的)专家注释的结肠镜检查视频数据以弱监督的方式训练过的深度神经网络分类器将结肠镜检查视频精确地分类到不同的严重程度类别。

[0007] 本发明的第一方面因此提供了一种分析结肠镜检查视频或其部分的方法，所述方法包括使用第一深度神经网络分类器将来自所述结肠镜检查视频或其部分的图像数据至少分类为第一严重程度类别和第二严重程度类别，所述第一严重程度类别所关联的内镜下病变比所述第二严重程度类别的更严重，其中所述第一深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练，其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧，其中一组中的帧具有相同的严重程度类别标签。有利地，所述内镜下病变可指示溃疡性结肠炎。在优选实施方案中，所述第一严重程度类别所关联的溃疡性结肠炎比所述第二严重程度类别的更严重。

[0008] 在本发明的上下文中，以“弱监督方式”训练分类器意味着训练数据包括不是“地面真值”数据的数据。地面真值训练数据是指其中为每一条训练数据(例如，视频中的每一帧)分配了被认为真实反映该条训练数据所属的类别的训练类别标签的训练数据。相比之下，本发明使用包括具有不确定的训练类别标签的数据的训练数据。例如，在训练视频中形成片段的多个帧可以被分配相同的类别标签，因为该片段总体上满足与所分配的类别标签相关联的判断标准。然而，不能确定特定片段中的每一帧都显示了导致该片段被分配所述类别标签的特征。因此，对于片段中的每个帧是否已被分配给正确的类别存在一些不确定性，并且所得到的训练数据仅能够对训练过程进行弱监督。作为另一示例，训练视频中的帧可以自动地被分配类别标签，例如使用先前训练过的分类器或其他机器学习算法。在这样的实施方案中，由于不期望分类器能够以100％的准确度预测先前未见数据的类别，因此对于是否将每一帧分配给了正确的类别也存在一些不确定性，并且所得到的训练数据仅能够对训练过程进行弱监督。以其最简单的形式，弱监督仅仅指的是使用被分配了具有(未知)不确定性水平的训练类别标签训练数据。在实施方案中，训练类别分配中的不确定性水平可以被量化(例如，估计或假设)，并且不确定性可以在训练中被考虑。

[0009] 在本发明的上下文中，一组连续帧是指一起形成一个视频片段的一组帧。因此，“一组连续帧”和“视频片段”的措词可互换地使用。在实践中，视频片段可能不包含原始视频的相应片段的每一个单个帧。实际上，可以在内容不可知的基础上选择帧，以减少来自视频的数据量。例如，这可以例如通过使用每秒24帧视频的每隔一帧(即每秒12帧)来执行。然而，在本公开文本的上下文中，形成视频片段的所有帧将具有相同的标签，因为标签被分配给片段，而不是通过单独分析每一帧。优选地，使用形成已与标签相关联的训练视频片段的所有连续帧。当不是所有帧都被使用时，那么数据缩减(帧选择)优选地是自动的(或基于完全自动化的方案)，而不是基于手动管理。用于数据缩减的完全自动化方案可以包括在每两帧中选择一帧、在每n帧中选择一帧(其中n可以是例如2、3、4等)、在每n帧中选择两帧、每秒中随机选择n帧等。

[0010] 在本发明的上下文中，术语“严重程度”是指通过内镜下病变的存在来评估的炎症性肠病、特别是UC的严重程度。内镜下病变可能包括红斑、血管模式减少或缺乏、质脆、糜

烂、出血和溃疡中的一种或多种。

[0011] 在实施方案中，所述方法进一步包括使用第二深度神经网络分类器至少以第一质量类别和第二质量类别对来自结肠镜检查视频或其部分的图像数据进行分类，其中所述第一质量类别所关联的图像的质量比所述第二质量类别的更好，其中所述第一质量类别中的图像数据被提供给所述第一深度神经网络分类器。

[0012] 本发明的发明人已经发现，在训练期间和评估受试者视频期间，使用单独的深度神经网络分类器对图像数据进行基于质量的过滤，在将原始视频用于训练和评估的情况中，显著提高了基于严重程度的分类的准确度。

[0013] 在实施方案中，所述第二深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练，其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧，其中一组中的帧具有相同的质量类别标签。

[0014] 在实施方案中，所述第二深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练，其中所述训练图像数据包括来自已被自动分配质量类别标签的所述多个训练结肠镜检查视频的多个帧。在一些这样的实施方案中，已使用一个或多个先前训练过的机器学习算法(例如，一个或多个分类器，诸如本文所述的深度神经网络分类器)自动地将所述质量类别标签分配给帧。

[0015] 本发明的发明人惊奇地发现，可以使用已经用弱(不确定的)类别标签标注质量的图像数据来训练深度神经网络分类器，所述深度神经网络分类器执行基于质量的过滤，从而提高基于严重程度的评估的准确度。弱类别标签可以通过在逐片段(即每组帧)的基础上注释图像数据来获取。通过在逐片段的基础上分配质量类别标签来注释视频是一个相对容易的任务，可以通过例如众包来规模化。然而，并不是一个片段中的所有单个帧都具有导致为该片段分配所述质量类别标签的特征。因此，分配给一个组的标签表示组成该组的各个帧的弱标记。类似地，使用以前训练过的机器学习算法自动注释视频非常容易，成本也很低，但并不是所有帧都被分配了正确的类别标签。鉴于这种不确定性，分配给每一帧的标签表示弱标记。然而，本发明的发明人已经发现，使用这种弱标签来训练用于基于质量的过滤的深度神经网络分类器来过滤数据，足以使得基于过滤的数据可以进行准确的基于严重程度的分类。

[0016] 在实施方案中，用于训练第二深度神经网络的多个训练结肠镜检查视频和用于训练第一深度神经网络的多个训练结肠镜检查视频可以部分重叠。例如，在实施方案中，用于训练第一深度神经网络的训练图像数据可以是用于训练第二深度神经网络的训练图像数据的子集。有利地，用于训练第一深度神经网络的训练图像数据可以包括被第二深度神经网络分类在第一类中的帧。

[0017] 在实施方案中，训练图像数据中的每组连续帧已经通过对包括相应组连续帧的视频片段的目视检查而被分配类别标签。

[0018] 在实施方案中，对于训练图像数据中的每组连续帧，如果通过对由一组连续帧组成的训练结肠镜检查视频片段的目视检查可以区分结肠壁和结肠血管，则该组连续帧被分配第一质量类别标签，否则被分配第二质量类别标签。

[0019] 本发明的发明人惊奇地发现，通过将结肠镜检查视频分割成(i)可见结肠壁和血管的区段和(ii)不可见结肠壁和血管的区段而对图像数据质量进行的粗略评估足以为用

于基于质量的过滤的深度神经网络分类器的训练提供信息。有利地，这样的粗略评估可以相对较快地获取，并且可以众包。

[0020] 有利地，如果训练结肠镜检查视频片段另外满足基于水、超反射区域、粪便和/或模糊的存在或不存在的一个或多个判断标准，则训练图像数据中的每组连续帧(视频片段)可以被分配第一质量类别标签，否则可被分配第二质量类别标签。在实施方案中，一个或多个判断标准包括任何水、粪便或超反射区域是否一起覆盖帧上可见区域的至多20％、至多15％、至多10％或至多5％，优选至多10％。在实施方案中，一个或多个判断标准包括任何水、粪便或超反射区域是否各自覆盖帧上可见区域的至多20％、至多15％、至多10％或至多5％，优选至多10％。在实施方案中，一个或多个判断标准包括视频片段是否被人类评估确定为模糊。

[0021] 使用基于不存在伪影的更严格的判断标准以将训练数据分配给质量好类别可以有助于提高用于基于严重程度的分类器的训练的图像的质量。如本领域技术人员所理解的，当增加基于质量的过滤的严格性时，在过滤数据的质量和通过过滤器并可用于训练的数据量之间存在权衡。本发明的发明人发现，使用上述判断标准组合(可见的壁和血管、可接受水平的一个或多个伪影)取得了良好的平衡，并能够提供准确的基于严重程度的评估。

[0022] 在实施方案中，训练图像数据中的每组帧中的帧对应于结肠的单个解剖学区段。换句话说，每一组帧可以被定义为将其限制在单个解剖学区段。完整的结肠镜检查视频可以包括多个这样的片段，每个片段探查一个区段，例如直肠、乙状结肠或降结肠。

[0023] 使用按照解剖学区段分割的训练数据可能特别有利，因为它可以为训练提供粒度更高的数据。此外，与结肠镜检查视频中所示的结肠解剖学区段有关的信息通常可作为来自例如临床试验的注释结肠镜检查视频数据的一部分而获得。

[0024] 在实施方案中，对于所述训练图像数据中的每组帧，如果目视检查将包括所述组帧的训练结肠镜检查视频片段与第一范围内的内镜下严重程度得分相关联，则所述组帧被分配第一严重程度类别标签。可选地，对于所述训练图像数据中的每组帧，如果两次独立的目视检查将包括所述组帧的训练结肠镜检查视频片段与第一范围内的内镜下严重程度得分相关联，则所述组帧可以被分配第一严重程度类别标签。

[0025] 在实施方案中，对于所述训练图像数据中的每组帧，如果目视检查将包括所述组帧的训练结肠镜检查视频片段与高于阈值的内镜下严重程度得分相关联，则所述组帧被分配第一严重程度类别标签，并且对于所述训练图像数据中的每组帧，如果目视检查将包括所述组帧的训练结肠镜检查视频片段与等于或小于阈值的内镜下严重程度得分相关联，则所述组帧被分配第二严重程度类别标签。在这样的实施方案中，第一深度神经网络分类器可以是二元分类器。

[0026] 内镜下严重程度得分可以是用于根据预定量表来量化内镜下病变严重程度的任何得分。在实施方案中，内镜下得分可与特定疾病(例如溃疡性结肠炎)相关联，其中内镜下病变的严重程度与疾病严重程度的临床评估相关联。有利地，内镜下严重程度得分可以是梅奥诊所内镜下子评分(MCES)。在一些这样的实施方案中，第一范围可以是MCES>1或MCES>2。

[0027] 本发明的发明人已经发现，使用基于所述“原始”结肠镜检查视频的弱监督方法训练的深度神经网络分类器能够将来自结肠镜检查视频的图像数据准确地分类为属于专家

会在标准内镜下严重程度得分量表(如梅奥内镜下子评分)上对其给出高评分的类别,以及专家将在该量表上对其给出较低范围上评分的类别。

[0028] 二元分类器比更复杂的分类器(例如3个或更多的类别)更容易训练。换句话说,这种分类器可以预期用相对较低数量或质量的训练数据来实现高准确度。相比之下,训练更复杂的分类器以获得类似的性能通常需要更多和/或更高质量的数据。本发明的发明人已经发现,通过使用两个类别限制训练问题的复杂性,同时使用“原始”注释的结肠镜检查视频放松对训练数据的要求,可以获取准确预测结肠镜检查视频的临床相关性质的分类器。因此,本发明的方法改进了实际适用性,因为它使用通常可用和/或容易获得的数据,并产生可靠的临床相关预测。

[0029] 梅奥诊所内镜下子评分(MCES)是用于评估溃疡性结肠炎分期的标准化量表,其完全基于内镜探查。在Rutgeerts P.等人(N Engl J Med.2005;353(23):2462-2476)中有描述。它包括四个级别,第一级别与正常黏膜/非活动性疾病相关联,第二级别与轻度疾病活动(红斑、血管模式减少、轻度质脆)相关联,第三级别与中度疾病活动(明显红斑、缺乏血管模式、质脆、糜烂)相关联,第四级别与严重疾病活动(自发出血、大溃疡)相关联。在本公开文本的上下文中,梅奥内镜下子评分量表中的第一级别被称为MCES(或Mayo)=0,梅奥内镜下子评分量表中的第二级别被称为MCES=1,梅奥内镜下子评分量表中的第三级别被称为MCES=2,并且梅奥内镜下子评分量表中的第四级别被称为MCES=3。

[0030] 使用将视频分类为属于Mayo>1的二元分类器可能特别有利,因为在临床试验中,Mayo得分<=1通常被定义为缓解期。使用将视频分类为属于Mayo>2的二元分类器可能特别有利,因为Mayo得分>2被定义为严重疾病。理想情况下,一个分类器应该至少能够识别那些显示严重疾病迹象的视频。

[0031] 如本领域技术人员所理解的,增加训练数据量可以在保持分类器性能的同时增加分类器的复杂性。特别地,具有例如对应于梅奥内镜下子评分量表的四个级别的四个类别的第一深度神经网络分类器可以被训练,并且在给定足够量的训练数据的情况下可以具有良好的性能。

[0032] 在实施方案中,使用有序分类模型训练分类器。当训练分类器预测有序变量时,有序分类模型可能特别合适。例如MCES量表的严重程度量表可以被认为代表这样的变量,因为该量表是任意的,并且只有值之间的顺序才有意义(即,设置在0、1、2、3的值除了1比0更严重、2比1更严重和3比2更严重之外没有其他意义)。因此,当定义两个或更多的严重程度类时,可以有利地使用有序分类模型,这些严重程度类旨在表示增加的严重程度级别。

[0033] 本发明的发明人已经发现,只要使用足够数量的训练数据和/或有序分类模型,即使使用“原始”注释的结肠镜检查视频对训练数据的质量的要求放宽,也可以获取可靠地预测结肠镜检查视频的临床相关性质(例如四级MCES)的分类器。因此,本发明的方法改进了实际适用性,因为它使用通常可用和/或容易获得的数据,并产生可靠的临床相关预测。例如,本发明的发明人已经发现,当使用大约100个视频作为训练数据时,可以使用“原始”结肠镜检查视频作为训练数据和受试者数据来获取可靠地预测结肠镜检查视频的临床相关性质(例如预测MCES>1或<=1,或预测MCES>2或<=2)的二元分类器。本发明的发明人还发现,当使用大约1000个视频作为训练数据时,可以使用“原始”结肠镜检查视频作为训练数据和受试者数据,获取可靠地预测结肠镜检查视频的临床相关性质(例如在整个四级量表

上预测MCES)的多类分类器。

[0034] 在实施方案中,可以通过训练第一深度神经网络分类器的多个实例来实现有序分类模型,其中第一深度神经网络分类器的每个实例是二元分类器,其计算图像数据属于第一严重程度类别或第二严重程度类别的概率。在这样的实施方案中,属于三个或更多严重程度类别中的每一个的概率(所述第一严重程度类别所关联的内镜下病变或溃疡性结肠炎比所述第二严重程度类别的更严重,并且所述第二严重程度类别所关联的内镜下病变或溃疡性结肠炎比所述第三严重程度类别的更严重,等等)可以基于第一深度神经网络分类器的多个实例的多个实例的组合输出来获取。例如,通过组合三个二元深度神经网络分类器的输出,可以获取预测图像数据属于四个类别(类别1到4,例如MCES的四个级别,其中1是最低严重程度,MCES=0,4是最高严重程度,MCES=3)之一的概率的分类器:

[0035] (i)提供图像数据属于除最低严重程度类别以外的任何严重程度类别的概率(即P(类别>1中的数据))和可选地图像数据属于第一类别的概率(P(类别1中的数据))的分类器,

[0036] (ii)提供图像数据属于第三或更高严重程度类别的概率(即P(类别>2中的数据))和可选地图像数据属于第二类别或更低类别的概率(P(类别1或类别2中的数据))的分类器,以及

[0037] (iii)提供图像数据属于第四严重程度类别的概率(即P(类别>3中的数据),相当于P(类别4中的数据))和可选地图像数据属于第三类别或更低类别的概率(即P(类别<4中的数据))的分类器。

[0038] 基于这些组合输出,可以计算图像数据属于第一(最低)严重程度类的概率为P(类别1中的数据)或1-P(类别>1中的数据)。类似地,可以将图像数据属于第四(最高)严重程度类别的概率计算为P(类别4中的数据)或1-P(类别<4中的数据)。图像数据属于第二严重程度类别的概率可以计算为P(类别>1中的数据)-P(类别>2中的数据)或1-P(类别>2中的数据)-P(类别1中的数据)。类似地,图像数据属于第三严重程度类别的概率可以计算为P(类别>2中的数据)-P(类别>3中的数据)或1-P(类别>3中的数据)-P(类别1或类别2中的数据)。

[0039] 在这样的实施方案中,可以同时训练第一深度神经网络分类器的多个实例,以便最大化使用第一深度神经网络分类器的多个实例的组合输出进行的预测的性能。

[0040] 在实施方案中,可以通过在输出层中用K-1个二元分类器训练单个CNN来实现具有K=1…K个类别的有序分类模型,其中每个二元分类器预测图像数据是否属于类别k>1,k>2…k>K-1,并且用于模型训练的损失函数适于使所有二元分类器的损失最小化,同时确保分类器一致性(即来自各个二元分类器的预测之间的一致性)。在实施方案中,所述第一深度神经网络分类器可以如在Cao等人(Rank-consistent Ordinal Regression for Neural Networks,2019,arXiv:1901.07884v4,可在https://arxiv.org/pdf/1901.07884.pdf获取)中所描述的来训练,所述文件的内容通过引用并入本文。

[0041] 在实施方案中,所述第一深度神经网络分类器以三个或更多个严重程度类别对图像数据进行分类,其中对于所述训练图像数据中的每组帧,如果目视检查将包括所述组帧的训练结肠镜检查视频片段与针对所述三个或更多个严重程度类别中的各严重程度类别的预定不同范围内的内镜下严重程度得分相关联,则所述组帧被分配三个或更多个严重程度类别标签中的一个。在一些这样的实施方案中,对于所述训练图像数据中的每组帧,如果

两次独立的目视检查将包括所述组帧的训练结肠镜检查视频片段与在与所述三个或更多个严重程度类别标签中的同一个严重程度类别标签相关联的范围内的内镜下严重程度得分相关联，则所述组帧被分配所述严重程度类别标签。

[0042] 在一些实施方案中，所述内镜下严重程度得分是梅奥诊所内镜下子评分，并且所述第一深度神经网络分类器将图像数据分类为四个严重程度类别，每个严重程度类别与不同的梅奥诊所内镜下子评分相关联。

[0043] 四级别梅奥内镜下子评分是一种广泛用于溃疡性结肠炎内镜下评估的量表。因此，能够将图像数据分类为对应于或可使其对应于梅奥内镜下子评分量表的类别的分类器可能特别有用，因为其输出可以由临床医生直接解释。此外，通过直接使用梅奥得分作为用于训练的类别标签，这种分类器可以使用已经用梅奥得分注释的现有结肠镜检查数据。

[0044] 在实施方案中，来自所述结肠镜检查视频或其部分的所述图像数据包括多个连续帧。

[0045] 本发明的发明人已经发现，本发明的方法能够分析结肠镜检查视频或其部分，并且即使在使用“原始”结肠镜检查数据(即，所述数据尚未被处理以选择提供特别有用信息的帧)时，也能够提供临床相关的严重程度评估。换句话说，本发明的方法即使在分类器已经使用“原始”结肠镜检查视频训练时也能够提供临床相关的严重程度评估，并且能够为“原始”结肠镜检查视频提供临床相关的评估。如技术人员所理解的，深度神经网络分类器通常为单个图像(即结肠镜检查视频的单个帧)产生输出。本发明的发明人已经发现，“原始”结肠镜检查视频可用于弱训练深度神经网络分类器，以预测共同形成原始结肠镜检查视频或其部分的一组连续帧中的多个帧的严重程度类别(其中，可使用第一个分类器评估一组连续帧的一些或全部帧，这取决于例如是否使用第二个分类器来确定哪些帧是低质量的并且不应由第一个分类器分类)。实际上，本发明的发明人已经发现，对于多个帧的预测可以组合成对于结肠镜检查视频或其部分的临床相关评估。

[0046] 在实施方案中，分析所述结肠镜检查视频或其部分包括使用所述第一深度神经网络分类器和可选的所述第二深度神经网络分类器对来自所述结肠镜检查视频或其部分的所述图像数据中的所述多个帧分别进行分类。

[0047] 在实施方案中，对各个帧进行分类包括为所述多个帧中的每一帧提供所述帧属于所述第一类别的概率和/或所述帧属于所述第二类别的概率。

[0048] 在实施方案中，如果某一帧属于第一质量类别的概率达到或超过阈值，则认为该帧被第二深度神经网络分类器分类在第一质量类别中。有利地，阈值可以在0.9到0.99之间。在实施方案中，阈值约为0.95。在实施方案中，动态地确定阈值，使得训练图像数据中的各组帧平均包含在20到40之间、优选大约30个被分类在第一质量类别中的帧。

[0049] 本发明的发明人令人惊讶地发现，在逐帧的基础上应用朴素截断来对图像数据进行质量过滤足以确保基于严重程度的分类为一组中的剩余帧产生准确的结果。如本领域技术人员所理解的，当增加基于质量的过滤的严格性时，在过滤数据的质量和通过过滤器并可用于训练的数据量之间存在权衡。本发明的发明人发现上述值在这方面取得了很好的平衡。

[0050] 在实施方案中，分析所述结肠镜检查视频或其部分进一步包括基于来自所述第一深度神经网络分类器的针对所述多个帧进行的各自分类为所述结肠镜检查视频或其部分

分配总体严重程度类别。

[0051] 本发明的发明人已经发现，可以基于来自各个帧的分类结果来获取结肠镜检查视频的临床相关总度量。特别地，这种总体度量可以准确地再现用于结肠镜检查视频的专家内镜评估度量，例如从例如临床试验中获得的那些。令人惊讶的是，尽管在多个帧的各自分类中存在可变性以及在训练图像数据中对各个帧可用的弱标记，情况也是如此。

[0052] 在实施方案中，对各个帧进行分类包括为所述多个帧中的每一帧提供所述帧属于所述第一严重程度类别的概率；并且基于针对所述多个帧进行的各自分类为所述结肠镜检查视频或其部分分配总体严重程度类别包括如果帧属于第一严重程度类别的概率的平均值(均值)高于阈值，则分配第一严重程度类别。

[0053] 类似地，对各个帧进行分类可包括为所述多个帧中的每一帧提供所述帧属于所述第二严重程度类别的概率，并且基于针对所述多个帧进行的各自分类为所述结肠镜检查视频或其部分分配总体严重程度类别包括如果帧属于第二严重程度类别的概率的平均值高于阈值，则分配第二严重程度类别。

[0054] 本发明的发明人已经发现，对一组中的各个帧属于第一严重程度类别的各自的概率进行平均提供了如同由专家分配给包括这些帧的视频的类别标签的准确预测。当由专家进行内镜下评估时，结肠镜检查视频或其部分通常被分配到视频中已确定的最严重类别(最高严重程度得分)。因此，可以预期视频中由分类器识别的最严重的类别可能最接近地再现了基于专家的评估过程。

[0055] 然而，本发明的发明人令人惊讶地发现，对视频中各个帧的预测进行平均导致了更准确的预测。例如总体预测器可能对离群值不太敏感，在使用弱监督训练的深度神经网络的情况中，离群值可能比例如使用人工管理的“地面真值”数据训练的深度神经网络更常见。

[0056] 在实施方案中，可使用其他总体度量来替代平均值或作为平均值的补充，包括例如集中趋势的其他统计度量，如修剪平均值、中值等。

[0057] 在一些这样的实施方案中，阈值在0.5到0.9之间，例如约0.5、约0.6、约0.7、约0.8或约0.9。优选地，阈值可以是大约0.5。在实施方案中，基于针对所述多个帧进行的各自分类为所述结肠镜检查视频或其部分分配总体严重程度类别包括：如果帧属于第一严重程度类别的概率的平均值高于阈值，则分配第一严重程度类别，其中所述阈值是基于从以下中选择的一个或多个判断标准动态确定的：预定可接受假阳性比例、预定可接受假阴性比例、预定最小准确度(真阳性和真阴性的组合比例)、预定最小准确度(真阳性与真阳性和假阳性的比率)和/或预定最小查全率(真阳性与真阳性和假阴性的比率)。当设置用于将数据分配给第一(较高)严重程度类别的阈值时，真阳性可指正确分配给第一严重程度类别的数据，真阴性可指正确分配给第二(较低)严重程度类别的数据，假阴性可指错误分配给第二(较低)严重程度类别的数据，假阳性可指错误分配给第一(较高)严重程度类别的数据。

[0058] 本发明的发明人惊讶地发现，即使是基于属于特定严重程度类别的概率超过0.5的朴素分类，也能够准确地再现基于专家的评估过程。

[0059] 在实施方案中，基于针对所述多个帧进行的分类为所述结肠镜检查视频或其部分分配总体严重程度类别包括：如果分配给第一严重程度类别的帧的比例高于阈值，则分配第一严重程度类别。在实施方案中，基于针对所述多个帧进行的分类为所述结肠镜检查视

频或其部分分配总体严重程度类别包括:分配在所述多个帧中出现最多的严重程度类别(即,所述多个帧的严重程度类别分配的众数)。

[0060] 在实施方案中,基于对多个帧的分类为结肠镜检查视频或其部分分配总体严重程度类别包括分配具有高于阈值的帧比例的最高严重程度类别。例如,如果被分配了第一严重程度类别的多个帧的比例高于第一阈值,则结肠镜检查视频或其部分的总体严重程度类别可被定义为第一(最高)严重程度类别,如果被分配了第二严重程度类别的多个帧的比例高于第二阈值(并且被分配给第一类别的帧的比例不高于第一阈值),则为第二严重程度类别,如果被分配了第三严重程度类别的多个帧的比例高于第三阈值(并且分别分配给第一和第二类别的帧的比例不高于第一和第二阈值),则可选地为第三严重程度类别,等等。阈值可能因类别而异。例如,第一严重程度类别的阈值可以高于第二和第三严重程度类别的阈值。

[0061] 在实施方案中,基于针对所述多个帧进行的分类为所述结肠镜检查视频或其部分分配总体严重程度类别包括:使用已经使用多示例学习训练的第一深度神经网络分类器。在一些这样的实施方案中,基于针对所述多个帧进行的分类为所述结肠镜检查视频或其部分分配总体严重程度类别包括:使用第一深度神经网络分类器来预测每一帧的类别,并且使用各个预测结果的加权平均值来组合这些预测结果,其中权重由神经网络确定(基于注意力的池)。例如,可以使用在Ilse等人(Attention-based Deep Multiple Instance Learning,2018,arXiv:1802.04712v4,可在https://arxiv.org/pdf/1802.04712.pdf获取)中所描述的多示例学习方法,所述文件通过引用并入本文。

[0062] 在实施方案中,所述方法进一步包括将结肠镜检查视频或其部分自动分离成片段,其中每个片段不包括由第一深度神经网络分类器以不同严重程度分类的帧。在一些这样的实施方案中,如果第一深度神经网络分类器输出某一帧属于某一特定严重程度类别的概率高于阈值,则认为所述帧被分类在所述严重程度类别中。有利地,所述方法能够自动检测显示出显示疾病迹象的结肠的一部分的视频片段。

[0063] 如本领域技术人员所理解的,不包括分类在不同严重程度类别中的帧的视频片段可以包括未被分配任何严重程度类别的一个或多个帧,例如,因为它们由于质量原因例如被第二深度神经网络分类器过滤掉,或者因为它们不能可靠地分配到特定严重程度类别。

[0064] 在实施方案中,所述方法进一步包括将总体严重程度类别标签分配给每个自动创建的片段,其中总体严重程度类别标签基于片段中分配给每个严重程度类别的帧的比例和/或每一帧属于每个严重程度类别的平均概率。在实施方案中,未被第一深度神经网络分类器分配严重程度类别标签的帧被分配属于每个严重程度类别相等的概率。

[0065] 将结肠镜检查视频分离成包括分配给相同类别的帧的片段,可使得能够识别显示与相似严重程度相关的内镜下病变的结肠区段(对应于原始结肠镜检查视频的连续的多个片段)。使用片段中分配给每个类别的帧的比例或每一帧属于每个类别的平均概率可以处理这样的情况,即片段包括没有被第一深度神经网络分类器分配类别标签的帧,因为它们被第二深度神经网络分类器过滤掉了,或者因为第一深度神经网络分类器输出的所述帧属于任何类别的概率没有超过选定的置信度阈值。

[0066] 在实施方案中,被第二深度神经网络分类器过滤掉的帧被赋予属于每个严重程度类的相等的概率。在实施方案中,由第一深度神经网络分类器输出的概率用于不能可靠地

分配给特定严重程度类别的帧，因为由第一深度神经网络分类器输出所述帧属于的任何类别的概率没有超过选定的置信度阈值。

[0067] 在获取总体严重程度类别的实施方案中，所述方法可以进一步包括为所述多个帧中的每一帧输出所述总体严重程度类别和/或来自所述第一深度神经网络分类器的分类。

[0068] 在实施方案中，所述方法进一步包括通过将解剖学区段分配给结肠镜检查视频或其部分中的每一帧，以及将结肠镜检查视频或其部分分割成不包括分配给不同解剖学区段的帧的片段，从而自动将结肠镜检查视频或其部分分割成片段。在实施方案中，来自第一深度神经网络分类器的任何总体严重程度类别标签都是使用来自单个片段的帧来获取的。

[0069] 在实施方案中，将解剖学区段分配给每一帧包括从帧中提取指示解剖学区段的图形标签。在实施方案中，从帧中提取图形标签包括使用第三深度神经网络分类器，所述第三深度神经网络分类器至少部分地基于帧上的图形标签中的信息将图像数据分类为多个解剖类别，每个解剖类别对应于一个解剖学区段。

[0070] 结肠镜检查视频可以用解剖学片段注释，例如由执行结肠镜检查的医生注释。关于在视频或其部分中可见的解剖片段的信息可以通过诸如嵌入图像中的文本串之类的图形标签被包括在视频的每一帧上。这样的图形标签可有利地用于自动地将受试者或训练结肠镜检查视频分离成对应于不同解剖学区段的片段。这些反过来可以与个性化的严重程度预测相关联。当可对受试者结肠镜检查视频执行基于解剖学区段的分割时，这可增加预测的准确度和/或其临床相关性。当可以对至少一些训练数据执行基于解剖学区段的分割时，这可以使第一深度神经网络分类器能够用可能具有更多临床相关意义的数据来训练。

[0071] 在实施方案中，所述第一、第二和/或第三深度神经网络分类器是卷积神经网络(CNN)。有利地，CNN可以是已经在不相关的图像数据上预先训练过的。在实施方案中，所述CNN是50层CNN。在实施方案中，所述CNN是已经使用深度残差学习框架预先训练的CNN。

[0072] 卷积神经网络已被证明在图像识别任务中表现特别好。此外，已经在诸如ImageNet数据库的大量图像数据集合上预先训练用于图像识别任务的CNN是可获得的。这些CNN可以在新数据上被部分地重新训练，例如通过“冻结”(即不重新训练)已被训练以识别图像中较低级别特征的较低级别层(例如卷积层)，并且仅重新训练较高级别层(例如分类层)以识别对当前分类问题特别有用的较高级别特征。这种部分再训练意味着有限数量的数据可以用于快速训练深度CNN，因为只有CNN的参数子集需要通过训练来确定。

[0073] 深度残差学习是一种开发用于图像识别的学习框架，旨在解决称为“退化”的问题(即随着网络深度的增加，准确度饱和然后迅速下降的观察结果)。关于使用深度残差学习训练CNN的更多细节可以在https://arxiv.org/pdf/1512.03385.pdf中查看，所述文件通过引用并入本文。

[0074] 在实施方案中，所述CNN是已使用深度残差学习(也称为ResNets)训练过的预训练网络。在实施方案中，所述CNN为ResNet50。ResNet50是已经利用来自ImageNet数据库的超过一百万张图像进行了训练的CNN，并且其可以以其原生形式(在再训练之前)将图像分类为1000个对象类别，包括例如键盘、铅笔、许多动物等。该CNN可以被重新训练以执行不同的图像分类任务。

[0075] 在实施方案中，训练图像数据包括至少50、至少60、至少70、至少80、至少90或至少100个单独的结肠镜检查视频。从不同受试者和/或在不同时间点获取各个结肠镜检查视

频。

[0076] 在实施方案中，训练图像数据包括来自结肠的一个或多个解剖学区段中的每个区段的至少50、至少60、至少70、至少80、至少90或至少100个单独的结肠镜检查视频。在实施方案中，训练图像数据包括来自结肠的三个解剖学区段中的每个区段的至少50、至少60、至少70、至少80、至少90或至少100个单独的结肠镜检查视频，优选地，其中所述区段为直肠、乙状结肠、降结肠。在实施方案中，训练图像数据包括来自结肠的一个或多个(例如，三个)解剖学区段中的每一个的至少50、至少60、至少70、至少80、至少90或至少100个单独的结肠镜检查视频，并且第一深度神经网络分类器是二元分类器。在实施方案中，训练图像数据包括来自结肠的一个或多个(例如，三个)解剖学区段中的每一个的至少500、至少600、至少700、至少800、至少900或大约1000个单独的结肠镜检查视频，并且第一深度神经网络分类器是多类分类器。优选地，第一深度神经网络分类器是多类分类器，其预测图像数据属于四个类别(类1至4，例如MCES的四个级别，其中1是最低严重程度，MCES＝0，4是最高严重程度，MCES＝3)之一的概率。

[0077] 在实施方案中，训练图像数据包括来自多个结肠镜检查视频中的每一个的平均30帧。

[0078] 在实施方案中，训练图像数据包括至少5000帧、至少6000帧、至少7000帧、至少8000帧或至少9000帧。

[0079] 在实施方案中，第一深度神经网络分类器能够以至少0.7的受试者工作特征曲线下面积(AUC)预测一组帧是属于第一严重程度类别还是属于第二严重程度类别。在一些这样的实施方案中，在用少于150个单独结肠镜检查视频训练之后，第一深度神经网络分类器能够以至少0.7的AUC预测一组帧是属于第一严重程度类别还是属于第二严重程度类别单独的结肠镜检查视频。

[0080] 二元分类器的性能可以通过量化受试者工作特性曲线(AUC)下面积来衡量。如技术人员所知，受试者工作特性曲线或ROC曲线说明了二元分类器的诊断能力。可以通过绘制不同阈值设定下的真阳性率(TPR)和假阳性率(FPR)的关系图来获取。例如，可以通过针对应用于属于第一严重程度类别的预测概率的阈值的不同值(例如，以0.05的步长取0到1之间的每一个值)绘制TPR与FPR的关系图来获取ROC曲线。在实施方案中，多类分类器的性能可以通过量化Cohen's kappa系数和/或预测类和真实类之间的百分比一致性来测量。优选地，通过量化Cohen's kappa系数来测量多类分类器的性能。如技术人员所知，Cohen's kappa可以计算为$(p_o-p_e/1-p_e)$，其中p_0是预测类别和真实类别之间的相对观测一致性，p_e是预测的和真实的类别偶然一致的概率(基于每个类别的数据量)。

[0081] 在实施方案中，可以通过执行交叉验证，例如5折或10折交叉验证，以及针对交叉验证的每个分裂(split)量化AUC和/或Cohen's kappa来测量性能。在实施方案中，交叉验证分裂被定义为使得来自同一个结肠镜检查视频的图像数据仅出现在其中一个集合中。在一些这样的实施方案中，交叉验证分裂被定义为使得来自同一患者的图像数据仅出现在集合中的一个中。这些可以防止不同分裂中的训练图像数据之间可能的依赖性虚假地改善交叉验证的结果。

[0082] 在实施方案中，第一深度神经网络分类器能够以至少0.75的AUC预测一组帧是属于第一严重程度类别还是属于第二严重程度类别，其中第一严重程度类别对应于Mayo得分

>1,第二严重程度类别对应于Mayo得分≤1。在一些这样的实施方案中,在用少于150个单独结肠镜检查视频训练之后,第一深度神经网络分类器能够以至少0.75的AUC预测一组帧是属于第一严重程度类别还是属于第二严重程度类别单独的结肠镜检查视频。

[0083] 在实施方案中,第一深度神经网络分类器能够以至少0.75、优选至少0.8的AUC预测一组帧是属于第一严重程度类别还是属于第二严重程度类别,其中第一严重程度类别对应于Mayo得分>2,第二严重程度类别对应于Mayo得分≤2。在一些这样的实施方案中,在用少于150个单独结肠镜检查视频训练之后,第一深度神经网络分类器能够以至少0.75的AUC预测一组帧是属于第一严重程度类别还是属于第二严重程度类别单独的结肠镜检查视频。

[0084] 根据第二方面,提供了一种评估受试者体内的溃疡性结肠炎(UC)严重程度的方法,所述方法包括使用第一方面的任一实施方案的方法分析来自所述受试者的结肠镜检查视频或其部分。

[0085] 如本领域技术人员所理解的,分析来自受试者的结肠镜检查视频的方法是计算机实现的方法。事实上,使用深度神经网络分类器的图像分析以及提供这种分类器的方法需要通过复杂的数学运算来分析大量数据,这些运算超出了脑力活动的范围。

[0086] 在第二方面的实施方案中,所述方法进一步包括针对UC对受试者进行治疗。在一些实施方案中,所述方法进一步包括如果第一深度神经网络将来自受试者的结肠镜检查视频的一组帧分类到第一严重程度类别中,则对受试者进行针对UC的治疗。在一些实施方案中,所述方法进一步包括如果来自结肠镜检查视频或其部分的一组帧的总体严重程度类别是第一严重程度类别,则对受试者进行针对UC的治疗。

[0087] 在一些实施方案中,所述第一深度神经网络分类器在对应于MCES>1的第一严重程度类别和对应于MCES≤1的第二严重程度类别之间对图像数据进行分类,并且所述方法进一步包括如果来自结肠镜检查视频或其部分的一组帧的总体严重程度类别是第一严重程度类别,则对受试者进行针对UC的治疗。在一些实施方案中,所述第一深度神经网络分类器在对应于MCES>2的第一严重程度类别和对应于MCES≤2的第二严重程度类别之间对图像数据进行分类,并且所述方法进一步包括如果来自结肠镜检查视频或其部分的一组帧的总体严重程度类别是第一严重程度类别,则对受试者进行针对UC的治疗。

[0088] 在一些实施方案中,分析来自所述受试者的结肠镜检查视频或其部分包括:(i)使用第一深度神经网络分类器分析所述结肠镜检查视频或其部分,所述第一深度神经网络分类器在对应于MCES>1的第一严重程度类别和对应于MCES≤1的第二严重程度类别之间对图像数据进行分类;(ii)使用第一深度神经网络分类器分析所述结肠镜检查视频或其部分,所述第一深度神经网络分类器在对应于MCES>2的第一严重程度类别和对应于MCES≤2的第二严重程度类别之间对图像数据进行分类;以及(iii)针对UC对所述受试者进行治疗,如果在步骤(ii)中所述视频的至少一个片段被分配了第一严重程度类别,则采用第一治疗,并且如果在步骤(ii)中所述视频没有片段被分配第一严重程度类别,但在步骤(i)中所述视频的至少一个片段被分配了第一严重程度类别,则采用第二治疗。

[0089] 在一些实施方案中,分析来自所述受试者的结肠镜检查视频或其部分包括:(i)使用第一方面的方法从结肠镜检查视频或其部分获取一组或多组帧的总体严重程度类别,其中所述第一深度神经网络分类器在对应于MCES>1的第一严重程度类别和对应于MCES≤1的第二严重程度类别之间对图像数据进行分类;(ii)使用第一方面的方法从结肠镜检查视频

或其部分获取一组或多组帧的总体严重程度类别，其中所述第一深度神经网络分类器在对应于MCES>2的第一严重程度类别和对应于MCES≤2的第二严重程度类别之间对图像数据进行分类；以及(iii)如果在步骤(ii)中获取的总体严重程度类别中的至少一个是第一严重程度类别，则采用第一治疗来为所述受试者进行针对UC的治疗；如果在步骤(ii)中获取的总体严重程度类别中没有一个是第一严重程度类别，但在步骤(ii)中获取的总体严重程度类别中的至少一个是第一严重程度类别，则用第二治疗来为所述受试者进行针对UC的治疗。

[0090] 根据第三方面，提供了一种提供用于分析结肠镜检查视频或其部分的工具的方法，所述方法包括：

[0091] 获取包括来自多个训练结肠镜检查视频的多组连续帧的训练图像数据，其中一组中的帧具有相同的严重程度类别标签，所述严重程度类别标签至少选自第一严重程度类别和第二严重程度类别，其中所述第一严重程度类别所关联的内镜下病变比所述第二严重程度类别的更严重，以及

[0092] 训练第一深度神经网络至少以第一严重程度类别和第二严重程度类别对图像数据进行分类，其中使用所述训练图像数据和所述严重程度类别标签以弱监督的方式执行所述训练。优选地，所述内镜下病变指示溃疡性结肠炎。换句话说，优选地，所述第一严重程度类别所关联的溃疡性结肠炎比所述第二严重程度类别的更严重。

[0093] 所述方法优选地是计算机实现的。如上所述，在任何实际应用中，至少训练深度神经网络分类器的步骤是计算机实现的。因此，所述方法的步骤可以包括处理器执行指令以执行所述步骤。例如，获取训练图像数据可以包括处理器执行指令以从数据源(例如，数据库、计算机存储器等)获取训练图像数据。类似地，训练第一深度神经网络可包括处理器执行指令以训练第一深度神经网络。

[0094] 在实施方案中，所述方法进一步包括使用第二深度神经网络分类器至少以第一质量类别和第二质量类别对包括多个帧的训练图像数据进行分类，其中所述第一质量类别所关联的图像的质量比所述第二类别的更好，并且其中训练所述第一深度神经网络是使用被所述第二深度神经网络分类器分类在所述第一质量类别中的所述训练图像数据来执行的。

[0095] 在实施方案中，如果某一帧属于第一质量类别的概率达到或超过阈值，则认为该帧被第二深度神经网络分类器分类在第一质量类别中。所述阈值可以有利地在0.9和0.99之间，优选约0.95。

[0096] 在实施方案中，所述方法进一步包括：

[0097] 获取包括来自多个训练结肠镜检查视频的多组连续帧的训练图像数据，其中训练图像数据中的一组连续帧中的帧具有相同的质量类别标签，所述质量类别标签至少选自第一质量类别和第二质量类别，其中所述第一质量类别所关联的图像的质量比所述第二质量类别的更好；以及

[0098] 训练第二深度神经网络至少以第一质量类别和第二质量类别对训练图像数据进行分类，其中使用所述训练图像数据和所述质量类别标签以弱监督的方式执行所述训练。

[0099] 在实施方案中，所述方法进一步包括：

[0100] 获取包括来自多个训练结肠镜检查视频的多个帧的训练图像数据，其中所述训练图像数据中的帧被自动分配质量类别标签，所述质量类别标签至少选自第一质量类别和第

二质量类别，其中所述第一质量类别所关联的图像的质量比所述第二质量类别的更好；可选地，其中已使用一个或多个先前训练过的机器学习算法自动地将所述质量类别标签分配给帧；以及

[0101] 训练第二深度神经网络至少以第一质量类别和第二质量类别对训练图像数据进行分类，其中使用所述训练图像数据和所述质量类别标签以弱监督的方式执行所述训练。

[0102] 在实施方案中，用于训练第一深度神经网络的训练图像数据是用于训练第二深度神经网络的训练图像数据的子集。

[0103] 在实施方案中，动态地确定用于确定帧是否被第二深度神经网络分类器认为被分类在第一类别中的阈值，使得训练图像数据中的各组帧平均包含20至40个、优选约30个过滤后的帧。

[0104] 在实施方案中，获取包括来自多个训练结肠镜检查视频的多组连续帧的训练图像数据，其中训练图像数据中的一组连续帧中的帧具有相同的质量类别标签，包括通过以下步骤将质量类别标签分配给训练图像数据中的每组连续帧：

[0105] 如果包括所述一组连续帧的所述视频片段在目视检查上满足一个或多个质量判断标准，则将所述一组连续帧分配给所述第一质量类别，以及

[0106] 否则，将一组连续帧分配给第二质量类别。

[0107] 在实施方案中，所述一个或多个质量判断标准选自：通过对由一组连续帧组成的训练结肠镜检查视频片段的目视检查可以区分结肠壁和结肠血管、基于水的存在或不存在的判断标准、基于高反射区域的存在或不存在的判断标准、基于大便的存在或不存在的判断标准以及基于模糊的判断标准。

[0108] 有利地，如果包括一组连续帧的训练结肠镜检查视频片段至少满足以下判断标准：在对由一组连续帧组成的视频片段的目视检查中可以区分结肠壁和结肠血管，则可将该组连续帧分配给第一质量类别。在实施方案中，如果包括一组连续帧的训练结肠镜检查视频片段附加地满足基于水、高反射区域、粪便和/或模糊的存在或不存在的一个或多个判断标准，则可将该组连续帧分配给第一质量类别。在实施方案中，如果包括一组连续帧的训练结肠镜检查视频片段附加地满足基于水、高反射区域、粪便和/或模糊的存在或不存在的全部判断标准，则可将该组连续帧分配给第一质量类别。

[0109] 在实施方案中，将质量类别标签分配给训练图像数据中的每组连续帧进一步包括基于所述多个训练结肠镜检查视频中的多个片段是否满足所述一个或多个质量判断标准来鉴别所述片段，以及将对应的质量类别标签分配给来自每个片段的所述组连续帧。

[0110] 在实施方案中，所述方法进一步包括可选地通过从所述多个训练结肠镜检查视频的每一帧中提取指示解剖学区段的图形标签，将解剖学区段标签分配给每一帧。可选地，从帧中提取图形标签可以包括使用第三深度神经网络分类器，所述第三深度神经网络分类器至少部分地基于帧上的图形标签中的信息将图像数据分类为多个解剖类别，每个类别对应于一个解剖学区段。在实施方案中，所述方法进一步包括训练第三深度神经网络分类器至少部分地基于帧上的图形标签中的信息将图像数据分类为多个解剖类别，每个类别对应于解剖学区段。

[0111] 在实施方案中，所述训练图像数据包括多个训练结肠镜检查视频，其中每个视频的片段已被分配内镜下严重程度得分，以及获取包括来自所述多个训练结肠镜检查视频的

多组连续帧的训练图像数据，其中一组中的帧具有相同的严重程度类别标签，所述严重程度类别标签至少选自第一严重程度类别和第二严重程度类别，包括：将第一严重程度类别分配给形成被分配了第一范围内的内镜下严重程度得分的片段的一部分的每组连续帧，以及将第二严重程度类别分配给形成被分配了第二范围内的内镜下严重程度得分的片段的一部分的每组连续帧。

[0112] 在实施方案中，所述训练图像数据包括多个训练结肠镜检查视频，其中每个训练结肠镜检查视频的片段已被分配至少两个独立获取的内镜下严重程度得分值，以及获取包括来自所述多个训练结肠镜检查视频的多组连续帧的训练图像数据，其中一组中的帧具有相同的严重程度类别标签，所述严重程度类别标签至少选自第一严重程度类别和第二严重程度类别，包括：针对所述内镜下严重程度得分的所有独立获取的值，将第一严重程度类别分配给形成被分配了第一范围内的内镜下严重程度得分的片段的一部分的每组连续帧，以及针对所述内镜下严重程度得分的所有独立获取的值，将第二严重程度类别分配给形成被分配了第二范围内的内镜下严重程度得分的片段的一部分的每组连续帧。

[0113] 在实施方案中，如果所述至少两个独立获取的内镜下严重程度得分值相同，则将训练图像数据中的一组连续帧分配给一个类别。在实施方案中，从训练图像数据中排除未被分配相同内镜下严重程度得分或相同范围内的内镜下严重程度得分的片段。

[0114] 在实施方案中，获取包括来自所述多个训练结肠镜检查视频的多组连续帧的训练图像数据，其中一组中的帧具有相同的严重程度类别标签，所述严重程度类别标签至少选自第一严重程度类别和第二严重程度类别，包括：将第一严重程度类别分配给形成被分配了高于阈值的内镜下严重程度得分的片段的一部分的每组连续帧，以及将第二严重程度类别分配给形成被分配了等于或小于阈值的内镜下严重程度得分的片段的一部分的每组连续帧。

[0115] 在实施方案中，所述内镜下严重程度得分是梅奥诊所内镜下子评分。有利地，所述第一范围可以是MCES>1或MCES>2。同样，所述阈值可以是梅奥诊所内镜下子评分1或2。

[0116] 在实施方案中，所述训练图像数据包括来自多个训练结肠镜检查视频的多组连续帧，其中每个视频的片段都被分配了内镜下严重程度得分；获取所述训练图像数据包括：对于每组连续帧，如果包括所述帧组的训练结肠镜检查视频片段已经被分配了针对三个或更多个严重程度类别中的任一个的预定不同范围内的内镜下严重程度得分，则将所述三个或更多个严重程度类别标签中的这个标签分配给该组连续帧；以及训练所述第一深度神经网络包括训练所述第一深度神经网络以所述三个或更多严重程度类别对图像数据进行分类。

[0117] 在实施方案中，所述内镜下严重程度得分是梅奥诊所内镜下子评分，并且所述第一深度神经网络分类器被训练用于将图像数据分类为四个严重程度类别，每个严重程度类别与不同的梅奥诊所内镜下子评分相关联。

[0118] 在实施方案中，所述第一、第二和/或第三深度神经网络分类器是卷积神经网络(CNN)。在一些这样的实施方案中，训练第一、第二和/或第三深度神经网络分类器包括获取经过预训练的CNN和使用训练图像数据部分地重新训练所述CNN。

[0119] 在实施方案中，部分地重新训练CNN包括固定CNN的一个或多个较低层的参数，以及确定CNN的剩余层(较高层)的参数。在实施方案中，部分地重新训练CNN包括确定CNN的最后5至10层(例如8层)的参数。在实施方案中，部分地重新训练CNN包括确定CNN的最后10％

至20％的层的参数(例如,对于50层CNN,可以重新训练最后5至10层)。

[0120] 在实施方案中,已经使用不相关的图像数据对CNN进行了预训练。在实施方案中,所述CNN是50层CNN。在实施方案中,所述CNN是已经使用深度残差学习框架预先训练的CNN。在实施方案中,所述CNN为ResNet50。

[0121] 在实施方案中,获取训练图像数据包括获取至少50、至少60、至少70、至少80、至少90或至少100个单独的结肠镜检查视频。

[0122] 在实施方案中,获取训练图像数据包括从结肠的一个或多个、优选三个(例如直肠、乙状结肠、降结肠)解剖学区段中的每一个获取至少50、至少60、至少70、至少80、至少90或至少100个单独的训练结肠镜检查视频。

[0123] 在实施方案中,获取训练图像数据包括从多个训练结肠镜检查视频中的每一个平均提取30帧。

[0124] 在实施方案中,获取训练图像数据包括从多个训练结肠镜检查视频中提取至少5000帧、至少6000帧、至少7000帧、至少8000帧或至少9000帧。

[0125] 在实施方案中,所述方法进一步包括评估所述第一深度神经网络的性能。在实施方案中,评估第一深度神经网络的性能包括使用验证图像数据量化接收器工作特性曲线(AUC)下的面积。在实施方案中,评估第一深度神经网络的性能包括使用验证图像数据计算Cohen's kappa。在实施方案中,验证图像数据包括训练图像数据的子集。

[0126] 在实施方案中,评估第一深度神经网络的性能包括使用训练图像数据执行交叉验证,例如5折或10折交叉验证,以及针对交叉验证的每个分裂量化AUC或Cohen's kappa。在一些这样的实施方案中,执行交叉验证包括将训练图像数据分离成多个分裂,其中来自同一个结肠镜检查视频的各组帧不会出现在一个以上的分裂中。

[0127] 本方面的方法可以进一步包括就第一方面所述的任何特征。特别地,在本方面的上下文中明确设想了与第一方面中的深度神经网络的训练相关的任何特征,包括与训练数据、类别或分类器的性质、以及分类器所做的预测相关的任何特征。

[0128] 根据第四方面,提供了一种评估受试者体内的溃疡性结肠炎(UC)严重程度的方法,所述方法包括:

[0129] 提供根据第三方面的任一实施方案的用于分析结肠镜检查视频或其部分的工具,以及

[0130] 使用所述用于分析结肠镜检查视频或其部分的工具分析来自所述受试者的结肠镜检查视频或其部分。

[0131] 根据第五方面,提供了一种评估受试者体内的溃疡性结肠炎(UC)严重程度的方法,所述方法包括:

[0132] 使用如第三方面的任何实施方案所述的方法训练第一深度神经网络分类器和可选地训练第二和/或第三深度神经网络分类器,以及

[0133] 使用第一方面的方法分析来自受试者的结肠镜检查视频或其部分。

[0134] 根据第六方面,提供了一种用于根据从受试者体内获取的结肠镜检查视频来评估所述受试者体内的溃疡性结肠炎严重程度的系统,所述系统包括:

[0135] 至少一个处理器;以及

[0136] 至少一个非暂时性计算机可读介质,所述至少一个非暂时性计算机可读介质包含

指令，当由所述至少一个处理器执行时，所述指令使所述至少一个处理器执行包括以下各项的操作：

[0137] 将来自所述结肠镜检查视频或其部分的图像数据至少分类为第一严重程度类别和第二严重程度类别，其中所述第一严重程度类别所关联的溃疡性结肠炎比所述第二严重程度类别的更严重，其中所述第一深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练，其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧，其中一组中的帧具有相同的严重程度类别标签。

[0138] 在实施方案中，当由所述至少一个处理器执行时，所述指令进一步使所述处理器执行关于第一方面的任何实施方案所描述的操作。

[0139] 根据第七方面，提供了一种非暂时性计算机可读介质，用于根据从受试者体内获取的结肠镜检查视频来评估受试者体内的溃疡性结肠炎严重程度，所述非暂时性计算机可读介质包括指令，当由所述至少一个处理器执行时，所述指令使所述至少一个处理器执行包括以下各项的操作：至少以第一严重程度类别和第二严重程度类别对来自所述结肠镜检查视频或其部分的图像数据进行分类，其中所述第一严重程度类别所关联的溃疡性结肠炎比所述第二严重程度类别的更严重，其中所述第一深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练，其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧，其中一组中的帧具有相同的严重程度类别标签。

[0140] 在实施方案中，当由所述至少一个处理器执行时，所述指令进一步使所述处理器执行关于第一方面的任何实施方案所描述的操作。

[0141] 根据另一方面，提供了一种用于分析从受试者体内获取的结肠镜检查视频的系统，所述系统包括：至少一个处理器；以及至少一个非暂时性计算机可读介质，所述至少一个非暂时性计算机可读介质包含指令，当由所述至少一个处理器执行时，所述指令使所述至少一个处理器执行包括以下各项的操作：

[0142] 至少以第一严重程度类别和第二严重程度类别对来自所述结肠镜检查视频或其部分的图像数据进行分类，其中所述第一严重程度类别所关联的内镜下病变(或者具体地，溃疡性结肠炎)比所述第二严重程度类别的更严重，其中所述第一深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练，其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧，其中一组中的帧具有相同的严重程度类别标签。

[0143] 根据又一个方面，提供了一种用于对从受试者体内获取的结肠镜检查视频进行分类的系统，所述系统包括：至少一个处理器；以及至少一个非暂时性计算机可读介质，所述至少一个非暂时性计算机可读介质包含指令，当由所述至少一个处理器执行时，所述指令使所述至少一个处理器执行包括以下各项的操作：

[0144] 至少以第一严重程度类别和第二严重程度类别对来自所述结肠镜检查视频或其部分的图像数据进行分类，其中所述第一严重程度类别所关联的内镜下病变(或者具体地，溃疡性结肠炎)比所述第二严重程度类别的更严重，其中所述第一深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练，其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧，其中一组中的帧具有

相同的严重程度类别标签。

[0145] 根据另一个方面,提供了一种用于分析从受试者体内获取的结肠镜检查视频的非暂时性计算机可读介质,所述非暂时性计算机可读介质包括指令,当由所述至少一个处理器执行时,所述指令使所述至少一个处理器执行包括以下各项的操作:至少以第一严重程度类别和第二严重程度类别对来自所述结肠镜检查视频或其部分的图像数据进行分类,其中所述第一严重程度类别所关联的内镜下病变(或者具体地,溃疡性结肠炎)比所述第二严重程度类别的更严重,其中所述第一深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练,其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧,其中一组中的帧具有相同的严重程度类别标签。

[0146] 根据另一个方面,提供了一种用于对从受试者体内获取的结肠镜检查视频进行分类的非暂时性计算机可读介质,所述非暂时性计算机可读介质包括指令,当由所述至少一个处理器执行时,所述指令使所述至少一个处理器执行包括以下各项的操作:至少以第一严重程度类别和第二严重程度类别对来自所述结肠镜检查视频或其部分的图像数据进行分类,其中所述第一严重程度类别所关联的内镜下病变(或者具体地,溃疡性结肠炎)比所述第二严重程度类别的更严重,其中所述第一深度神经网络已经至少部分地使用来自多个训练结肠镜检查视频的训练图像数据以弱监督的方式训练,其中所述训练图像数据包括来自所述多个训练结肠镜检查视频的多组连续帧,其中一组中的帧具有相同的严重程度类别标签。

[0147] 根据另一个方面,提供了一种治疗受试者体内的溃疡性结肠炎的方法,所述方法包括使用第一方面的任何实施方案的方法分析来自受试者的结肠镜检查视频或其部分。

[0148] 在一些实施方案中,所述方法进一步包括如果第一深度神经网络将来自受试者的结肠镜检查视频的一组帧分类到第一严重程度类别中,和/或如果来自结肠镜检查视频或其部分的一组帧的总体严重程度类别是第一严重程度类别,则对受试者进行针对UC的治疗。

[0149] 在实施方案中,所述方法包括:(i)使用第一深度神经网络分类器分析所述结肠镜检查视频或其部分,所述第一深度神经网络分类器在对应于MCES>1的第一严重程度类别和对应于MCES≤1的第二严重程度类别之间对图像数据进行分类;(ii)使用第一深度神经网络分类器分析所述结肠镜检查视频或其部分,所述第一深度神经网络分类器在对应于MCES>2的第一严重程度类别和对应于MCES≤2的第二严重程度类别之间对图像数据进行分类;以及(iii)针对UC对所述受试者进行治疗,如果在步骤(ii)中所述视频的至少一个片段被分配了第一严重程度类别,则采用第一治疗,并且如果在步骤(ii)中所述视频没有片段被分配第一严重程度类别,但在步骤(i)中所述视频的至少一个片段被分配了第一严重程度类别,则采用第二治疗。

附图说明

[0150] 图1示出了可以在其中使用本发明的实施方案的示例性计算系统。

[0151] 图2是示出提供用于分析结肠镜检查视频的工具的方法和分析结肠镜检查视频的方法的流程图。

[0152] 图3示意性地示出了提供用于分析结肠镜检查视频的工具的方法和分析结肠镜检

查视频的方法。

[0153] 图4图示了一种评估受试者体内的溃疡性结肠炎严重程度的方法。

[0154] 图5A及图5B示出了将来自结肠镜检查视频的图像数据分类在对应于MCES>1的第一严重程度类别和对应于MCES≤1的第二严重程度类别中的示例性深度神经网络(图5A)以及将来自结肠镜检查视频的图像数据分类在对应于MCES>2的第一严重程度类别和对应于MCES≤2的第二严重程度类别中的示例性深度神经网络(图5B)的受试者工作特征(ROC)曲线。

[0155] 在这里列出的附图示出了本发明的实施方案的情况下,这些实施方案不应被解释为对本发明的范围的限制。在适当的情况下,在不同的附图中使用相同的附图标记来表示所示实施方案的相同结构特征。

具体实施方式

[0156] 下面将参照附图描述本发明的具体实施方案。

[0157] 图1示出了可以在其中使用本发明的实施方案的示例性计算系统。计算设备1包括一个或多个处理器101(其可以是例如服务器的形式)和一个或多个存储器102。计算设备1配备有装置,以例如经由公共互联网与计算基础设施的其他元件通信。计算设备1还可以包括用户接口,其通常包括显示器。在所示实施方案中,计算设备1能够与一个或多个外部数据库202通信,例如以获取训练图像数据,接收待分析的结肠镜检查视频或其部分,和/或输出结肠镜检查视频的分析结果以用于存储。取决于特定的实现,数据库202实际上可以包括在数据库102中。在实施方案中,第一计算设备1被配置成提供用于分析结肠镜检查视频的工具,如下面将进一步描述的,通过使用从数据库202接收的训练数据。在这样的实施方案中,存储器101可以存储指令,当由处理器102执行时,所述指令使处理器执行提供用于分析结肠镜检查视频的工具的方法的步骤。在一些这样的实施方案中,所述工具一旦获取,可以以指令集的形式提供给第二计算设备(未示出),当由第二计算设备的处理器执行时,所述指令集使所述处理器执行如本文所述的分析结肠镜检查视频的方法的步骤。第二计算设备可以包括用户接口,例如显示器,通过该用户接口,第二计算设备可以将分析的结果输出给用户。在实施方案中,第一计算设备1被配置为分析(例如从数据库202接收的)结肠镜检查视频或其部分,和/或通过分析来自受试者的结肠镜检查视频或其部分来自动评估受试者体内的溃疡性结肠炎严重程度。在这样的实施方案中,存储器101可以存储指令,当由处理器102执行时,这些指令使处理器执行分析结肠镜检查视频的方法的步骤,并且可选地评估溃疡性结肠炎的严重程度,如本文所述。

[0158] 图2和图3示出了提供用于分析结肠镜检查视频的工具的方法和分析结肠镜检查视频的方法的大致实施方案。如下文将进一步解释的,提供用于分析结肠镜检查视频的工具的方法可包括步骤210/310(获取(训练)图像数据,包括步骤212(获取原始结肠镜检查视频)和可选的步骤214-218(用解剖学区段标签、质量类别标签和/或严重类别标签注释视频))和步骤240/340(训练严重程度评分网络,包括步骤242(获取每帧的严重程度类别预测结果)、246(将严重程度类别预测结果与严重程度类别标签进行比较)和可选的步骤244(获取总体严重程度类别预测结果))。可选地,这样的方法还可以包括步骤220/320(训练质量控制网络)和230/330(使用质量控制网络来过滤低质量帧)。一种分析结肠镜检查视频的方

法可以包括步骤210/310(获取用于分析的图像数据,包括步骤212(获取原始结肠镜检查视频)和可选的步骤218(用解剖学区段标签注释视频)),可选的步骤230/330(使用质量控制网络过滤低质量帧),和步骤250/350(使用严重程度评分网络获取严重程度类别预测结果,包括步骤252(获取每一帧的严重程度类别预测结果)和可选的步骤254(获取总体严重程度类别预测结果))。换句话说,训练步骤220/320(训练质量控制网络)和240/340(训练严重程度评分网络)以及用(用于训练的)类别标签注释214、216训练数据的相关步骤作为提供用于分析结肠镜检查视频的工具的方法的一部分来执行。相比之下,分析结肠镜检查视频的方法使用一个或多个分类器,如本文所述,这些分类器先前已经训练过,因此不一定需要训练步骤和训练数据的相关标记。

[0159] 在步骤210/310,获取用于训练(在这种情况下,图像数据被称为“训练图像数据”)或用于分析(在这种情况下,图像数据被称为来自受试者的结肠镜检查视频300或其部分)的图像数据。训练图像数据包括多个训练结肠镜检查视频300。每个结肠镜检查视频300(无论是在训练数据中还是在用于分析的数据中)包括多个连续帧300_1至300_n。在训练图像数据中,所述多个连续帧300_1至300_n形成已被分配了相同的严重程度类别标签S_1、S_2的组。至少使用两个严重程度类别,其中第一严重程度类别S_1所关联的内镜下病变和/或溃疡性结肠炎比第二严重程度类别S_2的更严重。在图3所示的实施方案中,共使用了两个严重程度类别,将被训练的分类器是二元分类器。然而,在其他实施方案中,可以使用3个或更多个类别,并且可以训练相应的多类分类器。获取210用于训练或分析的图像数据的步骤可以包括如图2所示的多个步骤,其中一些步骤可以是可选的。在最简单的情况下,获取210用于分析的图像数据可以简单地包括获取212来自受试者的待分析的原始结肠镜检查视频300。例如,这可以从数据库(例如数据库202)、从提交用于分析的数据的用户或从直接与内窥镜相关联的计算设备获取。可选地,用于分析的原始结肠镜检查视频300可被注释218以提供标签A_1、A_2,所述标签指示在视频的多个片段中的每一个片段中可见的解剖学区段。或者,解剖学区段标签A_1、A_2可以是已经与机器可读形式的原始结肠镜检查视频300一起提供。在实施方案中,可以用解剖学区段标签A_1、A_2手动注释视频的片段,每一帧300_1至300_n可以被分配与包括帧300_1-300_n的片段的解剖学区段标签相对应的解剖学区段标签。或者,可以从原始结肠镜检查视频300自动提取解剖学区段信息。在这样的实施方案中,用解剖学区段标签注释214视频片段300_1至300_n包括自动分配解剖学区段标签A_1、A_2给视频300的每一帧300_1至300_n。例如,原始结肠镜检查视频300可以被提供有指示在帧上可见的解剖学区段的图形标签,嵌入在每一帧中。这些可以基于例如医学专家在获取视频片段时对视频片段的手动注释。在这样的实施方案中,用解剖学区段标签注释214视频片段可以包括至少部分地基于嵌入在图像数据中的图形标签中的信息,使用已经被训练用于将图像数据(即帧的集合300_1至300_n)分类到多个解剖学类别(每个类别对应于一个解剖学区段)的深度神经网络分类器(本文中也称为第三深度神经网络分类器)来为每一帧分配解剖学标签A_1、A_2。对于深度神经网络来说,有限多样性的图形标签(例如,与结肠的部分(例如乙状结肠、直肠和降结肠)相对应的几个不同的文本串)是比较容易区分的模式,因此,这种分类的错误率预计非常低(例如,接近0%甚至0%的帧被错误分类)。在实施方案中,用于分析的原始结肠镜检查视频300可以涉及单个解剖学区段,或者可以被划分为涉及单个解剖学区段的多个部分,其中每个部分可以如本文所述单独分析。在这样的实施方案中,可以不使用具有解剖学区段标签

的注释，视频的每个部分可以形成其自己的片段。

[0160] 类似地，获取210/310训练图像数据包括获取212多个原始结肠镜检查视频300。这些视频例如可以从一个或多个数据库(例如数据库202)获取。用于训练的原始结肠镜检查视频300可选地被注释218以提供指示解剖学区段的解剖学标签A_1、A_2，如上文所述。训练视频300伴随着由专家提供并分配给视频片段的严重程度信息。将直接或间接地使用训练图像数据中的严重程度信息来训练基于严重程度的分类器(称为严重程度评分网络，SSN)，如下文将进一步描述的。实际上，视频片段300是一组帧300_1-300_n，因此一组中的所有帧都具有相同的严重程度信息。在图3所示的实施方案中，片段对应于解剖学区段，使得训练视频300的每个片段包括具有相同解剖学标签A_1、A_2和相同严重程度信息的帧。具体地，在图3所示的实施方案中，第一片段包括帧300_1至300_4，第二片段包括帧300_5至300_n。严重程度信息通常包括内镜下严重程度得分。或者，严重程度信息可以包括可被转换成严重程度得分的信息(例如，可使用预先建立的规则被转换成严重程度得分的自由文本)。例如，内镜下严重程度得分可以是梅奥诊所内镜下子评分(MCES)。MCES是通过对结肠镜检查视频的视觉评估来对内镜下病变进行评分的标准量表。通常提供MCES作为内窥镜视频的临床评估的一部分，例如在临床试验中。

[0161] 用于训练的原始结肠镜检查视频300可以任选地针对将构成训练数据一部分的每个视频片段用严重程度类别标签S_1、S_2来注释214。在实施方案中，这包括将严重程度信息转换为类别S_1、S_2。例如，当严重程度信息的格式与要训练的分类器不直接兼容时，这可能是有利的。这种情况例如包括严重程度信息不是离散类别或量表的形式、类别的数量不等于分类器将被训练来鉴别的类别的数量、和/或严重程度信息包括来自一个以上专家的评估。相反，如果严重程度信息已经是适合用于训练所选分类器的格式，则此步骤可能是不必要的。

[0162] 在实施方案中，训练图像数据中的一组帧(例如图3上的帧300_1至300_4)，可以通过以下方式用严重程度类别标签来注释214：如果目视检查将包括该组帧的训练结肠镜检查视频片段与第一范围内的内镜下严重程度得分相关联(即，如果严重程度信息包括第一范围内的严重程度得分)，则分配第一严重程度类别标签S_1。换句话说，在实施方案中，训练图像数据中的一组帧(例如图3上的帧300_1至300_4)，可以通过以下方式用严重程度类别标签来注释214：如果包括该组帧的训练结肠镜检查视频片段的严重程度信息包括第一范围内的内镜下严重程度得分(即，如果严重程度信息包括第一范围内的严重程度得分)，则分配第一严重程度类别标签S_1。在严重程度信息包括来自一个以上专家的评估的情况下，训练图像数据中的一组帧(例如图3上的帧300_1至300_4)，可以通过以下方式用严重程度类别标签来注释214：如果两次独立的目视检查将包括所述组帧的训练结肠镜检查视频片段与第一范围内的内镜下严重程度得分相关联，则分配第一严重程度类别标签。换句话说，在实施方案中，训练图像数据中的一组帧(例如图3上的帧300_1至300_4)，可以通过以下方式用严重程度类别标签来注释214：如果包括所述组帧的训练结肠镜检查视频片段的严重程度信息包括第一范围内的两个独立的内镜下严重程度得分，则分配第一严重程度类别标签。在实施方案中，训练图像数据中的一组帧(例如图3上的帧300_1至300_4)，可以通过以下方式用严重程度类别标签来注释214：如果两次独立的目视检查将包括所述组帧的训练结肠镜检查视频300的片段与相同的内镜下严重程度得分相关联，并且所述内镜下严重程度得分在第一范围内，则分配第一严重程度类别标签S_1。换句话说，在实施方案中，训练图像数据中的一组

帧(例如图3上的帧300_1至300_4),可以通过以下方式用严重程度类别标签来注释214:如果包括所述组帧的训练结肠镜检查视频片段的严重程度信息包括第一范围内的两个独立的内镜下严重程度得分,并且这两个独立的内镜下严重程度得分相同,则分配第一严重程度类别标签。可以从训练数据中排除训练图像数据中不满足将用严重程度类别标签注释的判断标准(例如,因为严重程度信息不包括一致(即相同或在相同范围内)的两个独立的严重程度得分)的帧集。对于进一步的严重程度类别标签和对应的严重程度得分范围,可以实现类似的方法。例如,如果严重程度信息包括第二范围内的内窥镜评分(或两个独立的内窥镜评分),则可被分配第二严重程度类别标签S_2。在将训练二元分类器的具体情况中,训练图像数据中的各组帧300_1至300_n可以通过以下方式用严重程度类别标签来注释214:如果目视检查将包括该组帧的训练结肠镜检查视频片段与高于(或至少等于)阈值的内镜下严重程度得分相关联(即,如果严重程度信息包括高于阈值的严重程度得分),则分配第一严重程度类别标签S_1。类似地,训练图像数据中的一组帧,可以通过以下方式用严重程度类别标签来注释214:如果目视检查将包括所述组帧的训练结肠镜检查视频片段与等于或低于(或低于,如果类别S_1包括内镜下严重程度得分至少与阈值一样高的数据)所述阈值的内镜下严重程度得分相关联,则分配第二严重程度类别标签S_2。在这种情况下,也可以选择使用针对多个得分之间的一致性的判断标准,如上文所解释的。例如,在内镜下严重程度得分为MCES的情况下,有用的阈值可以是MCES=1或MCES=2,相当于第一范围MCES>1或MCES>2。可以实现类似的方法,其中分类器将被训练用于以三个或更多个严重程度类别对图像数据进行分类。具体地,训练图像数据中的一组帧,可以通过以下方式用严重程度类别标签来注释214:如果目视检查将包括所述组帧的训练结肠镜检查视频片段与针对所述三个或更多个严重程度类别中的各严重程度类别的预定不同(即,不重叠)范围内的内镜下严重程度得分相关联,则分配所述三个或更多个严重程度类别标签中的一个标签。在这种情况下,也可以选择使用针对多个独立得分之间的一致性的判断标准,如上文所解释的。

[0163] 在实施方案中,严重程度信息是离散类的形式(例如,MCES量表的四个级别中的一个),并且分类器将被训练用于以对应于严重程度信息中使用的不同离散类别的类别(例如,四个严重程度类别,每个严重程度类别对应于不同的MCES)对图像数据进行分类。虽然在这样的实施方案中不需要聚合来将严重程度信息中的离散类别转换成适合于训练分类器的类别,但是针对将形成训练数据的一部分的视频300的每个视频片段用严重程度类别标签S_1、S_2注释214用于训练的视频300的步骤在一些实施方案中仍可执行。例如,在严重程度信息包括例如由独立专家提供的多个得分的情况下,可以从严重程度信息导出单个类别标签。在一些这样的实施方案中,训练图像数据中的一组帧,可以通过以下方式用严重程度类别标签来注释214:如果两次独立的目视检查将包括该组帧的训练结肠镜检查视频片段与相同的第一(相应地,第二、第三等)内镜下严重程度得分相关联,则分配第一(相应地,第二、第三等)严重程度类别标签。

[0164] 在可选步骤216,可以针对将形成训练数据的一部分的视频300的每个片段,用质量类别标签Q_1、Q_2注释用于训练的原始结肠镜检查视频300。在分类器(优选地是深度神经网络,本文中也称为质量控制网络(QCN)或第二深度神经网络分类器)被训练220并使用230以从将用于训练基于严重程度的分类器(SSN)的训练数据中过滤低质量帧时,这可能是有利的。用质量类别标签Q_1、Q_2标注训练图像数据(或从先前注释的数据中提取质量类别标签)使

得这种分类器能够被训练。步骤216可以通过基于一个或多个判断标准为训练数据中的视频片段手动分配质量类别标签Q_1、Q_2来执行。有利地，这些判断标准可应用于视频300的目视检查。例如，如果可以基于目视检查区分结肠壁和结肠血管，则可以为训练结肠镜检查视频300的片段分配第一(质量好)质量类别标签Q_1，否则分配第二质量类别标签。可选地，如果训练结肠镜检查视频300的片段另外满足一个或多个基于水的存在或不存在、超反射区域、大便和/或模糊的目视检查判断标准，则其可被分配第一质量类别标签Q_1，否则被分配第二质量类别标签Q_2。目视检查可以是众包的，和/或可以由非专家执行。为训练结肠镜检查视频300分配质量类别标签Q_1、Q_2可以使用能够注释视频文件的任何方式来执行。此外，单个质量注释可能就足够了。如本领域技术人员所理解的，当逐片段地执行质量注释时，片段包括多个连续的帧，并且每个这样的帧将具有它所属于的片段的标签。在执行多次独立质量注释的情况下，可以使用任何适当的方案在逐帧的基础上组合其结果，例如，为帧分配在所有独立质量注释中出现最多的质量类别标签、分配在所有独立质量注释中最低质量类别的标签、分配在所有独立质量注释中出现次数高于阈值的最低质量类别的标签等。如图3所示，通常不要求用相同质量类别标签Q_1、Q_2注释的片段对应用相同的严重程度类别标签S_1、S_2或解剖学标签A_1、A_2注释的片段。相比之下，具有相同解剖学标签A_1、A_2的训练视频300的片段可能具有相同的严重程度类别标签S_1、S_2，因为严重程度信息通常是在临床环境中以解剖学区段为基础提供的(例如，每个解剖学区段一个严重程度得分)。在其他实施方案中，可以针对将形成训练数据的一部分的视频300的每个片段或每个单独帧，用质量类别标签Q_1、Q_2自动注释用于训练的原始结肠镜检查视频300。在这样的实施方案中，可以使用先前训练的分类器(例如，深度神经网络分类器)来执行注释。

[0165] 作为步骤210的结果，获得包括来自多个结肠镜检查视频的多组帧300_1至300_n的训练图像数据，每一帧300_1至300_n与严重程度类别标签S_1、S_2相关联，并且可选地与质量类别标签Q_1、Q_2和/或解剖学标签A_1、A_2相关联。当训练数据中存在质量类别标签Q_1、Q_2时，可以在提供如本文所述的工具的方法中执行可选步骤220、320和230、330。在步骤220中，深度神经网络分类器(称为质量控制网络，QCN)可以被训练220以将帧分类到对应的质量类别中。随后可以使用230所述QCN来过滤图像数据(无论是在提供如本文所述的工具的方法中的训练图像数据，还是来自结肠镜检查视频的用于分析的数据)。训练220以弱监督的方式执行，因为质量类别标签Q_1、Q_2基于片段级注释被分配给帧和/或使用先前训练的分类器自动分配给片段或各个帧。因此，这些帧-质量类别配对不代表地面真值信息，因为与每个特定帧的质量类别分配相关的不确定性水平相对较高。实际上，并非片段中的所有帧都预期显示导致将质量标签分配给包含该帧的片段的特征，和/或任何先前训练的分类器都预期在分类先前未见的数据时准确度低于100％。一旦经过训练，在图像数据被严重程度评分网络分类之前，所述QCN可用于230、330对图像数据进行过滤，如下文将进一步描述的那样。具体地，在图3所示的实施方案中，QCN接受训练或分析视频300的各个帧300_1至300_n作为输入，并产生所述帧属于第一类别的概率$P(Q_1)^1$-$P(Q_1)^n$和/或所述帧属于第二类别的概率作为输出330A。离散分类标签330B是由此导出的，以便决定哪些帧属于质量较低的类别并且应该被过滤掉，如勾号和叉号所示。实际上，离散分类标签330B可以仅通过排除不满足用于确定哪些帧属于较高质量类别的判断标准的帧来隐式导出。例如，如果帧(例如帧300_1)属于第一质量类别的概率$P(Q_1)^1$达到或超过阈值，则所述帧可以被认为被QCN分类在第一质量类别

中，否则分类在第二质量类别中。阈值可以先验地或动态地选择(即，取决于QCN和数据的特定实例)。例如，可以动态地确定阈值，使得训练图像数据中的帧组(即，已被分配相同严重程度类别和可选地相同解剖学区段标签的连续帧组)平均包含20至40个、优选地约30个被分类在第一质量类别中的帧。阈值在0.9至0.99、优选0.95之间是合适的。

[0166] 在步骤240/340处，使用(可选地经过质量过滤的)训练数据来训练基于严重程度的深度神经网络(SSN，本文中也称为第一深度神经网络分类器)用之前获取的严重程度类别标签S_1、S_2将数据分类到各严重程度类别。训练240/340是以弱监督的方式执行的，因为严重程度类别标签S_1、S_2基于片段级注释分配给帧。同样地，这些帧-严重程度类别配对不表示地面真值信息，因为并非片段中的所有帧都预期会显示导致将严重程度类别标签分配给包含该帧的片段的特征。训练后的SSN可用于250/350分析结肠镜检查视频。SSN接受各个帧300_1-300_n作为输入，并产生242/252被分析的每一帧的严重程度类别预测结果作为输出。具体地，在图3所示的实施方案中，SSN接受训练或分析视频300的各个帧300_1-300_n(或各个帧300_1-300_3和300_{n-1}-300_n，如果使用基于质量的过滤)作为输入，并产生所述帧属于第一类别的概率$P(S_1)^1$至$P(S_1)^n$和/或所述帧属于第二类别的概率作为输出340A/350A。为了在训练期间评估SSN的性能，可从这些概率导出离散分类标签340B/350B，并在逐帧的基础上与严重程度类别注释进行比较246。离散分类标签340B/350B可以通过对由SSN输出的一个或多个概率应用阈值来获取。例如，如果帧(例如帧300_1)属于第一类别的概率(例如$P(S_1)^1$)高于阈值，则所述帧可以分配给第一类别。否则，所述帧可以被分配给第二类别，或者如果所述帧在第一类别中的概率在一个范围内，通常是大约0.5的范围内，则可以被认为具有未知的分类。类似地，对于多级分类器，帧可以被分配给具有最高概率的类别，前提是所述概率超过了阈值。或者，不管概率的实际值如何，都可以将帧分配给具有最高概率的类别。

[0167] 可另外实现可选步骤244和/或254，其中基于对构成视频片段(并且已经用SSN分析)的每个帧的预测结果(340A/350A或340B/350B)，获取视频片段的总体严重程度类别预测结果340C/350C。在图3所示的示例中，基于解剖学区段标签A_1将属于同一片段的帧分组在一起。由于训练数据的严重程度信息被分配给每个解剖学区段，在步骤246中，总体严重程度类别预测340C可以直接与原始严重程度信息进行比较，以便在训练240期间评估SSN的性能。此外，为了分析结肠镜检查视频的目的，获取254每个解剖学片段350C的总体严重程度类别预测也是有用的，因为它复制了常规临床评估中使用的粒度。可选地，为了分析结肠镜检查视频的目的，可以自动确定基于其获取总体严重程度类别预测结果340C/350C的片段，优选地不需要有关数据的先验信息。例如，片段可以被定义为一组连续帧，其中所有帧具有相同的离散类别分配350B，或者不具有离散类别分配350B(即，它是未知的或者该帧由于例如质量原因被排除)。

[0168] 可直接基于SSN输出的概率340A/350A或基于从这些概率导出的离散类别分配340B/350B来获取244/254包括多个帧的片段的总体严重程度类别340C/350C。在实施方案中，如果SSN输出的帧属于第一严重程度类别的概率340A/350A的平均值高于阈值，则通过将第一严重程度类别分配给片段，可以获取244/254片段的总体严重程度类别340C/350C。在图3所示实施方案中，如果SSN输出的帧300_1至300_3属于第一严重程度类别的概率($P(S_1)^1$至$P(S_1)^3$)340A/350A的平均值高于阈值，则通过将第一严重程度类别分配给片段，可以获取244/254包括帧300_1至300_3的片段的总体严重程度类别340C/350C。所述阈值可以是固定的。

可选地，阈值可基于从以下中选择的一个或多个判断标准来动态确定：预定可接受的假阳性比例、预定可接受的假阴性比例、预定最小准确度、预定最小准确度和/或预定最小查全率。阈值在0.5到0.9之间被认为是合适的。作为另一示例，如果分配给第一严重程度类别的帧的比例高于阈值，则通过将第一严重程度类别分配给片段，可以获取244/254片段的总体严重程度类别340C/350C。在图3所示的示例中，假设阈值为0.5，包括帧300_1至300_3的片段的总体严重程度类别340C/350C被分配为第二严重程度类别，因为帧300_1至300_3中的66％被分配给了第一严重程度类别。相比之下，包括帧300_{n-1}至300_n的片段的总体严重程度类别340C/350C被分配为第二严重程度类别，因为帧300_{n-1}至300_n中的50％被分配给了第一严重程度类别。可以使用其他方案，例如，在使用两个以上的类别的情况下。例如，通过将在片段的帧的离散类别分配340B/350C中出现次数高于相应严重程度类别阈值的最高严重程度类别(即，最严重程度类别)分配给片段，可获取片段244/254的总体严重程度类别340C/350C。对于较高的严重程度类别，阈值可以较低(即，要求被分配离散严重程度类别标签340B/350B的片段中较低比例的帧)。实际上，更高的严重程度类别更容易识别(因此，在更高的严重程度类别中的预测可能具有更高的置信度)。

[0169] 这里描述的所有深度神经网络分类器优选地是卷积神经网络(CNN)。有利地，所使用的CNN可以是已经在不相关的图像数据(例如来自ImageNet数据库(http://www.image-net.org)的图像数据)上预先训练过的。本发明的发明人已经发现50层CNN对于本用途是足够的，但是设想了包括例如附加层的替代实现。使用深度残差学习框架训练的CNN(He等人，Deep Residual Learning for Image Recognition,2015,arXiv:1512.03385，可在https://arxiv.org/pdf/1512.03385.pdf获取，并通过引用并入本文)已被发现是特别合适的。

[0170] 在步骤246中，将来自SSN的预测结果与对应的严重程度类别标签(从训练数据中的严重程度信息导出)进行比较，以便评估SSN的性能。在实施方案中，评估SSN(第一深度神经网络的性能)包括使用验证图像数据量化接收器工作特性曲线(AUC)下的面积。在实施方案中，评估SSN的性能包括使用验证图像数据计算Cohen's kappa。验证和训练图像数据可以形成相同数据的一部分，并且特别地具有如上所述的所有相同特征。这方面的特定示例是通过使用训练图像数据执行交叉验证来评估SSN，例如5折或10折交叉验证。在实施方案中，评估第一深度神经网络的性能包括使用训练图像数据执行5折或10折交叉验证，以及针对交叉验证的每个分裂量化AUC或Cohen's kappa。优选地，训练图像数据被分离成多个分裂以用于交叉验证，其中来自同一个结肠镜检查视频的各组帧不会出现在一个以上的分裂中。

[0171] 例如，SSN的输出可以使用显示器输出给用户，所述输出包括例如总体严重程度类别350C和/或每一帧的预测结果(属于一个或多个等级的概率350A或离散的等级分配350B)。这一信息可能有助于评估受试者溃疡性结肠炎的严重程度，特别是当SSN的输出与内镜下严重程度得分或评分范围(如MCES)相对应时。

[0172] 一种评估受试者溃疡性结肠炎严重程度的方法现在将通过参考图4来描述。如参照图2和图3所述的，分析例如经由用户接口从用户或从数据库202获取的来自受试者的结肠镜检查视频400(或其部分，例如，示出单个解剖学区段的部分、不包括所有原始数据以缩减其大小的部分等)。具体地，来自结肠镜检查视频400的图像数据(即结肠镜检查视频400

的所有帧)可选地被提供给QCN,以通过以至少第一质量类别和第二质量类别对每一帧进行分类来对帧应用430基于质量的过滤器,其中仅第一质量类别中的帧用于严重程度评估。然后,使用SSN对(可选地经过滤的)图像数据435(即,结肠镜检查视频400的所有或一些帧,取决于是否使用基于质量的过滤)评估溃疡性结肠炎的严重程度450A/450B,所述SSN将数据435分类为至少第一严重程度类别和第二严重程度类别,所述第一严重程度类别所关联的溃疡性结肠炎比所述第二严重程度类别的更严重。在图4所示的实施方案中,执行图像数据435的两次单独的严重程度评估450A/450B:在评估450A中使用第一SSN,以在对应于MCES>1的第一严重程度类别和对应于MCES≤1的第二严重程度类别之间对图像数据进行分类。在评估450B中使用第二SSN,以在对应于MCES>2的第一严重程度类别和对应于MCES≤2的第二严重程度类别之间对图像数据进行分类。其他配置也是可能的和可以设想的,例如使用具有不同SSN的两个以上单独的严重程度评估的配置、使用多类SSN的单个严重程度评估、使用SSN将数据分类为对应于一个或多个不同内镜下评分方案(包括但不限于MCES)的值的类的一个或多个单独的严重程度评估等。每个SSN产生数据435的每一帧或每个片段(数据中的帧集)属于第一严重程度类别的概率,或从这些概率导出的离散严重程度类别作为输出455A/455B。在步骤460/470中,这些输出455A/455B用于决定受试者的适当治疗过程。具体地,在步骤460,确定来自第二评估450B的输出455B是否指示MCES>2。例如,如果从对数据435的任何片段的SSN2的预测结果导出的总体类别预测是第一严重程度类别,那么来自第二评估450B的输出455B可被认为指示MCES>2。如果是这种情况,则选择465受试者进行第一治疗。如果不是这种情况,则在步骤470确定来自第一评估450A的输出455A是否指示MCES>1。例如,如果从对数据435的任何片段的SSN1的预测结果导出的总体类别预测是第一严重程度类别,那么来自第一评估450A的输出455A可被认为指示MCES>1。如果是这种情况,则选择475受试者进行第二治疗。选择受试者进行治疗可指推荐治疗和/或将该受试者识别为可能受益于该治疗和/或将该受试者识别为可能需要特定治疗。任选地,所述方法可以进一步包括根据所选择的治疗来治疗所述受试者。

[0173] 在其他实施方案中,可以执行单次严重程度评估,例如使用如图4所示的第二SNN(SSN2)。在这样的实施方案中,如果来自该评估的输出指示第一严重程度类别,则所述方法可以包括针对UC治疗该受试者,或者选择该受试者进行针对UC的治疗(即,推荐治疗,将该受试者标识为可能受益于治疗和/或需要治疗的受试者)。这可以是例如SSN将来自图像数据400的一组帧分类在第一严重程度类别中的情况,或者如果来自图像数据400的任何一组帧具有第一总体严重程度类别的情况。

[0174] 如本领域技术人员所理解的,使用深度神经网络来分类图像数据(基于严重程度或质量)的表述实际上可以包括使用多个深度神经网络并组合多个深度神经网络的预测结果。这样的多个深度神经网络中的每一个可以具有这里描述的性质。类似地,对训练深度神经网络的引用实际上可以包括如本文所述的多个深度神经网络的训练,其中的一些或全部随后可用于根据质量或严重程度来分类图像数据,这要视情况而定。

[0175] 实施例

[0176] 现在将描述提供用于分析结肠镜检查视频的工具的示例性方法。从HICKORY(NCT02100696)和LAUREL(NCT02165215)临床试验中获取了2000多个结肠镜检查视频--两者都是III期、双盲、安慰剂对照、多中心研究,研究依曲利组单抗(etrolizumab)治疗中重

度活动性溃疡性结肠炎(UC)的疗效和安全性。作为临床试验的一部分,每个视频都由胃肠病专家注释,以表明:(1)解剖学区段(直肠、乙状结肠、降结肠),(2)来自两个不同的识读者的对每个解剖学区段的MCES评估结果。通过筛选出两个读者对每个解剖学区段的MCES意见不一致的视频,以及读者标记质量问题的视频,如肠道准备不当、视频质量不佳等,共选择104个原始结肠镜检查视频。

[0177] 由非专家对104个原始结肠镜检查视频中的每一个进行手动质量注释,他们被要求定义被认为是"质量好"或"质量差"的视频片段。这是根据以下判断标准执行的:(i)相机离结肠壁足够远,以允许进行适当的评估,(ii)结肠壁及其血管可以通过目视检查进行评估,以及(iii)视觉伪影不存在或不占帧的约10%。观察到的视觉伪影包括:水,高反射区,大便,模糊。在实践中,这是由观看视频的非专家执行的,并在使用ELAN(https://tla.mpi.nl/tools/tla-tools/elan/;Brugman,H.,Russel,A.(2004).Annotating Multimedia/Multi-modal resources with ELAN.In:Proceedings of LREC 2004,Fourth International Conference on Language Resources and Evaluation)观察时突出标记质量好/差的片段。

[0178] 在这些临床试验中,解剖学区段注释作为图形标签包含在每个原始视频的每一帧上。训练深度神经网络(解剖学区段网络),通过聚焦于包括图形标签的帧的区域,将每个视频的每一帧分类到三个解剖学区段类别(直肠、乙状结肠、降结肠)中的一个。这是通过使用Keras(https://keras.io/)部分地再训练50层卷积神经网络ResNet50来实现的。特别地,使用Keras中实现的随机梯度下降(SGD)优化器对ResNet50的最后8层进行再训练。使用的学习率为0.001,动量为0.9。训练后的解剖学区段网络能够为每个视频的每一帧分配一个高置信度的解剖学区段。这个过程的结果是,对于104个视频中的每一个视频的每一帧,以下3个注释是可用的:解剖学区段(来自解剖学区段网络),质量类别(来自非专家片段级注释),MCES(来自专家片段级注释)。质量类别和MCES是弱标签,至少是因为它们是相对于整个片段提供的,其中形成一个片段的多个帧不太可能全部显示导致分配所述标签的视觉特征。特别地,对于MCES评分,结肠的解剖学区段被分配与区段中看到的最严重病变相对应的得分。换句话说,显示乙状结肠的整个视频片段将被分配MCES=3,如果两个读者都看到在这个解剖学区段的任何地方有中度疾病活动的迹象(明显的红斑、缺乏血管模式、质脆、糜烂)。然而,本区段中的有些帧可能没有显示这些迹象。根据两种不同的方案,将MCES得分转换为二元严重程度类别。第一方案将严重程度类别标签1分配给来自读取器的MCES评分>1的片段,否则将严重程度类别标签2分配给读取器的MCES评分>1的片段。第二个方案将严重程度类别标签2分配给来自读者的MCES得分>2的片段,否则为严重程度类别标签2

[0179] 所有104个视频每秒包括24帧,所有帧均被用于了质量控制网络(QCN)的训练。训练一个深度神经网络(质量控制网络)将104个原始视频中的每一帧分类为质量好的类别和质量差的类别。特别地,质量控制网络被训练为每一帧提供该帧属于"质量好"类别的概率。这是通过使用如上所述的Keras部分地再训练50层卷积神经网络ResNet50来实现的。特别地,最后8层使用SGD进行再训练,学习率为0.001,动量为0.9。如果属于"质量好"类别的帧的预测概率(P(好))超过0.97,则认为帧被归类为"质量好"。这个阈值导致每个原始结肠镜检查视频选择大约20帧。阈值为0.95,导致每个原始结肠镜检查视频选择约30帧(总共约9360帧)也进行了测试,结果相似(未显示)。此外,训练后的QCN的AUC为0.93±0.05。

[0180] 总共选择了大约6200个预测为“质量好”的帧(根据P(好)>0.97的阈值)。所有这些帧分别用于训练两个深度神经网络(严重程度评分网络,SSN):第一SSN使用根据上述第一二元方案(MCES>1,MCES≤1)的二元严重程度分类标签,第二SSN使用根据上述第二二元方案(MCES>2,MCES≤2)的二元严重程度分类标签。结果,第一SSN被训练用于为每一帧输出其属于第一严重程度类别MCES>1的概率,P(MCES>1)。类似地,第二SSN被训练用于为每一帧输出其属于第一严重程度类别MCES>2的概率,P(MCES>1)。两个SSN都是通过使用上面解释的方法部分再训练50层卷积神经网络ResNet50来训练的。当P(MCES>1)>0.5和P(MCES>2)>0.5时,第一和第二严重程度评分网络分别将一帧分配为第一严重程度类别。通过计算来自同一视频Y的同一解剖学区段A的所有帧的类1的平均概率,为每个视频的每个解剖学区段计算概要分类(使用来自解剖学区段网络的解剖学区段标签)。如果平均,则第一和第二严重程度评分网络将一片段视为被分配第一严重程度类别标签($P_{A,Y}$(MCES>1))>0.5,平均值($P_{A,Y}$(MCES>2))>0.5。

[0181] 通过使用用于训练的相同数据进行五折交叉验证,对两个SSN进行回顾性评估。具体地,根据以下方案将大约6200个质量选定帧分为训练集、调谐集和验证集:60%、20%、20%,附加的规则是来自同一病人视频的帧只出现在其中一组中。通过改变用于对第一严重程度类别中的片段进行分类的阈值,比较预测的类别分配和专家得出的类别分配(根据上述二元方案从专家注释导出的二元严重程度分类标签),并计算对应的假阳性率和假阴性率,在5折验证的每个迭代中计算ROC曲线。也为每条ROC曲线计算了相应的曲线下面积(AUC)。然后计算平均ROC曲线和对应的标准差,以及平均ROC曲线的AUC。

[0182] 这些分析的结果显示在图5A及5B中,分别针对第一SSN和第二SSN。这些图显示了针对每个交叉验证迭代(细曲线)获取的ROC曲线(真正作为假正的函数)、这些迭代的平均值(粗曲线)和±1标准差包络(灰色区域)。所示的ROC曲线对应于在0到1之间、以0.05为步长递增的上述阈值中的每个值的TPR和FPR值。针对第一SSN,AUC(平均值±标准差)为0.76±0.08。针对第二SSN,AUC(平均值±标准差)为0.81±0.06。这一数据表明,完全自动化的结肠镜检查视频分析方法能够使用原始内窥镜视频作为唯一的输入来提供对单个结肠区段的MCES值的准确预测。尽管训练是使用弱标签和相对少量的训练数据(104个原始视频)来执行的,但这是存在的。此外,在临床试验中,MCES≤1通常被归类为“缓解”,因此一种用于识别那些显示受试者处于缓解状态的结肠镜检查视频的值得信赖的自动方法可能特别有价值。

[0183] 如上所述,对由大约1000个视频组成的较大数据集进行了质量控制。所有通过质量控制检查的帧及其原始的MCES(0至3)注释被用来训练进一步的SSN。因此,SSN被训练输出每一帧属于对应于MCES量表的四个级别的四个类别之一的相应概率。在一些示例中,使用了如在Cao等人(Rank-consistent Ordinal Regression for Neural Networks,2019,arXiv:1901.07884v4)中所描述的有序分类模型。然后计算片段中各个帧的对应的平均概率。单个预测的MCES得分被分配给每个片段作为具有最高平均概率的MCES得分。这种SSN通过如上所述的交叉验证来评估,除了不计算ROC曲线和AUC,而是为交叉验证的每个迭代计算Cohen’s kappa系数。

[0184] 术语“计算机系统”包括用于根据上述实施方案具体化系统或执行方法的硬件、软件和数据存储设备。例如,计算机系统可以包括中央处理单元(CPU)、输入设备、输出设备和

数据存储，其可以实施为一个或多个连接的计算设备。优选地，计算机系统具有显示器或包括计算设备，所述计算设备具有用于提供视觉输出显示的显示器(例如，在商业过程的设计中)。数据存储器可以包含RAM、磁盘驱动器或其他计算机可读介质。计算机系统可以包括多个计算设备，其通过网络连接并且能够经由该网络相互通信。

[0185] 上文实施方案的方法可以作为计算机程序或作为计算机程序产品或携带计算机程序的计算机可读介质来提供，所述计算机程序在计算机上运行时配置为执行上述一种或多种方法。

[0186] 术语“计算机可读介质”包括但不限于任何一种或多种非暂时性介质，其可以由计算机或计算机系统直接读取和访问。所述介质可以包括但不限于磁性存储介质，如软盘、硬盘存储介质和磁带；光学存储介质，如光盘或CD-ROM；电存储介质，如存储器，包括RAM、ROM和闪速存储器；以及上述存储介质的混合物和组合，如磁性/光学存储介质。

[0187] 除非上下文另有规定，否则以上阐述的特征的描述和定义不限于本发明的任何特定方面或实施方案并且同样适用于所描述的所有方面和实施方案。

[0188] “和/或”在本文使用时，将被视为两个指定特征或组分中的每个与另一个或不与另一个一起的具体公开。例如，“A和/或B”将被视为(i)A、(ii)B和(iii)A和B中的每个的具体公开，如同每个在本文中单独阐述一样。

[0189] 必须注意的是，如在说明书和所附权利要求中所使用的，单数形式“一”、“一个”和“所述”包括复数对象，除非上下文另有明确规定。范围在本文可以表示为从“大约”一个特定值，和/或到“大约”另一个特定值。当表达这样的范围时，另一个实施方案包括从一个特定值和/或到另一个特定值。类似地，当通过使用先行词“大约”将值表示为近似值时，将理解特定值形成另一个实施方案。与数值相关的术语“大约”是可选的，并且例如表示+/-10％。

[0190] 在整个说明书中，包括随后的权利要求，除非上下文另有要求，否则词语“包括”和“包含”，以及诸如“其包括”、“包含了”和“包括了”的变化将被理解为暗示包括所述的整体或步骤或一组整体或步骤，但不排除任何其他整体或步骤或其他组整体或步骤。

[0191] 本发明的其他方面和实施方案提供了上面描述的方面和实施方案，其中术语“包括”被术语“由……组成”或“基本上由……组成”所取代，除非上下文另有规定。

[0192] 在前面的描述中、或者在下面的权利要求中、或者在附图中公开的以它们的特定形式或者根据用于执行所公开的功能的手段、或者用于获取所公开的结果的方法或过程来表达的特征在适当情况下可以单独地或者以这些特征的任意组合被用于以不同的形式实现本发明。

[0193] 虽然已经结合上述示例性实施方案描述了本发明，但是当给出本公开文本时，许多等同的修改和变化对于本领域技术人员来说是显而易见的。因此，上述本发明的示例性实施方案被认为是说明性的而非限制性的。在不脱离本发明的精神和范围的情况下，可以对所描述的实施方案进行各种改变。

[0194] 为了避免任何疑问，本文提供的任何理论解释都是为了改善读者的理解。本发明的发明人不希望受任何这些理论解释的约束。

[0195] 本文使用的任何章节标题仅用于组织目的，不应被解释为限制所描述的主题。

[0196] 本说明书中提及的所有文献通过引用以其整体并入本文。

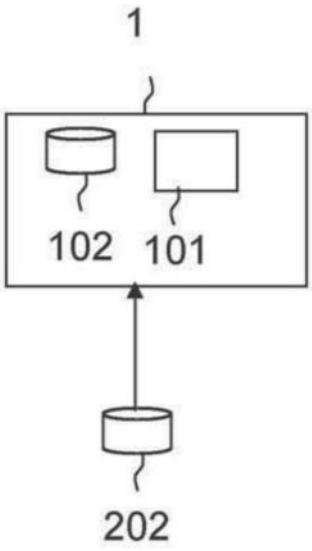


图1

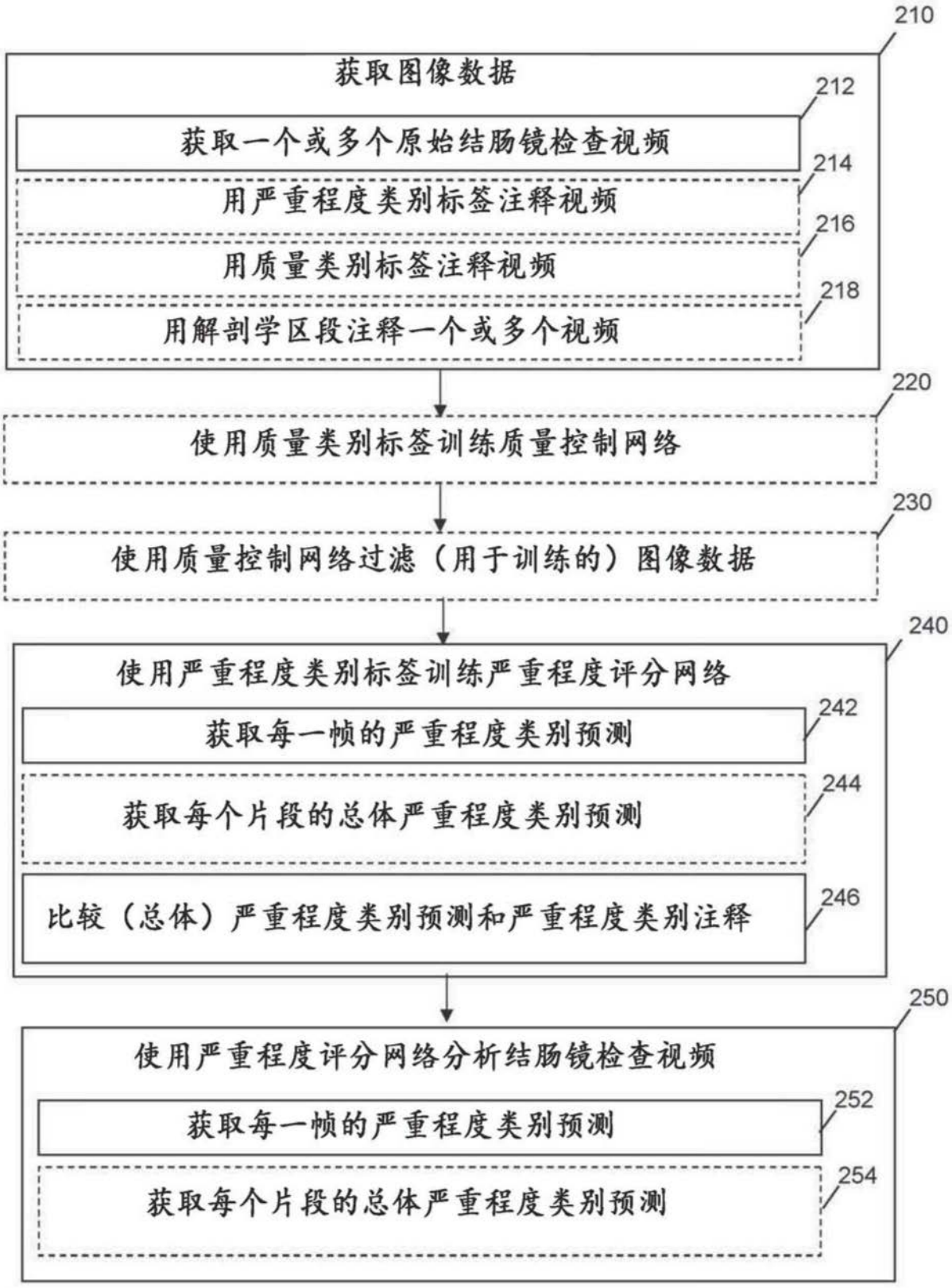

210
获取图像数据
212
获取一个或多个原始结肠镜检查视频
214
用严重程度类别标签注释视频
216
用质量类别标签注释视频
218
用解剖学区段注释一个或多个视频
220
使用质量类别标签训练质量控制网络
230
使用质量控制网络过滤（用于训练的）图像数据
240
使用严重程度类别标签训练严重程度评分网络
242
获取每一帧的严重程度类别预测
244
获取每个片段的总体严重程度类别预测
246
比较（总体）严重程度类别预测和严重程度类别注释
250
使用严重程度评分网络分析结肠镜检查视频
252
获取每一帧的严重程度类别预测
254
获取每个片段的总体严重程度类别预测

图2

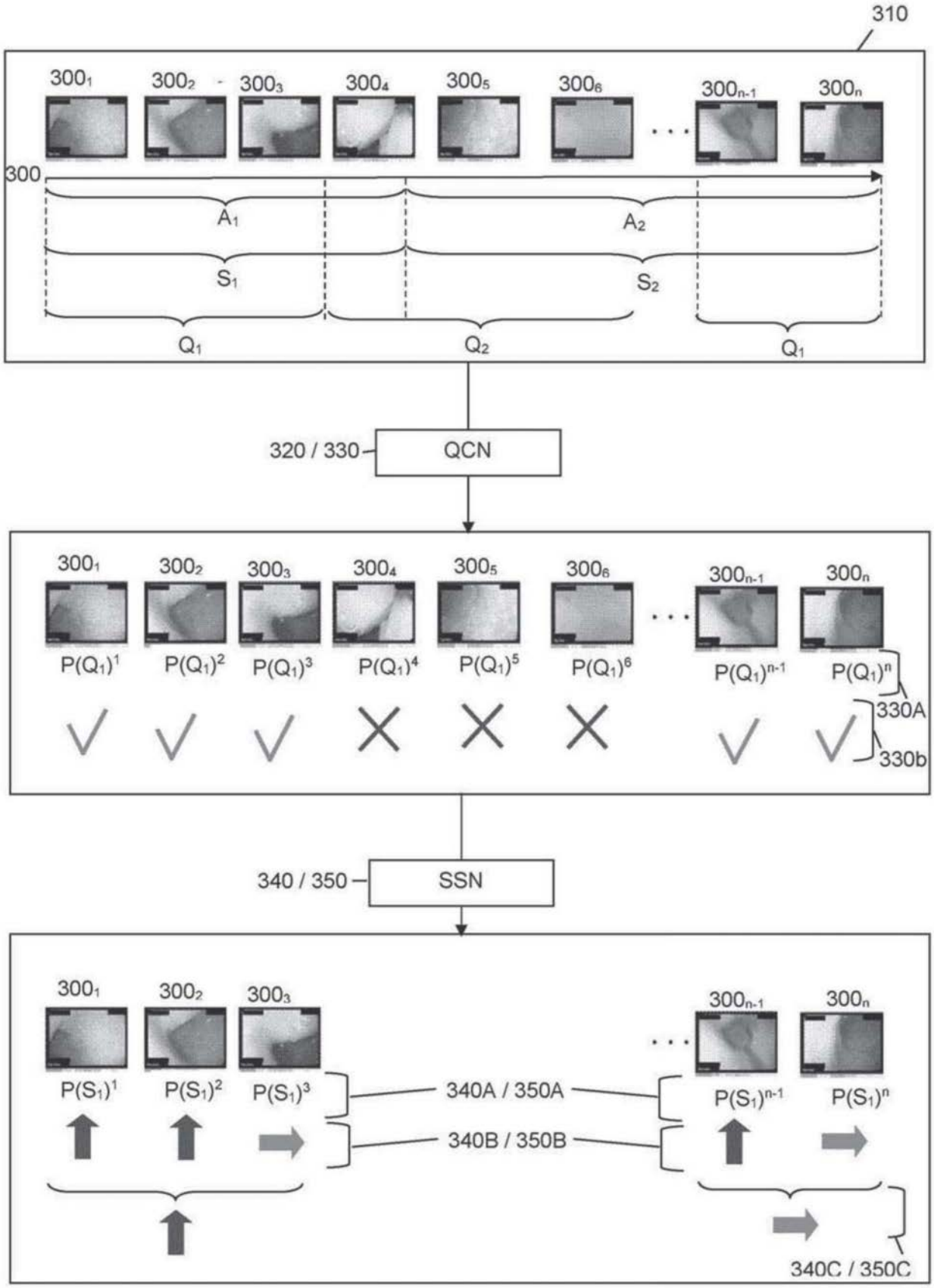

310
300
300_1 300_2 300_3 300_4 300_5 300_6 300_{n-1} 300_n
A_1 A_2
S_1 S_2
Q_1 Q_2 Q_1
320 / 330
QCN
$P(Q_1)^1$ $P(Q_1)^2$ $P(Q_1)^3$ $P(Q_1)^4$ $P(Q_1)^5$ $P(Q_1)^6$ $P(Q_1)^{n-1}$ $P(Q_1)^n$
330A
330b
340 / 350
SSN
$P(S_1)^1$ $P(S_1)^2$ $P(S_1)^3$ $P(S_1)^{n-1}$ $P(S_1)^n$
340A / 350A
340B / 350B
340C / 350C

图3

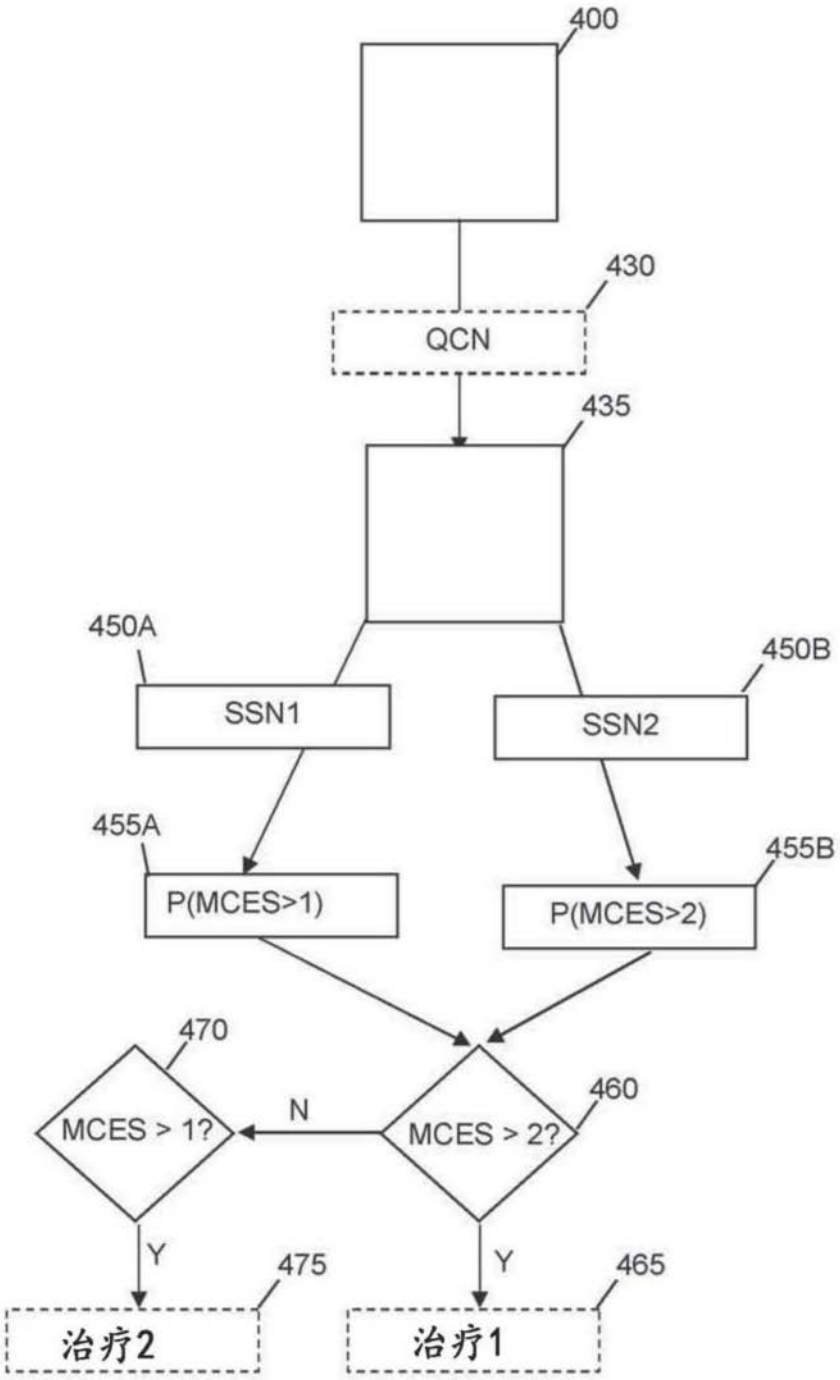

400
430
QCN
435
450A
SSN1
450B
SSN2
455A
P(MCES>1)
455B
P(MCES>2)
470
MCES > 1?
N
460
MCES > 2?
Y
475
治疗2
Y
465
治疗1

图4

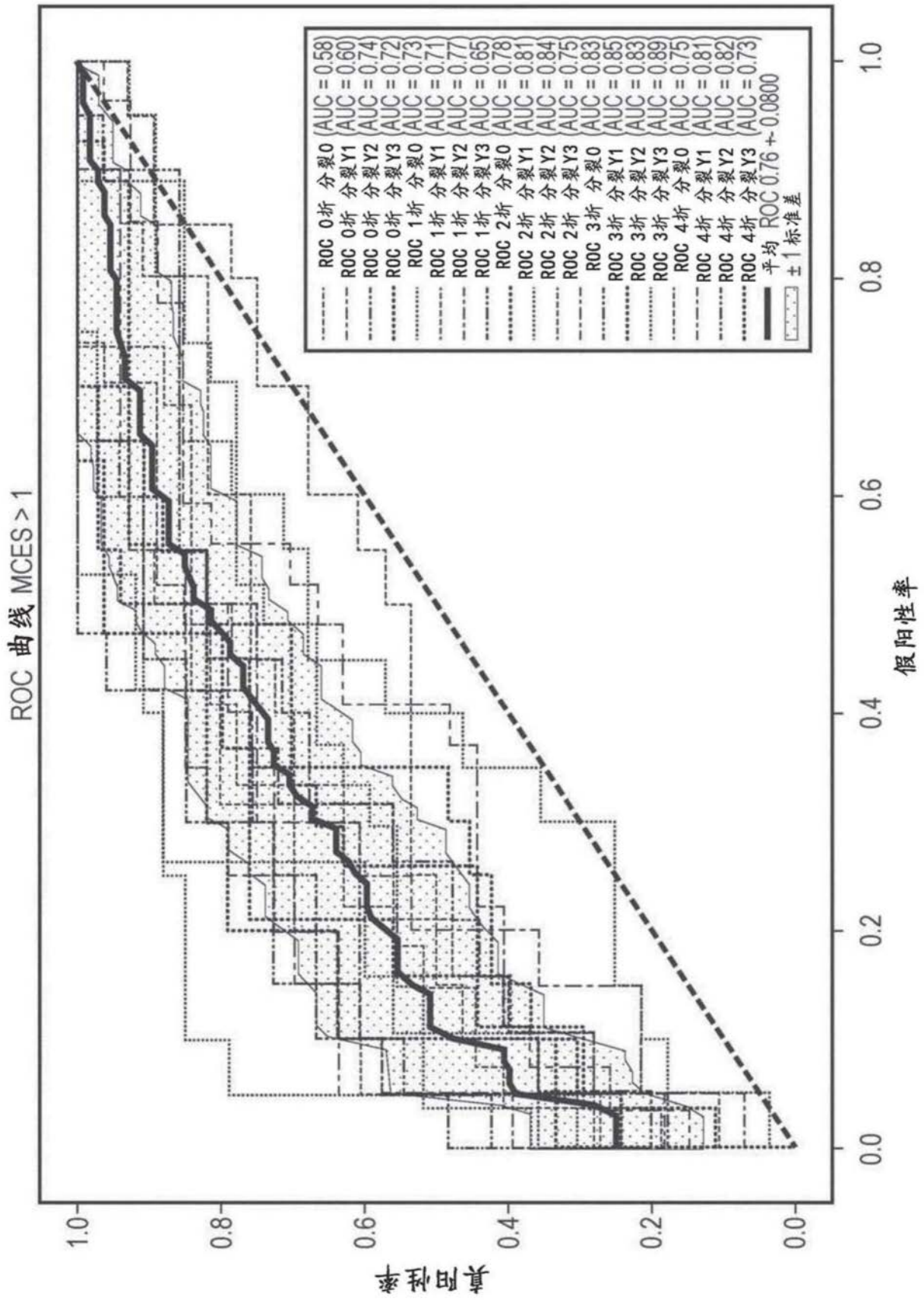

ROC 曲线 MCES > 1
真阳性率
假阳性率
0.0
0.2
0.4
0.6
0.8
1.0
ROC 0折 分裂0 (AUC = 0.58)
ROC 0折 分裂Y1 (AUC = 0.60)
ROC 0折 分裂Y2 (AUC = 0.74)
ROC 0折 分裂Y3 (AUC = 0.72)
ROC 1折 分裂0 (AUC = 0.73)
ROC 1折 分裂Y1 (AUC = 0.71)
ROC 1折 分裂Y2 (AUC = 0.77)
ROC 1折 分裂Y3 (AUC = 0.65)
ROC 2折 分裂0 (AUC = 0.78)
ROC 2折 分裂Y1 (AUC = 0.81)
ROC 2折 分裂Y2 (AUC = 0.84)
ROC 2折 分裂Y3 (AUC = 0.75)
ROC 3折 分裂0 (AUC = 0.83)
ROC 3折 分裂Y1 (AUC = 0.85)
ROC 3折 分裂Y2 (AUC = 0.83)
ROC 3折 分裂Y3 (AUC = 0.89)
ROC 4折 分裂0 (AUC = 0.75)
ROC 4折 分裂Y1 (AUC = 0.81)
ROC 4折 分裂Y2 (AUC = 0.82)
ROC 4折 分裂Y3 (AUC = 0.73)
平均 ROC 0.76 +- 0.0800
±1 标准差

图5A

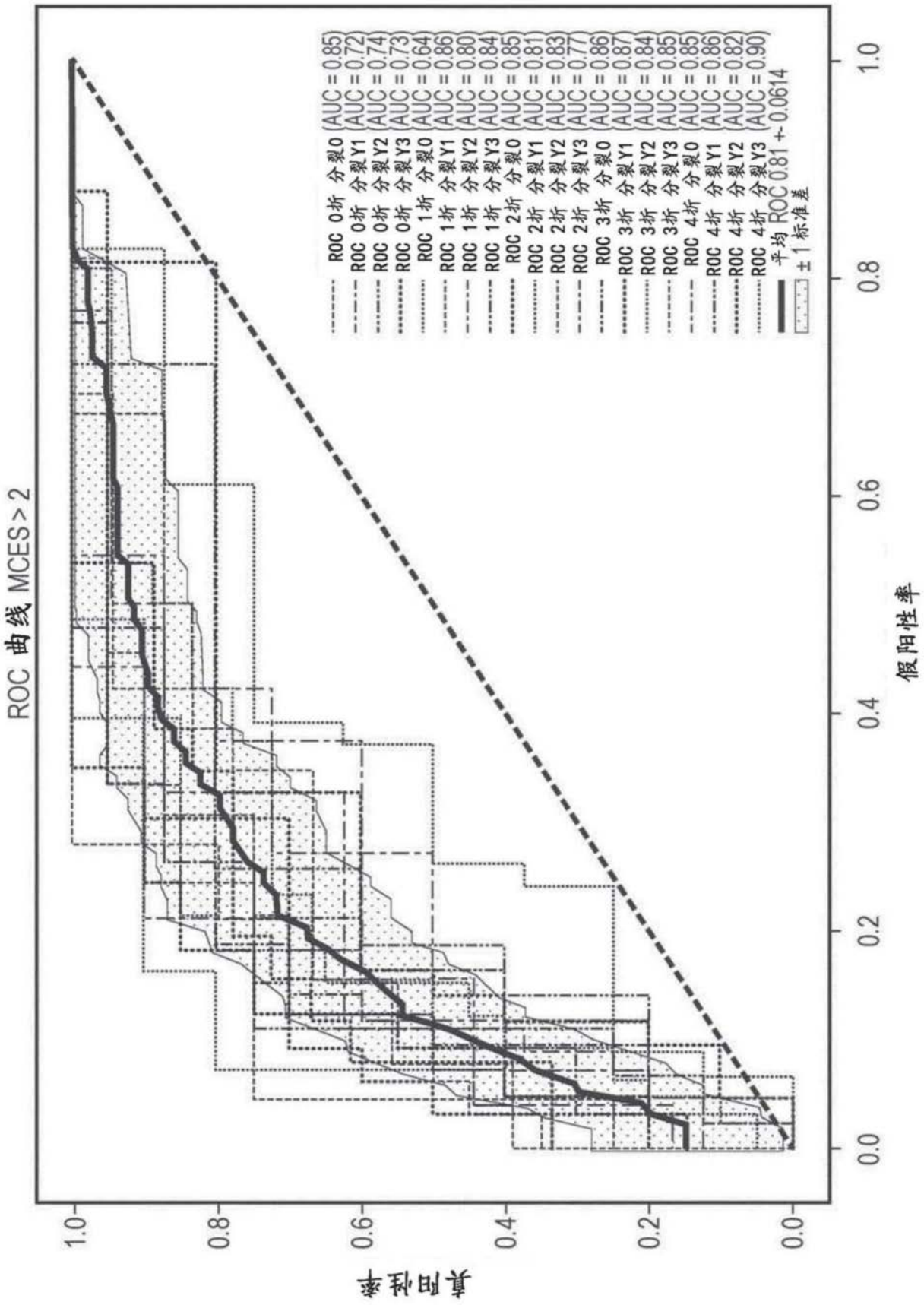

ROC 曲线 MCES > 2
真阳性率
假阳性率
0.0
0.2
0.4
0.6
0.8
1.0
ROC 0折 分裂0 (AUC = 0.85)
ROC 0折 分裂Y1 (AUC = 0.72)
ROC 0折 分裂Y2 (AUC = 0.74)
ROC 0折 分裂Y3 (AUC = 0.73)
ROC 1折 分裂0 (AUC = 0.64)
ROC 1折 分裂Y1 (AUC = 0.86)
ROC 1折 分裂Y2 (AUC = 0.80)
ROC 1折 分裂Y3 (AUC = 0.84)
ROC 2折 分裂0 (AUC = 0.85)
ROC 2折 分裂Y1 (AUC = 0.81)
ROC 2折 分裂Y2 (AUC = 0.83)
ROC 2折 分裂Y3 (AUC = 0.77)
ROC 3折 分裂0 (AUC = 0.86)
ROC 3折 分裂Y1 (AUC = 0.87)
ROC 3折 分裂Y2 (AUC = 0.84)
ROC 3折 分裂Y3 (AUC = 0.85)
ROC 4折 分裂0 (AUC = 0.85)
ROC 4折 分裂Y1 (AUC = 0.86)
ROC 4折 分裂Y2 (AUC = 0.82)
ROC 4折 分裂Y3 (AUC = 0.90)
平均 ROC 0.81 +- 0.0614
±1 标准差

图5B