



(12) 发明专利申请

(10) 申请公布号 CN 118591826 A

(43) 申请公布日 2024. 09. 03

(21) 申请号 202280085313.4

P • 瓦洛顿

(22) 申请日 2022.12.23

(74) 专利代理机构 北京市柳沈律师事务所

(30) 优先权数据

11105

21217712.5 2021.12.24 EP

专利代理师 张文辉

(85) PCT国际申请进入国家阶段日

(51) Int.Cl.

2024.06.21

G06V 10/25 (2006.01)

(86) PCT国际申请的申请数据

G06V 10/20 (2006.01)

PCT/EP2022/087779 2022.12.23

G06V 10/82 (2006.01)

(87) PCT国际申请的公布数据

G16H 50/20 (2006.01)

W02023/118586 EN 2023.06.29

G16H 30/40 (2006.01)

(71) 申请人 霍夫曼-拉罗奇有限公司

地址 瑞士巴塞尔

(72) 发明人 N • 布鲁恩格尔 P • 康韦

S • J • 戴维森 E • 德吉恩

J • 吉尔登布拉特 C • 萨吉夫

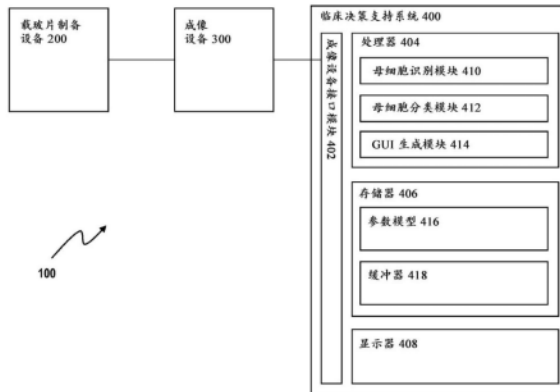
权利要求书2页 说明书10页 附图5页

(54) 发明名称

母细胞分类

(57) 摘要

一种区分淋巴母细胞和髓母细胞的计算机实现的方法,所述方法包括:接收(S30)包含一个或多个母细胞的数字图像;将参数模型分类器(412)应用(S34)于所述数字图像的一个或多个部分,所述一个或多个部分各自包含相应母细胞,参数模型(416)被配置成生成指示每个母细胞是淋巴母细胞还是髓母细胞的输出(S38)。也提供了训练参数模型(416)的计算机实现的方法,以及依赖于对母细胞进行分类的所述计算机实现的方法的临床决策支持系统(400)。



1. 一种区分淋巴母细胞和髓母细胞的计算机实现的方法,所述计算机实现的方法包括:

接收包含一个或多个母细胞的数字图像;

将图像分析应用于所述数字图像,图像分析算法被配置成:

检测所述数字图像中的一个或多个母细胞;

在所述一个或多个母细胞中的每一者周围生成边界框;并且

生成多个图像文件,每个图像文件包含所述一个或多个母细胞中的母细胞的数字图像,每个文件中所述图像的边界对应于相应边界框;

将深度神经网络分类器应用于所生成的图像文件中的每一者,所述深度神经网络分类器被配置成生成指示每个母细胞是淋巴母细胞还是髓母细胞的输出,所述输出包括在范围 $[0,1]$ 中的数值 x ,其中,以下任一者:

如果所述值 x 等于1,则以100%置信度将所述母细胞识别为淋巴母细胞,并且如果所述值 x 等于0,则以100%置信度将所述母细胞识别为髓母细胞;或

如果所述值 x 等于1,则以100%置信度将所述母细胞识别为髓母细胞,并且如果所述值 x 等于0,则以100%置信度将所述母细胞识别为淋巴母细胞;以及

基于针对所述数字图像中所有所述母细胞的相应输出值 x 来计算患者水平分数,其中所述患者水平分数包括:平均值;中值; x 的最大值;和/或 x 的最小值。

2. 根据权利要求1所述的计算机实现的方法,其中:

所述深度神经网络分类器是Resnet34分类器、Resnet50分类器或Resnet101分类器。

3. 根据权利要求1或权利要求2所述的计算机实现的方法,其中:

所述数字图像是包含已经经过罗曼诺夫斯基染色的一个或多个母细胞的明场显微镜图像。

4. 根据权利要求1至3中任一项所述的计算机实现的方法,其进一步包括:

基于参数模型分类器的输出来生成指令,所述指令被配置成使得计算系统的显示装置显示图库,所述图库包括:

第一多个图像,其示出以最高置信度被识别为淋巴母细胞的所述母细胞;和

第二多个图像,其示出以最高置信度被识别为髓母细胞的所述母细胞。

5. 根据权利要求1至4中任一项所述的计算机实现的方法,其中所述深度神经网络分类器使用计算机实现的方法生成,所述计算机实现的方法包括:

接收多对标记的训练数据,每对标记的训练数据包括:

输入数据,其包括来自已经被诊断患有急性髓性白血病或急性淋巴性白血病的患者的母细胞的数字图像;

输出数据,其包括对所述患者是患有急性髓性白血病还是急性淋巴性白血病的指示,所述输出包括所述范围 $[0,1]$ 中的数值 x ,其中,以下任一者:

如果所述值 x 等于1,则以100%置信度将所述母细胞识别为淋巴母细胞,并且如果所述值 x 等于0,则以100%置信度将所述母细胞识别为髓母细胞;或

如果所述值 x 等于1,则以100%置信度将所述母细胞识别为髓母细胞,并且如果所述值 x 等于0,则以100%置信度将所述母细胞识别为淋巴母细胞;以及

使用所述训练数据训练所述深度神经网络分类器。

6. 一种临床决策支持系统,其包括:

计算装置,其具有处理器,所述处理器被配置成通过进行根据权利要求1至4中任一项所述的计算机实现的方法来生成急性髓性白血病或急性淋巴性白血病的临时诊断;并且

其中所述处理器被进一步配置成:

基于根据所述深度神经网络分类器的所述输出计算的所述患者水平分数,确定其母细胞被示出在所述数字图像中的患者是遭受急性髓性白血病、急性淋巴性白血病还是两者都没有;以及

基于确定的结果来生成指令,所述指令被配置成使得计算系统的显示装置显示所述确定的所述结果。

母细胞分类

技术领域

[0001] 本发明涉及区分髓母细胞和淋巴母细胞的计算机实现的方法,以及用于基于分化的输出来诊断急性髓性白血病或急性淋巴性白血病的相关方法。还提供了用于训练深度神经网络的计算机实现的方法以及临床决策支持系统。

背景技术

[0002] 母细胞是成熟血细胞的前体,在人的血流中循环。通常,母细胞仅限于人的骨髓中。然而,当患者患有白血病时,异常母细胞在骨髓中不可控制地增殖到阻止对存活重要的其他细胞的产生的程度。此外,不可控制的增殖也导致异常母细胞泄漏到人的血流中。因此,可通过检测患者的血流中的这些异常母细胞来诊断白血病。

[0003] 急性白血病本身以包括急性髓性白血病(AML)和急性淋巴性白血病(ALL)的形式存在,该形式中的每一者具有多种亚型。为了确定存在哪种类型的白血病,有必要将异常母细胞分类为髓母细胞或淋巴母细胞。到目前为止,这种分类非常具有挑战性,因为细胞未成熟,并且缺乏细胞谱系分化。

[0004] 目前用于分类异常母细胞是髓母细胞还是淋巴母细胞的方法依赖于分化簇(CD)标志物评估的若干阶段。可使用的工作流程的实例如下。在第一步骤中,将获得全血细胞计数(CBC),作为常规筛查的一部分,或者由于白血病的可能症状的表达。如果CBC结果示出异常结果(即血液中母细胞的异常数量),则可能会进行血涂片并由血液学家进行检查。如果在血涂片的分析中确认血液中存在异常母细胞,则可进行重复采样以进行确认。然后,可进行包括CD标志物评估的进一步分析。CD标志物评估用于确定细胞谱系(即母细胞是髓母细胞还是淋巴母细胞)。在发生该确定之后,可实施特定针对骨髓细胞系或淋巴细胞系的CD标志物组,最终得出诊断。应当理解,这是一个漫长的过程,该漫长的过程需要CD标志物评估的若干阶段以便达到诊断。与CD标志物评估组合,可使用其他技术,该其他技术包括使用细胞遗传学、荧光原位杂交(FISH)或聚合酶链反应(PCR)技术来分析脑脊液(CSF)或骨髓样品。这些过程同样耗时,而且在许多情况下,成本高昂。

[0005] 本发明旨在通过提供一种用于区分髓母细胞和淋巴母细胞的计算机实现的方法来解决这个问题。

发明内容

[0006] 在高水平下,本发明提供了一种计算机实现的方法,其使用参数模型来区分淋巴母细胞和髓母细胞。换句话说,计算机实现的方法是确定数字图像中的母细胞是来自骨髓谱系还是淋巴谱系的方法。

[0007] 更具体地,本发明的第一方面提供了一种区分淋巴母细胞和髓母细胞的计算机实现的方法,该计算机实现的方法包括:接收包含一个或多个母细胞的数字图像;将参数模型分类器应用于该数字图像的一个或多个部分,该一个或多个部分各自包含相应母细胞,该参数模型分类器被配置成生成指示每个母细胞是淋巴母细胞还是髓母细胞的输出。

[0008] 通过在初始阶段处使用分类器,如根据本发明的第一方面的计算机实现的方法所要求的,可能减少或避免针对CD标志物分析的各个阶段(或其他冗长的技术)的需要,以便得出对AML或ALL的明确诊断。

[0009] 现在阐述本发明的第一方面的任选的特征。应当注意,任何或所有这些特征可彼此组合,除非当上下文另有指示时,或者当明确地声明某个特征不兼容或以其他方式不能与另一特征组合时。

[0010] 数字图像可从成像设备接收。并且,除了计算机实现的步骤之外,根据本发明的第一方面的方法可包括使用成像设备来捕获一个或多个母细胞的数字图像。成像设备可包括相机。成像设备可进一步包括显微镜。因此,数字图像可以是显微镜检查图像,诸如明场显微镜检查图像。另选地,然而,数字图像可基于相位对比成像、差分干涉对比显微镜检查或暗场显微镜检查。

[0011] 一个或多个母细胞的数字图像可以是载玻片上的样品的数字图像。

[0012] 除了计算机实现的步骤之外,根据本发明的第一方面的方法可进一步包括制备包含一个或多个母细胞的载玻片。优选地,该载玻片是使用允许一滴血在存在空气或在适度气流中干燥的方法并且使用在转移到载玻片的体积中产生所有细胞的单层的技术制备的。再次优选地,制备载玻片使得可对每个单个细胞进行计数并按类型区分。为了确保这一点,必须选择适当的稀释剂、稀释因子、通过改善载玻片的亲水性和/或通过机械地散布液滴来允许液滴散布在相对大的区域上的手段。参数和方法的适当选择产生非常薄的液体层,该液体层以细胞的呈现和保存模拟在常规血“涂片”中发现的细胞的呈现和保存的方式快速干燥。可制备有利的载玻片的方法的细节可在WO 2012/030313 A1中找到,其全部内容通过引用方式并入。关于载玻片的制备的附加公开内容可在US2009/0269799 A1、US 2010/284602A1、US2011/014645 A1和US2016/209320 A1中找到,其中的每一者的全部内容通过引用方式并入。

[0013] 如从上述参考文献中将要注意的,优选的是,对样品进行染色。因此,数字图像可以是载玻片上的染色样品的图像。可使用的染色剂包括:罗曼诺夫斯基染色、Giemsa染色、Jenner染色、Wright染色、Field染色、May-Grünwald染色和Leishman染色。应当理解,也可使用其他种类的染色剂。

[0014] 计算机实现的方法可进一步包括,在接收包含一个或多个母细胞的数字图像之后:识别数字图像中的一个或多个母细胞。换句话说,在将分类器应用于母细胞中的每一者之前,计算机实现的方法可进一步包括定位母细胞在图像中的位置。即使在患有AML或ALL的患者中,血液中母细胞的比例相对于例如红细胞仍然非常低。因此,在应用分类器之前识别数字图像内的母细胞是有益的,以确保分类器仅作用于识别的母细胞。识别母细胞可包括将图像分析算法应用于数字图像。图像分析算法优选地被配置成识别数字图像中的一个或多个母细胞周围的边界框。具体地,图像分析算法优选地被配置成识别数字图像中的一个或多个母细胞中的每一者周围的相应边界框。本文中,识别边界框应当理解为是指识别区域,优选地是包含单个母细胞的正方形或矩形区域。在一些情况下,图像分析算法可被配置成生成多个图像文件,每个图像文件包含一个或多个母细胞中的一个母细胞的数字图像。多个图像文件中的每个图像文件中的图像的边界可以是上述的边界框。在一些情况下,可在已经将图像分析算法应用于数字图像之后生成多个文件。在其中生成多个图像

文件的情况下,计算机实现的方法可包括将参数模型分类器应用于生成的图像文件中的每一者,并且被配置成生成指示每个图像中的相应母细胞是淋巴母细胞还是髓母细胞的相应输出。因此,计算机实现的方法的输出可包括多个输出,该多个输出各自指示一个或多个母细胞中的相应母细胞是淋巴母细胞还是髓母细胞。

[0015] 我们现在更详细地考虑参数模型分类器的输出的形式。针对数字图像中的每个母细胞,或者针对每个图像文件,参数模型分类器的输出可包括指示母细胞是淋巴母细胞还是髓母细胞的数值 x 。在一些情况下,输出(同样,针对数字图像中的每个母细胞,或者针对每个图像文件)可包括第一值和第二值,该第一值指示母细胞是淋巴母细胞(或图像文件包含淋巴母细胞的图像)的可能性,该第二值指示母细胞是髓母细胞(或图像文件包含髓母细胞的图像)的可能性。优选地,第一值和第二值是概率,并且优选地它们求和为1。值越极端,结果的置信度越高。换句话说,数值 x 可在 $[0,1]$ 的范围内;如果值 x 等于1,则可以100%置信度将该母细胞识别为淋巴母细胞;并且如果值 x 等于0,则可以100%的置信度将该母细胞识别为除淋巴母细胞以外的细胞类型。或者,数值 x 可在 $[0,1]$ 的范围内;如果值 x 等于1,则可以100%置信度将该母细胞识别为髓母细胞;并且如果值 x 等于0,则可以100%的置信度将该母细胞识别为除髓母细胞之外的细胞类型。众所周知,区分髓母细胞和淋巴母细胞可能具有挑战性,并且因此,可想象,在一些情况下(由于母细胞本身的性质,或例如,由于其在数字图像中示出的角度),不可能准确地对细胞进行分类。因此,在一些情况下,如果数值落入预先确定的范围内,则参数分类器的输出被配置成非决定性的或不确定的。此类结果可能被丢弃。预先确定的值可在0.1与0.9、在0.2与0.8、在0.3与0.7、在0.4与0.6、或者在0.45与0.55之间。

[0016] 每个输出可采用[第一值,第二值]的形式。考虑其中预先确定的阈值范围为0.4至0.6的情况:

[0017] -如果结果为 $[0.98,0.02]$,则可以98%的确定性得出结论,讨论中的母细胞是淋巴母细胞。

[0018] -如果结果为 $[0.43,0.57]$,则结果将被视为非决定性的,因为这些值落入0.4至0.6的范围内。

[0019] -如果结果为 $[0.22,0.78]$,则可以78%的确定性得出结论,讨论中的母细胞是髓母细胞。

[0020] 在这些情况下,假设母细胞是髓母细胞或淋巴母细胞。

[0021] 在迄今为止的整个申请中,我们称为“参数模型分类器”。在本申请的上下文中,参数模型是一种假设针对用于从输入数据(即,表示数字图像的数据)生成输出的函数的参数形式的模型,该函数包括固定数量的参数。顾名思义,参数形式依赖于多个参数,并且例如训练过程的目标就是用以识别那些参数。相比之下,非参数模型是无边界的,并且其复杂性也没有限制。参数模型分类器(相对于非参数分类器)的优点包括简单性、速度和它们可在较低数据量上产生可靠结果的事实。参数模型分类器优选地是基于机器学习的分类器。在优选的情况下,分类器基于卷积神经网络,诸如深度神经网络。合适的神经网络分类器的一个实例是残差神经网络分类器^{1 2 3}(本文中为“ResNet”)。ResNet是建立在从大脑皮层中的锥体细胞已知的结构之上的人工神经网络。ResNet基于跳过连接或快捷方式以跳过神经网络中的层的原理工作。典型的ResNet模型是利用包含非线性的双层或三层跳跃或/和之间

的批归一化来实现的。连接可被跳过存在两个原因:为了避免消失梯度的问题,或者为了减轻退化(精度饱和)问题,其中向适当深度模型添加更多的层导致更高的训练误差。在训练期间,权重适应使上游层静音并放大先前跳过的层。在最简单的情况下,仅适应针对相邻层的连接的权重,而没有针对上游层的显式权重。当跨过单个非线性层时或当中间层都是线性的时,这种方法效果最佳。

¹ He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian (2015-12-10). "Deep Residual Learning for Image Recognition". arXiv:1512.03385

² He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian (2016). "Deep Residual Learning for Image Recognition". *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*.

³ Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012): 1097-1105.

[0022] 如果不是,则可针对跳过的连接学习显式权重矩阵。ResNet是有利的,因为跳过有效地简化了网络,在初始训练阶段使用较少的层,从而通过减少消失梯度的影响来加速学习,由于存在较少的层要传播通过。然后网络在学习特征空间时逐渐恢复跳过的层。在训练结束时,当所有层都扩展时,它会更接近流形,并且因此学习得更快。合适的ResNet的实例包括:ResNet34、ResNet50和ResNet100,其中数字表示残差神经网络中存在的层数。卷积神经网络的其他类型诸如Inception⁴、VGG⁵和EfficientNet⁶也可用于本发明的具体实施中。

[0023] 另选地或附加地,可使用多示例学习来创建模型,例如使用Deep Attention MIL,如Ilse等人的(2018)⁷所述。

⁴ Szegedy et al. (2014) "Going Deeper with Convolutions" arXiv:1409.4842

⁵ Simonyan & Zisserman (2015) "Very Deep Convolutoinal Networks for Large-Scale Image Recognition" arXiv:1409.1556

[0024]

⁶ Tan & Le (2019) "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks" arXiv:1905.11946

⁷ Maximilian Ilse, Jakub M. Tomcak, Max Welling (2018) "Attention-based Deep Multiple Instance Learning". arXiv:1802.04712

[0025] 可使用Siamese模型,而不是涉及使用每个图像标记来训练模型的神经网络方法。Siamese模型需要两个图像,并且必须确定它们是否属于相同类别。单独地应用于完整载玻片或应用于每个细胞的标记可用于在每对由相同细胞类型(例如两个淋巴母细胞或两个髓母细胞)组成的情况下将每对识别为真,或者在图像来自不同类型的细胞的情况下识别为假。

[0026] 本发明的目的是用以提高诊断患者患有AML或ALL(还是两者都没有)的效率。因此,我们现在考虑如何使用参数模型分类器的输出。因此,计算机实现的方法可进一步包括基于针对数字图像中的所有母细胞的相应输出值x来计算患者水平分数。以这种方式,可基于多个母细胞而不是单个母细胞来做出诊断或临床决策。患者水平分数可包括以下中的一者或多者:平均值、中值、最大值或最小值。具体地,可基于表示母细胞是淋巴母细胞的可能性和/或母细胞是髓母细胞的可能性的所有x值来计算患者水平分数。如果患者水平分数的

数值落入预先确定的范围内,则参数分类器的输出被配置成非决定性的或不确定的。此类结果可能被丢弃

[0027] 在一些情况下,本发明的第一方面的计算机实现的方法可用作决策支持系统的一部分以使得临床医生能够选择适当的行动方案。在此类情况下,临床医生可能期望审查由参数模型分类器生成的结果。具体地,计算机实现的可进一步包括:基于参数模型分类器的输出来生成指令,该指令被配置成使得计算系统的显示装置显示图库,该图库包括:第一多个图像,其示出以最高置信度被识别为淋巴母细胞的母细胞;和第二多个图像,其示出以最高置信度被识别为髓母细胞的母细胞。在该上下文中,“以最高置信度”可理解为是指作为特定类型的母细胞(即,髓母细胞或淋巴母细胞)的概率最高的母细胞。这将使得临床医生能够审查图像,以便确定患者的预后,并且以便决定适当的行动方案。在一些情况下,第一多个图像包括与第二多个图像相同数量的图像。该数量可在1与100之间,更优选地在5与50之间,更优选地在10与25之间,并且最优选地约20。

[0028] 如本申请前面所讨论的,为了达到更明确的诊断,通常需要进行CD标志物筛查。可获得各种类型的CD标志物筛查,该各种类型的CD标志物筛查各自靶向特定的生物标志物。在一些情况下,临床医生可基于参数模型分类器的输出来识别CD标志物以用于进行筛查。然而,在其他情况下,计算机实现的方法可进一步包括基于参数模型分类器的输出从多个可用的CD标志物中选择一个或多个CD标志物。以这种方式,可基于参数模型的输出来自动地识别要用于后续测试的CD标志物。

[0029] 为了使参数模型提供准确结果,通常需要使用已知数据对该参数模型进行训练。因此,本发明的第二方面提供了一种生成参数模型分类器的计算机实现的方法,该参数模型分类器被配置成区分数字图像中的淋巴母细胞和髓母细胞,该计算机实现的方法包括:接收多对标记的训练数据,每对标记的训练数据包括:输入数据,其包括来自已经被诊断患有急性髓性白血病或急性淋巴性白血病的患者的母细胞的数字图像;和输出数据,其包括对患者是患有急性髓性白血病还是急性淋巴性白血病的指示;以及使用训练数据来训练参数模型分类器。在兼容的情况下,上文关于本发明的第一方面阐述的任选特征同样适用于本发明的第二方面。为了避免疑义,我们强调,在本发明的第二方面中,参数模型分类器可以是基于机器学习的分类器。在优选的情况下,分类器基于卷积神经网络,诸如深度神经网络。合适的神经网络分类器的一个实例是残差神经网络分类器(本文中为“ResNet”)。合适的ResNet的实例包括:ResNet34、ResNet50和ResNet100,其中数字表示残差神经网络中存在的层数。可使用Adam优化器⁸来训练卷积神经网络。可使用用于学习速率调度的一个周期策略来训练卷积神经网络,如Smith⁹中所述的。可用于卷积神经网络的训练的框架是fastai¹⁰,并且可使用Paszke等人¹¹的库。也可使用替代地依赖平坦学习速率和余弦退火的不同学习策略。除了卷积神经网络之外,也可使用视觉变换器或实验模型诸如胶囊。

⁸ “Adam: A Method for Stochastic Optimization” by Kingma et al <https://arxiv.org/abs/1412.6980>

⁹ “A disciplined approach to neural network hyper-parameters: Part I -- learning rate, batch size, momentum, and weight decay” by Smith
<https://arxiv.org/abs/1803.09820>

[0030]

¹⁰ “fastai: A Layered API for Deep Learning” by Jeremy Howard, Sylvain Gugger
<https://arxiv.org/abs/2002.04688>

¹¹ “PyTorch: An Imperative Style, High-Performance Deep Learning Library” by Paszke et al,
<https://arxiv.org/abs/1912.01703>

[0031] 对患者是患有急性髓性白血病还是急性淋巴性白血病的指示可包括数值。具体地,如果患者患有急性髓性白血病,则该数值可为1,并且否则可为0。相反地,如果患者患有急性淋巴性白血病,则该数值可为1,并且否则可为0。关于训练过程的更详细的信息将在本申请后面的“实验结果”章节中阐述。

[0032] 在本发明的第一方面的计算机实现的方法中,可使用本发明的第二方面的计算机实现的方法来训练参数模型分类器。

[0033] 本发明的第三方面提供了一种生成对急性髓性白血病或急性淋巴性白血病的临时诊断的计算机实现的方法,该计算机实现的方法包括:进行本发明的第一方面的计算机实现的方法;以及基于参数模型分类器的输出或基于参数模型分类器的输出计算的患者水平分数,确定其母细胞被示出在数字图像中的患者是遭受急性髓性白血病、急性髓性白血病还是两者都没有。如前所述,关于本发明的第一方面或本发明的第二方面阐述的所有任选特征同样适用于本发明的第三方面的计算机实现的发明。此外,基于确定的结果,本发明的第三方面的计算机实现的方法可进一步包括生成指令,该指令被配置成使得计算系统的显示装置显示确定的结果。

[0034] 本发明的第一方面至第三方面涉及计算机实现的方法。本发明的对应的第四方面至第六方面提供了计算机程序产品,该计算机程序产品当该程序由计算机或其他计算装置执行时,使得计算机分别实施本发明的第一方面至第三方面的计算机实现的方法。本发明的第七方面至第九方面分别提供了一种计算机可读数据载体,其上存储有本发明的第四方面至第六方面的计算机程序产品。

[0035] 如上所解释的,本发明的目的是用以提供一种临床决策支持系统,其辅助临床医生诊断急性髓性白血病和/或急性淋巴性白血病。因此,本发明的第十方面提供了一种临床决策支持系统,其包括具有处理器的计算装置,该处理器被配置为进行本发明的第一方面至第三方面中的任一者的计算机实现的方法。

[0036] 本发明的特别优选的第十一方面提供了一种区分淋巴母细胞和髓母细胞的计算机实现的方法,该计算机实现的方法包括:接收包含一个或多个母细胞的数字图像;将图像分析应用于数字图像,图像分析算法被配置成:检测数字图像中的一个或多个母细胞;在一个或多个母细胞中的每一者周围生成边界框;以及生成多个图像文件,每个图像文件包含一个或多个母细胞中的一个母细胞的数字图像,每个文件中的图像的边界对应于相应边界框;将深度神经网络分类器应用于生成的图像文件中的每一者,该深度神经网络分类器被配置成生成指示每个母细胞是淋巴母细胞还是髓母细胞的输出,该输出包括范围[0,1]中

的数值 x ,其中,以下任一者:如果值 x 等于1,则以置信度为100%将母细胞识别为淋巴母细胞,并且如果值 x 等于0,则以100%置信度将母细胞识别为髓母细胞;或如果值 x 等于1,则以100%置信度将母细胞识别为髓母细胞,并且如果值 x 等于0,则以100%置信度将母细胞识别为淋巴母细胞;以及基于针对数字图像中的所有母细胞的相应输出值 x 来计算患者水平分数,其中患者水平分数包括平均值;中值; x 的最大值;和/或 x 的最小值。除了技术上明显不兼容的情况或上下文明确指示其他情况之外,上文关于本发明的第一方面至第十方面阐述的任选特征同样适用于本发明的第十一方面。

[0037] 在一些情况下,使用计算机实现的方法来生成本发明的第十一方面的深度神经网络分类器,该计算机实现的方法包括:接收多对标记的训练数据,每对标记的训练数据包括:输入数据,其包括来自已经被诊断患有急性髓性白血病或急性淋巴性白血病的患者的母细胞的数字图像;输出数据,其包括对患者是患有急性髓性白血病还是急性淋巴性白血病的指示,该输出包括取值为0或1时的数值 x ,其中,以下任一者:如果值 x 等于1,则母细胞是淋巴母细胞,并且如果值 x 等于0,则母细胞是髓母细胞;或如果值 x 等于0,则母细胞是淋巴母细胞,并且如果值 x 等于1,则母细胞是髓母细胞;以及使用训练数据来训练深度神经网络分类器。

[0038] 本发明的第十二方面提供了一种临床决策支持系统,其包括:具有处理器的计算装置,该处理器被配置成通过进行本发明的第十一方面的计算机实现的方法来生成对急性髓性白血病或急性淋巴性白血病的临时诊断,并且其中该处理器被进一步配置成:基于根据深度神经网络分类器的输出计算的患者水平分数,确定其母细胞被示出在数字图像中的患者是遭受急性髓性白血病、急性淋巴性白血病还是两者都没有;以及基于确定的结果来生成指令,该指令被配置成使得计算系统的显示装置显示确定的结果。

附图说明

[0039] 现在将参考附图描述本发明的实施例,其中:

[0040] -图1示出了根据本发明的一些方面的可用于实现计算机实现的方法的系统。

[0041] -图2A和图2B示出了可通过图1所示的成像设备获得的包含母细胞的原始图像。

[0042] -图3是展示用于对母细胞进行分类的高水平方法的流程图。

[0043] -图4是可基于母细胞分类的输出在显示器上产生的视图的实例。

[0044] 附图详细说明

[0045] 现在将参考附图讨论本发明的方面和实施例。对于本领域技术人员来说,其他方面和实施例将是显而易见的。本文中提及的所有文件均通过引用并入本文。

[0046] 图1示出了系统100,该系统包括用于进行或使得能够进行根据本发明的各种方面的计算机实现的方法的各种部件。系统100包括载玻片制备设备200、成像设备300和临床决策支持系统400。载玻片制备设备200优选地被配置成制备含有一个或多个母细胞的载玻片。如本申请前面所讨论的,载玻片制备可使用WO 2012/030313 A1、US2009/0269799 A1、US 2010/284602A1、US2011/014645 A1和US2016/209320 A1中阐述的方法,其中的每一者的全部内容通过引用方式并入本文。载玻片制备在本申请的范围之外,并且将不进行任何更详细的讨论。一旦载玻片已经被制备,就使用成像设备300对其进行成像,该成像设备可包括例如显微镜和相机(未示出)。当载玻片制备设备200如上面引用的参考文献中那样制

备载玻片时,形成单层,该单层使得每个单独的细胞(包括红细胞、白细胞、血小板以及至关重要的母细胞)能够被可视化和计数。临床决策支持系统400然后从成像设备300接收包含母细胞的图像。可接收的图像种类的实例在图2A和图2B中示出。临床决策支持系统400然后可用于对图像中的母细胞进行识别和分类。图1所示的临床决策支持系统的实例包括成像设备接口模块402、处理器404、存储器406和显示器408。应当注意,在一些另选的布置中,显示器408可以是外部部件。在那些情况下,临床决策支持设备400的处理器404可进一步包括显示器模块接口模块(未示出)。临床决策支持系统400的处理器404可包括母细胞识别模块410、母细胞分类模块412和图形用户界面(GUI)生成模块414。在本文中,“模块”可以是物理模块或功能模块的形式,例如以软件模块的形式(即,以计算机可读代码)实现。存储器406可包括参数模型416,该参数模型可由母细胞分类模块412应用于包含母细胞的图像。存储器可进一步包括缓冲器418。

[0047] 我们现在解释临床决策支持设备400的操作以对母细胞进行识别和分类,并且以生成指令来使得各种图像显示在显示器408上。图3是示出计算机实现的方法的高水平步骤的流程图。

[0048] 在第一步骤S30中,在临床决策支持系统400处经由其成像设备接口模块402从成像设备300接收包含母细胞的图像。然后,在步骤S32中,由母细胞识别模块410识别图像内的母细胞。该过程的输出可以是例如若干图像文件(其可临时地存储在缓冲器418中或更永久地存储在存储器406中),该若干图像文件各自包含来自图像的单个母细胞的图像。另选地,不是生成多个图像文件,母细胞识别模块410可识别图像内的母细胞,并且限定区域的边界,该区域各自包含单个母细胞。母细胞识别模块410的输出可以是像素阵列列表的形式,每个像素阵列对应于包含单个母细胞的图像的区域。该列表也可存储在缓冲器418中或更永久地存储在存储器406中。在步骤S34中,由母细胞分类模块412将参数模型416应用于每个图像文件(或包含母细胞的图像的区域),以便确定母细胞是淋巴母细胞还是髓母细胞。如前面所讨论的,母细胞分类模块412的输出针对每个母细胞优选地是从0至1的数字,该数字表示所讨论的母细胞是髓母细胞或淋巴母细胞的概率或置信度水平。另选地,模型可返回两种概率,一种概率是母细胞是淋巴母细胞,并且一种概率是母细胞是髓母细胞。这些概率应当添加至1(或100%,或等效值)。然后,一旦已经由母细胞分类模块412使用参数模型416计算出多种概率,就可采取各种不同的步骤。在图3中,这些步骤被表示为流程图的不同分支,但是必须强调的是,这不应当被理解为意味着根据本发明的各种方面的计算机实现的方法不能包括不同分支上的步骤中的多于一个步骤。在步骤S36中,计算患者分数并将其存储在存储器406中(例如,在缓冲器418中)。在本文中,患者水平分数是表示针对(单个)母细胞的图像中的每一者计算的概率值的统计参数。如前面所讨论的,患者水平分数可采取各种形式。在步骤S38中,GUI生成模块414可被配置成基于母细胞分类模块412的输出来生成指令,该指令当被显示器408接收时使得显示器将结果呈现给临床决策支持系统400的用户。例如,显示器408可显示图库,如图4所示,该图库包括具有最高概率为淋巴母细胞的18个细胞和具有最高概率为髓母细胞的20个细胞。当然,显示器不需要是诸如这种的图库的形式—这仅仅是一种选择。在一些情况下,也可显示患者水平分数和针对每个母细胞的单独概率。

[0049] 在一些实例(未示出)中,该方法可进一步包括基于例如患者水平分数来确定用于

后续筛查步骤的适当的CD标志物的步骤。

[0050] 实验结果

[0051] 已经解释了本发明的计算机实现的方法,我们现在提出证明本发明的功效的一些实验数据。对不同数据集进行了两个实验,本文中被称为图卢兹数据集和波士顿数据集。

[0052] 从下文中应当理解,根据本发明的各种方面的计算机实现的方法在区分髓母细胞和淋巴母细胞方面是非常可靠的。

[0053] A. 图卢兹数据集

[0054] 在该实验中,使用了来自119名患者的血涂片的图像,这些患者中52名患者患有急性髓性白血病,并且67名患者患有急性淋巴性白血病(51名ALL-B,和16名ALL-T)。血涂片的图像是在CellaVision仪器上以高分辨率获取的。

[0055] 随机采样16张载玻片作为验证集,并且其余的载玻片用于训练神经网络,在这种情况下为ResNet50网络。

[0056] 在验证集上,每个细胞达到83.91%的AUC(接收者操作者图中的曲线下面积)。汇总结果后,即通过生成患者水平分数,AUC增加到95.31%。

[0057] 相同模型在来自相同医院的附加数据集上进行了测试,该附加数据集包含来自具有较低的Blast计数的AML/ALL病例的125张载玻片。在该数据集上,相同模型实现0.89963的AUC。

[0058] B. 波士顿数据集

[0059] 在该实验中,从AML或ALL患者获得了39张载玻片。使用本申请前面解释的方法对载玻片进行打印和成像。这些图像是使用高分辨率相机以20倍放大倍数获得的。在39张载玻片中,21张来自AML患者,并且18张来自ALL患者。神经网络(ResNet50网络)在29张载玻片上进行训练,并且保留10张载玻片(5张AML和5张ALL)作为验证集。根据选择哪些载玻片用于验证组,AUC为79.51%至91.55%。

[0060] C. 另选的训练/评估程序

[0061] 通过排除完整的载玻片,而不是从完整的数据集中随机地选择细胞,将数据集分为训练集和验证集。

[0062] 为了彻底地测试针对整个数据集的每张载玻片的性能,进行了每张载玻片的交叉验证,其中排除了一张载玻片,并在所有其他载玻片上训练了模型。性能是在排除的载玻片上评估的。针对所有载玻片重复此操作。

[0063] 使用了5个模型的集成(再次使用ResNet34)。因为来自各个模型的随机错误被校正,所以集成带来更好的性能。针对整个数据集,实现86.09%的总AUC。

[0064] 有关申请的一般说明

[0065] 前述说明书公开的特征,或以下权利要求中,以其特定形式表达或实施公开功能的方式方面,或用于获得公开结果的方法或过程,这些特征可以其多种形式适当地单独使用,或以任何组合的形式使用,以实现本发明。

[0066] 虽然已经结合上述示例性实施例描述了本发明,但是当给出本公开时,许多等效修改和变化对于本领域技术人员来说将是显而易见的。因此,上述本发明的示例性实施例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例进行各种改变。

[0067] 为避免任何疑问,本文提供的任何理论解释都是为了改进读者的理解。发明人不希望受任何这些理论解释的束缚。

[0068] 此处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。

[0069] 在整个说明书,包括随后的权利要求中,除非上下文另有要求,否则词语“包括(comprise)”和“包括(include)”以及诸如“包括(comprises和comprising)”和“包括(including)”的变体应被理解为暗示包括所陈述的整数或步骤或者整数或步骤的组,但不排除任何其他整数或步骤或者整数或步骤的组。

[0070] 必须注意的是,如在本说明书和所附权利要求中所用,单数形式“一个”、“一种”和“所述”包括复数指代物,除非上下文另外明确规定。范围在本文中 can 表示为从“约”一个特定值和/或到“约”另一个特定值。当表达这样的范围时,另一个实施例包括从一个特定值和/或到另一个特定值。类似地,当值表示为近似值时,通过使用先行词“约”,将理解特定值形成另一个实施例。与数值相关的术语“约”为任选的并且意指例如+/-10%。

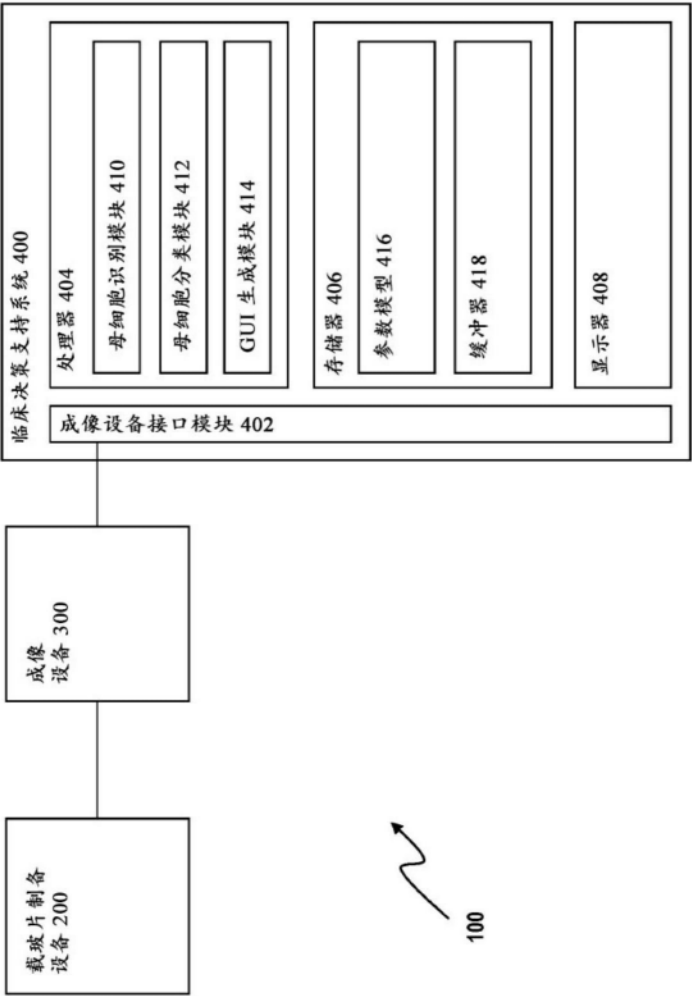


图1

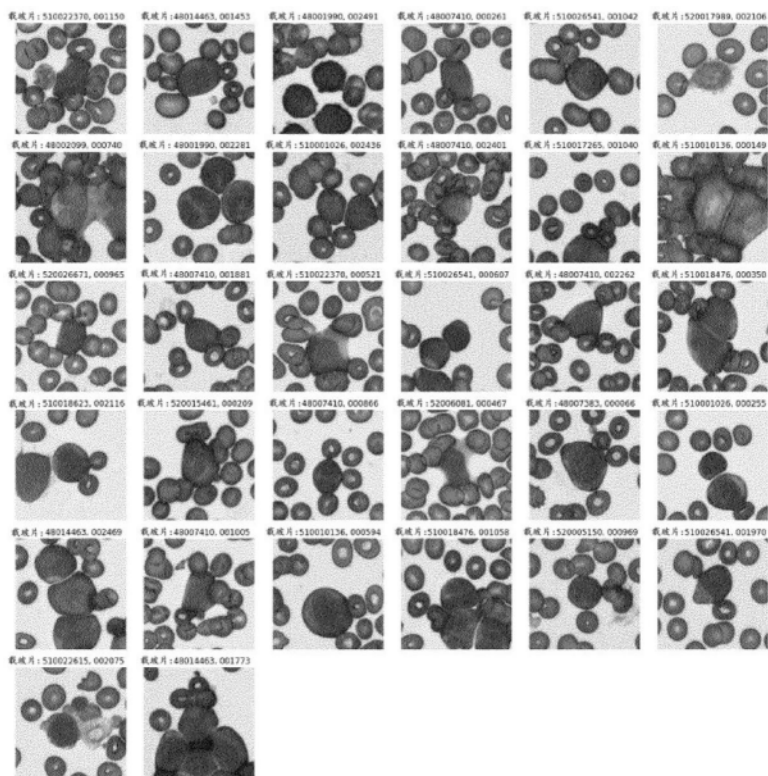


图2A

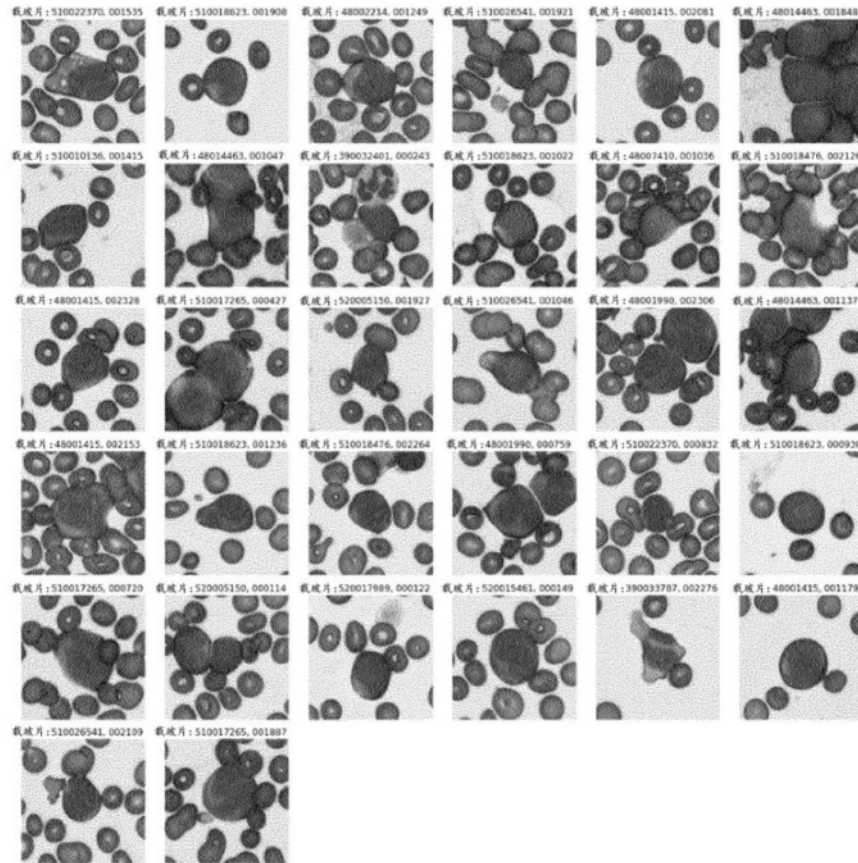


图2B

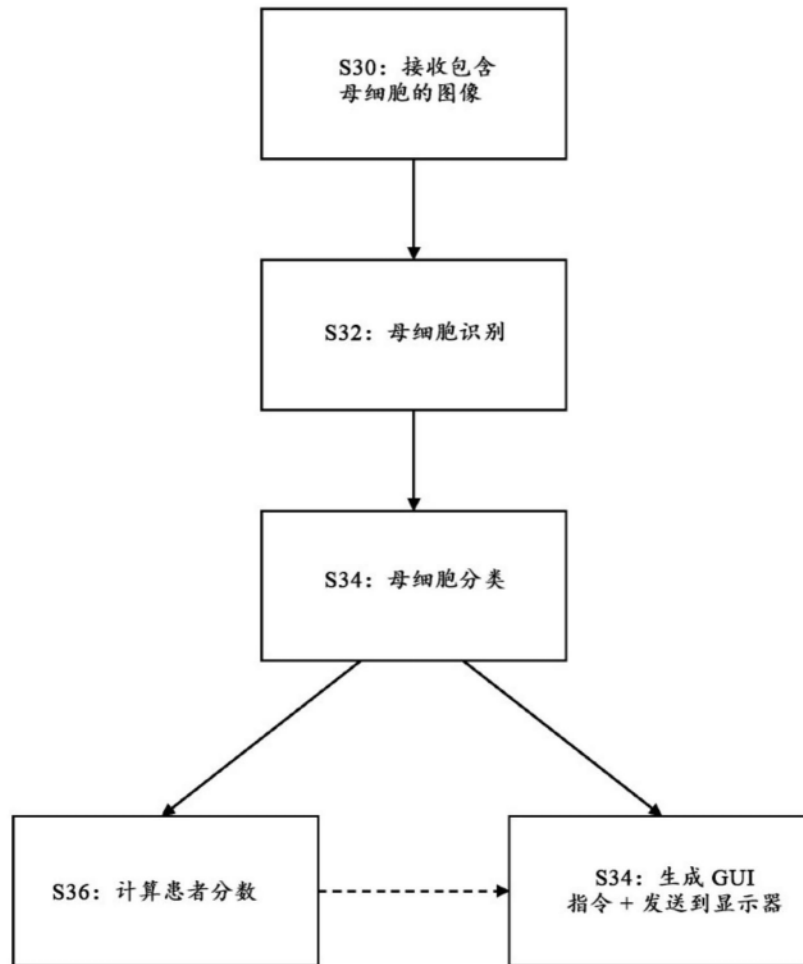


图3

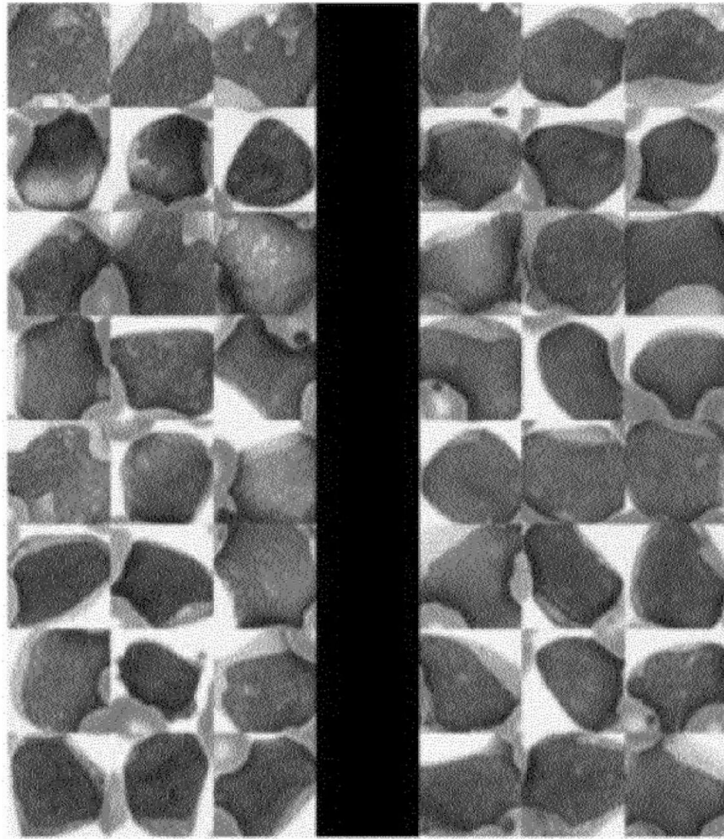


图4