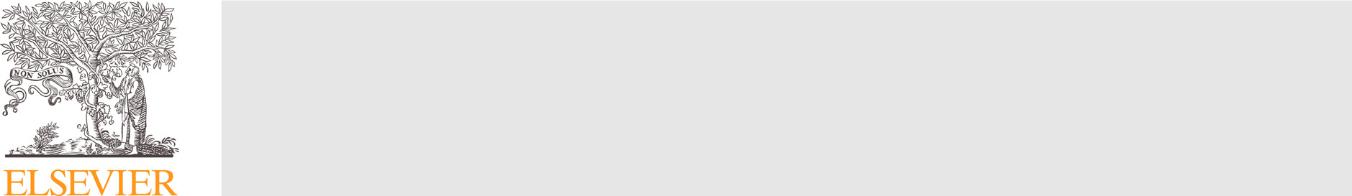
[医学图像分析59（2019）101561](https://doi.org/10.1016/j.media.2019.101561)



可在目录列表 [ScienceDirect](http://www.ScienceDirect.com)



医学图像分析

期刊主页： [www.elsevier.com/locate/media](http://www.elsevier.com/locate/media)

IDRiD：糖尿病视网膜病变 - 分割和分级挑战



人员Prasanna Porwal [一种，b，1，\*，](#page1) Samiksha Pachade [一种，b，1，](#page1) Manesh Kokare [一种，1，](#page1) 吉里什德希穆克 [C，1，](#page1)

Jaemin儿子 [d，](#page1) 雄裴 [d，](#page1) 刘利红 [Ë，](#page1) 剑宗王 [Ë，](#page1) 刘新会 [Ë，](#page1) 孙良高 [Ë，](#page1)

天博武 [Ë，](#page1) 晶科 [Ë，](#page1) 王峰岩 [F，](#page1) 尹宝才 [F，](#page1) 云芝王 [G，](#page1) Gopichandh Danala [G，](#page1)

林盛他 [G，](#page1) 尹浩彩 [H，](#page1) 永陈利 [H，](#page1) 桑赫荣 [H，](#page1) 中宇力 [一世，](#page1)

李晓丹穗 [Ĵ，](#page1) 吴张俊彦 [升，](#page1) 李枭龙 [米，](#page1) 周婷 [ñ，](#page1) 亚诺什·托特 [Ø，](#page1) 艾格尼丝·巴兰 [Ø，](#page1)

古里阿维纳什 [p，](#page1) 西Saketh Chennamsetty [p，](#page1) 穆罕默德萨夫万 [p，](#page1) Varghese表示亚历克斯 [p，](#page1)

Xingzheng吕 [q，[R，](#page1) 李成 [[R，d，](#page1) 秦浩楚 [小号，](#page1) 李彭城 [小号，](#page1) 新吉 [Ť，](#page1) 张三元 [q，](#page1)

沉雅欣 [ü，v，](#page1) 淩带 [ü，v，](#page1) Oindrila萨哈 [X，](#page1) Rachana Sathish所在 [X，](#page1) 塔尼亚梅洛 [ÿ，](#page1) 邓丽君阿劳霍 [ÿ，ž，](#page1)

鲍拉日Harangi [Ø，](#page1) 盛彬 [ü，v，](#page1) 若谷方 [w ^，](#page1) Debdoot表 [X，](#page1) 安德拉斯·豪伊杜 [Ø，](#page1)

郑渊洁郑 [Ĵ，](#page1) 安娜·玛丽亚·门多萨 [ÿ，ž，](#page1) 少汀张 [一世，](#page1) 奥雷里奥Campilho [ÿ，ž，](#page1) 郑斌 [G，](#page1) 顶岗沉 [ķ，](#page1) 卢卡Giancardo [b，1，](#page1) GwenoléQuellec [一种，1，](#page1) 法布里斯Mériaudeau [乙，C，1](#page1)

* *工程技术，楠德拉，印度的至尊宗师Gobind Singhji研究所*
* *得克萨斯大学健康科学中心的生物医学信息学院，大学在休斯敦，美国* C *眼科诊所，医院Sushrusha，楠德拉，印度马哈拉施特拉邦*

d *VUNO公司，首尔，韩国*

* *平安科技（深圳）有限公司，中国* F *科大讯飞的研究，中国合肥*
* *和计算机电气工程学院，美国俄克拉荷马州大学*
* *韩国成均馆大学三星综合技术院健康科学与技术（SAIHST），首尔，韩国*
* *北卡罗来纳大学夏洛特，美国计算机科学系*
* *山东师范大学信息科学与工程学院，中国学院*
* *放射学和金砖四国，北卡罗莱纳大学的系在教堂山，美国* 升 *Cleerly公司，纽约，美国*

米 *弗吉尼亚理工大学，弗吉尼亚州，美国*

ñ *布法罗大学，纽约州，美国*

* *德布勒森，信息学4002德布勒森学院，POB 400，匈牙利大学* p *个别研究员，印度*
* *浙江大学计算机科学与技术学院，杭州，中国的大学*
* *机器学习的生物影像分析小组，生物信息学研究所，A*\* *STAR，新加坡*
* *新加坡，新加坡国立大学计算机学院，* Ť *北京上工医疗科技有限公司，中国*
* *上海交通大学计算机科学与工程，中国系*
* *人工智能，人工智能研究所，上海交通大学，中国教育部重点实验室*
* *生物医学工程佛罗里达大学，美国J.克雷顿普鲁伊特家庭部，* X *技术克勒格布尔，印度的印度理工学院*
* *INESC TEC - 研究所系统与计算机工程，技术和科学，波尔图，葡萄牙*
* *FEUP - 波尔图大学的工程，波尔图，葡萄牙学院* 一种 *INSERM，UMR 1101，布雷斯特，法国*
* *和电气电子工程系，国油大学，马来西亚* C *ImViA / IFTIM，Üniversite电勃艮第法国第戎*
* *和计算机电气工程系，加拿大阿尔伯塔大学*

\* 在卓越中心的信号与图像处理，工程和技术，楠德拉（MS），印度的至尊宗师Gobind Singhji研究所通讯作者。

* 这些作者共同举办的挑战。所有其他的贡献文件中提出的算法（S）的结果

<https://doi.org/10.1016/j.media.2019.101561>

版权所有1361-8415 / 2019©爱思唯尔所有权利。

2

articl EI NFO

*文章历史：*

获得11月2019修订2019年9月9日接受2019年9月16日可在线2019年10月3日

*关键词：*

糖尿病性视网膜病变

视网膜图像分析

深度学习

挑战

*P. Porwal，S. Pachade和M. Kokare等。/医学图像分析59（2019）101561*

抽象

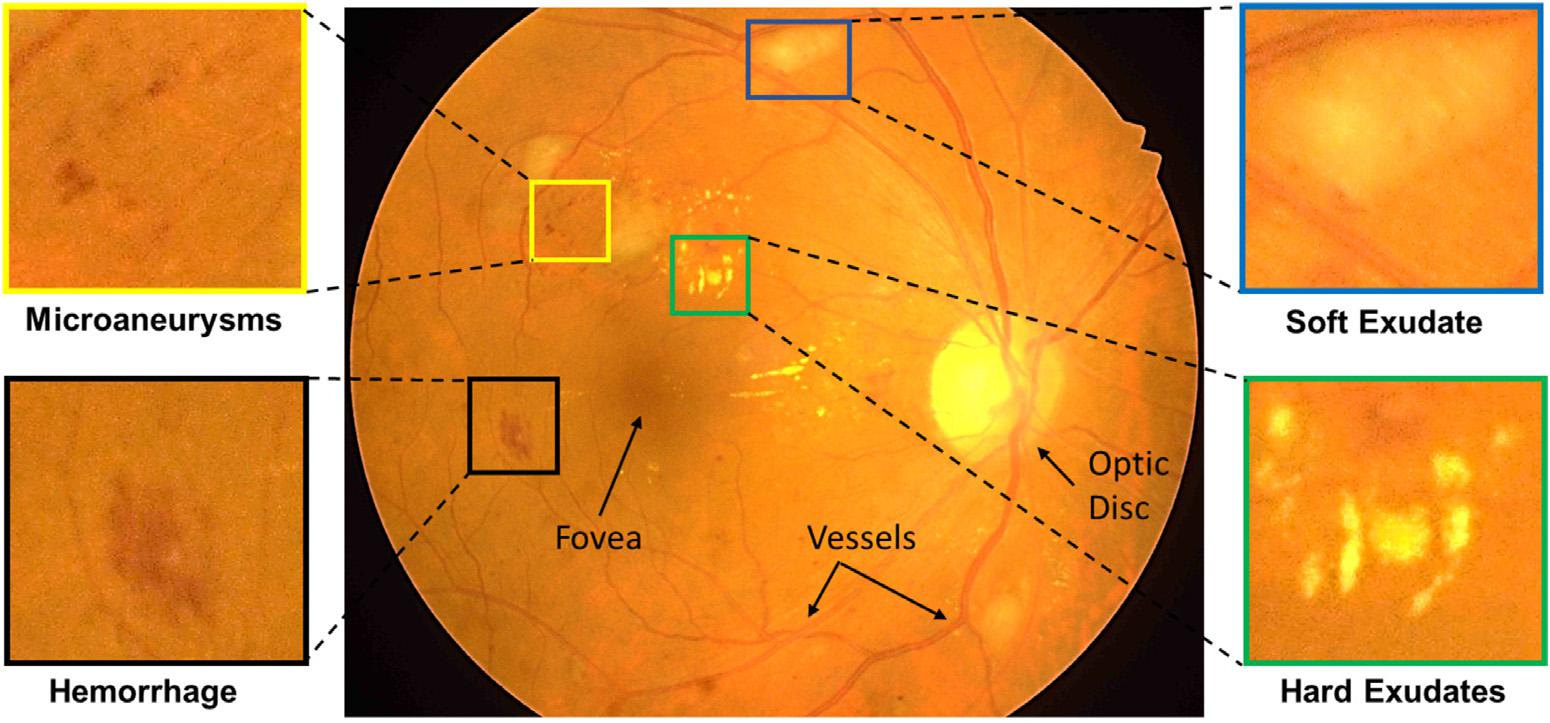
糖尿病视网膜病变（DR）是可以避免的视力丧失的最常见的原因，主要影响在全球范围内的劳动年龄人口。筛选DR，再加上及时的咨询和治疗，精神疾病，是全球值得信赖的政策，以避免视力减退。然而，实施DR筛查项目是具有挑战性，由于能够在为DR的风险筛选日益增长的全球糖尿病珀普-特征研医疗专业人员的稀缺性。在视网膜图像分析计算机辅助疾病诊断可以提供这样的大规模筛查努力为维持能的方法。最近在计算能力和机器学习的科学进步的方法提供了一个途径为生物医学科学家们实现这一目标。针对推进状态的最先进的自动诊断DR，关于“糖尿病性视网膜病变一个重大挑战

- 分割和分级”与IEEE国际研讨会生物医学成像（ - 2018 ISBI）共同举办。在本文中，我们报道了设置和这一挑战的结果主要是基于印度的糖尿病视网膜病变的图像数据集（IDRiD）。有三个主要子挑战：病变划分，疾病的严重程度分级和视网膜地标和分割的定位。在这个挑战这些多任务允许测试算法的普遍性，这是什么使得它从现有的不同。它收到了来自科学交通技术，无穷大的积极响应与495个登记148个提交了这场挑战有效的进入。本文从线的挑战，它的组织，使用的数据集，评价方法和表现最出色的解决方案参与的结果。的顶级表现的方法利用的临床信息，数据AUG-心理状态，以及模型的合奏的共混物。这些发现使视网膜图像分析和基于图像的DR特别是筛选新的发展潜力。

©2019爱思唯尔保留所有权利。

**1.简介**

糖尿病视网膜病变（DR）和糖尿病性黄斑水肿（DME）是由于引起视网膜微血管最常见的威胁视力的医疗状况的变化由二abetes触发（[赖歇尔和SALZ 2015](#page26)），主要影响的劳动年龄人口在世界上（[阿特拉斯，2017年](#page24)）。DR导致在血管结构的逐渐变化（包括血管TOR-tuosity，分支角度和口径）和所得abnormali-关系（微动脉瘤，出血和渗出物），而，DME的特征在于液体潴留，或黄斑肿胀[在DR的任何阶段，可能会出现（Bandello等人，2010; Ciulla等人，](#page24) [2003）。据国际糖尿病联盟（阿特拉斯，2017年）](#page24) 估计，目前，受糖尿病影响的个体的全球数量为425万美元，并可能上升到6.93亿由2045他们当中，一个三个人了，估计有某种形式的DR，而十分之一的是容易视觉的威胁DR（[ICO，2017年; Bourne等人，2013](#page25)）。DR是通过目测检查视网膜眼底图像的一个或多个视网膜病变的存在等微动脉瘤（MAS），出血（HES），软诊断



渗出物（SES）和硬性渗出（EXS）（[Wong等人，2016](#page26)），如图 [图。1](#page2)。

早期诊断和治疗DR可以防止视力下降。因此，糖尿病患者通常称为视网膜丝网[荷兰国际集团一次或每年两次（摩天，1993; Kollias和Ulbig，2010;](#page24) [汀等人，2016）。糖尿病眼部护理，主要是依赖于](#page24) 眼科医生和必要的保健infrastruc-TURE的数量（[琼斯和爱德华兹，2010; Lin等人2016](#page25)）。在印度，眼科医生与人口的比例为1：107000，然而，在城市地区这一比例为1：9000，而在农村地区存在对608000楼的居民只有一个眼科医生（[拉曼等人，2016](#page26)）。到2045年，仅印度预计将有大约151万狮患有糖尿病和三分之一的人预计将有DR（[阿特拉斯，2017年](#page24)）。计划到屏幕这么大的弹出ulation为与实施，人为agement，人类平地机的可用性和长期FINAN-官方可持续性DR对垒的问题。因此，计算机辅助诊断工具重新获得性筛选这样一个人口众多，需要CON-tinuous随访DR，并有效地促进减少[在眼科医生的负担（Jelinek和克里，2009年; 沃尔特](#page25) [等人，2002年）。这样的工具可以帮助临床医生鉴定，](#page25)

**图。1。** 扩大区域（在左侧）的MA，和的HE和（右）的SE，并EXS：视网膜图像（在中心）通过突出显示与DR相关的正常结构（血管，视盘和中央凹中心）和异常的图示。

|  |  |
| --- | --- |
| *P. Porwal，S. Pachade和M. Kokare等。/医学图像分析59（2019）101561* | 3 |

解释和视网膜异常的测量，并最终筛选和监测疾病。在计算能力和机器学习最新的科学进展方法提供途径，生物医学科学家，以满足[临床实践的迫切要求（Shortliffe和布卢瓦，2006; 巴顿](#page26) [等人，2006年）。为了满足这一需求，与沿前的原始图像](#page26)CISE像素或图像级别专家注释（又名基础事实）发挥了重要作用，促进了开发，验证，和DR病变区隔，重刑技术比较研究界（[特鲁科等人，2013](#page26)）。与DR如MA，请的HE，SES和EXS相关的损伤的精确像素级注释是用于评估的个体LE-锡永分割技术的精度非常宝贵的资源。这些精确的分段病变有助于确定疾病的严重程度，并进一步充当一个路线图，可以在后续程序协助疾病的自来水进展。同样，在另一方面，图像层次专家LA-贝尔为DR和DME的疾病严重程度在图像分析和检索算法的devel的-opment和评估很有帮助。这必然导致几个研究小组研制[共享视网膜图像数据集，即获月（Decencière等人，](#page24) [2014），Kaggle（夸德罗斯和Bresnick，2009年），ROC（Niemeijer等人，](#page25) [2010），E-Ophtha（Decencière等人，2013），DiaretDB（考皮等，](#page25) [2012）， 驾驶 （](#page25)[面包车Ginneken等人，2004年](#page26)[），凝视（胡佛，1975年）](#page25) ARIA（[派睿电子等人，2008年](#page24)）和HEI-MED（[Giancardo等人，2012](#page24)）。

此外，两个挑战在DR，即视网膜病变在线挑战赛（ROC）的背景下举办[2](#page3) 和Kaggle DR去tection挑战[3](#page3)。ROC与检测MA的目标组织。然而，Kaggle挑战的目的是获取阻止-DR挖掘的严重级别的解决方案。这些挑战促进科研界的来自世界各地的上有竞争力的参与在科学进步的同时建设性设置启用在该领域的进展。上一页EF-堡垒已经使用图像分类，模式识别和机器学习方面取得良好进展。经过近二十年的进展，已有若干研究进行了系统回顾[组（巴顿等人，2006年。络纱机等人，2009; 阿布拉莫夫等人，](#page25) [2010; Mookiah等人，2013a; Jordan等人，2017; Nørgaard和](#page25) [Grauslund，2018）。](#page25)

尽管许多人的努力在该领域已取得一自动化DR筛选过程，病变检测仍然是一个CHAL-有挑战性的任务，由于以下几个方面：（1）病变的结构复杂（形状，尺寸，强度），（B）检测在tessel-迟来的图像和在噪声（亮边界反射，脉冲噪声，光反射），（C）高类间的相似性（即，MA-HE和EX-SE之间），和（d）的外观的存在病变的不那么未常见的非病变结构（神经纤维反射，容器反射，玻璃疣）使得二FFI崇拜以建立病变分割一个灵活和强大的模型。据我们所知，这个挑战之前，有一个单一框架的发展段没有报告所有病变（MA，HE，SE和EX）SI-multaneously。也，有缺乏通用平台来测试的决定在同一组图像中的正常和异常的视网膜结构的方法的稳健性。此外，没有的像素级注释和simultane-OU中分级为DR和DME有限可用性（参见表在[附录A](#page23)）。

为了解决这些问题，我们引入了一种叫做印度的糖尿病视网膜病变的图像数据集（IDRiD）新的数据集（[Porwal等人，2018A](#page25)）。此外，它被用来作为一个基础数据集盛大挑战的“糖尿病视网膜病变 - 分割与分级”组织连同ISBI - 2018年IDRiD数据集提供了典型的DR病变和正常视网膜结构的专家标记。它还提供疾病

* <http://webeye.ophth.uiowa.edu/ROC/>
* <https://www.kaggle.com/c/diabetic-retinopathy-detection>

DR和DME的严重性级别为数据库中的每个图像。这种挑战汇集了计算机视觉和生物医学研究人员的最终目标，进一步激发和亲微尘研究，以及提供一个独特的平台，为实用的软件工具的开发，将支持电子FFI古老而精确的测量和视网膜图像分析这可能是在DR管理有用。最初，地面实况一起训练数据集提供给与会者的其算法的开发。后来，结果被判定在对测试数据集，这些算法的性能。成功是由算法的结果如何密切配合地面实况测量。有三个主要子挑战：病变划分，疾病的严重程度分级，和本地化和视网膜地标分割。在IDRiD挑战这些多任务允许测试的算法的普遍性，这是什么使得它从现有的不同。此外，这一挑战寻求的自动化解决方案来同时预测DR和DME的严重性。据预测作为一个单独的任务来增加这一挑战相比Kaggle DR挑战，即给定图像为一体的双FFI culty水平，对于DR和DME预测的严重程度应该是正确的计数得分的任务。

本文的其余部分的结构如下： [第2节](#page3) 给出了自动配对DR筛查的发展做前期工作的简短评论， [第3节](#page5) 提供参考数据集的细节， [第4节](#page6) 介绍 比赛通过VARI-OU的阶段和组织 [第5节](#page8) 详细介绍了表现最佳的竞争如此lutions。 [第6节](#page15) 礼物性能在这个挑战中使用的评估措施。然后，[第7节](#page16) 呈现结果，分析和相应的参赛队所有子挑战的排名。 [第8节](#page20) 提供 在简短的讨论结果，限制和 吸取这一挑战，并在最后总结的经验教训。除了本文中，[附录A](#page23) 包括提供与所述数据集IDRiD不同国家的最先进的可公开获得的数据库的COMPAR-ISON。

**2.审查用于检测DR的视网膜图像分析的**

自动图像处理已被证明是对视网膜眼底图像的分析，并将其应用到未来的眼部护理一个有前途的选择。引进的DR筛查项目的自动化技术和快速增长的深学习技术实现有趣的结果是成功STO-里斯和未来的潜力成就的例子。特别是，再搜索的后（[Krizhevsky等人，2012](#page25)）深学习基于模型显示在国家的最先进的ImageNet挑战显著的改善，出现了深度学习基于模型的医学图像分析激增。因此，我们决定基于他们是否在DR的上下文中使用深度学习分类展示最新的相关工作。

*2.1。非深学习方法*

用于基于通过TRA-ditional手工特征视网膜图像分析方法的总体框架涉及若干阶段，通常为：对比度增强或非均匀性均衡，图像分割，特征前牵引，并且分类的预处理阶段。特征提取的策略变化的AC-盘带客观地参与，即视网膜病变检测，DIS-便于筛选或地标定位。2006年，一个研究小组（[巴顿等人，2006年](#page25)）在其视网膜图像分析是基于并讨论用于检测与DR相关的视网膜标和病变的初始技术中概述的原则。后来，[复卷机等。（2009年）](#page26) 报道DR的自动配对分析工作的分析，1998年 - 2008年，他们归类文学为一系列的操作或步骤的预处理，

4 *P. Porwal，S. Pachade和M. Kokare等。/医学图像分析59（2019）101561*

脉管分割，本地化和视神经盘（OD）的分割，黄斑中心凹和的定位，检测 [和病变的分割。一些评论文章（阿布拉莫夫](#page24) [等人，2010; 乔丹等人，2017年）提供的简要介绍](#page24) 定量方法对眼底图像与视网膜病变和自动化技术的识别FO-CUS用于大规模筛选视网膜疾病的分析。

在文献中的尝试大部分是针对前clusive检测和/或一种类型的从图像病变（EI-疗法的MA，的HE，EXS或SES）的分割。一些COM-MON办法涉及的病变划分为mathemat-[iCal的形态（乔希和Karule，2019; Hatanaka等人，2008;](#page25) [张某等人。2014年），区域生长（Fleming等，2006; 李和](#page24) [Chutatape，2004年），和监督的方法（Wu等人，2017年。周](#page26) [等人，2017; Garcia等人，2009; 唐等人，2013）。除了这些](#page26) 模板匹配的方法，MAS中的情况下，大部分初始的研究表明EF-fectiveness（[Quellec等人，2008年](#page26)），熵阈值（[Das等人，2015年](#page24)），氡空间（[Giancardo等人，2011](#page24)），稀疏表示（[张等人，2012。Javidi等人，2017年](#page26)），HES-仙基于区域描述符（[阿达勒等人，2014](#page24)）和字典学习（[罗沙等人，2012](#page26)）。在另一方面，专用seg-[精神质的HE的，超象素的基于特征（Tang等，2013;](#page26) [罗梅罗 - Oraá等人，2019）被发现是有效的。这些红色LE-](#page26)sions（二者MAS和HES）也经常检测到一起我们-ING动态形状特征（[Seoud等人，2016](#page26)），滤波器响应和多个内核学习（[Srivastava等人，2017年](#page26)）和混合特征提取的方法（[Niemeijer等人，2005年](#page25)）。类似地，对于EXS，重新搜索依靠像聚类方法（[Osareh等人，2009](#page25)），基于模型的（[Sanchez等人，2009; Harangi和豪伊杜2014](#page26)），蚁群算法（ACO）（[Pereira等人，2015](#page25)）和在形成上下文（[Sanchez等人，2012](#page26)）。而对于社会企业的研究人员UTI-lized尺度不变特征变换（SIFT）（[纳克维等人，2018](#page25)），自适应阈值和ACO（[SRENG等人，2019](#page26)）。此外，SEV-全部擦除方法被设计成用于多种病变检测这样[作为多尺度幅度调制频率调制（阿古](#page24) [等人，2010），机器学习（](#page24)[Roychowdhury等人，2014](#page26)[），一个COM](#page24)海赛多尺度分析的bination，变分割和纹理特征（[菲格雷等人，2015年](#page24)）。这些技术被示出为通常涉及在检测到与病变检测anatom-的iCal结构（即OD和中央凹）的相互依存关系，而这又确定自动化DR筛选结果。

定位和OD的分割和中央凹视网膜促进病变的检测以及评估（基于这些病变的几何位置）的严重程度和DR和DME的周一itoring进展。因此，有几种方法已被提出用于OD的定位，且大多使用像强度，形状，颜色，纹理等的OD性质和许多其他显示数学MOR-phology的有效性（[莱斯等人，2013; 马林等人，2015年](#page25)），模板 [匹配（Giachetti等人，2014），可变形模型（Yu等人，](#page26) [2012; 吴等人。年，2016年）和强度分布分析（Kamble](#page25) [等人，2017; 乌里韦 - 瓦伦西亚和马丁内斯Carballido，2019）。毛皮-](#page25)疗法，方法用于OD分割是基于水平集（[Yu等人，2012](#page26)），阈值（[马林等人，2015年](#page25)），活性 [轮廓（玛丽等人，2015年）和形状建模（Cheng等，](#page24) [2015年），聚类（塔库尔和Juneja，2017年），以及混合（Bai等，](#page24) [2014）方法。类似地，当检测到中央凹大多采用](#page24) 通过的细胞形态与OD和容器的几何关系[逻辑（Welfer等人，2011），阈值（Gegundez-Arias等人，](#page24) [2013），模板匹配（](#page24)[Kao等人，2014](#page25)[）和强度亲](#page24)文件分析（[Kamble等人，2017年](#page25)）技术。在检测正常解剖结构的性能不佳可能病变的检测和筛查准确度产生不利影响。例如，基于CON-代尔数学形态学技术在2002年提出的（[Walter等人，2002](#page26)），2008年（[Sopharak等人，2008年](#page26)）和2014（[张某等人。2014年](#page26)）。这些作品展示了如何从形态

通过包括用于最终目标渗出物检测的多个步骤演进逻辑基于处理的办法。在最初的努力，[Walter等。（2002年）](#page26)设计 为OD和EXSSeğmen市-塔季翁的技术，随后除去OD获得EX候选。同样的，[Sopharak等。（2008年）](#page26)实现 同样的目标与OD和船只的DETEC，重刑和拆除。最近，一种方法通过预sented[张某等人。（2014）](#page26) 实现更好的结果，但它涉及到（a）中的空间校准，（b）中分别检测黑暗和明亮的解剖结构诸如血管和OD的，也（c）中亮的边界区域的实际提取之前检测candi-[日期。此外，还有基于纹理其他技术（莫拉莱斯](#page25) [等人，2017; Porwal等人，2018C）中级（皮雷斯等人，2017年）](#page25) 视网膜图像，对于DR放弃病变分割步骤筛选的功能。然而，大多数这些技术依赖于上面提到的中间步骤。在基于MA-茅根学习的方法（[Roychowdhury等人，2014](#page26)），作者检测出亮部和暗病变作为第一步和以后进行分层病变分类以生成用于DR一个严重性等级。同样的，[安塔尔和豪伊杜（2014）](#page24) 建议 涉及影像级战略 质量评估，预先筛选之后病变和anatom，iCal的特征提取，最后作出决定有关DR采用集成分类的存在。此外，对于DR的不同阶段，形态区域属性的标识（[Yun等人，2008年](#page26)），纹理参数（[阿查等人，2012; Mookiah等人，2013b](#page24)），高阶谱的非线性特性（[阿查里雅等人，2008年](#page24)） [混合（达拉等人，2015）和信息融合（Niemeijer等人，](#page25) [2009年）方法被发现是有用的。由于DME分级依据](#page25) [EXS从黄斑的位置，许多研究者（Giancardo](#page24) [等人，2012; Medhi和Dandapat 2014; 佩尔多莫等人，2016;](#page24) [马林等人，2018）提出了基于EXS功能，以确定](#page24) [DME的严重性。虽然几个人（迪帕克和Sivaswamy，](#page24) [2012; Mookiah等人，2015; 阿查等人，2017年）已经提出](#page24) 各种特征提取技术等级DME阶段没有分割EXS。主要用于本节中的方法，特征是基于颜色，亮度，尺寸，形状，边缘强度，纹理，并且在空间和/或变换域的像素簇的上下文信息。而分类是通过分类如K近邻（KNN），朴素贝叶斯，支持向量马中电信（SVM），人工神经网络（ANN），决策树等实现

这些病变检测或筛选技术被示出为通常涉及与检测其他土地标记的相互依存关系。然而，也缺乏一个单一的平台，以测试他们对每个目标的性能。对于这样的基于手工功能的方法，这一挑战提供了一个独特的平台，比较和对比算法的性能进行检测的解剖结构，病变以及DR和DME的筛选。

*2.2。深学习方法*

深度学习是一个通用术语来定义能够同时直接从数据学习低级别的代表性和更高级别的参数多层NEU-RAL网络。这种表示的学习能力大大减少了工程临时需要的功能，但是，全终端到高端培训的深度学习为基础的方法通常需要SAM-普莱斯的显著数量。它在最近一个时期的快速发展主要是由于数据的不断进步，在学习算法的计算能力和devel的-opments大量涌入是启用的建设[多层（多于两个）网络（欣顿，2018; Voulodimos](#page25) [等人，2018）。这一进展引起在创造利益](#page25) 基于健康信息的机器学习分析，数据驱动模型（[Ching等人，2018; Ravıet人，2017年](#page24)）。因此，它是EMERG-ING作为机器学习的有效工具，并承诺将重塑[自动化医学图像分析的未来（斯潘等人，](#page24) [2016年 Litjens等人，2017; 铃木，2017年; Shen等人，2017; Kim等人，](#page24)

|  |  |
| --- | --- |
| *P. Porwal，S. Pachade和M. Kokare等。/医学图像分析59（2019）101561* | 五 |

[2018; KER等人，2018](#page24)）。在各种方法变种 深度学习，卷积神经网络（细胞神经网络或ConvNets） [最流行的医学图像分析领域（灏昌](#page25) [等人，2016; 凯琳和Pencina，2018）。一些配置和](#page25) CNN的变体可在文献中，一些最 [流行的是AlexNet（Krizhevsky等人，2012），VGG（西蒙尼扬和](#page26) [Zisserman 2014），GoogLeNet（Szegedy等人，2015年）和RESNET](#page26) （[他等人。年，2016年](#page25)）。

深度学习也被广泛应用在视网膜图像分析，因为其独特的保护LO-CAL图像关系的特征。在文献中采用由利用“关闭的，现成的CNN”深学视网膜图像的方法多数设有作为补充信息信道到其它的手工制作的特征或局部显着性映射检测abnor-的[malities与DR（相关胡齐克等人，2018; 奥兰多等，](#page24) [2018; Dai等人，2018），OD的分割（Zilly等人，2017; 福](#page26) [等人，2018），和DR的检测（Rangrej和Sivaswamy，2017年）。](#page26) 作者（[Fu等人，2016](#page24)）雇用完全连接条件随机场与CNN沿着集成像素之间的区别VES-SEL概率图和远距离相互作用，以获得最终的二进制脉管系统。而一些方法INI-tialized与那些预训练模型（非医学图像）的参数，然后在“微调”（[塔杰巴赫什等人，2016](#page26)） [对于DR筛选网络参数（高尔杉等人，2016;](#page24) [卡森Lam等人，2018）。在使用另一种方法研究](#page24) 二维（2D）图像块作为输入，而不是全[对病变检测大小的图像（Tan等人，2017b; 面包车Grinsven](#page26) [等人，2016; Lam等，2018; 胡齐克等人，2018; Khojasteh](#page26) [等人，2018），以及OD和中央凹检测（Tan等人，2017A）。](#page26) [加西亚等人。（2017）](#page24)熟练 在“CNN从零开始”，并与基于其他两个现有的AR-chitectures微调结果COM-削减它。最近，[Shah等。（2018）](#page26) 证明了自动编码器可以刺激多样性的学习视觉内核检测异常的DIC-tionary是ensem-BLE培训。而[Giancardo等。（2017）](#page24)建议 一种新颖的方式来计算vascu-lature嵌入，它利用一个新的编码器增强CNN的内部表示，表明在DR CLASSI-fication和检索任务的改进。有使用CNN模型在最近的时间DR的自动识别显著的发展。定制的CNN（[Gargeya和冷，2017年](#page24)）提出了DR筛选和使用从EyePACS系统获得的75137幅图像（训练[夸德罗斯和Bresnick，2009年](#page24)），其中一个附加CLASSI-费里的CNN-导出的特征进一步用于阻止地雷如果图像是具有或不具有视网膜病变。同样，谷歌公司（[高尔杉等人，2016](#page24)）开发了用于图像分类而优化的网络（微调）中，其中一个是CNN由受过训练的UTI-LIZING由128175个图像的带标签的回顾性开发数据库。视网膜病变有一些混合算法，其中多个，半依赖性CNN的是基于训练的出现-ANCE（[阿布拉莫夫等人，2016; Quellec等人，2016](#page24)）。

* 另一步骤中，研究人员（[Quellec等人，2017年](#page26)）展示了基于CNN病变分割的能力训练图像级分类。然而，[Lynch等。（2017）](#page25) 证明了基于多个半依赖的细胞神经网络的混合算法可能会提供DR转诊筛查更稳健的选择，强调病变细分的重要性。对于进一步的细节，读者建议遵循检测EXU-日期的最新评论（[FRAZ等人，2018](#page24)），红色病变（[Biyani和Patre，2018](#page24)）和重点是DR的计算机辅助诊断系统综述（[Mookiah等人，2013a; Nørgaard和Grauslund，2018](#page25)）。

在人工智能这一目前的进展提供了一个oppor-tunity研究人员提高DR转诊系统的性能提升到一个更强大的诊断系统，可以提供针对多种疾病匹配的临床相关的国际标准泉titative信息。因此，所提出的挑战去符号提供一个途径，以评估精确DR严重性状态和OP-

portunity提供病灶准确的措施，这可能甚至在后续的研究帮助，观察视网膜图谱的变化。

**3.印度糖尿病性视网膜病变的图像数据集**

该IDRiD数据集（[Porwal等人，2018A](#page25)）从在位于Nanded的，（MS），印度眼科诊所获得的真实临床检查创建。的受糖尿病患者视网膜图片进行帽捕获的原始使用科瓦VX注重与黄斑- 10*α*眼底照相机。之前的图像采集，所有受试者的瞳孔与浓度为0.5％托吡卡胺的一滴扩张。捕获的IM-年龄有50个◦ 的视场，4288分辨率 ×2848点的像素，并且存储在jpg格式。最终的数据集是由516 IM-年龄分为五个DR*（*0 - 4*）* 三个DME *（*0 - 2*）*班，根据临床相关的国际斯坦 - dards定义良好的特性。它提供了典型的DR病变和正常视网膜结构的专家标记。它还提供了DR和DME的数据库中的每个图像的疾病严重程度。三种类型的地面真理的数据集中可用：

1. *像素级译注：* 这种类型的注释是有用的技术来定位图像内的各个病灶和SEG-彪出从背景感兴趣的区域。与DR的迹象八十一眼底彩色照片的像素级开发的均线，社企，EXS和的HE地面实况进行了注释。该二进制掩码（如在所示[图2](#page6)）对于每种类型的病变是亲vided在TIF文件格式。此外，OD也被注解在像素级和二进制掩模以相同的格式被提供所有81个图像。所有这些注解发挥研究至关重要的作用图像内的分割病变的计算分析。
2. *图片等级评分标准：* 它由意德隶与整个图像相关的整体风险因素的信息。两名医学专家提供的裁决一致等级的全套516个的图像与各种DR和DME的病理状态。分级所有图像是在CSV文件中。将二abetic视网膜图像被划分为根据单独的组[国际临床糖尿病性视网膜病变量表（Wu等人，](#page26) [2013），下观察限于图像，如图](#page26) [表格1](#page5)[。](#page26)

基于EXS的出现接近黄斑中心区（DME严重程度决定[Decencière等人，2014](#page24)），如图 [表2](#page5)。

1. *OD和中央窝中心坐标* 的OD和中央凹中心LO-阳离子被标记为所有516个图像和标记是可作为单独的CSV文件。

**表格1**

DR严重程度分级。NPDR：非增殖性DR和PDR：增殖DR

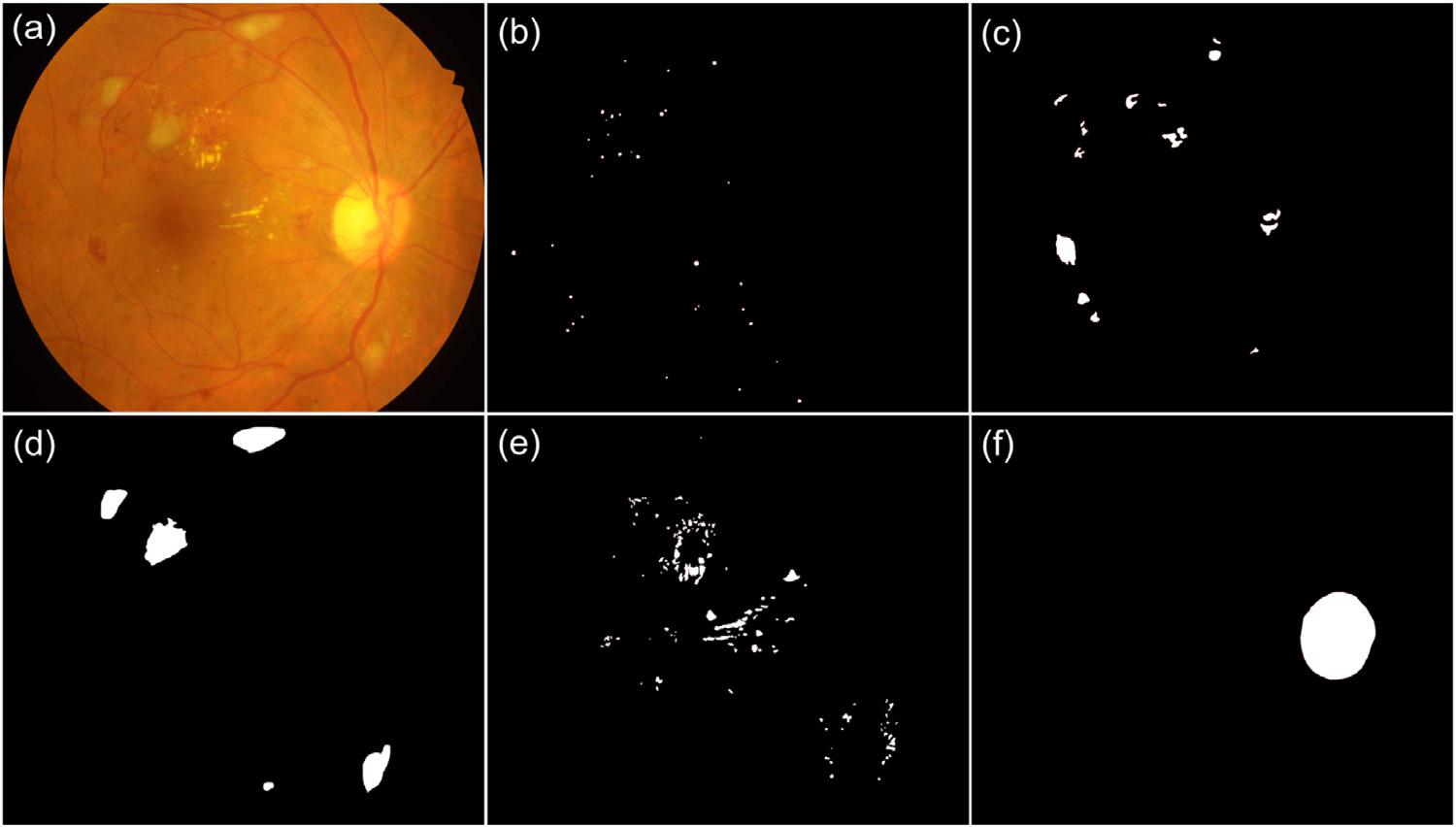
|  |  |  |
| --- | --- | --- |
| DR等级 | 发现 | |
|  |  |  |
| 0：无明显的视网膜病 | 异常的可见迹象 | |
| 1：轻度NPDR | 均线的存在只 | |
| 2：中度NPDR | 不仅仅是均线，但低于严重NPDR | |
| 3：严重NPDR | 以下任何方式： | |
|  | *• >*20点视网膜内的HE | |
|  | *•* 静脉串珠 | |
|  | *•* 视网膜微血管异常 | |
|  | *•* 没有PDR的迹象 | |
| 4：PDR | 任一或两个以下： | |
|  | 新生血管 | |
|  | 玻璃体/预视网膜HE | |
| **表2** |  |  |
| 风险二甲醚。 |  |  |
|  |  |  |
| DME GradeFindings |  |  |

* 没有明显的EX（S）
* 一个盘直径的从半径之外EX（一个或多个）的存在

黄斑中心

* 一个盘直径为黄斑中心的半径内EX（一个或多个）的存在

6 *P. Porwal，S. Pachade和M. Kokare等。/医学图像分析59（2019）101561*



**图2中。** 视网膜照片和不同的像素级别的注解：选自（a）样品的眼底图像的IDRiD数据集; 样品基础事实为（BF）的MA，的HE，社会企业，分别EXS和OD。

该IDRiD数据集可从IEEE数据端口库[4](#page6)下一个知识共享署名4.0许可。有关数据更详细的信息在数据描述符是可用的（[Porwal等人，2018B](#page25)）。 [表A.1](#page23) 和 [A2](#page23) 强调这个数据集的相对优势相对于现有的数据集。IDRiD是提供三种类型的注释以上男性，tioned唯一的数据集。这种精简的注解集将人低它在研究和带来更好的普及机型的发展，为图像分析利用，实现自动化DR诊断方面取得进一步进展。

**4.挑战组织**

在“糖尿病视网膜病变 - 分割和分级CHAL-lenge”是由成不同阶段，给人一种井井有条的工作过程中对大赛的成功增强。 [图3](#page7) 描述了整个挑战组织的工作流程。面临的挑战是ØFFI在ISBI cially公布 - 2018网站[五](#page6) 15*日* 2017年10月。

我们面临的挑战又分为三个子挑战FOL - 低点：

1. 病变分割：视网膜病变分割associ-ated与DR如MA，请的HE，EXS和SE。
2. 疾病评分标准：根据DR和DME的严重性级别眼底图像的分类。
3. OD检测与分割，以及中央窝检测：OD的自动马蒂奇定位和中央凹中心坐标，OD的分割。

我们面临的挑战涉及4个阶段，具体如下：

*阶段1。数据准备和分发：*The IDRiD dataset was adopted for this challenge, where experts verified that all images are of adequate quality, clinically relevant, that no image is dupli-cated and that a reasonable mixture of disease stratification repre-sentative of DR and DME is present. The dataset along with ground truths was separated into a training set and test set. For images with pixel-level annotations, data was separated as 2/3 for training (Set-A) and 1/3 for testing (Set-B) (See [Table 3](#page6)).

* <https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid>

5 <https://biomedicalimaging.org/2018/challenges/>

**Table 3**

Stratification of retinal images annotated at pixel level for different types of retinal lesions.

|  |  |  |
| --- | --- | --- |
| Lesion Type | Set - A Images | Set - B Images |
|  |  |  |
| MA | 54 | 27 |
| HE | 53 | 27 |
| SE | 26 | 14 |
| EX | 54 | 27 |
|  |  |  |

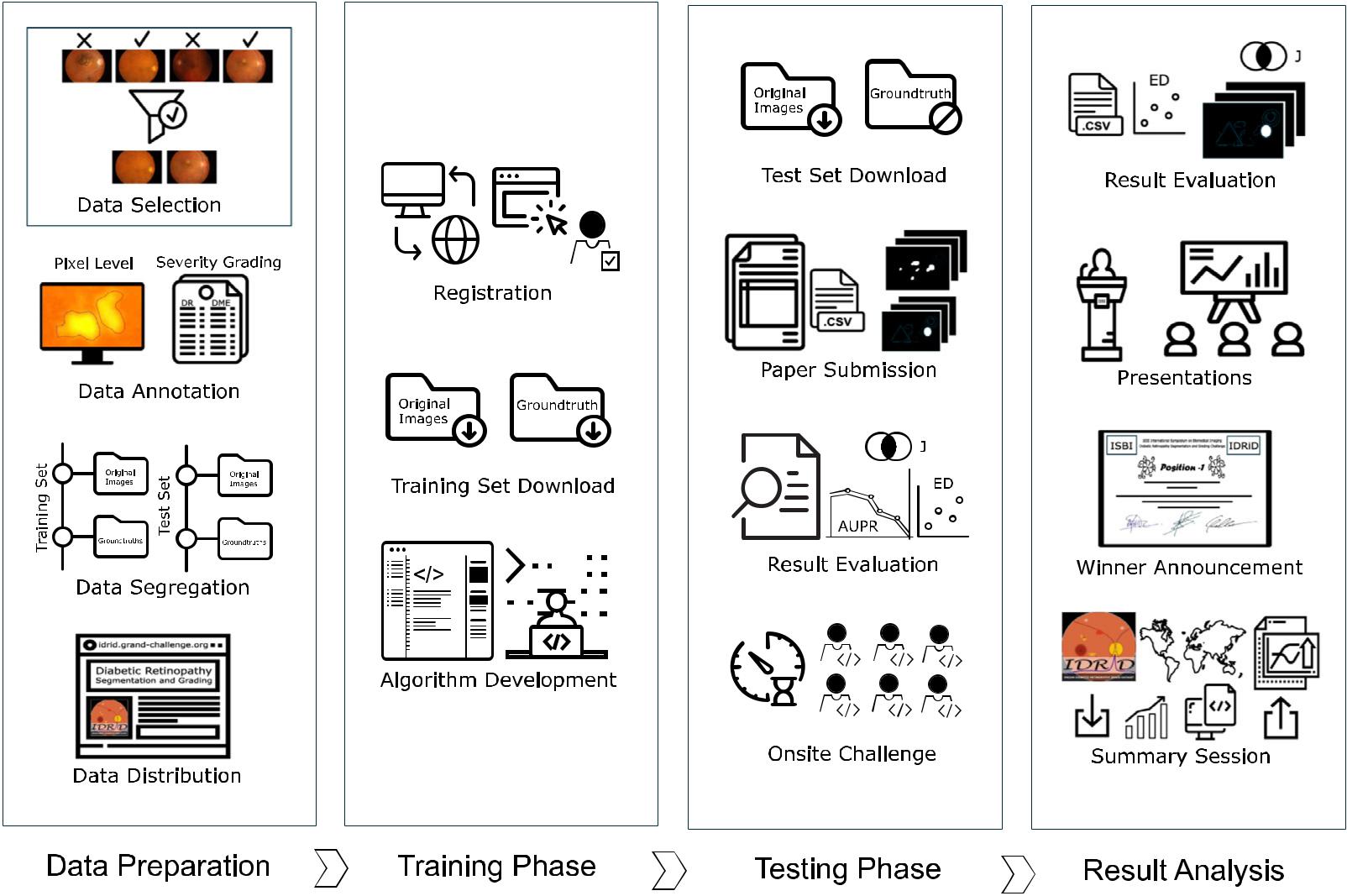
**Table 4**

Stratification of retinal images graded for DR and DME.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| DR Grade | Set-A | Set-B | DM Grade | Set-A | Set-B |
|  |  |  |  |  |  |
| 0 | 134 | 34 | 0 | 177 | 45 |
| 1 | 20 | 5 | 1 | 41 | 10 |
| 2 | 136 | 32 | 2 | 195 | 48 |
| 3 | 74 | 19 |  |  |  |
| 4 | 49 | 13 |  |  |  |
|  |  |  |  |  |  |

Similarly, data for OD segmentation (part of sub-challenge – 3) was divided in the same ratio into Set-A (54 images) and Set-B (27 images). Since the output of algorithms would be repre-sentative of learned perceptive patterns. The data for lesion and OD segmentation tasks were carefully split in such a way that it provides enough representative data to be learned and a holdout proportion that could be later used to gauge the algorithm per-formance. The percentage of images that should be in each sub-set for lesion and OD segmentation tasks (sub-challenge – 1 and part of sub-challenge – 3) was supported by the research out-come ([Dobbin and Simon, 2011](#page24)) which demonstrated that split-ting data into 2/3 (training): 1/3 (testing) is an optimal choice for the sample sizes from 50 to 200. For other sub-challenges (disease grading, and OD and fovea center locations), data was separated in 80 (Training set: Set-A): 20 (Testing set: Set-B) ratio. The per-centage of data split, in this case, is done to provide an adequate amount of data divided into different severity levels. Note that the dataset was stratified according to the DR and DME grades before splitting. A breakdown of the details of the dataset is shown in [Table 4](#page6).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561* | | | | | | | | | | | | | | | | | | | | | | | | 7 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |



**Fig. 3.** Workflow of the ISBI - 2018: Diabetic Retinopathy – Segmentation and Grading Challenge.

The challenge was hosted on Grand Challenges in Biomedical Imaging Platform[6](#page7), one of the popular platforms for biomedical imaging-related competitions. A challenge website was set up and launched on 25*th* October 2017 to disseminate challenge related in-formation. It was also used for registration, data distribution, sub-mission of results and paper, and communication between orga-nizers and participants.

*Stage-2. Registration and release of the training data:* Registra-tion of challenge for consideration to ISBI on-site contest was open from the launch of the grand-challenge website (i.e. 25*th* October 2017) till the deadline for submission of results (i.e. 11*th* March 2018). Interested research teams could register through challenge website for one or all sub-challenges. The first part of data, i.e., Set-A (images and ground truths) was made available to participants of challenge on 20*th* January 2018. Participants could download the dataset and start development or modification of their methods. Further, they were also allowed to use other datasets for the devel-opment of their methods, with a condition that external datasets should be publicly available.

*Stage-3. Release of test data:* Set-B (only images) for sub-challenge – 1 was released on 20*th* February, 2018. For other two sub-challenges, Set-B was released on 4*th* April which was part of ‘on-site’ challenge. Organizers refrained from an on-site evaluation of sub-challenge – 1 considering timing constraints in the evalua-tion of results for image segmentation tasks.

Submissions were sought for either of the following 8 different tasks corresponding to three sub-challenges (1 – Lesion Segmenta-tion, 2 – Disease Grading, 3 – OD and Fovea Detection) as follows:

1. Sub-challenge – 1: Lesion Segmentation

Task - 1: MA Segmentation

Task - 2: HE Segmentation

Task - 3: SE Segmentation

Task - 4: EX Segmentation

1. Sub-challenge – 2: Disease Grading

* <https://grand-challenge.org/>

Task - 5: DR and DME Grading

1. Sub-challenge – 3: OD and Fovea Detection

Task - 6: OD Center Localization

Task - 7: Fovea Center Localization

Task - 8: OD Segmentation

Challenge site was made open for submission from 12*th* Febru-ary and participants could submit their results and paper describ-ing their approach to the organizers till 11*th* March. Participants could submit up to three methods to be evaluated per team for each task, provided that there was a significant difference between the techniques, beyond a simple change or alteration of parame-ters. For tasks 1 to 4 (i.e. sub-challenge – 1) and task-8, teams were asked to submit output probability maps as grayscale images and for all other tasks, it was accepted in CSV format. The submitted results were evaluated by challenge organizers and their perfor-mance was displayed on the leaderboard of the challenge website. For sub-challenge – 1, teams were assessed based on the perfor-mance of results submitted on a test set, whereas for other two sub-challenges assessment was done using results on a training set obtained through leave one out cross-validation approach. In this phase, it received a very good response from the research com-munity with 148 submissions by 37 different teams, out of which 16 teams were shortlisted for participation to on-site challenge. Amongst invited, 13 teams confirmed their participation in the on-site challenge, whereas, two teams declined to participate due to other commitments and one team was not able to arrange finan-cial support in the limited time.

*Stage-4. ISBI Challenge Event:* The main challenge event was held in conjunction with ISBI - 2018 on April, 4*th* 2018. The Set-B (only images) for sub-challenge – 2 and 3 was made available to the par-ticipants via challenge website (on-line mode) as well as portable devices at the challenge site (off-line mode). Participants were asked to produce results for the respective challenge task within one hour. The participating teams could bring their own system or run the test through the remote system. Also, there was no restric-tion on the number of machines that could be used to produce the results. However, considering the timing constraints for processing,

8 *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561*

**Table 5**

List of all participating teams shortlisted and which participated in the ‘on-site’ challenge. All teams are color-coded for easier reference in all further listings. The DL denotes whether the submitted algorithm is based on deep learning. Where, sub-challenge – 1 (SC1) corresponds to lesion segmentation such as microaneurysms (MA), hemorrhages (HE), soft exudates (SE) and hard exudates (EX). Whereas, sub-challenge – 2 (SC2) denotes dis-ease severity grading corresponding to DR and DME. Similarly, sub-challenge – 3 (SC-3) deals with the optic disc detection (ODD), fovea detection (FD) and optic disc segmentation (ODS). Harangi et al. participated with two methods HarangiM1 and HarangiM2, for simplicity it is jointly represented as HarangiM1-M2 with a single color code. Similarly, Li et al. participated with two methods LzyUNCC (renamed in the text as LzyUNCC-I) and LzyUNCC\_Fusion (renamed in the text as LzyUNCC-II) that are jointly represented as LzyUNCC with same color code. However, these different methods are mentioned separately in the text wherever it was necessary. ∗ Team could not participate in ‘on-site’ challenge but later communicated the results to the organizers.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Team Name | | | Authors | DL | SC1 | |  |  |  | SC2 | SC3 | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | MA | HE | SE | EX |  | ODD | | FD | ODS |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | VRT | Jaemin Son et al. |  |  | |  |  |  |  |  | |  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  | iFLYTEK-MIG | Fengyan Wang et al. |  |  | |  |  |  | × | × | | × | × |  |  |
|  |  |  |  |
|  |  | PATech | Liu Lihong et al. |  |  | |  | × |  | × | × | | × | × |  |  |
|  |  |  |  |
|  |  | SOONER | Yunzhi Wang et al. |  |  | |  |  |  | × | × | | × | × |  |  |
|  |  |  |  |
|  |  | SAIHST | Yoon Ho Choi et al. |  | × | | × | × |  | × | × | | × | × |  |  |
|  |  |  |  |
|  |  | LzyUNCC | Zhongyu Li et al. |  | × | | × |  |  |  | × | | × | × |  |  |
|  |  |  |  |
|  |  | SDNU | Xiaodan Sui et al. |  |  | |  |  |  | × |  | |  |  |  |  |
|  |  |  |  |
|  |  | Mammoth | Junyan Wu et al. |  | × | | × | × | × |  | × | | × | × |  |  |
|  |  |  |  |
|  |  | HarangiM1-M2 | Balazs Harangi et al, |  | × | | × | × | × |  | × | | × | × |  |  |
|  |  |  |  |
|  |  | AVSASVA | Varghese Alex et al. |  | × | | × | × | × |  | × | | × | × |  |  |
|  |  |  |  |
|  |  | DeepDR | Ling Dai et al. |  | × | | × | × | × | × |  | |  | × |  |  |
|  |  |  |  |
|  |  | ZJU-Bll-SGEX | Xingzheng Lyu et al. |  | × | | × | × | × | × |  | |  |  |  |  |
|  |  |  |  |
|  |  | IITkgpKLIV | Oindrila Saha et al. |  | × | | × | × | × | × | × | | × |  |  |  |
|  |  |  |  |
|  |  | ∗ CBER | Ana Mendonça et al. | × | × | | × | × | × | × |  | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |



some teams which had previously entered with more than one so-lution decided to use only their best performing solution.

Further, the top three teams from sub-challenge – 1 were given the opportunity to present their work. During that time, some of the organizing team members compiled the results for sub-challenge – 2 and 3. The teams were given 7 minutes for presenta-tion of their approach and 3 minutes were reserved for question-answers. The first presentation session lasted for about 30 minutes and at the end of presentations of sub-challenge – 1 the results for sub-challenge – 2 and 3 were declared. Similarly, the top three performing teams from these sub-challenges gave short presenta-tions on their work. After the end of the on-site challenge event, on 6*th* April, the summary of challenge and analysis of results was presented, which included a final ranking of the competing solu-tions. This information is additionally accessible on the challenge website. It is important to note that many teams had participated in multiple sub-challenges as listed in [Table 5](#page8) and the remainder of this paper deals only with the methods that were selected for the challenge.

**5. Competing solutions**

Majority of participating teams proposed a CNN based approach for solving tasks in this challenge. This section details the ba-sic terminologies and abbreviations related to CNN and its vari-ants utilized by participating teams. Further, it summaries the so-lutions and related technical specifications. For the detailed de-scription of a particular approach, please refer to proceedings of [ISBI Grand Challenge Workshop at https://idrid.grand-challenge.](https://idrid.grand-challenge.org/Challenge_Proceedings/) [org/Challenge\_Proceedings/.](https://idrid.grand-challenge.org/Challenge_Proceedings/)

For the input image, CNN transforms raw image pixels on one end to generate a single differentiable score function at the other end. It exploits three mechanisms — sparse connections (a.k.a. local receptive field), weight sharing and invariant (or equivariant) rep-

[resentation — that makes it computationally eﬃcient (Shen et al.,](#page26) [2017). The CNN architecture typically consists of an input layer fol-](#page26)lowed by sequence of convolutional (CONV), subsampling (POOL), fully-connected (FC) layers and finally a Softmax or regression layer, to generate the desired output. Functions of all layers are de-tailed as follows:

CONV layer comprises of a set of independent filters (or ker-nels) that are utilized to perform 2D convolution with the input layer (I) to produce the feature (or activation) maps (A) that give the responses of kernels at every spatial position. Mathematically, for the input patch *(Ix,y* *)* centered at location (x, y) of *th* layer, the feature value in i*th* feature map, A*x,y,i,* is obtained as:

|  |  |
| --- | --- |
| *Ax,y,i* = *f ((wi )T Ix,y* + *bi )* = *f (Cx,y,i )* | (1) |
| Where the parameters w*i* and b*i* | are weight vector and bias |

term of i*th* filter of *th* layer, and f( · ) is a nonlinear activation func-tion such as sigmoid, rectified linear unit (ReLU) or hyperbolic tan-gent (tanh). It is important to note that the kernel w*i* that gen-erates the feature map C:*,*:*,i* is shared, reducing model complexity and making network easier to train.

POOL layer aims to achieve translation-invariance by reducing the resolution of feature maps. Each unit in a feature map of POOL layer is derived using a subset of units within sparse connections from a corresponding convolutional feature map. The most com-mon pooling operations are average pooling and max pooling. It performs downsampling operation and is usually placed between two CONV layers to achieve a hierarchical set of image features. The kernels in initial CONV layers detect low-level features such as edges and curves, while the kernels in higher layers are learned to encode more abstract features. The sequence of several CONV and POOL layers gradually extract higher-level feature representation.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561* | | | 9 |  |
| FC layer aims to perform higher-level reasoning by computing | | | | | | | | formation from a shallow, fine layer to produce accurate and de- | |  |
| the class scores. Each neuron in this layer is connected to all neu- | | | | | | | | tailed segmentations. |  |  |
| rons in the previous layer to generate global semantic information. | | | | | | | | For lesion segmentation task, most of the participating teams | |  |
| The last layer of CNN’s is an output layer (O), here the Softmax | | | | | | | | exploit U-Net architecture ([Ronneberger et al., 2015](#page26)). The main | |  |
| operator is commonly used for classification tasks. The optimum | | | | | | | | idea in U-Net architecture is to supplement the usual contract- | |  |
| parameters (*θ* , a common notation for both w and b) for a partic- | | | | | | | | ing network through a symmetric expansive path by addition of | |  |
| ular task can be determined by minimizing the loss function (L) | | | | | | | | successive layers, where upsampling (via deconvolution) is per- | |  |
| defined for the task. Mathematically, for N input-output relations | | | | | | | | formed instead of the pooling operation. The upsampling part con- | |  |
| {(I*n, On*); n ∈ [1, , N]} and corresponding labels G*n* | | | | | | | the loss can | sists of a large number of feature channels, that allow the net- | |  |
| be derived as: | | |  |  |  |  |  | work to propagate context information to higher-resolution layers. | |  |
|  | 1 *N* | |  |  |  |  |  | The high-resolution features from the contracting path are merged | |  |
|  | *n* |  | *n* |  |  | with the upsampled output and fed to soft-max classifier for pixel- | |  |
| *L* = |  | *n*=1 *ln(θ* ; *G* |  | *, O* |  | *)* | (2) |  | |  |
| *N* |  |  | wise classification. This network works with very few training im- | |  |
|  | | |  |  |  |  |  | ages and enables the seamless segmentation of high-resolution im- | |  |
| Where N denotes the number of training images, I*n, On* and | | | | | | | | ages by means of an overlap-tile strategy. Other similar architec- | |  |
| *Gn* correspond to n*th* training image. Here, a critical challenge in | | | | | | | | ture SegNet ([Badrinarayanan et al., 2015](#page24)) was opted by a team; it | |  |
| training CNN’s arises from the limited number of training sam- | | | | | | | | consists of an encoder and decoder network, where the encoder | |  |
| ples as compared to the number of learnable parameters that need | | | | | | | | network is topologically identical to CONV layers in VGG16 and in | |  |
| to be optimized for the task at hand. Recent studies have devel- | | | | | | | | which FC layer is replaced by a Softmax layer. Whereas, the de- | |  |
| oped some key techniques to better train and optimize the deep | | | | | | | | coder network comprises a hierarchy of decoders, one correspond- | |  |
| models such as data augmentation, weight initialization, Stochastic | | | | | | | | ing to each encoder. The decoder uses max-pooling indices for up- | |  |
| Gradient Descent (SGD), batch normalization, shortcut connections, | | | | | | | | sampling its encoder input to produce sparse feature maps. Later, | |  |
| and regularization. For more understanding related to advances in | | | | | | | | it convolves the sparse feature maps with a trainable filter bank to | |  |
| [CNN’s, the reader is recommended to refer a paper by Gu et al.](#page24) | | | | | | | | densify them. At last, decoder output is fed to a soft-max classifier | |  |
| [(2018).](#page24) | | |  |  |  |  |  | for the generation of segmentation map. One team choose Mask R- | |  |
| The growing use of CNN’s as the backbone of many visual | | | | | | | | CNN ([He et al., 2017](#page25)), a technique primarily based on a Region Pro- | |  |
| tasks, ready for different purposes (such as segmentation, classi- | | | | | | | | posal Network (RPN) that shares convolutional features of an entire | |  |
| fication or localization) and the available data, has made architec- | | | | | | | | image with the detection network, thus enabling region proposals | |  |
| ture search a primary channel in solving the problem. | | | | | | |  | to localize and further segment normal and abnormal structures in | |  |
| In this challenge, mainly for disease severity grading problem, | | | | | | | | the retina. RPN is a fully convolutional network that contributes to | |  |
| participants either directly utilized existing variants of CNN’s or | | | | | | | | concurrently predicting object bounds and “objectness” scores at | |  |
| ensembled them to demarcate the input image to one of the | | | | | | | | each position. |  |  |
| classes mentioned in [Table 4](#page6). Several configurations and variants | | | | | | | | Following subsections present the solutions designed by partic- | |  |
| of CNN’s are available in the literature; some of the most popular | | | | | | | | ipating teams with respect to three sub-challenges. [Table 6](#page10) sum- | |  |
| [are AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisser-](#page26) | | | | | | | | marizes data augmentation, normalization and preprocessing tasks | |  |
| [man, 2014), GoogLeNet (Szegedy et al., 2015) and ResNet (He et al.,](#page25) | | | | | | | | performed by each team. |  |  |
| [2016) due to their superior performance on different benchmarks](#page25) | | | | | | | |  |  |  |
| for object recognition tasks. A typical trend with the evolution | | | | | | | |  |  |  |
| of these architectures is that the networks have gotten deeper, | | | | | | | | *5.1. Sub-challenge – 1: Lesion segmentation* |  |  |
| e.g., ResNet is about 19, 8 and 7 times deeper than AlexNet, VG- | | | | | | | |  |  |  |
| GNet and GoogLeNet respectively. While the increasing depth im- | | | | | | | | For a given image, this task seeks to get the probability of a | |  |
| proves feature representation and prediction performance, it also | | | | | | | | pixel being a lesion (either MA, HE, EX or SE). Although different | |  |
| increases complexity, making it diﬃcult to optimize and even be- | | | | | | | | retinal lesions have distinct local features, for instance, MA, HE, | |  |
| comes prone to overfitting. Further, the increasing number of lay- | | | | | | | | EX, SE have a different shape, color and distribution characteris- | |  |
| ers (i.e., network depth) lead to vanishing gradient problems as | | | | | | | | tics, these lesions share similar global features. Hence, the major- | |  |
| a result of a large number of multiplication operations. Hence, | | | | | | | | ity of participating teams built a general framework that would be | |  |
| many teams chose the DenseNet ([Iandola et al., 2014](#page25)) which con- | | | | | | | | suitable for the segmentation of different lesions, summarized as | |  |
| nects each layer to every other layer in a feed-forward fashion, re- | | | | | | | | follows: |  |  |
| ducing the number of training parameters and alleviates the van- | | | | | | | |  |  |  |
| ishing gradient problem. DenseNet exhibits *(* + 1*)/*2 connections | | | | | | | |  |  |  |
| in layer network, instead of only , as in the networks men- | | | | | | | | *5.1.1. VRT (Jaemin Son et al.)* |  |  |
| tioned above. This enables feature reuse throughout the network | | | | | | | | Son et al. modified U-Net ([Ronneberger et al., 2015](#page26)) in such a | |  |
| that leads to more compact internal representations and in turn, | | | | | | | | way that upsampling layers have the same number of feature maps | |  |
| enhances its prediction accuracy. Another opted approach, Deep | | | | | | | | with layers concatenated. It was based on the motivation that fea- | |  |
| Layer Aggregation (DLA) structures ([Yu et al., 2017](#page26)), extends the | | | | | | | | tures in initial layers and upsampled layers are equally important | |  |
| “shallow” skip connections in DenseNet to incorporate more depth | | | | | | | | to segmentation. Additionally, they adjusted the number of max- | |  |
| and sharing of the features. DLA uses two structures – iterative | | | | | | | | pooling so that the radius of the largest lesion spans a pixel in the | |  |
| deep aggregation (IDA) and hierarchical deep aggregation (HDA) | | | | | | | | coarsest layer. In case of EX and HE, max-pooling is done six times, | |  |
| that iteratively and hierarchically fuse the feature hierarchies (i.e. | | | | | | | | whereas for SE and MA it is done four times and twice. Further, | |  |
| semantic and spatial) to make networks work with better accu- | | | | | | | | for dealing with MA’s, they used inverse pixel shuﬄing to convert | |  |
| racy and fewer parameters. Recent Fully Convolutional Network | | | | | | | | a 1280 × 1280 × 3 pixels image to 640 × 640 × 12 for network in- | |  |
| (FCN) ([Long et al., 2015](#page25)) adapt and extend deep classification ar- | | | | | | | | put and pixel shuﬄing ([Shi et al., 2016](#page26)) to convert 640 × 640 × 4 | |  |
| chitectures (VGG and GoogLeNet) into fully convolutional networks | | | | | | | | segmentation map into 1280 × 1280 × 1 pixels. Later, the pairs of | |  |
| and transfer their learned representations by fine-tuning to the | | | | | | | | a normalized fundus image and reference ground truths were fed | |  |
| segmentation task. It defines a skip architecture that combines se- | | | | | | | | to the network to generate segmentation result in the range [0, 1]. | |  |
| mantic information from a deep, coarse layer with appearance in- | | | | | | | | They used weighted binary cross entropy ([Murphy, 2012](#page25)) as loss | |  |

10 *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561*

**Table 6**

Summary of data augmentation, normalization and pre-processing in the competing solutions. Where, RF, RR, RS, RT, RC represent random flip, rotation, scaling, translation and crop respectively.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Task | Team name | | |  |  |  | Data augmentation | |  |  | Data normalization | Data preprocessing |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | RF | | RR | RS | RT | RC | Other |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | VRT |  | |  |  |  |  | shear |  | FOV cropping, division |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | by 255 then mean |  |
|  |  |  |  |  |  |  |  |  |  |  |  | subtraction |  |
| Sub-challenge - 1 |  |  | iFLYTEK |  | |  |  |  |  | × |  | lesion patch extraction |  |
|  |  |  |
|  |  |  |
|  |  |  | PATech |  | |  | × |  | × | color[a](#page10) |  | RGB to LUV, contrast |  |
|  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | adjustment |  |
|  |  |  | SDNU |  | |  | × | × | × | × | – | – |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  | SOONER |  | |  | × | × |  | × |  | mean subtraction, |  |
|  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | lesion patch extraction |  |
|  |  |  | LzyUNCC |  | | × | × | × |  | stochastic and | – | FOV cropping, image |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | photo-metric[b](#page10) |  | enhancement |  |
|  |  |  | SAIHST |  | |  | × | × | × | × |  | CLAHE, Gaussian |  |
|  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | smoothing |  |



|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | LzyUNCC |  | × | × | × |  | color[a ,](#page10) |  |
|  |  |  |
|  |  |  |  |  |  |  |  | stochastic and |  |
|  |  |  |  |  |  |  |  | photo-metric[b](#page10) |  |
| Sub-challenge - 2 |  | VRT | × | × | × | × | × | × |  |
|  |  |
|  |  | Mammoth |  |  |  |  | × | color |  |
|  |  |  |



– FOV cropping, image

enhancement

mean subtraction

× morphological opening

and closing

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | AVASAVA |  | × | × | × |  | × |  | intensity scaling |  |
|  |  |  |
|  |  | HarangiM1 | × | × | × | × | × | × |  | FOV cropping |  |
|  |  |  |
|  |  | HarangiM2 | × | × | × | × | × | × |  | – |  |
|  |  |  |
| Sub-challenge - 3 |  | DeepDR | × | × | × | × |  | OD, fovea |  | FOV cropping, mean |  |
|  |  |
|  |  |
|  |  |  |  |  |  |  |  | region |  | subtraction |  |
|  |  | VRT |  |  |  |  |  | shear and |  | FOV cropping, contrast |  |
|  |  |  |
|  |  |  |
|  |  |  |  |  |  |  |  | cropped OD |  | adjustment |  |
|  |  | ZJU-BII-SGEX | × | × | × | × | × | × |  | FOV cropping |  |
|  |  |  |
|  |  |  |
|  |  | SDNU |  | × |  | × | × | × | – | – |  |
|  |  |  |
|  |  | IITkgpKLIV |  |  | × | × | × | × |  | – |  |
|  |  |  |
|  |  | CBER | × | × | × | × | × | × | – | – |  |
|  |  |  |



* Reference: [Krizhevsky et al. (2012)](#page25) b Reference: [Howard (2013)](#page25)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| function given by | | | | | |  |  |  |  | **Table 7** |  |  |  |  |  |
|  | 1 *N* | | | |  |  |  |  |  | *γ* values in[Eq. (4).](#page10) | | |  |  |  |
|  | − *αGn* log *On* − *(*1 − *Gn* *)* log*(*1 − *On* *)* |  |  |  |  |  |  |  |  |  |  |
| *L* = |  | (3) | |  | EXs | SEs | HEs | MAs |  |  |
| *N* | |  |  |  |  |  |
| *n*=1 | |  |  |  | 64 | | 512 | 8 | 32 |  |  |
|  |  |  |  |  |  |  |  |  |
|  | | | | |  | *n* | and O | *n* |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| where N denotes the number of the pairs in a batch, G | | | | | |  |  |  |  |  |  |  |  |  |
| represent true segmentation and predicted segmentation for n*th* | | | | | | | | |  |  |  |  |  |  |  |
| image. The value of *α* was determined as follows: | | | | | |  |  |  | with tensorflow backend 1.4.0 using a server with 8 TITAN X (pas- | | | | | |  |
|  |  | *Bi* | | |  |  |  |  | [cal). The source code is available at https://bitbucket.org/woalsdnd/](https://bitbucket.org/woalsdnd/isbi_2018_fundus_challenge) | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | | 0 | |  |  |  |  |  | [isbi\_2018\_fundus\_challenge.](https://bitbucket.org/woalsdnd/isbi_2018_fundus_challenge) | |  |  |  |  |  |
| *α* = *γ F*1*i* | | | | |  |  | (4) | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

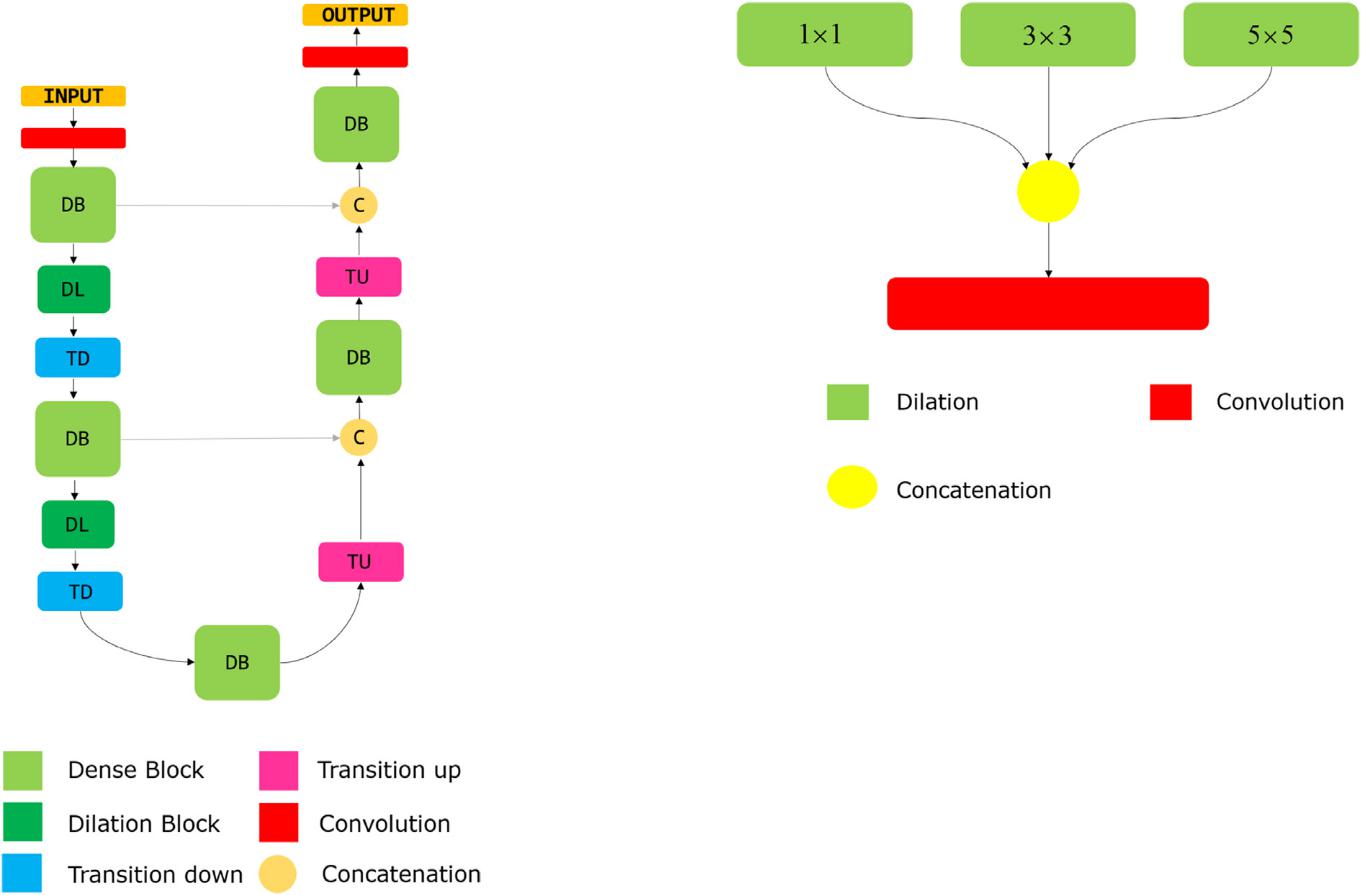
where B*n*0 and F1n denote the number of background and foreground pixels in n*th* image. Since background overwhelms foreground in lesion segmentation task, this loss function was designed to penal-ize false negatives in order to boost sensitivity, an important fac-tor in detecting lesions. Also, *γ* was left as a hyper-parameter and chosen out of {0.25, 0.5, 1, 2, 4, 8, 16, 32, 64, 256, 512} to yield the highest AUPR on validation set. The final selected *γ* values for different lesions are summarized in [Table 7](#page10).

They trained the network over 300 epochs using Adam op-timizer ([Kingma and Ba, 2014](#page25)) with hyper-parameters of *β*1 = 0*.*5*,* *β*2 = 0*.*999 and learning rate of 2e−4 until 250 epochs and 2e−5 until the end. All implementation was done by Keras 2.0.8

*5.1.2. IFLYTEK-MIG (Fengyan Wang et al.)*

Wang et al. proposed a novel cascaded CNN based approach for retinal lesion segmentation with U-Net ([Ronneberger et al., 2015](#page26)) as a base model. It consists of three stages, the first stage is a coarse segmentation model to get initial segmentation masks, then the second stage is a cascade classifier which was designed for false-positive reduction, at last, a fine segmentation model was used to refine results from previous stages. First stage model was trained using the patches of size 256 × 256 pixels centered on a particular lesion amongst MA, HE or EX and 320 × 320 pixels for SE, resulting in the coarse segmentation outcome. Results of the previous stage are coarse due to the fact that non-focus regions

|  |  |
| --- | --- |
| *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561* | 11 |



**Fig. 5.** Architecture for dilation block.

**Fig. 4.** Proposed architecture for lesion segmentation.

(non-target lesions) were not utilized in the learning process lead-ing to high false-positive count. In the second stage, unlike the first segmentation model which used a lesion centered sample from in-put dataset pool, candidate regions were extracted using probabil-ity maps from the previous stage. Here, the input size fed to model for SE was 320 × 320 × 3 pixels, for HE and EX it was 256 × 256 × 3 pixels, and for MA it was modified to 80 × 80 × 3 pixels consider-ing its small appearance. In this step, a candidate region was re-garded as a positive sample if its intersection-over-union with the ground truth was greater than the given threshold (i.e. 0.5). In this way, most trivial non-focus regions were effectively rejected. How-ever, it was identified in the test that a small proportion of false positives still exist, so an additional model was introduced to re-fine the segmentation results. In the last stage, candidate regions survived from the second stage were utilized as the input patches resulting in more accurate segmentation results. For the first and third stage, they used binary cross-entropy or dice loss function (multi-model training), whereas, for the second stage, they used only binary cross-entropy as a loss function. The first, second and third stage models were trained for 100, 300 and 100 epochs re-spectively with the momentum of 0.9. In which, the initial learning rate for the first and third stage was set 0.1 and is reduced by 10 times every 30 epochs, and for the second stage it was set to 0.001 reduced by 10 times every 80 epochs. MXNET platform was used for training the models.

*5.1.3. PATech (Liu lihong et al.)*

Lihong et al. developed a novel patch-based CNN model (as shown in [Fig. 4](#page11)) in which they innovatively combined the DenseNets ([Iandola et al., 2014](#page25)) and dilation block with U-Net ([Ronneberger et al., 2015](#page26)) to capture more context information and multi-scale features.

The model is composed of a down-sampling path with 4 Tran-sitions Down (TD), 4 Dilation Block (DL) and an up-sampling path

with 4 Transitions Up (TU). To capture multi-scale features, DL (see [Fig. 5](#page11)) is used with dilation rate of 1, 3 and 5 are concatenated for the convolution. The dense block (DB) is constructed by four lay-ers. The idea behind novel combination of dilation convolution is to better deal with the lesions appearing at different scales, where small dilation rate pay closer attention to the characteristics of tiny lesions, larger dilation rate focus on large lesions. On the other hand, use of DB’s enabled a deeper and more eﬃcient network.

Initially, they extracted regions within FOV from the images and then normalized them to eliminate local contrast differences and uneven illumination. Later, they used small patches 256 × 256 pix-els at a stride of 64 (128 for MA) to generate the training samples (only patches that overlap with the lesion ground truth) followed by data augmentation before feeding to the model. To deal with highly imbalanced spread of data, they designed a loss function that is a combination of dice function ([Sudre et al., 2017](#page26)) and 2D cross Entropy as follows:

*L* = −*mean(w*10∗ *G* ∗ *log(O)*

+*w*11 ∗ *(*1 − *G)* ∗ *log(*1 − *O)* (5)

+*w*2 ∗ *dice(G))*

where w10 and w11 are the factors utilized to keep a balance be-tween the positive and negative pixels, and w2 is the factor uti-lized to control significance between dice and cross entropy loss. The values of w10, w11 and w2 were empirically set to 0.7, 0.3 and 0.4 respectively. The models were trained using Adam optimizer with default parameters, *β*1 = 0*.*9 and *β*2 = 0*.*999. The initial learn-ing rate was set to 2 × 10−4*,* and then divided by 20 in every 20 epochs. This model was implemented with PyTorch1.12 and Tesla M60 platform was utilized for training on the CentOS 7.2 operat-ing system.

*5.1.4. SOONER (Yunzhi Wang et al.)*

Wang et al. adopted U-Net ([Ronneberger et al., 2015](#page26)) architec-ture for solving retinal lesion segmentation problem. The network takes a 380 × 380 pixels fundus image patch as input and pre-dicts the binary mask of the retinal lesion within 196 × 196 pixels central region of an input patch. They pre-processed fundus im-ages by subtracting the local mean of each color channel and per-formed random flipping for data augmentation. Batch normaliza-tion was utilized to improve training eﬃciency and all convolution operations adopted ‘valid’ paddings. For training, they followed a three-stage process for each type of lesions (i.e. MA, HE, EX and SE). For the first stage, they extracted positive image patches from the training set according to given ground truth mask, and ran-domly extracted negative image patches from fundus images with

12 *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561*

and without apparent retinopathy. The objective function was a summation of cross-entropy loss functions for MA, HE, EX and SE. Adam algorithm was employed to optimize the parameters. In the second stage, they fine-tuned U-Net using the extracted patches for each lesion type. Subsequently, they applied optimized U-Net on fundus images in the training set and extracted false-positive patches generated by U-Net. They further fine-tuned U-Net using positive image patches together with false-positive patches (hard negative patches) as a third stage. In the testing phase, they ex-tracted overlapped image patches using a sliding window and fed these patches into the network to get corresponding probability maps. The initial learning rate was set to e−4 and the fixed number of steps was used as a stopping criterion. They implemented U-Net architecture based on TensorFlow library with Nvidia GeForce GTX 1080Ti GPU.

*5.1.5. LzyUNCC (Zhongyu Li et al.)*

Li et al. developed a method based on FCN by embedding DLA structure ([Yu et al., 2017](#page26)) for segmentation of EX’s and SE’s. As the lesions are located dispersively and irregularly, the embedding of DLA structure with FCN enables better aggregation of semantic and spatial information from local and global level provides a boost in recognizing their presence. They used retinal images with pixel-level ground truth annotations from both IDRiD and E-Ophtha database. They first adopted a series of methods for data prepro-cessing and augmentation. Subsequently, considering the correla-tion between EX’s and SE’s, they first trained an initial model for segmentation of EX. They chose a smaller model, i.e., DLA-34 to train the segmentation network with binary cross-entropy as a loss function. At last, the trained deep model was fine-tuned for seg-mentation of SE. While the model training of EX segmentation, a trade-off parameter (penalty) was assigned in loss function to con-trol the weights of foreground pixels, and tried different penalty value from 1 to 16. At last, these segmentation results were fused to adaptively compute the best performance. They adopted original DLA cityscapes segmentation experimental settings and trained the model for 100 epochs with batch size 4, where the poly learning

rate was *(*1 − *epoch*−1 *)*0*.*9 with the momentum of 0.9. The initial *totalepoch*

learning rate was set to 0.01.

*5.1.6. SAIHST (Yoon Ho Choi et al.)*

Choi et al. proposed a model for segmentation of EX based on U-Net ([Ronneberger et al., 2015](#page26)), in which CONV layers of encoder path are replaced with DB’s. Whereas, the decoder path of their model was kept identical to that of general U-Net. They built DB with a growth factor of 12 and 3 × 3 CONV layers, batch normal-ization, and ReLU activation. The last layer generates a pixel level prediction map for EXs through the sigmoid activation function. For training, they utilized only the green channel of fundus im-age and enhanced it using Contrast Limited Adaptive Histogram Equalization (CLAHE). Later, each image was padded to a size of 4352 × 3072 pixels and cropped into 204 patches of 512 × 512 pix-els. These patches are further augmented and used for training. The losses were calculated by binary cross-entropy. The model was trained for 20 epochs with a mini-batch size of 10 and they used Adam optimizer with an initial learning rate of 2e−4*,* *β*1 of 0.9 and

* 2 of 0.999. The model was programmed in Keras 2.1.4 served with TensorFlow 1.3.0 backend.

*5.1.7. SDNU (Xiaodan sui et al.)*

Sui et al. proposed a method based on Mask R-CNN structure to segment lesions from the fundus image. They adopted the im-plementation of Mask R-CNN from [Abdulla (2017)](#page24) for solving the problem. This method could detect different objects while simulta-neously generating instance segmentation mask. Network training precedes the data augmentation process and binary cross-entropy

was used as a loss function. The initial learning rate was set to 0.02 with a momentum of 0.9. They chose ResNet-101 as a back-bone. They implemented an algorithm in Keras with Tensorflow as backend and processed on 8 NVIDIA TITAN Xp GPUs. The experi-mental environment was built under Ubuntu 16.06.

*5.2. Sub-challenge – 2: Disease grading*

For a given image, this task seeks to get a solution to produce severity grade of the diseases i.e. DR (5 class problem) and DME (3 class problem). The summary of participating solutions is as fol-lows:

*5.2.1. LzyUNCC (Zhongyu Li et al.)*

Li et al. developed a method based on ResNet by embedding DLA structure for automated grading of DR and DME. For this work, they used IDRiD and Kaggle dataset. Initially, for the given training images, they perform data preprocessing and data aug-mentation. Subsequently, based on the designed ResNet with DLA structure, initial models are trained using 35,000 retinal images from Kaggle dataset. Later, they fine-tuned the model using IDRiD dataset through 5 fold cross-validation technique. Finally, the five outputs are ensembled together as final grades for input images. It is important to note that networks for grading of DR and DME were trained separately. The training was performed by SGD with a mini-batch size of 64, while the learning rate starts from 0.001 and it is then divided by 10 every 20 epochs, for 30 epochs in total. The other hyper-parameters are fixed to settings of original DLA ImageNet classification ([Yu et al., 2017](#page26)).

*5.2.2. VRT (Jaemin Son et al.)*

Son et al. used network ([Son et al., 2018](#page26)) for DR grading. Kag-gle dataset was initially used to pre-train the network and then the model was fine-tuned using IDRiD dataset. The penultimate layer was Global Average Pooled (GAP) and connected with FC layer. The entire output is a single value from which L2 loss was calculated against the true label. SGD was used with Nesterov momentum of 0.9 as an optimizer. Learning rate was set to 10−3. The model was trained for 100 epochs. Fundus image was normalized in the range [0, 1] and the mean was subtracted channel-wise. For grading of DME, segmented EXs (using the segmentation network proposed in sub-challenge – 1), localized fovea and segmented OD (using the segmentation network proposed in sub-challenge – 3) were utilized for making the final decision. With this information, the semi-major axis of segmented OD (r) was estimated. Further, the fundus image was divided into three regions as macular region: x − *c < r,* near macular region: r *< x* − *c <* 2r and remaining region: 2r *<* x − c . where x denotes a point in the image. Fur-thermore, several features such as sum of intensity for segmented EX, the number of pixels above threshold (178 in the [0, 255] scale), the number of pixels for smallest and largest blob, mean of the number of pixels for blobs are extracted for each area, and binary flag that indicates whether the OD is segmented. Now, fea-tures with high importance were selected among numerous fea-tures in the initial training due to gradient boosting (for instance, XGBoost) was likely to overfit when provided with overly redun-dant features. Messidor dataset was added to the given data and out of which 10% of images were left as the validation set. Set of hyper-parameters were searched by grid-search approach. The combination of hyper-parameters that yielded the highest accu-racy in the validation set was min child weight: 2, subsample: 0.2, colsample by tree: 0.2, *λ*: 9.0, *α*: 1.0, and depth: 6. Other hyper-parameters are set to default values. All implementations were done by PyTorch v0.4.1 using a server with 8 TITAN X (pascal). [The source code is available at https://bitbucket.org/woalsdnd/](https://bitbucket.org/woalsdnd/isbi_2018_fundus_challenge) [isbi\_2018\_fundus\_challenge.](https://bitbucket.org/woalsdnd/isbi_2018_fundus_challenge)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561* | | | | | | | | | |  |  |  |  | 13 |  |
| *5.2.3. Mammoth (Junyan Wu et al.)* | | | | | | | | | | | | | | |  |  |  |  | For the grading of DME, the final layers of member CNNs con- | | | | | | | | | |  |
|  | Wu et al. proposed a unified framework that combines deep | | | | | | | | | | | | | | | | | | sist of 3 neurons, and weight matrices A1, A2 were 3 × 3, initialized | | | | | | | | | |  |
| feature extractor and statistical feature blending to automatically | | | | | | | | | | | | | | | | | | | as |  |  |  |  |  |  |  |  |  |  |
| predict the DR and DME severity scores. For DME, they used | | | | | | | | | | | | | | | | | | |  | 1*/*3 | 0 | 0 |  |  |  |  |  |  |  |
| DenseNet ([Iandola et al., 2014](#page25)) to directly predict the severity | | | | | | | | | | | | | | | | | | |  | 0 | 1*/*3 | 0 |  |  |  |  |  |  |  |
| score. Whereas for DR, Kaggle training dataset was used to pre- | | | | | | | | | | | | | | | | | | | *A*1= *A*2=0 | | 0 | 1*/*3 | |  |  |  |  | (9) |  |
| train the DenseNet model through a dynamic sampling mechanism | | | | | | | | | | | | | | | | | | | For training, they merged IDRiD and Kaggle training set. The | | | | | | | | | |  |
| to balance the training instances and later fine-tuned using IDRiD | | | | | | | | | | | | | | | | | | |  |
| parameters of architectures were found by SGD algorithm in 189 | | | | | | | | | |  |
| dataset. Initially, the background of all images was cropped and re- | | | | | | | | | | | | | | | | | | |  |
| and 50 epochs respectively for DR and DME classification tasks. | | | | | | | | | |  |
| sized to 512 × 512 pixels. Later, morphological opening and closing | | | | | | | | | | | | | | | | | | |  |
| Learning rate was set to 0.0001. Training times required on the | | | | | | | | | |  |
| are utilized to preserve bright and dark regions. For instance, the | | | | | | | | | | | | | | | | | | | datasets for DR and DME were 96.6 (189 epochs) and 23.4 (50 | | | | | | | | | |  |
| morphological opening can erase the EXs and highlight the MAs. | | | | | | | | | | | | | | | | | | |  |
| epochs) hours respectively. Implementation of this work was done | | | | | | | | | |  |
| Whereas, the closing operation can remove MAs and preserve EXs. | | | | | | | | | | | | | | | | | | |  |
| in MATLAB 2017b. The training was performed using an NVIDIA | | | | | | | | | |  |
| These operations can be used to denoise specific levels of classi- | | | | | | | | | | | | | | | | | | |  |
| TITAN X GPU card with 7 TFlops of single-precision performance, | | | | | | | | | |  |
| fications, for example, the risk of DME only depends on the loca- | | | | | | | | | | | | | | | | | | |  |
| 336.5 GB/s of memory bandwidth, 3072 CUDA cores, and 12 GB | | | | | | | | | |  |
| tion of the EXs. Further, several standard data augmentation meth- | | | | | | | | | | | | | | | | | | |  |
| memory. |  |  |  |  |  |  |  |  |  |  |
| ods (as shown in [Table 6](#page10)) are also employed. Mean Squared Error | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| (MSE) and cross-entropy with five classes were the loss functions | | | | | | | | | | | | | | | | | | | *5.2.5. AVSASVA (Varghese Alex et al.)* | | | | |  |  |  |  |  |  |
| employed to train the network and SGD for optimization. The ini- | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |
| Alex | et al. used ensembles of pre-trained | | | | | | CNNs (on | | Im- |  |
| tial learning rate was set to 0.0005 with a decrement of 0.1 after | | | | | | | | | | | | | | | | | | |  |
| ageNet | dataset), | namely, | | ResNets | ([He](#page25) | [et](#page25) | [al.,](#page25) | [2016](#page25)) | and |  |
| every 30 epochs. The initial training was done by 200 epochs and | | | | | | | | | | | | | | | | | | |  |
| DenseNets ([Iandola et](#page25) | | | [al.,](#page25) | [2014](#page25)) for | the | task | of | disease grad- | |  |
| fine-tuning by 50 epochs. Afterward, the last layer was removed | | | | | | | | | | | | | | | | | | |  |
| ing. For DR grading, two ensembles of CNN’s namely “primary” | | | | | | | | | |  |
| before the final prediction, and its statistical features were aggre- | | | | | | | | | | | | | | | | | | |  |
| and “expert” classifiers | | | were used. The | | | primary | | classifier | was |  |
| gated together into a boosting tree. Specifically, 50 pseudo-random | | | | | | | | | | | | | | | | | | |  |
| trained to classify a fundus image as one of the 4 classes viz; Nor- | | | | | | | | | |  |
| augmentations were performed to get 50 outputs from last sec- | | | | | | | | | | | | | | | | | | |  |
| mal, Mild NPDR, Moderate NPDR or S-(N)-PDR, a class formed by | | | | | | | | | |  |
| ond FC layer (size of 4096), then the mean and standard deviation | | | | | | | | | | | | | | | | | | |  |
| clubbing Severe NPDR and PDR. The expert classifier was trained | | | | | | | | | |  |
| of 50 feature vectors for each image was computed, and both vec- | | | | | | | | | | | | | | | | | | |  |
| exclusively on Severe NPDR or PDR images and was utilized to | | | | | | | | | |  |
| tors were then concatenated together for training in LightGBM. The | | | | | | | | | | | | | | | | | | |  |
| demarcate the input image as one of the aforementioned classes. | | | | | | | | | |  |
| output from the second last layer of fine-tuning experiments was | | | | | | | | | | | | | | | | | | |  |
| During inference, each fundus image was resized to a dimension of | | | | | | | | | |  |
| used to train a blending model, strategy adopted from team o\_O’s | | | | | | | | | | | | | | | | | | |  |
| 256 × 256 pixels. For the task of grading of DR in fundus images, | | | | | | | | | |  |
| solution | | | | | of Kaggle | | | | | DR challenge. Finally, | | | | | | for | the | disease grading |  |
| prediction, gradient boosting tree model was built on a combined | | | | | | | | | | | | | | | | | | | they used test time augmentation through the “Ten Crop” function | | | | | | | | | |  |
| defined | in PyTorch. The images were | | | | first passed through | | | | the |  |
| second last layer from the pre-trained network and fine-tuned net- | | | | | | | | | | | | | | | | | | |  |
| primary | classifier | and then | | through the expert | | | classifier, only if | | |  |
| work. | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | the image was classified as S-(N)-PDR by the primary classifier. | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | The final prediction was achieved by using a majority voting | | | | | | | | | |  |
| *5.2.4. Harangim1 (Balazs Harangi et al.)* | | | | | | | | | | | | | | |  |  |  |  | scheme. |  |  |  |  |  |  |  |  |  |  |
|  | Harangi et al. proposed an approach for the classification of | | | | | | | | | | | | | | | | | | For DME grading, two ensembles were trained in a one versus | | | | | | | | | |  |
| [retinal images via the fusion of two AlexNet (Krizhevsky et al.,](#page25) | | | | | | | | | | | | | | | | | | | rest approach. Ensemble 1 was trained to classify the input as ei- | | | | | | | | | |  |
| [2012) and GoogLeNet (](#page25)[Szegedy et al., 2015](#page26)[). For this aim, they re-](#page25) | | | | | | | | | | | | | | | | | | | ther “image with no apparent EXs” (Grade 0) or “presence of EXs | | | | | | | | | |  |
| moved FC and classification layers and interconnect them by in- | | | | | | | | | | | | | | | | | | | in image” (Grade 1 & Grade 2), while the Ensemble 2 was trained | | | | | | | | | |  |
| serting a joint FC layer followed by the classic softmax/ classifi- | | | | | | | | | | | | | | | | | | | to classify an image as “Grade 2” DME or not (Grade 0 & Grade 1). | | | | | | | | | |  |
| cation layers for the final prediction. In this way, single network | | | | | | | | | | | | | | | | | | | During inference, the resized images were fed to both ensembles, | | | | | | | | | |  |
| architecture was created which allows to train the member CNN’s | | | | | | | | | | | | | | | | | | | and the final prediction was obtained by combining the two pre- | | | | | | | | | |  |
| simultaneously. For each I(n), let us denote the outputs of the fi- | | | | | | | | | | | | | | | | | | | dictions by utilizing a set of user-defined rules. Briefly, the user- | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ˆ *(n)* | | ˆ *(n)* | . The FC layer of | defined rules were: an image was classified as Grade 0 DME if | | | | | | | | | |  |
| nal FC layers of the member CNN’s by O1 | | | | | | | | | | | | | | | | *, O*2 | | ensemble 1 and ensemble 2 predict the absence of EXs and the | | | | | | | | | |  |
| their ensemble aggregates them via | | | | | | | | | | | | | | |  |  |  |  |  |
|  |  |  |  | absence of grade 2 DME respectively. A scenario wherein ensem- | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ´ *(n)* | |  |  |  | ˆ *(n)* | | |  |  | ˆ *(n)* |  |  |  |  |  |  |  |  | ble 2 predicts the presence of grade 2 DME, images were classified | | | | | | | | | |  |
| *O* |  | = *A*1*O*1 | | | | |  | + *A*2*O*2 | | |  |  |  |  |  |  |  | (6) | under category “Grade 2 DME” irrespective of the prediction from | | | | | | | | | |  |
| where weight matrices A1, A2 were of size 5 × 5 and initialized as | | | | | | | | | | | | | | | | | | | ensemble 1. Lastly, images were classified as Grade 1 DME if none | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | of the above conditions were satisfied. | | | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  | 1*/*5 | | | 0 | 0 | |  | 0 | 0 |  |  |  | Both models for DR and DME were initialized with the pre- | | | | | | | | | |  |
|  |  |  |  |  |  | 0 | |  | 1*/*5 | 0 | |  | 0 | 0 |  |  | trained weights and the parameters of networks were optimized | | | | | | | | | |  |
| *A*1 | = | | *A*2 | |  | 0 | |  | 0 | 1*/*5 | | | 0 | 0 |  | (7) | by reducing cross-entropy loss with ADAM as an optimizer. The | | | | | | | | | |  |
|  |  |  | = | | 0 | |  | 0 | 0 | |  | 1*/*5 | 0 |  |  |  | learning rate was initialized to 10−3 for DR and 10−4 for DME. For | | | | | | | | | |  |
|  |  |  |  |  |  |  | 0 | |  | 0 | 0 | |  | 0 | 1*/*5 | |  |  | DR, the learning rate was reduced by a factor of 10% every instance | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | when the validation loss failed to drop. Each network was trained | | | | | | | | | |  |
|  | The last two layers of the ensemble were a Softmax and a clas- | | | | | | | | | | | | | | | | | | for 30 epochs and the model parameters that yielded the lowest | | | | | | | | | |  |
| sification one. Let O*SM(n)* be an output of a former layer, the MSE was | | | | | | | | | | | | | | | | | | |  |
| validation loss were used for inference. For DME, the learning rate | | | | | | | | | |  |
| used for optimization as a loss function: | | | | | | | | | | | | | | | |  |  |  | was annealed step-wise with a step size of 10 and the multiplica- | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | 1 |  | *N* |  | ´ | *(n)* | *(n)* |  | 2 |  |  |  |  |  | tive factor of learning rate decay value of 0.9. | | | | | | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *MSE* = | | | |  | | *n*=1 | | | *(OSM* − *O* | |  | *)* |  |  |  |  |  | (8) |  |  |  |  |  |  |  |  |  |  |  |
| 2N | |  |  |  |  |  |  | *5.2.6. HarangiM2 (Balazs Harangi et al.)* | | | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | During the training phase, back-propagation is applied to min- | | | | | | | | | | | | | | | | | | Harangi et al. combined self-extracted, CNN-based features with | | | | | | | | | |  |
| imize the loss via adjusting all parameters of member CNNs and | | | | | | | | | | | | | | | | | | | traditional, handcrafted ones for disease classification. They mod- | | | | | | | | | |  |
| weight matrices A1, A2. | | | | | | | | | | |  |  |  |  |  |  |  |  | ified AlexNet ([Krizhevsky et al., 2012](#page25)) to allow the embedding of | | | | | | | | | |  |

14 *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561*

handcrafted features via the FC layer. In this way, they created a network architecture that could be trained in the usual way and additionally uses domain knowledge. They extended the FC layer, to get FC*fuse*, originally containing 4096 neurons of AlexNet by adding 68-dimensional vector containing handcrafted features.

Then, the 4164 × 5 (or 4164 × 3 for DME) layer FC*class* was consid-ered for DR (or DME) classification task. In this way, both final

weighings FC*class* of handcrafted features were obtained and the 4096 AlexNet features were trained by backpropagation.

To obtain 68 handcrafted features used by CNN, they employed one image level and two lesion-specific methods. The amplitude-frequency modulation (AM-FM) method extracts information from an image by decomposing its green channel at different scales into AM-FM components ([Havlicek, 1996](#page25)). As a result, a 30-element feature vector was obtained, which reflects the intensity, geom-[etry, and texture of structures contained in the image (Agurto](#page24) [et al., 2010). Whereas to extract features related to the lesions MA](#page24) [and EX, they employed two detector ensembles (Antal and Hajdu,](#page24) [2012; Nagy et al., 2011), which consist of a set of *<* preprocessing](#page24) method (PP), candidate extractor (CE) *>* pairs organized into a vot-ing system. Such a *<* PP, CE *>* pair was formed by applying PP to the retinal image and CE to its output. This way, a *<* PP, CE *>* pair extracts a set of lesion candidates from the input image, acting as a single detector algorithm. They used the output of these ensem-bles to obtain 38 features related to the number and size of MA’s and EX’s. Parameters of the architectures were optimized by SGD algorithm in 85 and 50 epochs for DR and DME respectively. Train-ing times were 83.1 (85 epochs) and 46.2 (50 epochs) hours on the datasets for DR and DME. Implementation of this work was done in MATLAB 2017b. Training has been performed using an NVIDIA TITAN X GPU card with 7 TFlops of single-precision, 336.5 GB/s of memory bandwidth, 3072 CUDA cores, and 12 GB memory.

*5.3. Sub-challenge – 3: Optic disc and fovea detection*

For a given image, this task seeks to get a solution to localize the OD and Fovea. Further, it seeks to get the probability of a pixel being OD (OD segmentation). Summary of approaches is detailed as follows:

*5.3.1. Deepdr (Ling Dai et al.)*

Dai et al. proposed a novel deep localization method, which allows coarse-to-fine feature encoding strategy for capturing the global and local structures in fundus images, to simultaneously model two-task learning problem of the OD and fovea localization. They took advantage of prior knowledge such as the number of landmarks and their geometric relationship to reliably detect the OD and fovea. Specifically, they first designed a global CNN en-coder (with a backbone network of ResNet-50 ([He et al., 2016](#page25))) to localize the OD and fovea centers as a whole by solving a re-gression task. All max-pooling layers were replaced with average pooling layers as compared to original ResNet architecture, due to the fact that max-pooling could lose some useful pixel-level in-formation for regression to predict the coordinates. This step was used to simultaneously perform the two detection tasks, because of the geometric relationship between OD and fovea, the perfor-mance of multi-task learning is better than a single task. The pre-dicted output coordinates of this global CNN encoder component were used for detecting the bounding boxes of the target OD and fovea. Then the current center coordinates are refined through a [local encoder (with a backbone network of VGG-16 (Simonyan and](#page26) [Zisserman, 2014)) which only localizes the OD center or fovea cen-](#page26)ter of their related bounding boxes. During the training stage, they designed an effective data augmentation scheme to solve the prob-lem of insuﬃcient training data. In particular, to build the train-ing set of a local encoder, bounding boxes were randomly selected

based on the ground truth, for each object several bounding boxes of different positions and scales were cropped. The local encoder can be reused multiple times to approximate the target coordi-nates. The local encoder was iterated twice for refining centers comprehensively. All three models were initialized from the pre-trained ImageNet network and replaced the network’s last FC layer and Softmax layer by the center coordinates regressor. The regres-sion loss for the central location was the Euclidean loss. The mod-ified loss function for global and local encoders was 0*.*045*(LOD* +

* *f ovea )* and 0.045(L*OD*/L*fovea*) respectively. Where *LOD* and *Lfovea* arelosses for OD and fovea, and scaling factor was introduced since the original Euclidean distance is too large in practice to converge. The proposed learning model was implemented in Caffe framework and trained using SGD with momentum. The FC layers for center regression were initialized from zero-mean Gaussian distributions with standard deviations 0.01 and 0.001. Biases were initialized to 0. The global encoder was trained for 200 epochs, local encoders (OD and fovea both) for 30 epochs respectively. The batch size for global encoder was 16, and 64 for the other two local encoders. The learning rate was set as 0.01 and was divided by 10 when the error plateaus.

*5.3.2. VRT (Jaemin Son et al.)*

Son et al. proposed an OD segmentation model consisting of U-Net ([Ronneberger et al., 2015](#page26)) and CNN that takes a vessel im-age and outputs 20 × 20 activation map whose penultimate layer is concatenated to the bottleneck layer of U-Net. Initially, original images were cropped (3500 × 2848 pixels), padded (3500 × 3500 pixels) and then resized (640 × 640 pixels). Each image was stan-dardized with its mean and standard deviation (SD). When cal-culating the mean and SD, values less than 10 (usual artifacts in the black background) are ignored. Vessel images were prepared with an external network ([Son et al., 2017](#page26)). Pixel values in a ves-sel image range from 0 to 1. It uses external datasets DRIONS-DB ([Carmona et al., 2008](#page24)) and DRIVE ([van Ginneken et al., 2004](#page26)) available with OD and vessel ground truths respectively. For aug-mentation, the fundus images were aﬃne-transformed and addi-tionally OD was cropped and randomly placed on the image for a random number of times (0 to 5). This augmentation was done to prevent the network from segmenting OD solely by brightness. Pairs of a fundus image and vessel segmentation were provided as input and OD segmentations in the resolution of 640 × 640 and 20 × 20 pixels are given as the ground truth. Binary cross-entropy is used as a loss function for both U-Net and vessel net-

work with the loss of L*total* = L*U*−*Net* + 0*.*1 ∗ L*vessel* . Total 800 epochs were trained via Adam optimizer and decreasing learning rate with

hyper-parameters of *β*1 = 0*.*5*,* *β*2 = 0*.*999. The learning rate was 2e−4 until 400 epochs and 2e−5 until the end. Weights and biases [were initialized with Glorot initialization method (Glorot and Ben-gio, 2010).](#page24)

They also proposed a four branch model in which two branches were dedicated to the prediction of locations for OD and fovea from vessels (vessel branches) and the other two branches aim to predict the locations from both fundus and vessels (main branches). Similar to OD segmentation, penultimate layers of ves-sel branches were depth-concatenated to the main branches. Af-ter deriving an activation map that represents the probability of containing an anatomical landmark, a hard-coded matrix was mul-tiplied to yield co-ordinates. Original images were cropped as in the segmentation task and standardized with an identical method and later augmented by flip and rotation to ease implementation efforts. Mean absolute error was used as loss function for both

outputs with the loss of L*total* = L*main* + 0*.*3 ∗ L*vessel* . SGD was used with Nesterov momentum of 0.9 as an optimizer. Learning rate was set to 10−3 from 1*st* to 500*th* epochs and 10−4 from 501*th* to 1000*th* epochs. All implementation was done in Keras 2.0.8

|  |  |
| --- | --- |
| *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561* | 15 |

with TensorFlow backend 1.4.0 using a server with 8 TITAN X [(pascal). Source code is available at https://bitbucket.org/woalsdnd/](https://bitbucket.org/woalsdnd/isbi_2018_fundus_challenge) [isbi\_2018\_fundus\_challenge.](https://bitbucket.org/woalsdnd/isbi_2018_fundus_challenge)

*5.3.3. ZJU-BII-SGEX (Xingzheng Lyu et al.)*

Lyu et al. utilized Mask R-CNN ([He et al., 2017](#page25)) to localize and segment OD and fovea simultaneously. It scans the image and gen-erates region proposals by 2D bounding boxes. Then the proposals were classified into different classes and compute a binary mask for each object. They firstly preprocessed the original retinal im-age into fixed dimensions as network input. A feature extractor (ResNet-50) with feature pyramid networks (FPN) generates fea-ture maps at different scales, which could be used for regions of interest (ROI) extraction. Then a region proposal network (RPN) scans over the feature maps and locates regions that contain ob-jects. Finally, a ROI head network (RHN) is employed to obtain the label, mask, and refined bounding box for each ROI. They also in-corporated prior knowledge of retinal image as a post-processing step to improve the model performance. They used IDRiD dataset and two subsets in RIGA dataset ([Almazroa et al., 2018](#page24)) (Messidor and BinRushed, 605 images) with OD mask provided. They applied the transfer learning technique to train the model. They firstly trained the RHN network by freezing all the layers of FPN and RPN networks and then fine-tuned all layers. The model was imple-mented in TensorFlow 1.3 and Python 3.4 (source code was modi-fied from [Abdulla (2017)](#page24)). The learning rate started from 0.001 and a momentum of 0.9 was used. The network was trained on one GPU (Tesla K80) with 20 epochs.

*5.3.4. IITkgpKliv (Oindrila Saha et al.)*

Saha et al. used SegNet ([Badrinarayanan et al., 2015](#page24)) for seg-mentation of lesions and OD. OD was added as an additional class in the same problem as lesion segmentation so that the model could better differentiate EXs and OD which have similar bright-ness levels. However, in contrast to original SegNet, the final de-coder output is fed to a sigmoid layer to produce class probabilities for each pixel independently in 7 channels. Each channel has the same size as input image: 536 × 356 pixels and consists of activa-tions in the range [0, 1] where 0 corresponds to background and 1 to the presence of a corresponding class. Apart from 5 classes i.e. MA, HE, SE, EX and OD, two additional classes: (i) retinal disk excluding the lesions and OD, and (ii) black background form the 7 channels. Images were downsampled to 536 × 356 pixels, pre-[serving the aspect ratio. Additionally, Drishti-GS (Sivaswamy et al.,](#page26) [2014) dataset was used for data augmentation to account for the](#page26) case of absence of lesions. Further, horizontal, vertical and 180◦ flipped versions of the original images were taken. The network was trained using binary cross-entropy loss function and Adam op-timizer with learning rate 10−3 and *β* = 0*.*9. Early stopping of the training based on the validation loss is adopted to prevent overfit-ting. It was observed that the validation loss started to increase af-ter 200 epochs. One more softmax layer is introduced after the Sig-moid layer for normalizing the value of a pixel for each class across channels. The segmented output is finally upsampled for each class to 4288 × 2848 pixels. All implementations were done in PyTorch using 2x Intel Xeon E5 2620 v3 processor with GTX TITAN X GPU 12 GB RAM and 64 GB System RAM.

*5.3.5. SDNU (Xiaodan Sui et al.)*

Sui et al. used Mask R-CNN ([He et al., 2017](#page25)) for solving all tasks in this sub-challenge. Mask R-CNN could realize accurate tar-get detection based on proposed candidate object bounding boxes of RPN to achieve the objective of OD and Fovea localization. At the same time, it could also get the OD segment at the mask predicting branch. The head architecture of Mask R-CNN (ResNet-101 as a backbone) consists of three parallel branches for clas-

sification, bounding-box regression, and predicting mask. By this method, the localization of OD and fovea, and segmentation of OD could be achieved directly. They retrained the network to get the new weight parameter of the framework. During the training phase, the dataset of this challenge was augmented by flipping, re-sizing and trained by 10-fold cross-validation. After training 2000 epochs, the last trained model is obtained. They implemented this algorithm in TensorFlow and it is processed on 8 NVIDIA TI-TAN Xp GPUs. The experiment environment is built under Ubuntu 16.06.

*5.3.6. CBER (Ana Mendonça et al.)*

Mendonça et al. proposed hand-crafted features based approach for the localization and segmentation tasks in this sub-challenge. Distinct methodologies have been developed for detecting and seg-menting these structures, mainly based on color and vascular infor-mation. The methodology proposed in the context of this challenge includes three inter-dependent modules. Each module performs a single task: OD localization, OD segmentation or fovea localization. While the modules responsible for the OD localization and seg-mentation were an improved version of two methods previously published ([Mendonça et al., 2013; Dashtbozorg et al., 2015](#page25)), the method proposed for fovea localization was completely new. Ini-tially, the module associated with the OD localization receives a fundus image and segments the retinal vasculature. Afterward, the entropy of the vessel directions is computed and combined with the image intensities in order to find the OD center coordinates. For OD segmentation, the module responsible for this task uses the position of the OD center for defining the region where the sliding band filter ([Pereira et al., 2007; Esteves et al., 2012](#page25)) is applied. The positions of the support points which give rise to the maximum filter response were found and used for delineating the OD bound-ary. Since a relation between the fovea-OD distance and the OD diameter was known ([Jonas et al., 2015](#page25)), the module responsible for the fovea localization begins by defining a search region from the OD position and diameter. The fovea center is then assigned to the darkest point inside that region.

**6. Evaluation measures**

The performance of each sub-challenge was evaluated based on different evaluation metrics. Following evaluation measures were used for different sub-challenges:

*6.1. Sub-challenge – 1*

In this sub-challenge, the performance of algorithms for lesion segmentation tasks was evaluated using submitted grayscale im-ages and available binary masks. As in the lesion segmentation task(s) background overwhelms foreground, a highly imbalanced scenario, the performance of this task was measured using area under precision (a.k.a. Positive Predictive Value (PPV)) recall (a.k.a. Sensitivity (SN)) curve (AUPR) ([Saito and Rehmsmeier, 2015](#page26)).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *SN* = |  | *True Positives* | | | | | | (10) |  |
|  | |  |  |  | |  |  |
| *True Positives* | | + | | *False Negatives* | | |  |
|  |  |  |  |  |  |  |  |
| *PPV* = | | *True Positives* | | | | | | (11) |  |
|  | |  |  |  | |  |
| *True Positives* | | + | | *False Positives* | |  |
|  |  |  |  |  |  |  |  |

The curve was obtained by thresholding the results at 33 equally spaced instances i.e. [0, 8, 16, , 256] in gray levels or [0, 0.03125, 0.0625 , 1] in probabilities. The AUPR provides a single-figure measure (a.k.a. mean average precision (mAP)), computed over the Set-B, was used to rank the participating methods. This performance metric was used for object detection in The PASCAL Visual Object Classes (VOC) Challenge ([Everingham et al., 2010](#page24)).

16 *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561*

[The AUPR measure is more realistic (Boyd et al., 2013; Saito and](#page24) [Rehmsmeier, 2015) for the lesion segmentation performance over](#page24) the area under Receiver Operating Characteristic (ROC) curve.

*6.2. Sub-challenge – 2*

Let the expert labels for DR and DME be represented by DR*G*(n) and DME*G*(n). Whereas DR*O*(n) and DME*O*(n) are the predicted re-sults, then correct instance is the case when the expert label for DR and DME matches with the predicted outcomes for both DR and DME. This was done since, even with the presence of some exudation that may be categorized as mild DR, its location on the retina is also an important governing factor (to check DME) to de-cide the overall grade of disease. For instance, EXs presence in the macular region can affect the vision of patient to a greater ex-tent and hence, it should be dealt with priority for referral (that may otherwise be missed or cause a delay in treatment with the present convention of only DR grading) in the automated screen-ing systems. Hence, disease grading performance accuracy for this sub-challenge, from the results submitted in CSV format for test images (i.e. N = 103), is obtained by [algorithm 1](#page16) as follows:

**Algorithm 1:** Computation of disease grading accuracy.

**Data**: Method Results and Labels with DR and DME Grading

**Result**: Average disease grading accuracy for DR and DME

* **for** *n*=1*,*2*,*· · ·*,**N* **do 2** Correct = 0;

**3if** *(DRO**(n)*==*DRG**(n))**and**(DMEO**(n)*==*DMEG**(n))* **then**

**4**Correct = Correct + 1;

* **end**

**6 end**

**7** Average Accuracy =Correct

*N*

*6.3. Sub-challenge – 3*

For the given retinal image, the objective of sub-challenge – 3 (task - 6 and 7) was to predict the OD and fovea center co-ordinates. The performance of results submitted in CSV format was evaluated by computing the Euclidean distance (in pixels) between manual (ground truth) and automatically predicted center location. Lower Euclidean distance indicates better localization. After deter-mining these distances for each image in the Set-B, i.e. for 103 images, the average distance representing the whole dataset was computed and used to rank the participating methods.

The optic disc segmentation (task - 8) performance is evaluated using Jaccard index (J) ([Jaccard, 1908](#page25)). It represents the proportion of overlapping area between the segmented OD (O) and the ground truth (G).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *J* | = | |*O* ∩ *G*| | (12) |  |
|  | |*O* ∪ *G*| |  |  |

Higher J indicates better segmentation. For the segmented results, images in range [0, 255], it was computed at 10 different equally spaced thresholds [0, 0.1, , 0.9] and averaged to obtain final score.

**7. Results**

This section reports and discusses the results of all sub-challenges. Performance of all competing solutions on the Set-B for all eight subtasks are divided into three sub-challenge categories and discussed including their leaderboard rank.

*7.1. Sub-challenge – 1*

In this section, we present the performance of all competing solutions for the lesion segmentation task. All results received from the participating teams were analyzed using the validation measure given in [Section 6.1](#page15). This measure generated a set of precision-recall curves for each of the different techniques. Out of the total 37 teams that participated in the challenge, 22 teams par-ticipated (a complete list is available on the challenge website) in the sub-challenge–1 whose results were evaluated and ranked us-ing the AUPR values. Amongst them, 7 teams (see [Table 5](#page8)) hav-ing performance within top 4 positions in either of lesion segmen-tation task were invited for the challenge workshop and 3 teams having overall better performance, i.e. solutions developed by the teams that ranked amongst the top three for at least three differ-ent lesion segmentation tasks, presented their work at ISBI.

[Table 8](#page17) summarizes the individual performance (Off-site evalu-ation) of each solution listed in order of their final placement for each subtask. It also contains various approaches followed and ex-ternal dataset (if any) used for training the models. A higher rank indicates more favorable performance for the individual task(s). The top-3 entries according to the individual lesion segmentation task are VRT, iFLYTEK-MIG and PATech. Some sample lesion seg-mentation results illustrated in [Fig. 6](#page18) and their corresponding over-all evaluation score from [Table 8](#page17) gives a better idea of how the evaluation scores correlate with the quality of segmentation. [Fig. 7](#page19) summarizes the performance of top-4 teams per lesion segmen-tation task. The different curves represent the performance of the participating methods for various lesions (MAs, HEs, SEs and EXs). Team VRT achieved highest AUPR score for HE and SE segmenta-tion task. Whereas, team PATech and iFLYTEK-MIG obtained best score for EX and MA segmentation task respectively.

*7.2. Sub-challenge – 2*

This section presents the results achieved (On-site evaluation) by participating teams for the DR and DME grading task. It is important to note that this task was evaluated for simultaneous grading of DR and DME using the validation algorithm outlined in [Section 6.2](#page16) on the Set-B. This algorithm produced an average grad-ing accuracy of joint DR and DME on all images. [Table 9](#page17) summa-rizes the result of teams for the on-site challenge along with the approach followed and external dataset used for training the model by respective team.

The top-performing solution at the “on-site” challenge was pro-posed by team LzyUNCC followed by team VRT and team Mam-moth. [Fig. 8](#page19) shows the average accuracy of competing solutions for the individual as well as simultaneous grading of DR and DME. Teams are observed to perform poorly in the DR grading task that reduced the overall accuracy for simultaneous grading of DR and DME. Major reason seems to be the diﬃcult test set, diﬃculty in accurately discriminating the DR severity grades.

*7.3. Sub-challenge – 3*

This section presents an evaluation of “On-site” results for par-ticipating teams in the sub-challenge – 3, for all three subtasks. The results for subtasks of OD and Fovea center localization were evaluated by computing Euclidean distance, whereas OD segmen-tation results were evaluated and ranked using Jaccard similarity score as outlined in [Section 6.3](#page16). Results from the on-site evalua-tions are reported in [Table 10](#page17) and [Table 11](#page18) that summarises the performance of all participating algorithms for all three subtasks.

The winning methods for localization tasks were developed by team DeepDR and team VRT, with DeepDR performing best in both

|  |  |
| --- | --- |
| *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561* | 17 |

**Table 8**

Sub-challenge – 1 “Off-site” leaderboard highlighting top 4 teams from each lesion (MAs, HEs, SEs and EXs) segmentation task on the testing dataset. It details the approach followed by respective team and external dataset used for training their model (if any).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Lesion | Team name | | | | | AUPR | Approach | Ensemble | Input Size (Pixels) | External dataset |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Microaneurys |  |  |  |  | iFLYTEK | 0.5017 | Cascaded CNN |  | 320 × 320 | × |  |
|  |  |  |  |  |
|  |  |  |
|  |  |  |  |  | VRT | 0.4951 | U-Net | × | 1280 × 1280 | × |  |
|  |  |  |  |  |
|  |  |  |  |  | PATech | 0.4740 | DenseNet+U-Net |  | 256 × 256 | × |  |
|  |  |  |  |  |  |
|  |  |  |  |  | SDNU | 0.4111 | Mask R-CNN | × | 3584 × 2380 | × |  |
|  |  |  |  |  |  |
| Hemorrhages |  |  |  |  | VRT | 0.6804 | U-Net | × | 640 × 640 | × |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  | PATech | 0.6490 | DenseNet+U-Net |  | 256 × 256 | × |  |
|  |  |  |  |  |  |
|  |  |  |  |  | iFLYTEK | 0.5588 | Cascaded CNN |  | 320 × 320 | × |  |
|  |  |  |  |  |  |
|  |  |  |  |  | SOONER | 0.5395 | U-Net | × | 380 × 380 | × |  |
|  |  |  |  |  |  |
| Soft Exudates |  |  |  |  | VRT | 0.6995 | U-Net | × | 640 × 640 | × |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  | LzyUNCC-I | 0.6607 | FCN+DLA | × | 1024 × 1024 | E-ophtha |  |
|  |  |  |  |  |  |
|  |  |  |  |  | iFLYTEK | 0.6588 | Cascaded CNN |  | 320 × 320 | × |  |
|  |  |  |  |  |  |
|  |  |  |  |  | LzyUNCC-II | 0.6259 | FCN+DLA | × | 1024 × 1024 | E-ophtha |  |
|  |  |  |  |  |  |
| Hard Exudates |  |  |  |  | PATech | 0.8850 | DenseNet+U-Net |  | 256 × 256 | × |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  | iFLYTEK | 0.8741 | Cascaded CNN |  | 320 × 320 | × |  |
|  |  |  |  |  |  |
|  |  |  |  |  | SAIHST | 0.8582 | U-Net | × | 512 × 512 | × |  |
|  |  |  |  |  |  |
|  |  |  |  |  | LzyUNCC-I | 0.8202 | FCN+DLA | × | 1024 × 1024 | E-ophtha |  |
|  |  |  |  |  |  |



**Table 9**

Sub-challenge – 2 “On-site” leaderboard highlighting performance of top 6 teams for DR and DME grading on the test dataset. It details the approach followed by respective team and external dataset used for training their model.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Team Name | | | Accuracy | Approach | Ensemble | Input Size (Pixels) | External Dataset |  |
|  |  |  |  |  |  |  |  |  |
|  |  | LzyUNCC | 0.6311 | Resnet + DLA | 5 | 896 × 896 | Kaggle |  |
|  |  |  |
|  |  |  |
|  |  | VRT | 0.5534 | CNN | 10 | 640 × 640 | Kaggle, Messidor |  |
|  |  |  |
|  |  | Mammoth | 0.5146 | DenseNet |  | 512 × 512 | Kaggle |  |
|  |  |  |
|  |  | HarangiM1 | 0.4757 | AlexNet + GoogLeNet | 2 | 224 × 224 | Kaggle |  |
|  |  |  |
|  |  | AVSASVA | 0.4757 | ResNet + DenseNet | DR-8, DME-5 | 224 × 224 | DiaretDB1 |  |
|  |  |  |
|  |  | HarangiM2 | 0.4078 | AlexNet + Handcrafted features | 2 | 224 × 224 | Kaggle |  |
|  |  |  |



**Table 10**

“On-site” leaderboard highlighting performance of top 5 teams in OD and fovea localization task on the test dataset. It highlights the approach followed by respective team and external dataset used for training their model (if any). ED: Euclidean distance.

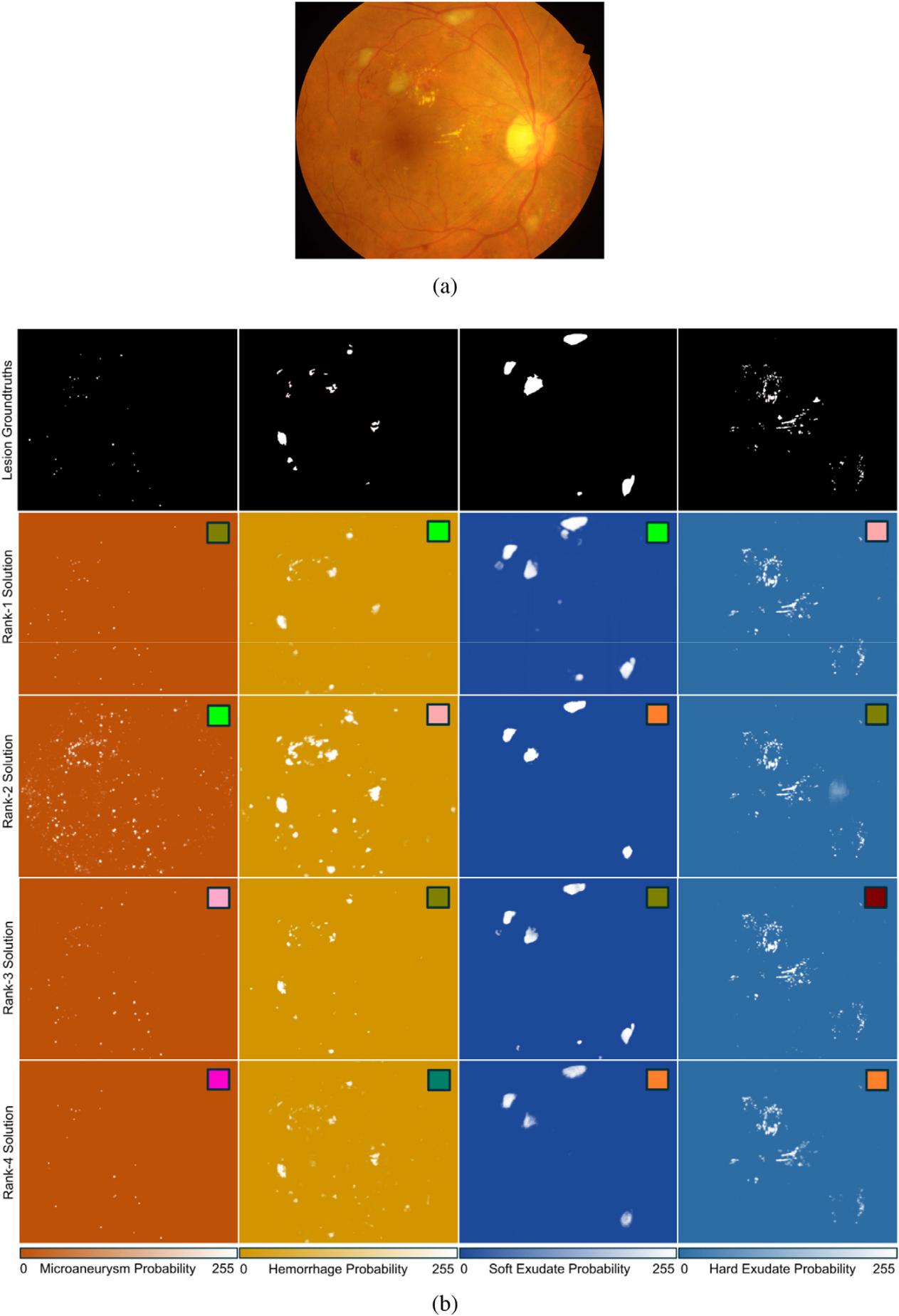
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Localize | Team Name | | | | ED (Pixels) | Rank | Approach | Input Size (Pixels) | | External Dataset |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| Optic Disc |  |  |  | DeepDR | 21.072 | 1 | ResNet + VGG | 224 | × 224, 950 × 950 | – |  |
|  |  |  |  |
|  |  |  |
|  |  |  |  | VRT | 33.538 | 2 | U-Net | 640 | × 640 | DRIVE |  |
|  |  |  |  |  |
|  |  |  |  | ZJU-BII-SGEX | 33.875 | 3 | Mask R-CNN | 1024 × 1024 | | RIGA |  |
|  |  |  |  |  |
|  |  |  |  | SDNU | 36.220 | 4 | Mask R-CNN | 1984 × 1318 | | – |  |
|  |  |  |  |  |
|  |  |  |  | CBER | 29.183 | – | Handcrafted Features | 536 | × 356 | – |  |
|  |  |  |  |  |
| Fovea |  |  |  | DeepDR | 64.492 | 1 | ResNet + VGG | 224 | × 224, 950 × 950 | – |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  | VRT | 68.466 | 2 | U-Net | 640 | × 640 | DRIVE |  |
|  |  |  |  |  |
|  |  |  |  | SDNU | 85.400 | 3 | Mask R-CNN | 1984 × 1318 | |  |  |
|  |  |  |  |  |  |
|  |  |  |  | ZJU-BII-SGEX | 570.133 | 4 | Mask R-CNN | 1024 × 1024 | | RIGA |  |
|  |  |  |  |  |
|  |  |  |  | CBER | 59.751 | – | Handcrafted Features | 536 | × 356 | – |  |
|  |  |  |  |  |



OD and Fovea detection tasks. But the winning entries for OD seg-mentation task were from teams ZJU-BII-SGEX, VRT and IITKgp-KLIV. Some sample OD segmentation results from these teams are illustrated in [Fig. 9](#page20).

[Fig. 10](#page20) shows box-plots illustrating the range of Euclidean dis-tances from the center of (a) OD and (b) fovea as well as (c) spread of Jaccard index for OD segmentation.

18 *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561*



**Fig. 6.** Illustration of lesion segmentation results: (a) sample image and (b) segmentation outcome of top-4 teams (from left to right) (i) MAs, (ii) HEs, (iii) SEs, and (iv) EXs in retinal fundus images. Top row corresponds to ground truths, second row to entry from top performing team, similarly, third, fourth and fifth rows correspond to entries from other three teams respectively. The lesion segmentation entries are colored for better illustration and separation from each type of lesion.

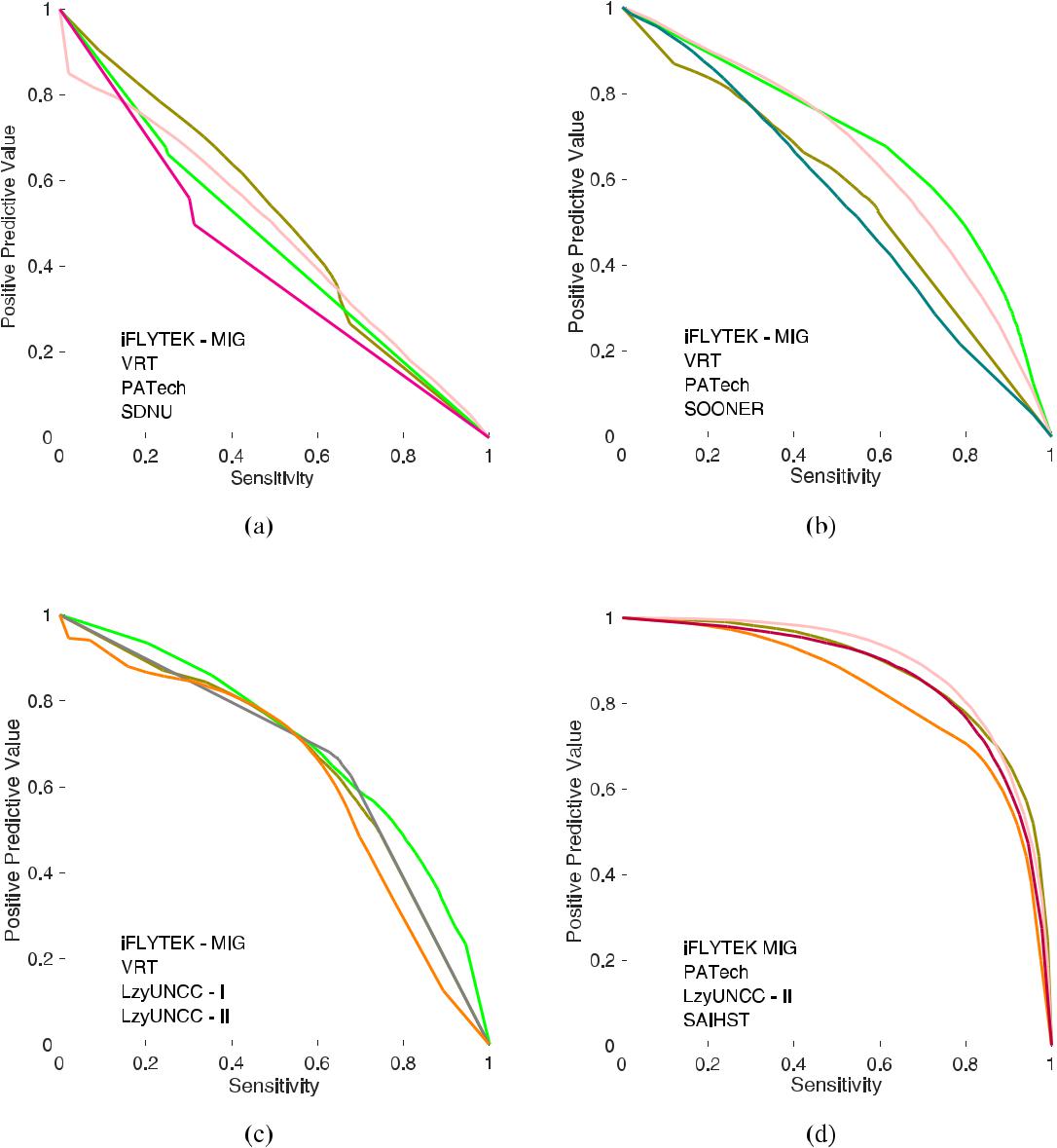
**Table 11**

“On-site” leaderboard highlighting performance of top 5 teams in OD segmentation task on the test dataset. It details the approach followed by respective team and external dataset used for training their model (if any). J: Jaccard index.

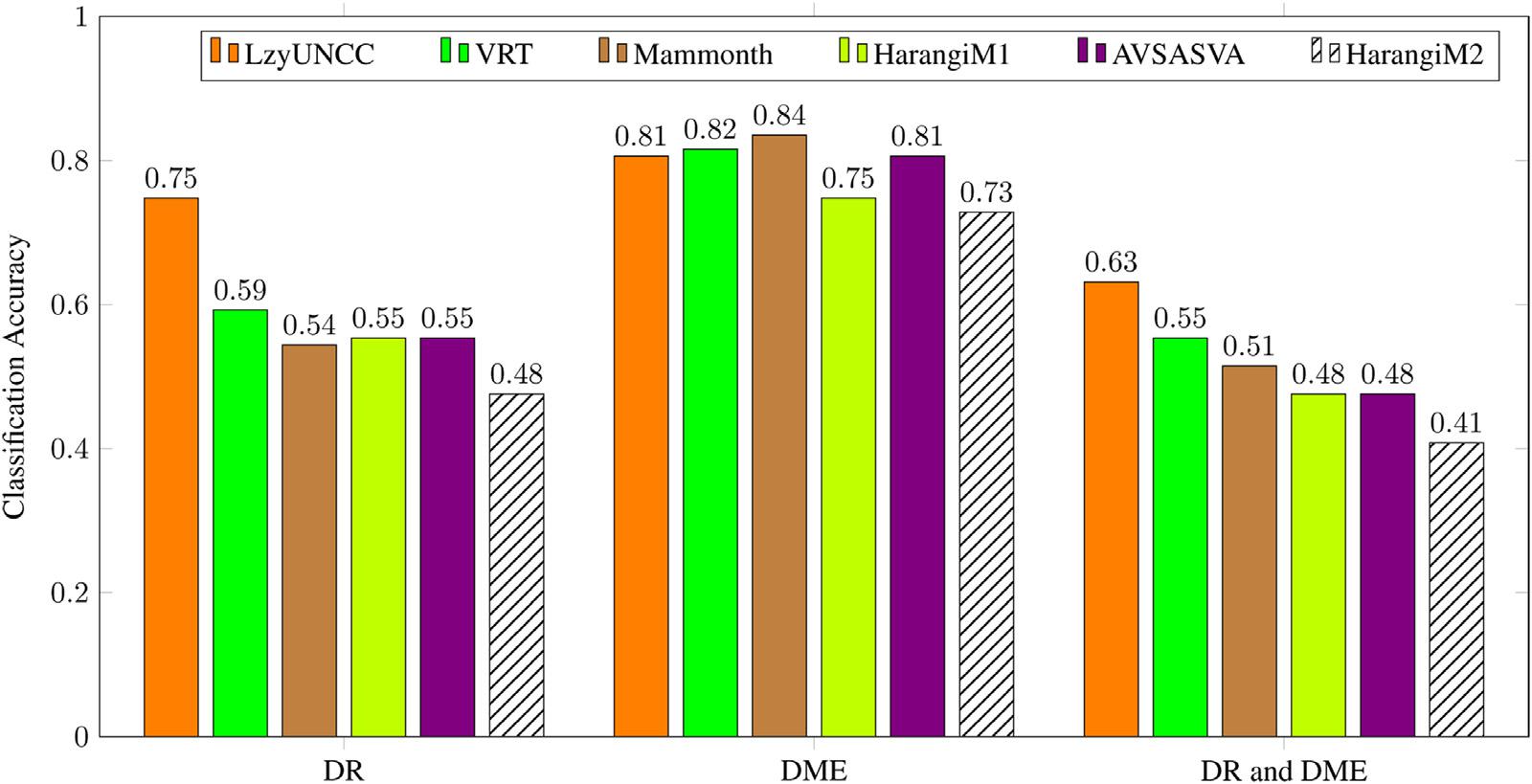
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Team name | | | J | Rank | Approach | Input size (Pixels) | External dataset |  |
|  |  |  |  |  |  |  |  |  |
|  |  | ZJU-BII-SGEX | 0.9338 | 1 | Mask R-CNN | 1024 × 1024 | RIGA |  |
|  |  |  |
|  |  |  |
|  |  | VRT | 0.9305 | 2 | U-Net | 640 × 640 | DRIVE, DRIONS-DB |  |
|  |  |  |
|  |  | IITKgpKLIV | 0.8572 | 3 | SegNet | 536 × 356 | Drishti-GS |  |
|  |  |  |
|  |  | SDNU | 0.7892 | 4 | Mask R-CNN | 1984 × 1318 | – |  |
|  |  |  |
|  |  | CBER | 0.8912 | – | Handcrafted Features | 536 × 356 | – |  |
|  |  |  |



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561* | | | | | | | | | | | | | | | 19 | |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

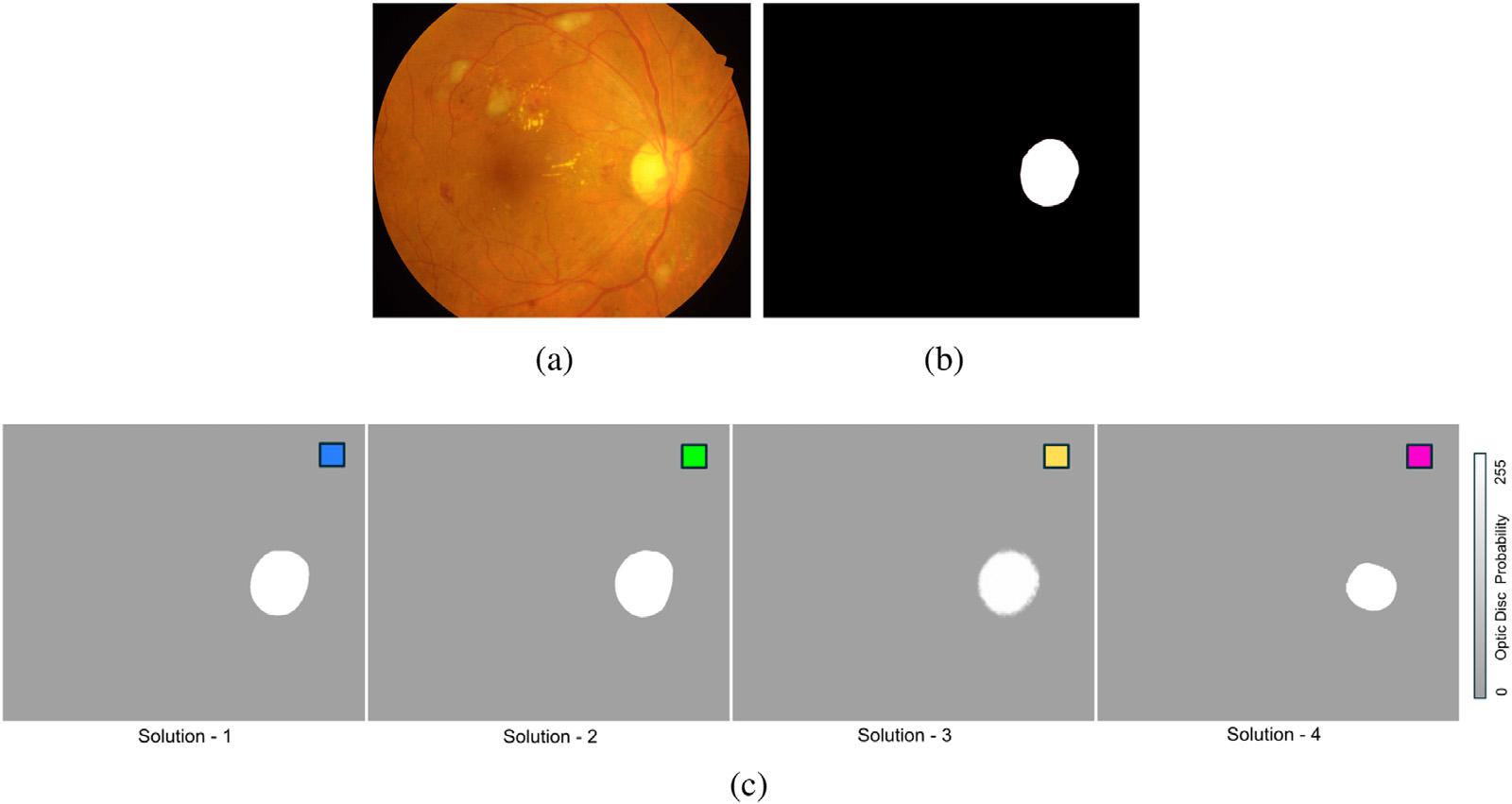


**Fig. 7.** The AUPR curves for the four top performing individual methods on the test dataset. These curves plot the sensitivity versus the positive predictive values for the different lesions, namely, (a) MAs, (b) HEs, (c) SEs, and (d) EXs.

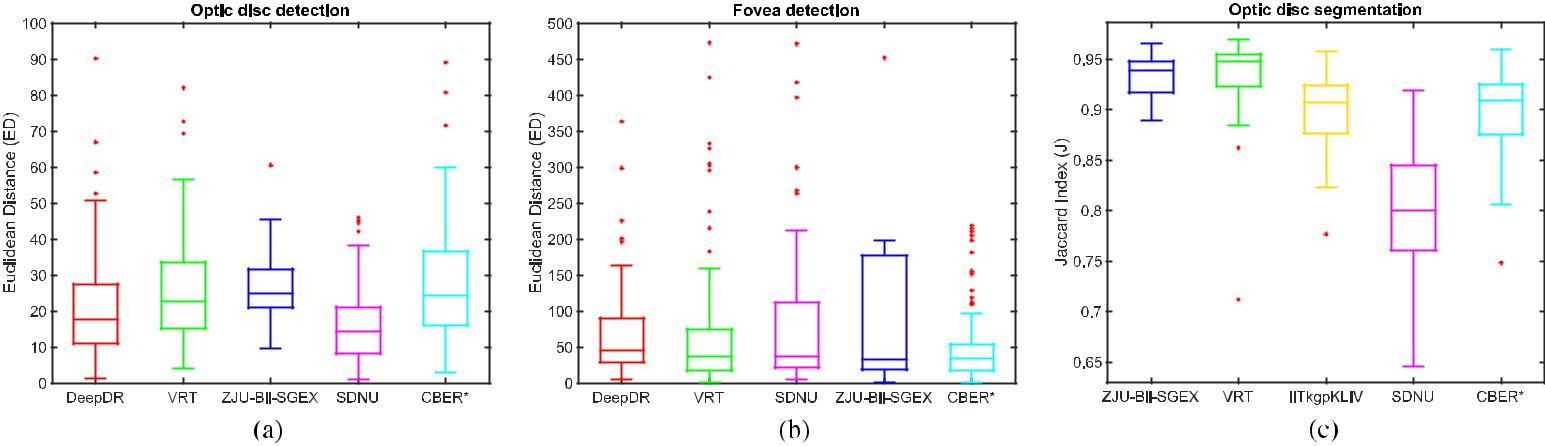


**Fig. 8.** Barplots showing separate and simultaneous classification accuracy of solutions developed by top - 6 teams for grading of DR and DME.

20 *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561*



**Fig. 9.** Illustration of OD segmentation results: (a) sample image, (b) OD ground truth and (c) segmentation outcome of top-4 teams (from left to right).



**Fig. 10.** Boxplots (a,b) showing dispersion of Euclidean distance for individual methods for OD and fovea and (c) showing the dispersion of Jaccard index for OD segmentation task. Boxplots show quartile ranges of the scores on the test dataset; plus sign indicate outliers (full range of data is not shown).

**8. Discussion and conclusion**

In this paper, we have presented the details of IDRiD challenge including information about the data, evaluation metrics, an orga-nization of the challenge, competing solutions and final results for all sub-tasks, i.e., lesion segmentation, disease grading and local-ization and segmentation of other normal retinal structures. Given the significant number of participating teams (37) and results ob-tained, we believe this challenge was a success. To the organiza-tional end, efforts have been made in creating a relevant, stimu-lating and fair competition, capable of advancing collective knowl-edge in the research community. This section presents a discussion, limitations, and lessons learned from this challenge.

The first sub-challenge was conducted in an off-site mode in which 22 teams participated with their lesion segmentation meth-ods. The results of these methods on the Set-B were evaluated by the organizers and amongst them, top-4 performing methods per lesion segmentation task are included in this paper. The computed AUPR values ranged between 0.4111 (for MAs) and 0.885 (for EXs). When the performance of top solutions was analyzed by comput-ing the area under ROC curve (AUC) at the pixel level, in threshold range [0:0.01:1], it resulted in AUC of 0.8263, 0.9716, 0.9540 and 0.9883 for MA, HE, SE and EX respectively. The best approach for lesion segmentation used U-Net, with data augmentation and ad-

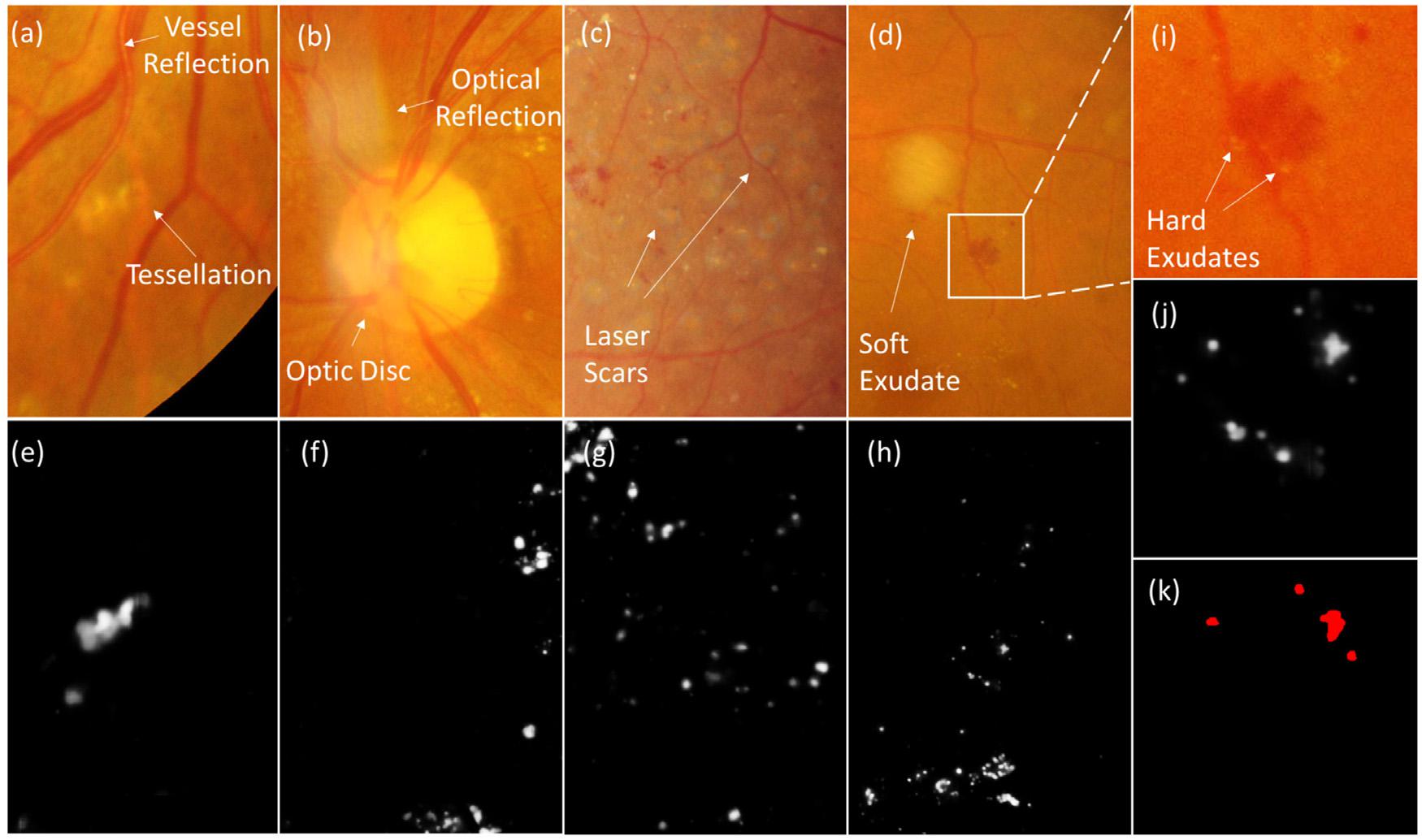
dition of dense block to extract features eﬃciently, boosting results significantly. [Fig. 11](#page21) highlights the performance of top solution for EX that performs significantly well in the presence of normal reti-nal structures and different challenging circumstances.

From the top-performing approaches, it is evident that solving the data imbalance problem improves the model performance sig-nificantly. Since background overwhelms foreground i.e. there are more background pixels than lesion pixels (see [Fig. 6](#page18)), the loss dur-ing training is more effectively back-propagated than that of the foreground that penalizes false negatives, boosting the sensitivity of lesion segmentation. In general, the architectural modifications to U-Net-based networks provided widely varying results for the different types of lesion.

For instance, the cascaded CNN approach yielded the best score for MAs segmentation, as it adds modules to reduce false positives. This approach dramatically impacts MA segmentation performance due to the class imbalance of the task. Further, [Fig. 12](#page21) shows that some false positives detected by participating solutions are due to noise, predominantly for MA and HE. This indicates that there is still room for improvement for lesion segmentation tasks with cur-rent fundus cameras.

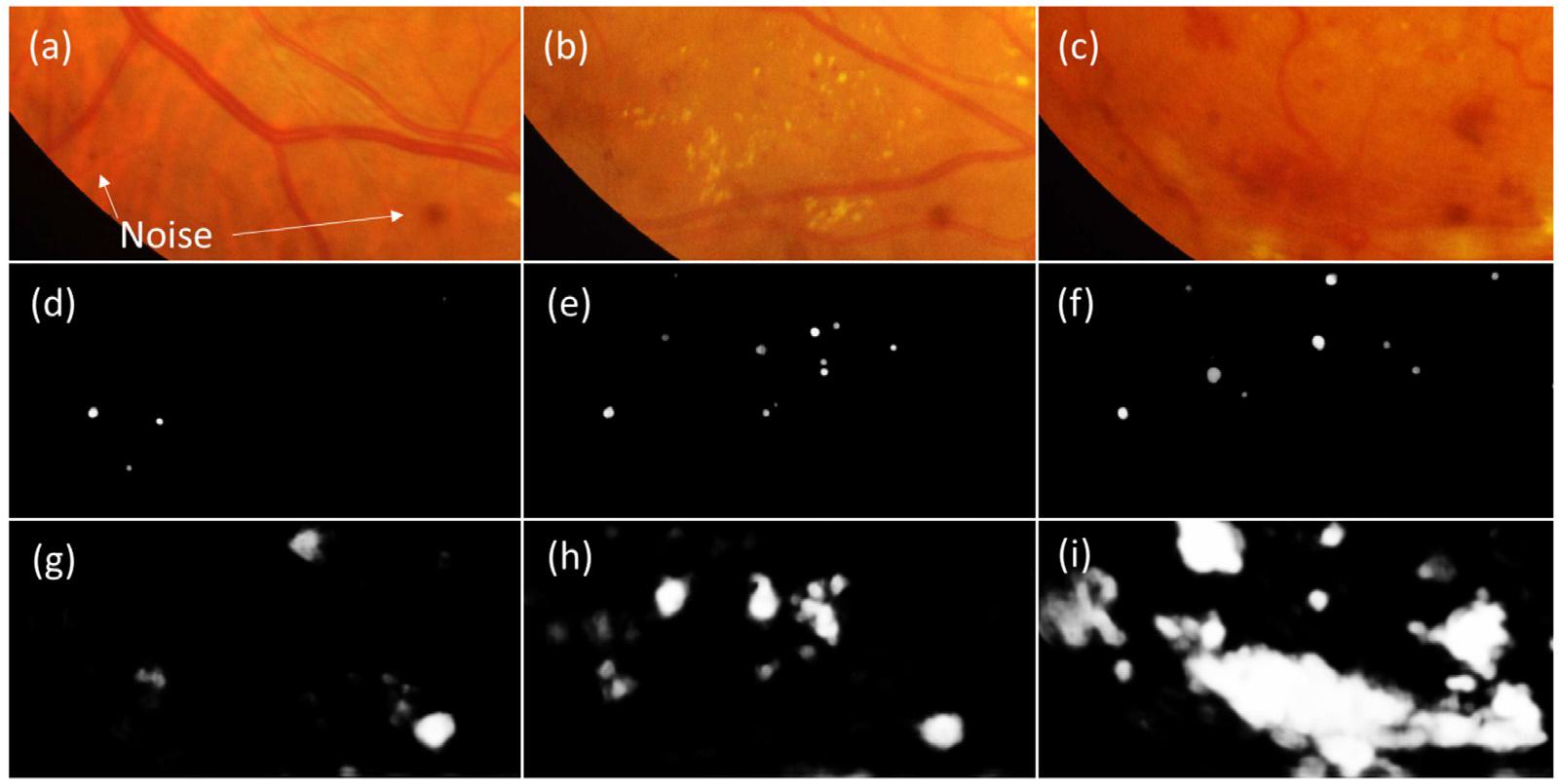
In the on-site disease-grading task, six methods were compared and contrasted. When assessed using the test data set hidden from the participants, the grading accuracy ranged between 0.4078 and

|  |  |
| --- | --- |
| *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561* | 21 |



**Fig. 11.** Illustration of (a-d) different challenging circumstances for segmentation of EXs, (e-h) segmentation results (probability map) of the top-performing team for EXs,

1. enlarged part of Fig. (d), and (j) depicts its performance to be better than (k) the human annotator (The annotator tool had a limitation of the markup capability when there is an overlap of multiple types of lesion. In this case, EXs and HE).



**Fig. 12.** Illustration of results by top performing solutions for (a-c) different images with noise causing most common false positives in the segmentation of (d-f) MAs, and (g-i) HEs respectively.

0.6311 as shown in [Table 9](#page17). Notably, all teams except AVASAVA used the external Kaggle DR dataset for pre-training their models. This dataset contains a large number of retina images annotated with the disease level, in contrast, team AVASAVA pre-trained their model on ImageNet, a dataset containing natural images and ob-ject annotations, effectively showing the network a much smaller number of retina images at the training stage, approximately 1% compared to the other teams. This indicates that in the presence of a limited number of labeled data, transfer learning approaches along with the good model pruning could yield comparable and competitive results. However, while the models do determine the variability of performance, the number, type, and quality of train-ing data is a crucial factor for a fair comparison of competing so-lutions. There is still work needed on simultaneous grading of DR

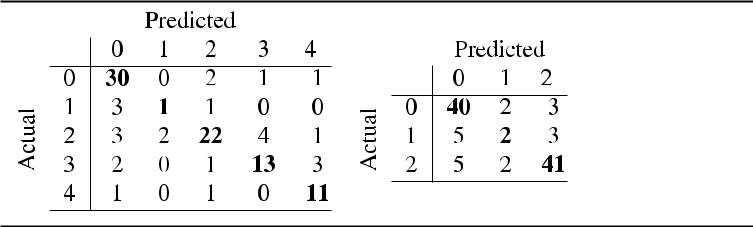
and DME as the reported results do not yet reach the performance needed for a clinically viable automatic screening. Considering the misclassified instances in confusion matrices shown in [Table 12](#page22), along with the lesion information, it is essential to give attention towards characterization of intra-retinal microvascular abnormali-ties (IRMA’s) and venous beading for improvement in the overall grading results.

In the sub-challenge – 3, another on-site challenge, four teams were evaluated for the task of OD/fovea localization and OD seg-mentation. For the task of OD localization, the Euclidean distance varied between 21.072 and 36.22 (lower values indicate better per-formance). However, for Fovea localization task the same perfor-mance metric ranged between 64.492 and 570.133. This massive variation is due to outliers, e.g. team ZJU-BII-SGEX had 23 outliers

22 *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561*

**Table 12**

Confusion matrix of retinal images predicted by top performing solution for DR (5 class) and DME (3 class).



whose Euclidean distance exceeded 700. In the OD segmentation task, the average Jaccard similarity index score amongst the par-ticipants ranged between 0.7892 and 0.9338. The top-performing solutions developed by DeepDR and VRT leveraged prior clinical knowledge, such as the number of landmarks and their geometric relationship to detect another retinal landmark. It is also observed that data augmentation and ensemble of models yield substantial improvements in terms of accuracy. Considering the clinical sig-nificance of OD diameter while DME severity grading, we further compute the average OD diameter (in pixels) for each image of the test set. The average diameter of OD ground truth is 516.61 pixels, whereas, corresponding values for the results of solutions devel-oped by the teams ZJU-BII-SGEX, VRT, IITKgpKLIV, CBER and SDNU are 514.25, 519.21, 513.48, 508.04 and 460.19 pixels respectively. Team CBER submitted their results after the competition and they were not included in the leaderboard.

As expected, we found that image resolution is a vital factor for the model performance, especially for the task of segmentation of small objects such as MAs or EXs. In fact, the top-performing approaches processed the images patch-wise, which allow mod-els to have a local high-resolution image view or directly with the high-resolution image as a whole. This is essential as MAs or small EXs lesions span very few pixels in some cases, and reduc-ing the original image size would prevent an accurate segmenta-tion. Similarly, image resolution plays a very important role for disease classification task (see [Table 9](#page17)), the most likely reason is that presence of the disease is determined by the presence of le-sions in the image, including the small ones that might be invis-ible at low resolution. This is corroborated by the confusion ma-trices in [Table 12](#page22) which show misclassified instances in DR (par-ticularly, grade 1 and 2) as well as DME (5 images each belong-ing to grade 1 and 2 are predicted as grade 0). For the localiza-tion tasks, all participants were asked to identify retinal structures with coordinates at full image resolution. Most of them performed these tasks by scaling image to the smaller size and then con-verted their predictions in the original image space. Comparative analysis indicates that the input image resolution has limited ef-fect on the results of the localization problem. For instance, in the case of OD localization, the top-performing team utilized two image resolutions, one (224 × 224 pixels) for approximate location prediction and other (cropped ROIs 950 × 950 pixels) for refining that estimate. Similarly, teams CBER and VRT resized the image to 536 × 356 pixels and 640 × 640 pixels respectively to get an ap-proximate center location whereas the team SDNU utilized the in-put size of 1984 × 1318 pixels. Considering the OD average diame-ter of approximately 516 pixels, limited performance variation (10 to 15 pixels) is observed as compared to the top-performing solu-tion for huge variation (multiple times) in input resolutions (see [Table 10](#page17)). This is because the retinal structures to be identified, OD and fovea, are very unlikely to disappear due to a reduction of image resolution and they have clear geometrical constraints.

[As confirmed by recent studies (Krause et al., 2018; Son et al.,](#page25) [2019), we hypothesized that algorithms developed using images](#page25) with fine visibility and images having high resolution with ad-

judicated consensus grades yield better performance when com-pared to datasets consisting of poor-quality (non-gradable) images and images captured in varied acquisition settings. Therefore, this challenge provides data collected in the routine clinical practice using an acquisition protocol consistent for all images. The data was acquired after pupil dilation with the same camera at the same resolution, ensuring consistent quality. This dataset did not include non-gradable images and images with substantial disagree-ment amongst the expert annotators. Even after these efforts to provide the best possible data, the annotation process is still in-herently subjective, and the annotator judgment is a limiting factor for the method performance which is mostly trained and evaluated in a supervised manner. We also note that images captured with different retinal cameras or with different diseases would have al-lowed for a better estimation of the generalization ability of the proposed methods since they might be more representative from clinical settings. Further, while we believe that data challenges like ours foster “methodology diversity”, the majority of competing so-lutions used deep convolutional networks. These approaches are comparably easier to implement than approaches based on fea-ture engineering and do generalize well to multiple medical imag-ing domains, which in turn, dramatically reduces the need for spe-cialized task knowledge. Notably, amongst the competing solutions in this challenge that utilized the deep learning approach along with the task-relevant subject knowledge have demonstrated su-perior performance. However, it seems there might be some im-pact of challenge duration, apart from the number of submissions, on the quality of developed solutions. Considering the time span from data availability to deadline of results submission, about one and a half month, was considerably tight for managing all tasks at the same time. For the team VRT who had been working on an-alyzing fundus images for more than a year when participated in the competition that attempting all tasks were possible, still, it was challenging for them to commit all the tasks. However, it would be highly challenging for a newcomer to succeed in multiple tasks. In that sense, the competition period was not suﬃcient for perfecting all tasks. However, it would be enough for a competent participant, e.g. new entrants in the field as team SAIHST, to finish one task if the participant can focus on the competition completely. Also, in this challenge, the results were evaluated all at once after the result submission deadline. However, a continuous on-line assess-ment of participating solutions would have facilitated the submis-sion procedure by providing real-time feedback to the teams per-formance. This would have enabled a maximum number of sub-missions during the challenge period, probably boosting the final count of submissions. However, this would have introduced a risk of overfitting the test data by continuous submissions based on the system’s performance on the test set.

This challenge led to the development of a variety of new ro-bust solutions for lesion segmentation, detection, and segmenta-tion of retinal landmarks and disease severity grading. Despite the complexity of the tasks, less than one-and-a-half month time for development, it received a very positive response, and the top-performing solutions were able to achieve results close to the hu-man annotators. Still, there is room for improvement, especially in the lesion segmentation and disease-grading tasks. Though the competition is now completed, the dataset has been made publicly available for research purposes to attract newcomers to the prob-lem and to encourage the development of novel solutions to meet current and future clinical standards.

**Declaration of Competing Interest**

The authors have no conflicts of interest to declare.

|  |  |
| --- | --- |
| *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561* | 23 |

**Table A.1**

Summary of technical specifications and hardware used in different databases.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name of Database |  | Number of Images | |  |  | Technical Details | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | Image Size(s) | | | | |  |  | FOV | |  |  | Camera | |  |  |  | NMY | Format |  |
|  |  |  |  |  |  | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ARIA | 212 | |  |  | 768 × 576 | | | |  |  |  |  | 50 | |  |  | Zeiss F F 450+ | |  |  |  |  | TIFF |  |
| DIARETDB | 130+89 | | |  | 1500 × | | | 1152 |  |  |  |  | 50 | |  |  | Zeiss F F 450+ | |  |  |  |  | PNG |  |
| DRIVE | 40 | |  |  | 768 × 584 | | | |  |  |  |  | 45 | |  |  | Canon CR5 | |  |  |  |  | JPEG |  |
| E-Ophtha |  | 47EX+35H 148MA+233H | | | 1440 × | | | 960 - 2048 × 1360 (4) | | | |  | 45 | |  |  | Canon CR − DGI & Topcon T RC − NW 6 | | | | |  | JPEG |  |
| HEIMED | 169 | |  |  | 2196 × | | | 1958 |  |  |  |  | 45 | |  |  | Zeiss Visucam PRO | | |  |  |  | JPEG |  |
| Kaggle | 88,702 | | |  | 433 × 289 - 3888 × 2592 | | | | | |  |  | Varying | | | | Any camera (EyePACS Platform) | | | | | – | TIFF |  |
| MESSIDOR | 800 | | MY+ 400 NMY+ 1756 | | 1440 × 960, 2240 × 1488, 2304 × 1536 | | | | | | | | 45 | |  |  | 3CCD/ Topcon TRC NW6 | | |  |  | Both | TIFF |  |
| ROC | 100 | |  |  | 768 × 576, 1058 × 1061, 1389 × 1383 | | | | | | | | 45 | |  |  | Topcon NW100 & NW200 Canon CR5 − 45NM | | | | |  | JPEG |  |
| STARE | 397 | |  |  | 605 × 700 | | | |  |  |  |  | 35 | |  |  | Topcon T RV − 50 | | |  |  | × | PPM |  |
| **IDRiD** | 516 | | (81 with LA) |  | 4288 × | | | 2848 |  |  |  |  | 50 | |  |  | Kowa V X − 10*α* | | |  |  |  | JPG |  |
| EX - Hard Exudate, MA - Microaneurysms, H - Healthy, MY - Mydriatic, NMY - Non-Mydriatic, FOV - Field of View, LA - Lesion Annotation. | | | | | | | | | | | | | | | | | | | | | |  |  |  |
|  | **Table A.2** | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Comparison of different databases with the IDRiD database. | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | |  | | | | |  |  | |  |  |  |  |  | |  |  |  |  |  |  |  |
|  |  | Name of database | | Normal fundus structuresAbnormalities | | | | | | | |  |  |  |  | Multiple experts | | | DR grading | DME grading | |  |  |  |
|  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |
|  |  |  |  | OD | VS | | FA |  |  | MA | HE | EX | SE | | Yes/No# | | | |  |  |  |  |  |  |
|  |  |  |  |  |  | |  |  | | |  |  |  |  |  | |  |  |  |  |  |  |  |  |
|  |  | ARIA |  |  |  | |  | × | | | × | × | × |  |  | | 2 |  | × | × |  |  |  |  |
|  |  | DIARETDB1 | | × | × | | × |  |  |  |  |  |  | |  | | 4 |  | × | × |  |  |  |  |
|  |  | DRIVE | | × |  | | × | × | | | × | × | × |  |  | | 3 |  | × | × |  |  |  |  |
|  |  | E-Optha | | × | × | | × |  |  |  | × |  | × |  |  | | 2 |  | × | × |  |  |  |  |
|  |  | HEIMED | | × | × | | × | × | | | × |  |  | | × | | 1 |  | × |  | |  |  |  |
|  |  | Kaggle | | × | × | | × | × | | | × | × | × |  |  | | 2 |  |  | × |  |  |  |  |
|  |  | MESSIDOR | | × | × | | × | × | | | × | × | × |  | × | | 1 |  |  |  | |  |  |  |
|  |  | ROC |  | × | × | | × |  |  |  | × | × | × |  |  | | 4 |  | × | × |  |  |  |  |
|  |  | STARE | |  |  | | × | × | | | × | × | × |  |  | | 2 |  | × | × |  |  |  |  |
|  |  | **IDRiD** | |  | × | |  |  |  |  |  |  |  | |  | | 2 |  |  |  | |  |  |  |

OD - Optic Disc, VS - Vessels, FA - Fovea, MA - Microaneurysms, HE - Hemorrhage, EX - Hard Exudate, SE - Soft Exudate, # - Number of Experts

**Acknowledgments**

This work is sponsored by the Shri Guru Gobind Singhji In-stitute of Engineering and Technology, Nanded (M.S.), INDIA. The authors would like to thank the following people for their help in various aspects of organizing the ISBI-2018 Diabetic Retinopa-thy Segmentation and Grading Challenge: Prof. Emanuele Trucco (University of Dundee, Scotland), Tom MacGillivray (University of Edinburgh, Scotland), Ravi Kamble (SGGS Institute of Engineering and Technology, Nanded), Prof. Vivek Sahasrabuddhe (Government Medical College, Nanded) and Désiré Sidibé (Université de Bour-gogne, France). We would also like to thank Prof. Jorge Cuadros, University of California, Berkeley (Organizer of Kaggle Diabetic Retinopathy challenge) for his kind permission for reporting the results of the models trained on their dataset. VRT: This study was supported by the Research Grant for Intelligence Information [Service Expansion Project, which is funded by.National IT Industry](https://doi.org/10.13039/501100003665) [Promotion Agency (](https://doi.org/10.13039/501100003665)[NIPA-C0202-17-1045) in South Korea. *DeepDR*:](#page23) [This work was supported in part by the National Natural Science](https://doi.org/10.13039/501100001809) [Foundation of China under Grant Grant](https://doi.org/10.13039/501100001809) [61872241, Grant 61572316,](#page23) [in part by the National Key Research and Development Program of](https://doi.org/10.13039/501100013290) [China under Grant](https://doi.org/10.13039/501100013290) [2016YFC1300302 and Grant 2017YFE0104000,](#page23) [in part by the Science and Technology Commission of Shanghai](https://doi.org/10.13039/501100003399) [Municipality under Grant](https://doi.org/10.13039/501100003399) [16DZ0501100 and Grant 17411952600.](#page23) *HarangiM1-M2*: Research was supported in part by the Janos Bolyai

Research Scholarship of the Hungarian Academy of Sciences and the project EFOP-3.6.2-16-2017-00015 supported by the European Union and the State of Hungary, co-financed by the European So-cial Fund. ZJU-BII-SGEX: This work is supported by Beijing Shang-gong Medical Technology Co., Ltd., which provided ocular health-care solutions in China. This research is partially supported by the A∗ STAR A1818g0022 grant of Singapore. Many thanks to the la-beled images from Image Annotation Group of Beijing Shanggong Medical Technology. Team CBER (A.M. Mendonça, T. Melo, T. Araújo and A. Campilho) is financed by the ERDF European Regional De-velopment Fund through the Operational Programme for Competi-tiveness and Internationalisation - COMPETE 2020 Programme, and by National Funds through the FCT Fundação para a Ciência e a Tecnologia ([Portuguese Foundation for Science and Technology](https://doi.org/10.13039/501100001871)) within project [CMUP-ERI/TIC/0028/2014](#page23). Teresa Araújo is funded by the FCT grant SFRH/BD/122365/2016. SDNU: This study was sup-ported by the National Natural Science Foundation of China (Grant No. 61572300).

**Appendix A. Comparison of Publicly Available Retinal Image Databases**

[Table A.1](#page23) and [Table A.2](#page23) provides the summary of technical spec-ifications and available ground truths in several existing datasets and the IDRiD dataset.

24 *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561*

**Supplementary material**

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.media.2019.101561](https://doi.org/10.1016/j.media.2019.101561)

**References**

Abdulla, W., 2017. Mask r-CNN for object detection and instance segmentation on keras and tensorflow. [https://github.com/matterport/Mask\_RCNN.](https://www.github.com/matterport/Mask_RCNN)

[Abràmoff, M.D., Garvin, M.K., Sonka, M., 2010. Retinal imaging and image analysis.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0001)

[IEEE Rev. Biomed. Eng. 3, 169–208.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0001)

[Abràmoff, M.D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J.C., Niemeijer, M.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0002) [2016. Improved automated detection of diabetic retinopathy on a publicly avail-able dataset through integration of deep learning. Investigat. Ophthalmol. Vis.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0002) [Sci. 57 (13), 5200–5206.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0002)

[Acharya, R., Chua, C.K., Ng, E., Yu, W., Chee, C., 2008. Application of higher order](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0003) [spectra for the identification of diabetes retinopathy stages. J. Med. Syst. 32 (6),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0003) [481–488.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0003)

[Acharya, U.R., Mookiah, M.R.K., Koh, J.E.W., Tan, J.H., Bhandary, S.V., Rao, A.K., Hagi-wara, Y., Chua, C.K., Laude, A., 2017. Automated diabetic macular edema (DME)](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0004) [grading system using DWT, DCT features and maculopathy index. Comput. Biol.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0004) [Med. 84, 59–68.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0004)

[Acharya, U.R., Ng, E.Y.-K., Tan, J.-H., Sree, S.V., Ng, K.-H., 2012. An integrated index](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0005) [for the identification of diabetic retinopathy stages using texture parameters. J.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0005) [Med. Syst. 36 (3), 2011–2020.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0005)

[Adal, K.M., Sidibé, D., Ali, S., Chaum, E., Karnowski, T.P., Mériaudeau, F., 2014. Au-tomated detection of microaneurysms using scale-adapted blob analysis and](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0006) [semi-supervised learning. Comput. Method. Progr.. Biomed. 114 (1), 1–10.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0006)

[Agurto, C., Murray, V., Barriga, E., Murillo, S., Pattichis, M., Davis, H., Russell, S., Abrà-moff, M., Soliz, P., 2010. Multiscale AM-FM methods for diabetic retinopathy le-sion detection. IEEE Trans. Med. Imag. 29 (2), 502–512.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0007)

[Almazroa, A., Alodhayb, S., Osman, E., Ramadan, E., Hummadi, M., Dlaim, M.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0008) [Alkatee, M., Raahemifar, K., Lakshminarayanan, V., 2018. Retinal fundus images](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0008) [for glaucoma analysis: the RIGA dataset. In: Medical Imaging 2018: Imaging In-formatics for Healthcare, Research, and Applications, Vol. 10579. International](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0008) [Society for Optics and Photonics, p. 105790B.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0008)

[Antal, B., Hajdu, A., 2012. Improving microaneurysm detection using an optimally](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0009) [selected subset of candidate extractors and preprocessing methods. Pattern](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0009) [Recognit. 45 (1), 264–270.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0009)

[Antal, B., Hajdu, A., 2014. An ensemble-based system for automatic screening of](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0010) [diabetic retinopathy. Knowl.-Based Syst. 60, 20–27.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0010)

Atlas, I.D.F.D., 2017. Brussels, belgium: international diabetes federation. Int. Diabet. Federat. (IDF). <http://diabetesatlas.org/resources/2017-atlas.html>

Badrinarayanan, [V., Kendall, A., Cipolla, R., 2015. Segnet: a deep convolu-tional encoder-decoder architecture for image segmentation. arXiv preprint](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0012) [arXiv:1511.00561.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0012)

[Bai, J., Miri, M.S., Liu, Y., Saha, P., Garvin, M., Wu, X., 2014. Graph-based optimal mul-ti-surface segmentation with a star-shaped prior: application to the segmenta-tion of the optic disc and cup. In: 2014 IEEE 11th International Symposium on](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0013) [Biomedical Imaging (ISBI). IEEE, pp. 525–528.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0013)

[Bandello, F., Parodi, M.B., Lanzetta, P., Loewenstein, A., Massin, P., Menchini, F.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0014) [Veritti, D., 2010. Diabetic macular edema. In: Macular Edema, Vol. 47. Karger](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0014) [Publishers, pp. 73–110.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0014)

[Biyani, R.S., Patre, B.M., 2018. Algorithms for red lesion detection in diabetic](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0015) [retinopathy: a review. Biomed. Pharmacother. 107, 681–688.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0015)

[Bourne, R.R.A., Stevens, G.A., White, R.A., Smith, J.L., Flaxman, S.R., Price, H.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0016) [Jonas, J.B., Keeffe, J., Leasher, J., Naidoo, K., et al., 2013. Causes of vision](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0016) [loss worldwide, 1990–2010: a systematic analysis. Lancet Global Health 1 (6),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0016) [e339–e349.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0016)

[Boyd, K., Eng, K.H., Page, C.D., 2013. Area under the precision-recall curve: point](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0017) [estimates and confidence intervals. In: Joint European Conference on Machine](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0017) [Learning and Knowledge Discovery in Databases. Springer, pp. 451–466.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0017)

[Carin, L., Pencina, M.J., 2018. On deep learning for medical image analysis. JAMA](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0018) [320 (11), 1192–1193.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0018)

[Carmona, E.J., Rincón, M., García-Feijoó, J., Martínez-de-la Casa, J.M., 2008. Identifi-cation of the optic nerve head with genetic algorithms. Artif. Intell. Med. 43 (3),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0019) [243–259.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0019)

[Carson Lam, D.Y., Guo, M., Lindsey, T., 2018. Automated detection of diabetic](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0020) [retinopathy using deep learning. AMIA Summit. Translat. Sci. Proc. 2017, 147.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0020)

[Cheng, J., Yin, F., Wong, D.W.K., Tao, D., Liu, J., 2015. Sparse dissimilarity-constrained](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0021) [coding for glaucoma screening. IEEE Trans. Biomed. Eng. 62 (5), 1395–1403.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0021)

[Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0022) [Ferrero, E., Agapow, P.-M., Zietz, M., Hoffman, M.M., et al., 2018. Opportunities](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0022) [and obstacles for deep learning in biology and medicine. J. R. Soc. Interface 15](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0022) [(141), 20170387.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0022)

[Chudzik, P., Majumdar, S., Calivá, F., Al-Diri, B., Hunter, A., 2018. Microaneurysm](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0023) [detection using fully convolutional neural networks. Comput. Method. Progr.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0023) [Biomed. 158, 185–192.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0023)

[Ciulla, T.A., Amador, A.G., Zinman, B., 2003. Diabetic retinopathy and diabetic mac-ular edema: pathophysiology, screening, and novel therapies. Diabetes Care 26](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0024)

[(9), 2653–2664.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0024)

[Cuadros, J., Bresnick, G., 2009. EyePACS: an adaptable telemedicine system for dia-betic retinopathy screening. J. Diabetes Sci. Technol. 3 (3), 509–516.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0025)

[Dai, L., Fang, R., Li, H., Hou, X., Sheng, B., Wu, Q., Jia, W., 2018. Clinical report guided](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0026) [retinal microaneurysm detection with multi-sieving deep learning. IEEE Trans.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0026) [Med. Imag. 37 (5), 1149–1161.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0026)

[Das, V., Puhan, N.B., Panda, R., 2015. Entropy thresholding based microaneurysm](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0027) [detection in fundus images. In: 2015 Fifth National Conference on Computer](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0027) [Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG). IEEE,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0027) [pp. 1–4.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0027)

[Dashtbozorg, B., Mendonça, A.M., Campilho, A., 2015. Optic disc segmentation using](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0028) [the sliding band filter. Comput. Biol. Med. 56, 1–12.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0028)

[Decencière, E., Cazuguel, G., Zhang, X., Thibault, G., Klein, J.-C., Meyer, F., Mar-cotegui, B., Quellec, G., Lamard, M., Danno, R., et al., 2013. Teleophta: machine](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0029) [learning and image processing methods for teleophthalmology. IRBM 34 (2),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0029) [196–203.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0029)

[Decencière, E., Zhang, X., Cazuguel, G., Laÿ, B., Cochener, B., Trone, C., Gain, P., Or-dóñez Varela, J.-R., Massin, P., Erginay, A., et al., 2014. Feedback on a publicly](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0030) [distributed image database: the messidor database. Image Anal. Stereol. 33 (3),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0030) [231–234.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0030)

[Deepak, K.S., Sivaswamy, J., 2012. Automatic assessment of macular edema from](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0031) [color retinal images. IEEE Trans. Med. Imag. 31 (3), 766–776.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0031)

[Dhara, A.K., Mukhopadhyay, S., Bency, M.J., Rangayyan, R.M., Bansal, R., Gupta, A.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0032) [2015. Development of a screening tool for staging of diabetic retinopathy in](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0032) [fundus images. In: Medical Imaging 2015: Computer-Aided Diagnosis, Vol. 9414.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0032) [International Society for Optics and Photonics, p. 94140H.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0032)

[Dobbin, K.K., Simon, R.M., 2011. Optimally splitting cases for training and testing](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0033) [high dimensional classifiers. BMC Med. Genom. 4 (1), 31.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0033)

[Esteves, T., Quelhas, P., Mendonça, A.M., Campilho, A., 2012. Gradient convergence](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0034) [filters and a phase congruency approach for in vivo cell nuclei detection. Mach.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0034) [Vis. Appl. 23 (4), 623–638.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0034)

[Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The pas-cal visual object classes (voc) challenge. Int. J. Comput. vis. 88 (2), 303–338.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0035)

[Farnell, D.J.J., Hatfield, F.N., Knox, P., Reakes, M., Spencer, S., Parry, D., Harding, S.P.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0036) [2008. Enhancement of blood vessels in digital fundus photographs via the ap-plication of multiscale line operators. J. Franklin Inst. 345 (7), 748–765.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0036)

[Ferris, F.L., 1993. How effective are treatments for diabetic retinopathy? JAMA 269](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0037) [(10), 1290–1291.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0037)

[Figueiredo, I.N., Kumar, S., Oliveira, C.M., Ramos, J.a.D., Engquist, B., 2015. Automated](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0038) [lesion detectors in retinal fundus images. Comput. Biol. Med. 66, 47–65.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0038)

[Fleming, A.D., Philip, S., Goatman, K.A., Olson, J.A., Sharp, P.F., 2006. Automated mi-croaneurysm detection using local contrast normalization and local vessel de-tection. IEEE Trans. Med. Imag. 25 (9), 1223–1232.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0039)

[Fraz, M.M., Badar, M., Malik, A.W., Barman, S.A., 2018. Computational methods for](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0040) [exudates detection and macular edema estimation in retinal images: a survey.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0040) [Arch. Comput. Method. Eng. 1–28.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0040)

[Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X., 2018. Joint optic disc and](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0041) [cup segmentation based on multi-label deep network and polar transformation.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0041) [arXiv preprint arXiv:1801.00926.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0041)

[Fu, H., Xu, Y., Wong, D.W.K., Liu, J., 2016. Retinal vessel segmentation via deep](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0042) [learning network and fully-connected conditional random fields. In: Biomedical](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0042) [Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE, pp. 698–701.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0042)

[García, G., Gallardo, J., Mauricio, A., López, J., Del Carpio, C., 2017. Detection of di-abetic retinopathy based on a convolutional neural network using retinal fun-dus images. In: International Conference on Artificial Neural Networks. Springer,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0043) [pp. 635–642.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0043)

[Garcia, M., Sanchez, C.I., Poza, J., López, M.I., Hornero, R., 2009. Detection of hard](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0044) [exudates in retinal images using a radial basis function classifier. Annal. Biomed.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0044) [Eng. 37 (7), 1448–1463.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0044)

[Gargeya, R., Leng, T., 2017. Automated identification of diabetic retinopathy using](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0045) [deep learning. Ophthalmology 124 (7), 962–969.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0045)

[Gegundez-Arias, M.E., Marin, D., Bravo, J.M., Suero, A., 2013. Locating the fovea cen-ter position in digital fundus images using thresholding and feature extraction](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0046) [techniques. Comput. Med. Imag. Graph. 37 (5–6), 386–393.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0046)

[Giachetti, A., Ballerini, L., Trucco, E., 2014. Accurate and reliable segmentation of the](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0047) [optic disc in digital fundus images. J. Med. Imag. 1 (2), 024001.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0047)

[Giancardo, L., Meriaudeau, F., Karnowski, T.P., Li, Y., Garg, S., Tobin, K.W., Chaum, E.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0048) [2012. Exudate-based diabetic macular edema detection in fundus images using](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0048) [publicly available datasets. Med. Image Anal. 16 (1), 216–226.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0048)

[Giancardo, L., Meriaudeau, F., Karnowski, T.P., Li, Y., Tobin, K.W., Chaum, E., 2011.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0049) [Microaneurysm detection with radon transform-based classification on retina](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0049) [images. In: 2011 Annual International Conference of the IEEE Engineering in](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0049) [Medicine and Biology Society. IEEE, pp. 5939–5942.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0049)

[Giancardo, L., Roberts, K., Zhao, Z., 2017. Representation learning for retinal vas-culature embeddings. In: Fetal, Infant and Ophthalmic Medical Image Analysis.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0050) [Springer, pp. 243–250.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0050)

[Glorot, X., Bengio, Y., 2010. Understanding the diﬃculty of training deep feedfor-ward neural networks. In: Proceedings of the Thirteenth International Confer-ence on Artificial Intelligence and Statistics, pp. 249–256.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0051)

[Greenspan, H., Van Ginneken, B., Summers, R.M., 2016. Guest editorial deep learning](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0052) [in medical imaging: overview and future promise of an exciting new technique.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0052) [IEEE Trans. Med. Imag. 35 (5), 1153–1159.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0052)

[van Grinsven, M.J., van Ginneken, B., Hoyng, C.B., Theelen, T., Sánchez, C.I., 2016. Fast](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0053) [convolutional neural network training using selective data sampling: application](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0053) [to hemorrhage detection in color fundus images. IEEE Trans. Med. Imag. 35 (5),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0053) [1273–1284.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0053)

[Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0054) [Cai, J., et al., 2018. Recent advances in convolutional neural networks. Pattern](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0054) [Recognit. 77, 354–377.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0054)

|  |  |
| --- | --- |
| *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561* | 25 |

[Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venu-gopalan, S., Widner, K., Madams, T., Cuadros, J., et al., 2016. Development and](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0055) [validation of a deep learning algorithm for detection of diabetic retinopathy in](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0055) [retinal fundus photographs. JAMA 316 (22), 2402–2410.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0055)

[Harangi, B., Hajdu, A., 2014. Detection of exudates in fundus images using a marko-vian segmentation model. In: 2014 36th Annual International Conference of the](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0056) [IEEE Engineering in Medicine and Biology Society. IEEE, pp. 130–133.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0056)

[Hatanaka, Y., Nakagawa, T., Hayashi, Y., Kakogawa, M., Sawada, A., Kawase, K.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0057) [Hara, T., Fujita, H., 2008. Improvement of automatic hemorrhage detection](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0057) [methods using brightness correction on fundus images. In: Medical Imaging](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0057) [2008: Computer-Aided Diagnosis, Vol. 6915. International Society for Optics and](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0057) [Photonics, p. 69153E.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0057)

[Havlicek, J.P., 1996. Am-Fm Image Models. University of Texas at Austin.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0058)

[He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: Computer Vision](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0059) [(ICCV), 2017 IEEE International Conference on. IEEE, pp. 2980–2988.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0059)

[He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0060) [In: Proceedings of the IEEE conference on computer vision and pattern recogni-tion, pp. 770–778.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0060)

[Hinton, G., 2018. Deep learninga technology with the potential to transform health](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0061) [care. JAMA 320 (11), 1101–1102.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0061)

[Hoo-Chang, S., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0062) [Summers, R.M., 2016. Deep convolutional neural networks for computer-aided](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0062) [detection: CNN architectures, dataset characteristics and transfer learning. IEEE](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0062) [Trans. Med. Imag. 35 (5), 1285.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0062)

[Hoover, A., 1975. Stare Database. Available. Available: http://www.ces.clemson.edu/](http://www.ces.clemson.edu/ahoover/stare) [ahoover/stare](http://www.ces.clemson.edu/ahoover/stare)

[Howard, A.G., 2013. Some improvements on deep convolutional neural network](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0064) [based image classification. arXiv preprint arXiv:1312.5402.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0064)

[Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K., 2014.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0065) [Densenet: implementing eﬃcient convnet descriptor pyramids. arXiv preprint](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0065) [arXiv:1404.1869.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0065)

[ICO, 2017. Guidelines for diabetic eye care, 2nd edn. Int. Council Ophthalmol. (ICO).](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0066)

[Jaccard, P., 1908. Nouvelles recherches sur la distribution florale. Bull. Soc. Vaud. Sci.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0067)

[Nat. 44, 223–270.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0067)

[Javidi, M., Pourreza, H.-R., Harati, A., 2017. Vessel segmentation and microaneurysm](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0068) [detection using discriminative dictionary learning and sparse representation.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0068) [Comput. Method. Progr. Biomed. 139, 93–108.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0068)

[Jelinek, H., Cree, M.J., 2009. Automated Image Detection of Retinal Pathology. Crc](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0069) [Press.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0069)

[Jonas, R.A., Wang, Y.X., Yang, H., Li, J.J., Xu, L., Panda-Jonas, S., Jonas, J.B., 2015. Optic](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0070) [disc-fovea distance, axial length and parapapillary zones. the beijing eye study](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0070) [2011. PloS One 10 (9), e0138701.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0070)

[Jones, S., Edwards, R.T., 2010. Diabetic retinopathy screening: a systematic review of](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0071) [the economic evidence. Diabet. Med. 27 (3), 249–256.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0071)

[Jordan, K.C., Menolotto, M., Bolster, N.M., Livingstone, I.A.T., Giardini, M.E., 2017. A](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0072) [review of feature-based retinal image analysis. Expert Rev. Ophthalmol. 12 (3),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0072) [207–220.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0072)

[Joshi, S., Karule, P.T., 2019. Mathematical morphology for microaneurysm detection](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0073) [in fundus images. Eur. J. Ophthalmol.. 1120672119843021](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0073)

[Kamble, R., Kokare, M., Deshmukh, G., Hussin, F.A., Mériaudeau, F., 2017. Localization](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0074) [of optic disc and fovea in retinal images using intensity based line scanning](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0074) [analysis. Comput. Biol. Med. 87, 382–396.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0074)

[Kao, E.-F., Lin, P.-C., Chou, M.-C., Jaw, T.-S., Liu, G.-C., 2014. Automated detection of](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0075) [fovea in fundus images based on vessel-free zone and adaptive gaussian tem-plate. Comput. Method. Progr. Biomed. 117 (2), 92–103.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0075)

[Kauppi, T., Kamarainen, J.-K., Lensu, L., Kalesnykiene, V., Sorri, I., Uusitalo, H.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0076) [Kälviäinen, H., 2012. A framework for constructing benchmark databases and](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0076) [protocols for retinopathy in medical image analysis. In: International Confer-ence on Intelligent Science and Intelligent Data Engineering. Springer, pp. 832–](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0076) [843.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0076)

[Ker, J., Wang, L., Rao, J., Lim, T., 2018. Deep learning applications in medical image](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0077) [analysis. IEEE Access 6, 9375–9389.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0077)

[Khojasteh, P., Júnior, L.A.P., Carvalho, T., Rezende, E., Aliahmad, B., Papa, J.a.P., Ku-mar, D.K., 2018. Exudate detection in fundus images using deeply-learnable fea-tures. Comput. Biol. Med..](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0078)

[Kim, J., Hong, J., Park, H., Kim, J., Hong, J., Park, H., 2018. Prospects of deep learning](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0079) [for medical imaging. Precis. Future Med. 2 (2), 37–52.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0079)

[Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. arXiv](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0080) [preprint arXiv:1412.6980.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0080)

[Kollias, A.N., Ulbig, M.W., 2010. Diabetic retinopathy: early diagnosis and effective](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0081) [treatment. Deutsch. Arzteblatt Int. 107 (5), 75.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0081)

[Krause, J., Gulshan, V., Rahimy, E., Karth, P., Widner, K., Corrado, G.S., Peng, L., Web-ster, D.R., 2018. Grader variability and the importance of reference standards](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0082) [for evaluating machine learning models for diabetic retinopathy. Ophthalmol-ogy 125 (8), 1264–1272.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0082)

[Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0083) [convolutional neural networks. In: Advances in neural information processing](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0083) [systems, pp. 1097–1105.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0083)

[Lam, C., Yu, C., Huang, L., Rubin, D., 2018. Retinal lesion detection with deep learning](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0084) [using image patches. Invest. Ophthalmol. Vis. Sci. 59 (1), 590–596.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0084)

[Li, H., Chutatape, O., 2004. Automated feature extraction in color retinal images by](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0085) [a model based approach. IEEE Trans. Biomed. Eng. 51 (2), 246–254.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0085)

[Lin, S., Ramulu, P., Lamoureux, E.L., Sabanayagam, C., 2016. Addressing risk factors,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0086) [screening, and preventative treatment for diabetic retinopathy in developing](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0086) [countries: a review. Clin. Exper. Ophthalmol. 44 (4), 300–320.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0086)

[Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0087) [Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0087) [medical image analysis. Med. Image Anal. 42, 60–88.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0087)

[Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0088) [segmentation. In: Proceedings of the IEEE conference on computer vision and](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0088) [pattern recognition, pp. 3431–3440.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0088)

[Lynch, S.K., Shah, A., Folk, J.C., Wu, X., Abramoff, M.D., 2017. Catastrophic failure](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0089) [in image-based convolutional neural network algorithms for detecting diabetic](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0089) [retinopathy. Investigat. Ophthalmol. Vis. Sci. 58 (8). 3776–3776](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0089)

[Marin, D., Gegundez-Arias, M.E., Ponte, B., Alvarez, F., Garrido, J., Ortega, C.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0090) [Vasallo, M.J., Bravo, J.M., 2018. An exudate detection method for diagnosis risk](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0090) [of diabetic macular edema in retinal images using feature-based and supervised](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0090) [classification. Med. Biol. Eng. Comput. 56 (8), 1379–1390.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0090)

[Marin, D., Gegundez-Arias, M.E., Suero, A., Bravo, J.M., 2015. Obtaining optic disc](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0091) [center and pixel region by automatic thresholding methods on morphologically](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0091) [processed fundus images. Comput. Method. Progr. Biomed. 118 (2), 173–185.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0091)

[Mary, M.C.V.S., Rajsingh, E.B., Jacob, J.K.K., Anandhi, D., Amato, U., Selvan, S.E., 2015.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0092)

[An empirical study on optic disc segmentation using an active contour model.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0092)

[Biomed. Signal Process. Control 18, 19–29.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0092)

[Medhi, J.P., Dandapat, S., 2014. Analysis of maculopathy in color fundus images. In:](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0093)

[2014 Annual IEEE India Conference (INDICON). IEEE, pp. 1–4.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0093)

[Mendonça, A.M., Sousa, A., Mendonça, L., Campilho, A., 2013. Automatic localiza-tion of the optic disc by combining vascular and intensity information. Comput.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0094) [Med. Imag. Graph. 37 (5–6), 409–417.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0094)

[Mookiah, M.R.K., Acharya, U.R., Chandran, V., Martis, R.J., Tan, J.H., Koh, J.E.W.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0095) [Chua, C.K., Tong, L., Laude, A., 2015. Application of higher-order spectra for](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0095) [automated grading of diabetic maculopathy. Med. Biol. Eng. Comput. 53 (12),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0095) [1319–1331.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0095)

[Mookiah, M.R.K., Acharya, U.R., Chua, C.K., Lim, C.M., Ng, E., Laude, A., 2013. Com-puter-aided diagnosis of diabetic retinopathy: a review. Comput. Biol. Med. 43](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0096) [(12), 2136–2155.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0096)

[Mookiah, M.R.K., Acharya, U.R., Martis, R.J., Chua, C.K., Lim, C.M., Ng, E., Laude, A.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0097) [2013. Evolutionary algorithm based classifier parameter tuning for automatic di-abetic retinopathy grading: a hybrid feature extraction approach. Knowl.-Based](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0097) [Syst. 39, 9–22.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0097)

[Morales, S., Engan, K., Naranjo, V., Colomer, A., 2017. Retinal disease screening](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0098) [through local binary patterns. IEEE J. Biomed. Health Informat. 21 (1), 184–192.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0098)

[Morales, S., Naranjo, V., Angulo, J., Alcañiz, M., 2013. Automatic detection of optic](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0099) [disc based on PCA and mathematical morphology. IEEE Trans. Med. Imag. 32](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0099)

[(4), 786–796.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0099)

[Murphy, K.P., 2012. Machine learning: a probabilistic perspective. MIT press.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0100)

[Nagy, B., Harangi, B., Antal, B., Hajdu, A., 2011. Ensemble-based exudate detection in](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0101) [color fundus images. In: Image and Signal Processing and Analysis (ISPA), 2011](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0101) [7th International Symposium on. IEEE, pp. 700–703.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0101)

[Naqvi, S.A.G., Zafar, H.M.F., et al., 2018. Automated system for referral of cotton-wool](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0102) [spots. Current Diabetes Rev. 14 (2), 168–174.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0102)

[Niemeijer, M., Abramoff, M.D., Van Ginneken, B., 2009. Information fusion for di-abetic retinopathy CAD in digital color fundus photographs. IEEE Trans. Med.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0103) [Imag. 28 (5), 775–785.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0103)

[Niemeijer, M., Van Ginneken, B., Cree, M.J., Mizutani, A., Quellec, G., Sánchez, C.I.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0104) [Zhang, B., Hornero, R., Lamard, M., Muramatsu, C., et al., 2010. Retinopathy on-line challenge: automatic detection of microaneurysms in digital color fundus](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0104) [photographs. IEEE Trans. Med. Imag. 29 (1), 185–195.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0104)

[Niemeijer, M., Van Ginneken, B., Staal, J., Suttorp-Schulten, M.S.A., Abràmoff, M.D.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0105) [2005. Automatic detection of red lesions in digital color fundus photographs.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0105) [IEEE Trans. Med. Imag. 24 (5), 584–592.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0105)

[Nørgaard, M.F., Grauslund, J., 2018. Automated screening for diabetic retinopathy–a](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0106) [systematic review. Ophthalmic Res..](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0106)

[Orlando, J.I., Prokofyeva, E., del Fresno, M., Blaschko, M.B., 2018. An ensemble deep](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0107) [learning based approach for red lesion detection in fundus images. Comput.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0107) [Method. Progr. Biomed. 153, 115–127.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0107)

[Osareh, A., Shadgar, B., Markham, R., 2009. A computational-intelligence-based ap-proach for detection of exudates in diabetic retinopathy images. IEEE Trans. Inf.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0108) [Technol. Biomed. 13 (4), 535–545.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0108)

[Patton, N., Aslam, T.M., MacGillivray, T., Deary, I.J., Dhillon, B., Eikelboom, R.H., Yoge-san, K., Constable, I.J., 2006. Retinal image analysis: concepts, applications and](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0109) [potential. Progress Retinal Eye Res. 25 (1), 99–127.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0109)

Perdomo, O., Otalora, S., Rodríguez, F., Arevalo, J., González, F. A., 2016. A novel ma-chine learning model based on exudate localization to detect diabetic macular edema.

[Pereira, C., Gonçalves, L., Ferreira, M., 2015. Exudate segmentation in fundus images](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0110) [using an ant colony optimization approach. Inf. Sci. 296, 14–24.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0110)

[Pereira, C.S., Mendonça, A.M., Campilho, A., 2007. Evaluation of contrast enhance-ment filters for lung nodule detection. In: International Conference Image Anal-ysis and Recognition. Springer, pp. 878–888.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0111)

[Pires, R., Avila, S., Jelinek, H.F., Wainer, J., Valle, E., Rocha, A., 2017. Beyond le-sion-based diabetic retinopathy: a direct approach for referral. IEEE J. Biomed.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0112) [Health Informat. 21 (1), 193–200.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0112)

Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Meriaudeau, F., 2018. Indian diabetic retinopathy image dataset (IDRid). IEEE Dataport doi:[10.21227/H25W98.](https://doi.org/10.21227/H25W98)

Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Meriaudeau, F., 2018. Indian diabetic retinopathy image dataset (IDRid): a [database for diabetic retinopathy screening research. Data 3 (3/25). doi:10.3390/](https://doi.org/10.3390/data3030025) [data3030025.](https://doi.org/10.3390/data3030025)

26 *P. Porwal, S. Pachade and M. Kokare et al. / Medical Image Analysis 59 (2019) 101561*

[Porwal, P., Pachade, S., Kokare, M., Giancardo, L., Mériaudeau, F., 2018. Retinal image](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0115) [analysis for disease screening through local tetra patterns. Comput. Biol. Med.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0115) [102, 200–210.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0115)

[Quellec, G., Charrière, K., Boudi, Y., Cochener, B., Lamard, M., 2017. Deep image min-ing for diabetic retinopathy screening. Med. Image Anal. 39, 178–193.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0116)

Quellec, [G., Lamard, M., Erginay, A., Chabouis, A., Massin, P., Cochener, B.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0117) [Cazuguel, G., 2016. Automatic detection of referral patients due to retinal](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0117) [pathologies through data mining. Med. Image Anal. 29, 47–64.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0117)

[Quellec, G., Lamard, M., Josselin, P.M., Cazuguel, G., Cochener, B., Roux, C., 2008.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0118) [Optimal wavelet transform for the detection of microaneurysms in retina pho-tographs. IEEE Trans. Med. Imag. 27 (9), 1230–1241.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0118)

[Raman, R., Gella, L., Srinivasan, S., Sharma, T., 2016. Diabetic retinopathy: an epi-demic at home and around the world. Indian J. Ophthalmol. 64 (1), 69.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0119)

[Rangrej, S.B., Sivaswamy, J., 2017. Assistive lesion-emphasis system: an assistive sys-tem for fundus image readers. J. Med. Imag. 4 (2), 024503.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0120)

[Ravı, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., Yang, G.-Z.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0121) [2017. Deep learning for health informatics. IEEE J. Biomed. Health Informat. 21](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0121)

[(1), 4–21.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0121)

[Reichel, E., Salz, D., 2015. Diabetic retinopathy screening. In: Managing Diabetic Eye](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0122) [Disease in Clinical Practice. Springer, pp. 25–38.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0122)

[Rocha, A., Carvalho, T., Jelinek, H.F., Goldenstein, S., Wainer, J., 2012. Points of inter-est and visual dictionaries for automatic retinal lesion detection. IEEE Transac-tions on biomedical engineering 59 (8), 2244–2253.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0123)

[Romero-Oraá, R., Jiménez-García, J., García, M., López-Gálvez, M.I., Oraá-Pérez, J.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0124) [Hornero, R., 2019. Entropy rate superpixel classification for automatic red lesion](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0124) [detection in fundus images. Entropy 21 (4), 417.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0124)

[Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0125) [biomedical image segmentation. In: International Conference on Medical image](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0125) [computing and computer-assisted intervention. Springer, pp. 234–241.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0125)

[Roychowdhury, S., Koozekanani, D.D., Parhi, K.K., 2014. Dream: diabetic retinopa-thy analysis using machine learning. IEEE J. Biomed. Health Informat. 18 (5),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0126) [1717–1728.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0126)

[Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0127) [the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0127) [One 10 (3), e0118432.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0127)

[Sánchez, C.I., García, M., Mayo, A., López, M.I., Hornero, R., 2009. Retinal image anal-ysis based on mixture models to detect hard exudates. Med. Image Anal. 13 (4),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0128) [650–658.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0128)

[Sánchez, C.I., Niemeijer, M., Išgum, I., Dumitrescu, A., Suttorp-Schulten, M.S.A., Abrà-moff, M.D., van Ginneken, B., 2012. Contextual computer-aided detection: im-proving bright lesion detection in retinal images and coronary calcification](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0129) [identification in CT scans. Med. Image Anal. 16 (1), 50–62.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0129)

[Seoud, L., Hurtut, T., Chelbi, J., Cheriet, F., Langlois, J.M.P., 2016. Red lesion detection](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0130) [using dynamic shape features for diabetic retinopathy screening. IEEE Trans.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0130) [Med. Imag. 35 (4), 1116–1126.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0130)

[Shah, M.P., Merchant, S.N., Awate, S.P., 2018. Abnormality detection using deep neu-ral networks with robust quasi-norm autoencoding and semi-supervised learn-ing. In: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0131) [on. IEEE, pp. 568–572.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0131)

[Shen, D., Wu, G., Suk, H.-I., 2017. Deep learning in medical image analysis. Annu.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0132)

[Rev. Eng. 19, 221–248.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0132)

[Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0133) [2016. Real-time single image and video super-resolution using an eﬃcient sub--pixel convolutional neural network. In: Proceedings of the IEEE Conference on](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0133) [Computer Vision and Pattern Recognition, pp. 1874–1883.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0133)

[Shortliffe, E.H., Blois, M.S., 2006. The computer meets medicine and biology: emer-](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0134)

[gence of a discipline. In: Biomedical Informatics. Springer, pp. 3–45.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0134)

[Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0135) [image recognition. arXiv preprint arXiv:1409.1556.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0135)

[Sivaswamy, J., Krishnadas, S.R., Joshi, G.D., Jain, M., Tabish, A.U.S., 2014. Drishti-gs:](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0136) [Retinal image dataset for optic nerve head (onh) segmentation. In: 2014 IEEE](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0136) [11th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 53–56.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0136)

[Son, J., Bae, W., Kim, S., Park, S.J., Jung, K.-H., 2018. Classification of findings with](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0137) [localized lesions in fundoscopic images using a regionally guided CNN. In:](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0137) [Computational Pathology and Ophthalmic Medical Image Analysis. Springer,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0137) [pp. 176–184.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0137)

[Son, J., Park, S.J., Jung, K.-H., 2017. Retinal vessel segmentation in fundoscopic im-ages with generative adversarial networks. arXiv preprint arXiv:1706.09318.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0138)

[Son, J., Shin, J.Y., Kim, H.D., Jung, K.-H., Park, K.H., Park, S.J., 2019. Development and](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0139) [validation of deep learning models for screening multiple abnormal findings in](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0139) [retinal fundus images. Ophthalmology.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0139)

[Sopharak, A., Uyyanonvara, B., Barman, S., Williamson, T.H., 2008. Automatic de-tection of diabetic retinopathy exudates from non-dilated retinal images us-ing mathematical morphology methods. Comput. Med. Imag. Graph. 32 (8),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0140) [720–727.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0140)

[Sreng, S., Maneerat, N., Hamamoto, K., Panjaphongse, R., 2019. Cotton wool spots](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0141) [detection in diabetic retinopathy based on adaptive thresholding and ant colony](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0141) [optimization coupling support vector machine. IEEJ Trans. Electr. Electron. Eng..](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0141)

[Srivastava, R., Duan, L., Wong, D.W.K., Liu, J., Wong, T.Y., 2017. Detecting retinal mi-croaneurysms and hemorrhages with robustness to the presence of blood ves-sels. Comput. Method. Progr. Biomed. 138, 83–91.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0142)

[Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B., 2004.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0143)

[Ridge based vessel segmentation in color images of the retina. IEEE Trans. Med.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0143)

[Imag. 23 (4), 501–509.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0143)

[Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J., 2017. Generalised dice](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0144) [overlap as a deep learning loss function for highly unbalanced segmentations.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0144) [In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clin-ical Decision Support. Springer, pp. 240–248.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0144)

[Suzuki, K., 2017. Overview of deep learning in medical imaging. Radiol. Phys. Tech-nol. 10 (3), 257–273.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0145)

[Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Van-houcke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceed-ings of the IEEE Conference on Computer vVsion and Pattern Recognition,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0146) [pp. 1–9.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0146)

[Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0147) [2016. Convolutional neural networks for medical image analysis: full training or](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0147) [fine tuning? IEEE Trans. Med. Imag. 35 (5), 1299–1312.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0147)

[Tan, J.H., Acharya, U.R., Bhandary, S.V., Chua, K.C., Sivaprasad, S., 2017. Segmentation](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0148) [of optic disc, fovea and retinal vasculature using a single convolutional neural](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0148) [network. J. Comput. Sci. 20, 70–79.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0148)

[Tan, J.H., Fujita, H., Sivaprasad, S., Bhandary, S.V., Rao, A.K., Chua, K.C., Acharya, U.R.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0149) [2017. Automated segmentation of exudates, haemorrhages, microaneurysms us-ing single convolutional neural network. Inf. Sci. 420, 66–76.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0149)

[Tang, L., Niemeijer, M., Reinhardt, J.M., Garvin, M.K., Abramoff, M.D., 2013. Splat fea-ture classification with application to retinal hemorrhage detection in fundus](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0150) [images. IEEE Trans. Med. Imag. 32 (2), 364–375.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0150)

[Thakur, N., Juneja, M., 2017. Clustering based approach for segmentation of optic](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0151) [cup and optic disc for detection of glaucoma. Current Med. Imag. Rev. 13 (1),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0151) [99–105.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0151)

[Ting, D.S.W., Cheung, G.C.M., Wong, T.Y., 2016. Diabetic retinopathy: global preva-lence, major risk factors, screening practices and public health challenges: a re-view. Clin. Exper. Ophthalmol. 44 (4), 260–277.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0152)

[Trucco, E., Ruggeri, A., Karnowski, T., Giancardo, L., Chaum, E., Hubschman, J.P.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0153) [Al-Diri, B., Cheung, C.Y., Wong, D., Abramoff, M., et al., 2013. Validating retinal](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0153) [fundus image analysis algorithms: issues and a proposal. Invest. Ophthalmol.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0153) [Visual Sci. 54 (5), 3546–3559.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0153)

[Uribe-Valencia, L.J., Martínez-Carballido, J.F., 2019. Automated optic disc region lo-cation from fundus images: using local multi-level thresholding, best channel](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0154) [selection, and an intensity profile model. Biomed. Signal Processing and Control](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0154) [51, 148–161.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0154)

[Voulodimos, A., Doulamis, N., Bebis, G., Stathaki, T., 2018. Recent Developments in](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0155) [deep learning for engineering applications. Comput. Intell. Neurosci. 2018.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0155)

[Walter, T., Klein, J.-C., Massin, P., Erginay, A., 2002. A contribution of image pro-cessing to the diagnosis of diabetic retinopathy-detection of exudates in color](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0156) [fundus images of the human retina. IEEE Trans. Med. Imag. 21 (10), 1236–1243.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0156)

[Welfer, D., Scharcanski, J., Marinho, D.R., 2011. Fovea center detection based on the](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0157) [retina anatomy and mathematical morphology. Comput. Method. Progr. Biomed.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0157) [104 (3), 397–409.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0157)

[Winder, R.J., Morrow, P.J., McRitchie, I.N., Bailie, J.R., Hart, P.M., 2009. Algorithms for](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0158) [digital image processing in diabetic retinopathy. Comput. Med. Imag. Graph. 33](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0158)

[(8), 608–622.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0158)

[Wong, T.Y., Cheung, C.M.G., Larsen, M., Sharma, S., Simó, R., 2016. Diabetic retinopa-thy. Nature Rev. Disease Primers.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0159)

[Wu, B., Zhu, W., Shi, F., Zhu, S., Chen, X., 2017. Automatic detection of microa-neurysms in retinal fundus images. Comput. Med. Imag. Graph. 55, 106–112.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0160)

[Wu, L., Fernandez-Loaiza, P., Sauma, J., Hernandez-Bogantes, E., Masis, M., 2013.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0161) [Classification of diabetic retinopathy and diabetic macular edema. World J. Dia-betes 4 (6), 290.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0161)

[Wu, X., Dai, B., Bu, W., 2016. Optic disc localization using directional models. IEEE](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0162) [Trans. Image Process. 25 (9), 4433–4442.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0162)

[Yu, F., Wang, D., Shelhamer, E., Darrell, T., 2017. Deep layer aggregation. arXiv](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0163) [preprint arXiv:1707.06484.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0163)

[Yu, H., Barriga, E.S., Agurto, C., Echegaray, S., Pattichis, M.S., Bauman, W., Soliz, P.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0164) [2012. Fast localization and segmentation of optic disk in retinal images using](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0164) [directional matched filtering and level sets. IEEE Trans. Inf. Technol. Biomed. 16](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0164)

[(4), 644–657.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0164)

[Yun, W.L., Acharya, U.R., Venkatesh, Y.V., Chee, C., Min, L.C., Ng, E.Y.K., 2008. Identi-fication of different stages of diabetic retinopathy using retinal optical images.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0165) [Inf. Sci. 178 (1), 106–121.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0165)

[Zhang, B., Karray, F., Li, Q., Zhang, L., 2012. Sparse representation classifier for mi-croaneurysm detection and retinal blood vessel extraction. Inf. Sci. 200, 78–90.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0166)

[Zhang, X., Thibault, G., Decencière, E., Marcotegui, B., Laÿ, B., Danno, R., Cazuguel, G.,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0167) [Quellec, G., Lamard, M., Massin, P., et al., 2014. Exudate detection in color reti-nal images for mass screening of diabetic retinopathy. Med. Image Anal. 18 (7),](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0167) [1026–1043.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0167)

[Zhou, W., Wu, C., Yi, Y., Du, W., 2017. Automatic detection of exudates in digital](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0168) [color fundus images using superpixel multi-feature classification. IEEE Access 5,](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0168) [17077–17088.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0168)

[Zilly, J., Buhmann, J.M., Mahapatra, D., 2017. Glaucoma detection using entropy sam-pling and ensemble learning for automatic optic cup and disc segmentation.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0169) [Comput. Med. Imag. Graph. 55, 28–41.](http://refhub.elsevier.com/S1361-8415(19)30103-3/sbref0169)