



IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge

Prasanna Porwal^{a,b,1,*}, Samiksha Pachade^{a,b,1}, Manesh Kokare^{a,1}, Girish Deshmukh^{c,1}, Jaemin Son^d, Woong Bae^d, Lihong Liu^e, Jianzong Wang^e, Xinhui Liu^e, Liangxin Gao^e, TianBo Wu^e, Jing Xiao^e, Fengyan Wang^f, Baocai Yin^f, Yunzhi Wang^g, Gopichandh Danala^g, Linsheng He^g, Yoon Ho Choi^h, Yeong Chan Lee^h, Sang-Hyuk Jung^h, Zhongyu Liⁱ, Xiaodan Sui^j, Junyan Wu^l, Xiaolong Li^m, Ting Zhouⁿ, Janos Toth^o, Agnes Baran^o, Avinash Kori^p, Sai Saketh Chennamsetty^p, Mohammed Safwan^p, Varghese Alex^p, Xingzheng Lyu^{q,r}, Li Cheng^{r,d}, Qin hao Chu^s, Pengcheng Li^s, Xin Ji^t, Sanyuan Zhang^q, Yaxin Shen^{u,v}, Ling Dai^{u,v}, Oindrila Saha^x, Rachana Sathish^x, Tânia Melo^y, Teresa Araújo^{y,z}, Balazs Harangi^o, Bin Sheng^{u,v}, Ruogu Fang^w, Debdoot Sheet^x, Andras Hajdu^o, Yuanjie Zheng^j, Ana Maria Mendonça^{y,z}, Shaoting Zhangⁱ, Aurélio Campilho^{y,z}, Bin Zheng^g, Dinggang Shen^k, Luca Giancardo^{b,1}, Gwennolé Quéllec^{A,1}, Fabrice Mériaudeau^{B,C,1}

^a Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India

^b School of Biomedical Informatics, University of Texas Health Science Center at Houston, USA

^c Eye Clinic, Sushrusha Hospital, Nanded, Maharashtra, India

^d VUNO Inc., Seoul, Republic of Korea

^e Ping An Technology (Shenzhen) Co., Ltd, China

^f iFLYTEK Research, Hefei, China

^g School of Electrical and Computer Engineering, University of Oklahoma, USA

^h Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Seoul, Republic of Korea

ⁱ Department of Computer Science, University of North Carolina at Charlotte, USA

^j School of Information Science and Engineering, Shandong Normal University, China

^k Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA

^l Cleerly Inc., New York, United States

^m Virginia Tech, Virginia, United States

ⁿ University at Buffalo, New York, United States

^o University of Debrecen, Faculty of Informatics 4002 Debrecen, POB 400, Hungary

^p Individual Researcher, India

^q College of Computer Science and Technology, Zhejiang University, Hangzhou, China

^r Machine Learning for Bioimage Analysis Group, Bioinformatics Institute, A*STAR, Singapore

^s School of Computing, National University of Singapore, Singapore

^t Beijing Shangong Medical Technology Co., Ltd., China

^u Department of Computer Science and Engineering, Shanghai Jiao Tong University, China

^v MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, China

^w J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, USA

^x Indian Institute of Technology Kharagpur, India

^y INESC TEC - Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal

^z FEUP - Faculty of Engineering of the University of Porto, Porto, Portugal

^A INSERM, UMR 1101, Brest, France

^B Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Malaysia

^C ImViA/IFTIM, Université de Bourgogne, Dijon, France

^D Department of Electric and Computer Engineering, University of Alberta, Canada

* Corresponding author at Center of Excellence in Signal and Image Processing, Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded (M.S.), India.

E-mail address: porwal.prasanna@sngs.ac.in (P. Porwal).

¹ These authors co-organized the challenge. All others contributed results of their algorithm(s) presented in the paper

ARTICLE INFO

Article history:

Received 11 January 2019

Revised 9 September 2019

Accepted 16 September 2019

Available online 3 October 2019

Keywords:

Diabetic Retinopathy

Retinal image analysis

Deep learning

Challenge

ABSTRACT

Diabetic Retinopathy (DR) is the most common cause of avoidable vision loss, predominantly affecting the working-age population across the globe. Screening for DR, coupled with timely consultation and treatment, is a globally trusted policy to avoid vision loss. However, implementation of DR screening programs is challenging due to the scarcity of medical professionals able to screen a growing global diabetic population at risk for DR. Computer-aided disease diagnosis in retinal image analysis could provide a sustainable approach for such large-scale screening effort. The recent scientific advances in computing capacity and machine learning approaches provide an avenue for biomedical scientists to reach this goal. Aiming to advance the state-of-the-art in automatic DR diagnosis, a grand challenge on “Diabetic Retinopathy – Segmentation and Grading” was organized in conjunction with the IEEE International Symposium on Biomedical Imaging (ISBI - 2018). In this paper, we report the set-up and results of this challenge that is primarily based on Indian Diabetic Retinopathy Image Dataset (IDRiD). There were three principal sub-challenges: lesion segmentation, disease severity grading, and localization of retinal landmarks and segmentation. These multiple tasks in this challenge allow to test the generalizability of algorithms, and this is what makes it different from existing ones. It received a positive response from the scientific community with 148 submissions from 495 registrations effectively entered in this challenge. This paper outlines the challenge, its organization, the dataset used, evaluation methods and results of top-performing participating solutions. The top-performing approaches utilized a blend of clinical information, data augmentation, and an ensemble of models. These findings have the potential to enable new developments in retinal image analysis and image-based DR screening in particular.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Diabetic Retinopathy (DR) and Diabetic Macular Edema (DME) are the most common sight-threatening medical conditions caused due to retinal microvascular changes triggered by diabetes (Reichel and Salz, 2015), predominantly affecting the working-age population in the world (Atlas, 2017). DR leads to gradual changes in vasculature structure (including vascular tortuosity, branching angles and calibers) and resulting abnormalities (microaneurysms, hemorrhages and exudates), whereas, DME is characterized by retention of fluid or swelling of macula that may occur at any stage of DR (Bandello et al., 2010; Ciulla et al., 2003). According to International Diabetes Federation (Atlas, 2017) estimates, presently, the global number of individuals affected with diabetes is 425 million, and it may rise to 693 million by 2045. Amongst them, one out of three individuals is estimated to have some form of DR, and one in ten is prone to vision-threatening DR (ICO, 2017; Bourne et al., 2013). DR is diagnosed by visually inspecting retinal fundus images for the presence of one or more retinal lesions like microaneurysms (MAs), hemorrhages (HEs), soft

exudates (SEs) and hard exudates (EXs) (Wong et al., 2016) as shown in Fig. 1.

Early diagnosis and treatment of DR can prevent vision loss. Hence, diabetic patients are typically referred for retinal screening once or twice a year (Ferris, 1993; Kollias and Ulbig, 2010; Ting et al., 2016). The diabetic eye care is mainly reliant on the number of ophthalmologists and necessary health care infrastructure (Jones and Edwards, 2010; Lin et al., 2016). In India, ophthalmologist to population ratio is 1:107,000, however, in urban regions this ratio is 1:9000 whereas in rural parts there is only one ophthalmologist for 608,000 inhabitants (Raman et al., 2016). By 2045, India alone is projected to have approximately 151 million people with diabetes and one-third of them are expected to have DR (Atlas, 2017). Programs to screen such a large population for DR confront issues related to implementation, management, availability of human graders, and long-term financial sustainability. Hence, computer-aided diagnosis tools are required for screening such a large population that requires continuous follow-up for DR and to effectively facilitate in reducing the burden on ophthalmologists (Jelinek and Cree, 2009; Walter et al., 2002). Such a tool would help clinicians in identification,

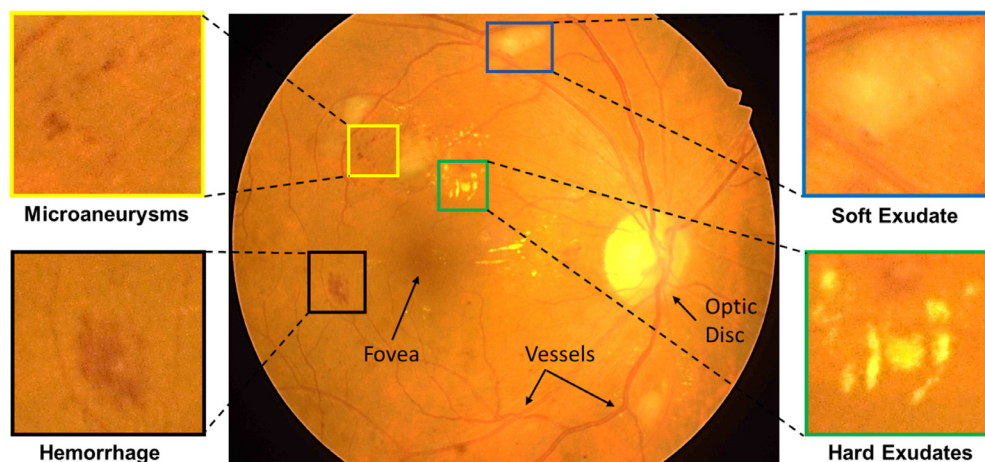


Fig. 1. Illustration of retinal image (in center) by highlighting normal structures (blood vessels, optic disc and fovea center) and abnormalities associated with DR: Enlarged regions (in left) MAs, and HEs and (in right) SEs, and EXs.

interpretation, and measurements of retinal abnormalities, and ultimately in screening and monitoring of the disease. Recent scientific advances in computing capacity and machine learning approaches provide an avenue to biomedical scientists to meet desideratum of clinical practice (Shortliffe and Blois, 2006; Patton et al., 2006). To meet this need, raw images along with the precise pixel or image-level expert annotations (*a.k.a.* ground truths) play an important role to facilitate the research community for the development, validation, and comparison of DR lesion segmentation techniques (Trucco et al., 2013). Precise pixel-level annotations of lesions associated with DR such as MAs, HEs, SEs and EXs are invaluable resources for evaluating the accuracy of individual lesion segmentation techniques. These precisely segmented lesions help in determining disease severity and further act as a roadmap that can assist to tap progression of disease during follow-up procedures. Similarly, on the other hand, image-level expert labels for disease severity of DR and DME are helpful in the development and evaluation of image analysis and retrieval algorithms. This necessity has led several research groups to develop and share retinal image datasets, namely Messidor (Decencière et al., 2014), Kaggle (Cuadros and Bresnick, 2009), ROC (Niemeijer et al., 2010), E-Ophtha (Decencière et al., 2013), DiaretDB (Kauppi et al., 2012), DRIVE (van Ginneken et al., 2004), STARE (Hoover, 1975), ARIA (Farnell et al., 2008) and HEI-MED (Giancardo et al., 2012).

Further, two challenges were organized in the context of DR, namely Retinopathy Online Challenge (ROC)² and Kaggle DR detection challenge³. ROC was organized with the goal of detecting MAs. Whereas, Kaggle challenge aimed to get solution for determining the severity level of DR. These challenges enabled advances in the field by promoting the participation of scientific research community from all over the globe on a competitive at the same time constructive setting for scientific advancement. Previous efforts have made good progress using image classification, pattern recognition, and machine learning. The progress through the last two decades has been systematically reviewed by several research groups (Patton et al., 2006; Winder et al., 2009; Abràmoff et al., 2010; Mookiah et al., 2013a; Jordan et al., 2017; Nørgaard and Grauslund, 2018).

Although lots of efforts have been made in the field towards automating DR screening process, lesion detection is still a challenging task due to the following aspects: (a) Complex structures of lesions (shape, size, intensity), (b) detection of lesions in tessellated images and in presence of noise (bright border reflections, impulsive noise, optical reflections), (c) high inter-class similarity (*i.e.* between MA-HE and EX-SE), and (d) appearance of not so uncommon non-lesion structures (nerve fiber reflections, vessel reflections, drusen) makes it difficult to build a flexible and robust model for lesion segmentation. To the best of our knowledge, prior to this challenge, there were no reports on the development of a single framework to segment all lesions (MA, HE, SE, and EX) simultaneously. Also, there was a lack of common platform to test the robustness of approaches that determine normal and abnormal retinal structures on the same set of images. Furthermore, there was limited availability of pixel-level annotations and simultaneous gradings for DR and DME (see Tables in Appendix A).

In order to address these issues, we introduced a new dataset called Indian Diabetic Retinopathy Image Dataset (IDRiD) (Porwal et al., 2018a). Further, it was used as a base dataset for the organization of grand challenge on “Diabetic Retinopathy – Segmentation and Grading” in conjunction with ISBI - 2018. The IDRiD dataset provides expert markups of typical DR lesions and normal retinal structures. It also provides disease

severity level of DR and DME for each image in the database. This challenge brought together computer vision and biomedical researchers with an ultimate aim to further stimulate and promote research, as well as to provide a unique platform for the development of a practical software tool that will support efficient and accurate measurement and analysis of retinal images that could be useful in DR management. Initially, a training dataset along with the ground truth was provided to participants for the development of their algorithms. Later, the results were judged on the performance of these algorithms on the test dataset. Success was measured by how closely the algorithmic outcome matched the ground truth. There were three principal sub-challenges: lesion segmentation, disease severity grading, and localization and segmentation of retinal landmarks. These multiple tasks in IDRiD challenge allow to test the generalizability of the algorithms, and this is what makes it different from the existing ones. Further, this challenge seeks an automated solution to predict the severity of DR and DME simultaneously. It was projected as an individual task to increase the difficulty level of this challenge as compared to the Kaggle DR challenge *i.e.* for a given image, the predicted severity for both DR and DME should be correct to count for scoring the task.

The rest of the paper is structured as follows: Section 2 gives a short review of previous work done in the development of automated DR screening, Section 3 provides details of reference dataset, Section 4 describes the organization of competition through various phases and Section 5 details the top-performing competing solutions. Section 6 presents performance evaluation measures used in this challenge. Then, Section 7 presents the results, analysis and corresponding ranking of participating teams for all sub-challenges. Section 8 provides a brief discussion on results, limitations, and lessons learnt from this challenge and at last conclusion. Along with this paper, Appendix A is included that provides a comparison of different state-of-the-art publicly available databases with the IDRiD dataset.

2. Review of retinal image analysis for the detection of DR

Automatic image processing has proven to be a promising choice for analysis of retinal fundus images and its application to future eye care. The introduction of automated techniques in DR screening programs and interesting outcomes achieved by rapidly growing deep learning technology are examples of success stories and potential future achievements. Particularly, after the researcher's (Krizhevsky et al., 2012) deep learning based model showed significant improvements over the state-of-the-art in the ImageNet challenge, there was a surge of deep learning based models in medical image analysis. Hence, we decided to present the most recent relevant works with a classification based on whether or not they used deep learning in the context of DR.

2.1. Non-deep learning methods

A general framework for retinal image analysis through traditional handcrafted features based approaches involve several stages, typically: a preprocessing stage for contrast enhancement or non-uniformity equalization, image segmentation, feature extraction, and classification. Feature extraction strategy varies according to the objective involved, *i.e.* retinal lesion detection, disease screening or landmark localization. In 2006, one research group (Patton et al., 2006) outlined principles upon which retinal image analysis is based and discussed initial techniques used to detect retinal landmarks and lesions associated with DR. Later, Winder et al. (2009) reported an analysis of work in automated analysis of DR during 1998 - 2008. They categorized the literature into a series of operations or steps as preprocessing,

² <http://webeye.ophth.uiowa.edu/ROC/>

³ <https://www.kaggle.com/c/diabetic-retinopathy-detection>

vasculature segmentation, localization, and segmentation of the optic disk (OD), localization of the macula and fovea, detection and segmentation of lesions. Some of the review articles (Abramoff et al., 2010; Jordan et al., 2017) provide a brief introduction to quantitative methods for the analysis of fundus images with a focus on identification of retinal lesions and automated techniques for large scale screening for retinal diseases.

Majority of attempts in the literature are directed towards exclusive detection and/or segmentation of one type of lesions (either MAs, HEs, EXs or SEs) from an image. Some of the common approaches involved for lesion segmentation are mathematical morphology (Joshi and Karule, 2019; Hatanaka et al., 2008; Zhang et al., 2014), region growing (Fleming et al., 2006; Li and Chutatape, 2004), and supervised methods (Wu et al., 2017; Zhou et al., 2017; Garcia et al., 2009; Tang et al., 2013). Apart from these approaches, in case of MAs, most initial studies have shown effectiveness of template matching (Quellec et al., 2008), entropy thresholding (Das et al., 2015), radon space (Giancardo et al., 2011), sparse representation (Zhang et al., 2012; Javidi et al., 2017), Hessian based region descriptors (Adal et al., 2014) and dictionary learning (Rocha et al., 2012). On the other hand, for exclusive segmentation of HEs, super-pixel based features (Tang et al., 2013; Romero-Oraá et al., 2019) were found to be effective. These red lesions (both MAs and HEs) are also frequently detected together using dynamic shape features (Seoud et al., 2016), filter response and multiple kernel learning (Srivastava et al., 2017) and hybrid feature extraction approach (Niemeijer et al., 2005). Similarly, for EXs, researchers relied on approaches like clustering (Osareh et al., 2009), model-based (Sánchez et al., 2009; Harangi and Hajdu, 2014), ant colony optimization (ACO) (Pereira et al., 2015) and contextual information (Sánchez et al., 2012). Whereas for SEs researchers utilized Scale Invariant Feature Transform (SIFT) (Naqvi et al., 2018), adaptive thresholding and ACO (Sreng et al., 2019). Further, several approaches were devised for multiple lesion detection such as multiscale amplitude-modulation-frequency-modulation (Agurto et al., 2010), machine learning (Roychowdhury et al., 2014), a combination of Hessian multiscale analysis, variational segmentation and texture features (Figueiredo et al., 2015). These techniques are shown to usually involve interdependence on detection of anatomical structures (i.e. OD and fovea) with lesion detection, and that in turn determines automated DR screening outcome.

Localization and segmentation of OD and fovea facilitate the detection of retinal lesions as well as the assessment (based on the geometric location of these lesions) of the severity and monitoring progression of DR and DME. Hence, several approaches have been proposed for localization of OD, and most of them utilized the OD properties like intensity, shape, color, texture, etc. and many others showed effectiveness of mathematical morphology (Morales et al., 2013; Marin et al., 2015), template matching (Giachetti et al., 2014), deformable models (Yu et al., 2012; Wu et al., 2016) and intensity profile analysis (Kamble et al., 2017; Uribe-Valencia and Martínez-Carballido, 2019). Further, approaches utilized for OD segmentation are based on level set (Yu et al., 2012), thresholding (Marin et al., 2015), active contour (Mary et al., 2015) and shape modeling (Cheng et al., 2015), clustering (Thakur and Juneja, 2017), and hybrid (Bai et al., 2014) approaches. Similarly, the fovea is detected mostly using a geometric relationship with OD and vessels through morphological (Welfer et al., 2011), thresholding (Gegundez-Arias et al., 2013), template matching (Kao et al., 2014) and intensity profile analysis (Kamble et al., 2017) techniques. Poor performance on the detection of normal anatomical structures could adversely affect lesion detection and screening accuracy. For instance, consider mathematical morphology based techniques presented in 2002 (Walter et al., 2002), 2008 (Sopharak et al., 2008) and 2014 (Zhang et al., 2014). These works demonstrate how morpho-

logical processing-based approaches evolved by including multiple steps for the final objective of exudate detection. In initial efforts, Walter et al. (2002) devised a technique for OD and EXs segmentation, afterward removed OD to obtain EX candidates. Similarly, Sopharak et al. (2008) achieved the same objective with the detection and removal of OD and vessels. Recently, an approach presented by Zhang et al. (2014) achieved much better results, but it involved (a) spatial calibration, (b) detection of dark and bright anatomical structures such as vessels and OD respectively, also (c) bright border regions detection before actual extraction of candidates. Also, there are other techniques based on textural (Morales et al., 2017; Porwal et al., 2018c) and mid-level (Pires et al., 2017) features of retinal images that forgo lesion segmentation step for DR screening. However, most of these techniques depend on the intermediate steps mentioned above. In an approach based on machine learning (Roychowdhury et al., 2014), authors detected bright and dark lesions as a first step and later performed hierarchical lesion classification to generate a severity grade for DR. Similarly, Antal and Hajdu (2014) proposed a strategy involving image-level quality assessment, pre-screening followed by lesion and anatomical features extraction to finally decide about the presence of DR using ensemble of classifiers. Further, for identification of different stages of DR, morphological region properties (Yun et al., 2008), texture parameters (Acharya et al., 2012; Mookiah et al., 2013b), non-linear features of higher-order spectra (Acharya et al., 2008), hybrid (Dhara et al., 2015) and information fusion (Niemeijer et al., 2009) approaches were found useful. As DME is graded based on the location of EXs from the macula, many researchers (Giancardo et al., 2012; Medhi and Dandapat, 2014; Perdomo et al., 2016; Marin et al., 2018) proposed EXs based features to determine the severity of DME. While several others (Deepak and Sivaswamy, 2012; Mookiah et al., 2015; Acharya et al., 2017) have proposed various feature extraction techniques to grade DME stages without segmenting EXs. Mainly for approaches in this section, features are based on color, brightness, size, shape, edge strength, texture, and contextual information of pixel clusters in spatial and/or transform domain. Whereas classification is achieved through classifiers such as K Nearest Neighbors (KNN), Naive Bayes, Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Trees, etc.

These lesion detection or screening techniques are shown to usually involve interdependence with detection of other landmarks. However, there was a lack of a single platform to test their performance for each objective. For such handcrafted features based approaches, this challenge provided a unique platform to compare and contrast the algorithm's performance for detection of anatomical structures, lesions as well as the screening of DR and DME.

2.2. Deep learning methods

Deep learning is a general term to define multi-layered neural networks able to concurrently learn a low-level representation and higher-level parameters directly from data. This representation learning capability drastically reduces the need for engineering ad-hoc features, however, full end-to-end training of deep learning-based approaches typically require a significant number of samples. Its rapid development in recent times is mostly due to a massive influx of data, advances in computing power and developments in learning algorithms that enabled the construction of multi-layer (more than two) networks (Hinton, 2018; Voulodimos et al., 2018). This progress has induced interests in the creation of analytical, data-driven models based on machine learning in health informatics (Ching et al., 2018; Raviet et al., 2017). Hence, it is emerging as an effective tool for machine learning, promising to reshape the future of automated medical image analysis (Greenspan et al., 2016; Litjens et al., 2017; Suzuki, 2017; Shen et al., 2017; Kim et al.,

2018; Ker et al., 2018). Among various methodological variants of deep learning, Convolutional Neural Networks (CNNs or ConvNets) are most popular in the field of medical image analysis (Hoo-Chang et al., 2016; Carin and Pencina, 2018). Several configurations and variants of CNN's are available in the literature, some of the most popular are AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2015) and ResNet (He et al., 2016).

Deep learning has also been widely utilized in retinal image analysis because of its unique characteristic of preserving local image relations. Majority of approaches in literature employ deep learning to retinal images by utilizing “off-the-shelf CNN” features as complementary information channels to other hand-crafted features or local saliency maps for detection of abnormalities associated with DR (Chudzik et al., 2018; Orlando et al., 2018; Dai et al., 2018), segmentation of OD (Zilly et al., 2017; Fu et al., 2018), and detection of DR (Rangrej and Sivaswamy, 2017). The authors (Fu et al., 2016) employ fully connected conditional random fields along with CNN to integrate discriminative vessel probability map and long-range interactions between pixels to obtain final binary vasculature. Whereas some approaches initialized the parameters with those of pre-trained models (on non-medical images), then “fine-tuned” (Tajbakhsh et al., 2016) the network parameters for DR screening (Gulshan et al., 2016; Carson Lam et al., 2018). In another approach researchers used two-dimensional (2D) image patches as an input instead of full-sized images for lesion detection (Tan et al., 2017b; van Grinsven et al., 2016; Lam et al., 2018; Chudzik et al., 2018; Khojasteh et al., 2018), and OD and fovea detection (Tan et al., 2017a). García et al. (2017) trained the “CNN from scratch” and compared it with fine-tuning results based on two other existing architectures. Recently, Shah et al. (2018) demonstrated that ensemble training of auto-encoders stimulates diversity in learning dictionary of visual kernels for detection of abnormalities. Whereas Giancardo et al. (2017) proposed a novel way to compute vasculature embedding that leverages internal representation of a new encoder-enhanced CNN, demonstrating improvement in DR classification and retrieval task. There is a significant development in the automated identification of DR using CNN models in recent time. A customized CNN (Gargeya and Leng, 2017) was proposed for DR screening and trained using 75,137 images obtained from EyePACS system (Cuadros and Bresnick, 2009), where an additional classifier was further employed on the CNN-derived features to determine if an image is with or without retinopathy. Similarly, Google Inc. (Gulshan et al., 2016) developed a network optimized (fine-tuning) for image classification, in which a CNN is trained by utilizing a retrospective development database consisting of 128,175 images with labels. There are some hybrid algorithms, in which multiple, semi-dependent CNN's are trained based on the appearance of retinal lesions (Abràmoff et al., 2016; Quellec et al., 2016). A step further, researchers (Quellec et al., 2017) demonstrated an ability of lesion segmentation based on CNN trained for image-level classification. However, Lynch et al. (2017) demonstrated that hybrid algorithms based on multiple semi-dependent CNNs might offer a more robust option for DR referral screening, stressing the importance of lesion segmentation. For further details, readers are recommended to follow recent reviews for detection of exudates (Fraz et al., 2018), red lesions (Biyani and Patre, 2018) and a systematic review with a focus on the computer-aided diagnosis of DR (Mookiah et al., 2013a; Nørgaard and Grauslund, 2018).

This current progress in artificial intelligence provides an opportunity to researchers for enhancing the performance of DR referral system to a more robust diagnosis system that can provide quantitative information for multiple diseases matching international standards of clinical relevance. Thus, the presented challenge design offers an avenue to gauge precise DR severity status and op-

portunity to deliver accurate measures for lesions, that could even help in follow-up studies to observe changes in the retinal atlas.

3. Indian diabetic retinopathy image dataset

The IDRiD dataset (Porwal et al., 2018a) was created from real clinical exams acquired at an eye clinic located in Nanded, (M.S.), India. Retinal photographs of people affected by diabetes were captured with focus on macula using Kowa VX – 10α fundus camera. Prior to image acquisition, pupils of all subjects were dilated with one drop of tropicamide at 0.5% concentration. The captured images have 50° field of view, resolution of 4288 × 2848 pixels and are stored in *jpg* format. The final dataset is composed of 516 images divided into five DR (0 – 4) and three DME (0 – 2) classes with well-defined characteristics according to international standards of clinical relevance. It provides expert markups of typical DR lesions and normal retinal structures. It also provides disease severity level of DR and DME for each image in the database. Three types of ground-truths are available in the dataset:

1. *Pixel Level Annotation*: This type of annotation is useful in techniques to locate individual lesions within an image and to segment out regions of interest from the background. Eighty-one color fundus photographs with signs of DR were annotated at the pixel-level for developing ground truth of MAs, SEs, EXs and HES. The binary masks (as shown in Fig. 2) for each type of lesion are provided in tif file format. Additionally, OD was also annotated at the pixel-level and binary masks for all 81 images are provided in the same format. All of these annotations play a vital role in research for computational analysis of segmenting lesions within the image.

2. *Image Level Grading*: It consists of information meant to describe the overall risk factor associated with an entire image. Two medical experts provided adjudicated consensus grades to the full set of 516 images with a variety of pathological conditions of DR and DME. Grading for all images is available in the CSV file. The diabetic retinal images were classified into separate groups according to the International Clinical Diabetic Retinopathy Scale (Wu et al., 2013), confined to image under observation, as shown in Table 1.

The DME severity was decided based on occurrences of EXs near to macula center region (Decencièrre et al., 2014) as shown in Table 2.

3. *OD and Fovea center co-ordinates* The OD and fovea center locations are marked for all 516 images and the markup is available as a separate CSV file.

Table 1

DR Severity Grading. NPDR: Non-proliferative DR and PDR: Proliferative DR

DR Grade	Findings
0: No apparent retinopathy	No visible sign of abnormalities
1: Mild NPDR	Presence of MAs only
2: Moderate NPDR	More than just MAs but less than severe NPDR
3: Severe NPDR	Any of the following: <ul style="list-style-type: none"> • >20 intraretinal HES • Venous beading • Intraretinal microvascular abnormalities • no signs of PDR
4: PDR	Either or both of the following: <ul style="list-style-type: none"> Neovascularization Vitreous/pre-retinal HE

Table 2

Risk of DME.

DME Grade	Findings
0	No Apparent EX(s)
1	Presence of EX(s) outside the radius of one disc diameter from the macula center
2	Presence of EX(s) within the radius of one disc diameter from the macula center

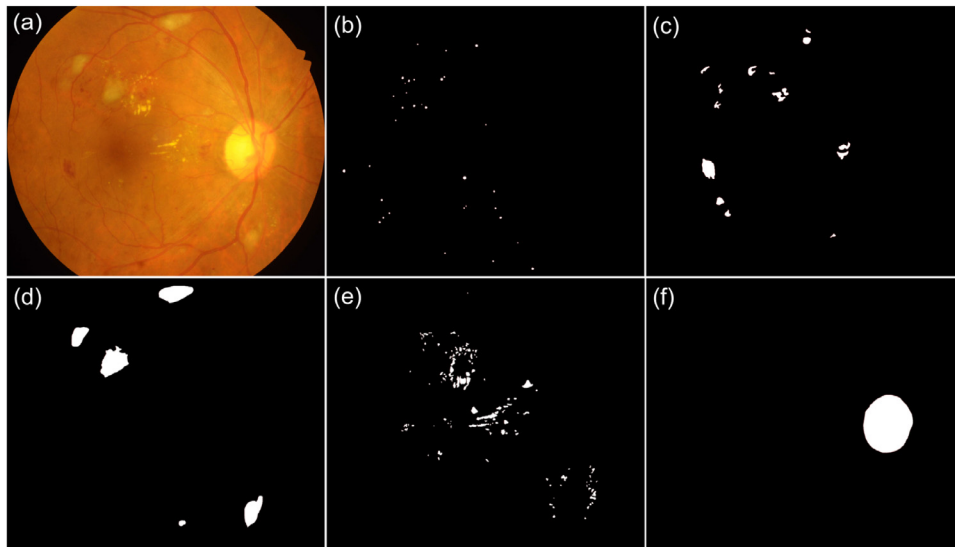


Fig. 2. Retinal photograph and different pixel-level annotations: (a) sample fundus image from the IDRiD dataset; sample ground truths for (b-f) MAs, HES, SEs, EXs and OD respectively.

The IDRiD dataset is available from IEEE Dataport Repository⁴ under a Creative Commons Attribution 4.0 license. More detailed information about the data is available in the data descriptor (Porwal et al., 2018b). Tables A.1 and A.2 highlight a comparative strength of this dataset with respect to existing datasets. IDRiD is the only dataset that provides all three types of annotations mentioned above. This streamlined collection of annotations would allow it to be utilized in research and lead to the development of better generalizable models for image analysis, enabling further progress in automated DR diagnosis.

4. Challenge organization

The “Diabetic Retinopathy – Segmentation and Grading Challenge” was composed into various stages, giving a well-organized work process to potentiate success of the contest. Fig. 3 depicts the work-flow of the overall challenge organization. The challenge was officially announced at the ISBI - 2018 website⁵ on 15th October 2017.

The challenge was subdivided into three sub-challenges as follows:

1. Lesion Segmentation: Segmentation of retinal lesions associated with DR such as MAs, HES, EXs and SEs.
2. Disease Grading: Classification of fundus images according to the severity level of DR and DME.
3. OD Detection and Segmentation, and Fovea Detection: Automatic localization of OD and fovea center coordinates, and segmentation of OD.

The challenge involved 4 stages, as detailed below:

Stage-1. Data Preparation and Distribution: The IDRiD dataset was adopted for this challenge, where experts verified that all images are of adequate quality, clinically relevant, that no image is duplicated and that a reasonable mixture of disease stratification representative 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).

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 representative of learned perceptual 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 performance. The percentage of images that should be in each subset for lesion and OD segmentation tasks (sub-challenge – 1 and part of sub-challenge – 3) was supported by the research outcome (Dobbins and Simon, 2011) which demonstrated that splitting 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 percentage 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.

⁴ <https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid>

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

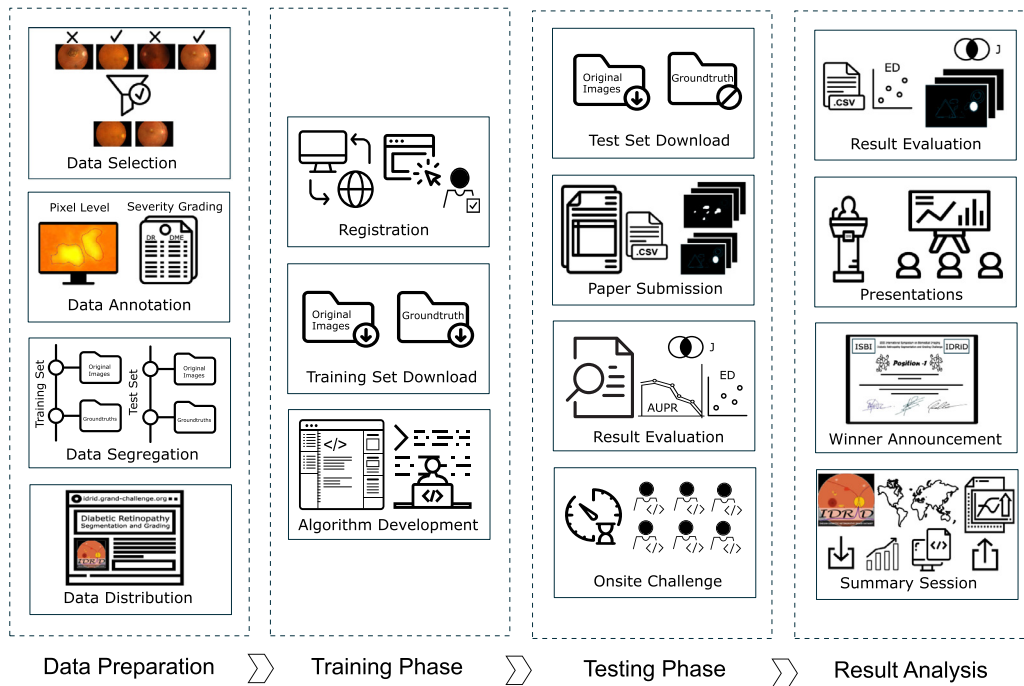


Fig. 3. Workflow of the ISBI - 2018: Diabetic Retinopathy – Segmentation and Grading Challenge.

The challenge was hosted on *Grand Challenges in Biomedical Imaging Platform*⁶, one of the popular platforms for biomedical imaging-related competitions. A challenge website was set up and launched on 25th October 2017 to disseminate challenge related information. It was also used for registration, data distribution, submission of results and paper, and communication between organizers and participants.

Stage-2. Registration and release of the training data: Registration of challenge for consideration to ISBI on-site contest was open from the launch of the grand-challenge website (i.e. 25th October 2017) till the deadline for submission of results (i.e. 11th 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 20th 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 development 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 20th February, 2018. For other two sub-challenges, Set-B was released on 4th April which was part of 'on-site' challenge. Organizers refrained from an on-site evaluation of sub-challenge – 1 considering timing constraints in the evaluation of results for image segmentation tasks.

Submissions were sought for either of the following 8 different tasks corresponding to three sub-challenges (1 – Lesion Segmentation, 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

2. Sub-challenge – 2: Disease Grading

Task - 5: DR and DME Grading

3. 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 12th February and participants could submit their results and paper describing their approach to the organizers till 11th 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 parameters. 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 performance was displayed on the leaderboard of the challenge website. For sub-challenge – 1, teams were assessed based on the performance 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 community 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 financial support in the limited time.

Stage-4. ISBI Challenge Event: The main challenge event was held in conjunction with ISBI - 2018 on April, 4th 2018. The Set-B (only images) for sub-challenge – 2 and 3 was made available to the participants 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 restriction on the number of machines that could be used to produce the results. However, considering the timing constraints for processing,

⁶ <https://grand-challenge.org/>

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 disease severity grading corresponding to DR and DME. Similarly, sub-challenge – 3 (SC3) 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-BII-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 solution 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 presentation 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 presentations on their work. After the end of the on-site challenge event, on 6th April, the summary of challenge and analysis of results was presented, which included a final ranking of the competing solutions. 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 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 basic terminologies and abbreviations related to CNN and its variants utilized by participating teams. Further, it summarizes the solutions and related technical specifications. For the detailed description of a particular approach, please refer to proceedings of ISBI Grand Challenge Workshop at 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., 2017). The CNN architecture typically consists of an input layer followed 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 detailed as follows:

CONV layer comprises of a set of independent filters (or kernels) 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 ($I_{x,y}^\ell$) centered at location (x, y) of ℓ^{th} layer, the feature value in i^{th} feature map, $A_{x,y,i}^\ell$, is obtained as:

$$A_{x,y,i}^\ell = f((w_i^\ell)^T I_{x,y}^\ell + b_i^\ell) = f(C_{x,y,i}^\ell) \quad (1)$$

Where the parameters w_i^ℓ and b_i^ℓ are weight vector and bias term of i^{th} filter of ℓ^{th} layer, and $f(\cdot)$ is a nonlinear activation function such as sigmoid, rectified linear unit (ReLU) or hyperbolic tangent (tanh). It is important to note that the kernel w_i^ℓ that generates the feature map $C_{x,y,i}^\ell$ 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 common 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.

FC layer aims to perform higher-level reasoning by computing the class scores. Each neuron in this layer is connected to all neurons in the previous layer to generate global semantic information.

The last layer of CNN's is an output layer (O), here the Softmax operator is commonly used for classification tasks. The optimum parameters (θ , a common notation for both w and b) for a particular task can be determined by minimizing the loss function (L) defined for the task. Mathematically, for N input-output relations $\{(I^n, O^n); n \in [1, \dots, N]\}$ and corresponding labels G^n the loss can be derived as:

$$L = \frac{1}{N} \sum_{n=1}^N \ln(\theta; G^n, O^n) \quad (2)$$

Where N denotes the number of training images, I^n , O^n and G^n correspond to n^{th} training image. Here, a critical challenge in training CNN's arises from the limited number of training samples as compared to the number of learnable parameters that need to be optimized for the task at hand. Recent studies have developed some key techniques to better train and optimize the deep models such as data augmentation, weight initialization, Stochastic Gradient Descent (SGD), batch normalization, shortcut connections, and regularization. For more understanding related to advances in CNN's, the reader is recommended to refer a paper by Gu et al. (2018).

The growing use of CNN's as the backbone of many visual tasks, ready for different purposes (such as segmentation, classification or localization) and the available data, has made architecture search a primary channel in solving the problem.

In this challenge, mainly for disease severity grading problem, participants either directly utilized existing variants of CNN's or ensembled them to demarcate the input image to one of the classes mentioned in Table 4. Several configurations and variants of CNN's are available in the literature; some of the most popular are AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2015) and ResNet (He et al., 2016) due to their superior performance on different benchmarks for object recognition tasks. A typical trend with the evolution of these architectures is that the networks have gotten deeper, e.g., ResNet is about 19, 8 and 7 times deeper than AlexNet, VGGNet and GoogLeNet respectively. While the increasing depth improves feature representation and prediction performance, it also increases complexity, making it difficult to optimize and even becomes prone to overfitting. Further, the increasing number of layers (i.e., network depth) lead to vanishing gradient problems as a result of a large number of multiplication operations. Hence, many teams chose the DenseNet (Iandola et al., 2014) which connects each layer to every other layer in a feed-forward fashion, reducing the number of training parameters and alleviates the vanishing gradient problem. DenseNet exhibits $\ell(\ell + 1)/2$ connections in ℓ layer network, instead of only ℓ , as in the networks mentioned above. This enables feature reuse throughout the network that leads to more compact internal representations and in turn, enhances its prediction accuracy. Another opted approach, Deep Layer Aggregation (DLA) structures (Yu et al., 2017), extends the "shallow" skip connections in DenseNet to incorporate more depth and sharing of the features. DLA uses two structures – iterative deep aggregation (IDA) and hierarchical deep aggregation (HDA) that iteratively and hierarchically fuse the feature hierarchies (i.e. semantic and spatial) to make networks work with better accuracy and fewer parameters. Recent Fully Convolutional Network (FCN) (Long et al., 2015) adapt and extend deep classification architectures (VGG and GoogLeNet) into fully convolutional networks and transfer their learned representations by fine-tuning to the segmentation task. It defines a skip architecture that combines semantic information from a deep, coarse layer with appearance in-

formation from a shallow, fine layer to produce accurate and detailed segmentations.

For lesion segmentation task, most of the participating teams exploit U-Net architecture (Ronneberger et al., 2015). The main idea in U-Net architecture is to supplement the usual contracting network through a symmetric expansive path by addition of successive layers, where upsampling (via deconvolution) is performed instead of the pooling operation. The upsampling part consists of a large number of feature channels, that allow the network to propagate context information to higher-resolution layers. The high-resolution features from the contracting path are merged with the upsampled output and fed to soft-max classifier for pixel-wise classification. This network works with very few training images and enables the seamless segmentation of high-resolution images by means of an overlap-tile strategy. Other similar architecture SegNet (Badrinarayanan et al., 2015) was opted by a team; it consists of an encoder and decoder network, where the encoder network is topologically identical to CONV layers in VGG16 and in which FC layer is replaced by a Softmax layer. Whereas, the decoder network comprises a hierarchy of decoders, one corresponding to each encoder. The decoder uses max-pooling indices for up-sampling its encoder input to produce sparse feature maps. Later, it convolves the sparse feature maps with a trainable filter bank to densify them. At last, decoder output is fed to a soft-max classifier for the generation of segmentation map. One team choose Mask R-CNN (He et al., 2017), a technique primarily based on a Region Proposal Network (RPN) that shares convolutional features of an entire image with the detection network, thus enabling region proposals to localize and further segment normal and abnormal structures in the retina. RPN is a fully convolutional network that contributes to concurrently predicting object bounds and "objectness" scores at each position.

Following subsections present the solutions designed by participating teams with respect to three sub-challenges. Table 6 summarizes data augmentation, normalization and preprocessing tasks performed by each team.

5.1. Sub-challenge – 1: Lesion segmentation

For a given image, this task seeks to get the probability of a pixel being a lesion (either MA, HE, EX or SE). Although different retinal lesions have distinct local features, for instance, MA, HE, EX, SE have a different shape, color and distribution characteristics, these lesions share similar global features. Hence, the majority of participating teams built a general framework that would be suitable for the segmentation of different lesions, summarized as follows:

5.1.1. VRT (Jaemin Son et al.)

Son et al. modified U-Net (Ronneberger et al., 2015) in such a way that upsampling layers have the same number of feature maps with layers concatenated. It was based on the motivation that features in initial layers and upsampled layers are equally important to segmentation. Additionally, they adjusted the number of max-pooling so that the radius of the largest lesion spans a pixel in the coarsest layer. In case of EX and HE, max-pooling is done six times, whereas for SE and MA it is done four times and twice. Further, for dealing with MA's, they used inverse pixel shuffling to convert a $1280 \times 1280 \times 3$ pixels image to $640 \times 640 \times 12$ for network input and pixel shuffling (Shi et al., 2016) to convert $640 \times 640 \times 4$ segmentation map into $1280 \times 1280 \times 1$ pixels. Later, the pairs of a normalized fundus image and reference ground truths were fed to the network to generate segmentation result in the range [0, 1]. They used weighted binary cross entropy (Murphy, 2012) as loss

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		
Sub-challenge - 1	 VRT	✓	✓	✓	✓	✓	shear	✓	FOV cropping, division by 255 then mean subtraction
	 iFLYTEK	✓	✓	✓	✓	✓	×	✓	lesion patch extraction
	 PATech	✓	✓	×	✓	×	color ^a	✓	RGB to LUV, contrast adjustment
	 SDNU	✓	✓	×	×	×	×	–	–
	 SOONER	✓	✓	×	×	✓	×	✓	mean subtraction, lesion patch extraction
	 LzyUNCC	✓	×	×	×	✓	stochastic and photo-metric ^b	–	FOV cropping, image enhancement
	 SAIHST	✓	✓	×	×	×	×	✓	CLAHE, Gaussian smoothing
Sub-challenge - 2	 LzyUNCC	✓	×	×	×	✓	color ^a , stochastic and photo-metric ^b	–	FOV cropping, image enhancement
	 VRT	×	×	×	×	×	×	✓	mean subtraction
	 Mammoth	✓	✓	✓	✓	×	color	×	morphological opening and closing
	 AVASAVA	✓	×	×	×	✓	×	✓	intensity scaling
	 HarangiM1	×	×	×	×	×	×	✓	FOV cropping
	 HarangiM2	×	×	×	×	×	×	✓	–
Sub-challenge - 3	 DeepDR	×	×	×	×	✓	OD, fovea region	✓	FOV cropping, mean subtraction
	 VRT	✓	✓	✓	✓	✓	shear and cropped OD	✓	FOV cropping, contrast adjustment
	 ZJU-BII-SGEX	×	×	×	×	×	×	✓	FOV cropping
	 SDNU	✓	×	✓	×	×	×	–	–
	 IITkgpKLIV	✓	✓	×	×	×	×	✓	–
	 CBER	×	×	×	×	×	×	–	–

^a Reference: Krizhevsky et al. (2012)

^b Reference: Howard (2013)

function given by

$$L = \frac{1}{N} \sum_{n=1}^N \left[-\alpha G^n \log O^n - (1 - G^n) \log(1 - O^n) \right] \quad (3)$$

where N denotes the number of the pairs in a batch, G^n and O^n represent true segmentation and predicted segmentation for n^{th} image. The value of α was determined as follows:

$$\alpha = \frac{B_0^i}{\gamma F_1^i} \quad (4)$$

where B_0^n and F_1^n 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 penalize false negatives in order to boost sensitivity, an important factor 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.

They trained the network over 300 epochs using Adam optimizer (Kingma and Ba, 2014) with hyper-parameters of $\beta_1 = 0.5$, $\beta_2 = 0.999$ and learning rate of $2e^{-4}$ until 250 epochs and $2e^{-5}$ until the end. All implementation was done by Keras 2.0.8

Table 7

γ values in Eq. (4).

EXs	SEs	HEs	MAAs
64	512	8	32

with tensorflow backend 1.4.0 using a server with 8 TITAN X (pascal). The source code is available at https://bitbucket.org/woalsdnd/isbi_2018_fundus_challenge.

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) 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

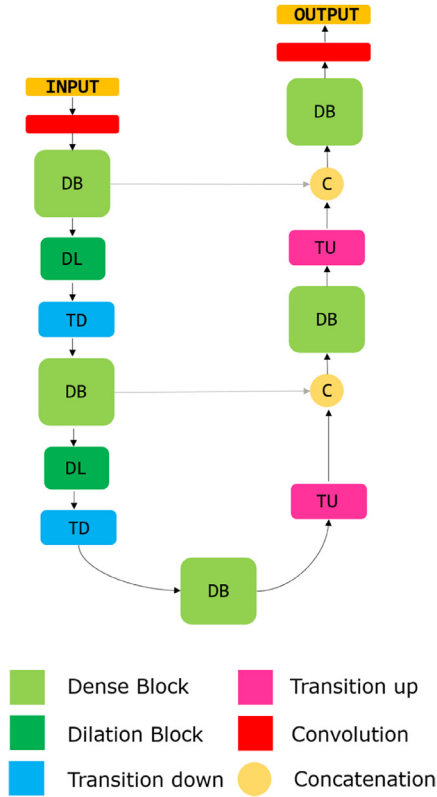


Fig. 4. Proposed architecture for lesion segmentation.

(non-target lesions) were not utilized in the learning process leading to high false-positive count. In the second stage, unlike the first segmentation model which used a lesion centered sample from input dataset pool, candidate regions were extracted using probability maps from the previous stage. Here, the input size fed to model for SE was $320 \times 320 \times 3$ pixels, for HE and EX it was $256 \times 256 \times 3$ pixels, and for MA it was modified to $80 \times 80 \times 3$ pixels considering its small appearance. In this step, a candidate region was regarded 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. However, it was identified in the test that a small proportion of false positives still exist, so an additional model was introduced to refine 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 respectively 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) in which they innovatively combined the DenseNets (Iandola et al., 2014) and dilation block with U-Net (Ronneberger et al., 2015) to capture more context information and multi-scale features.

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

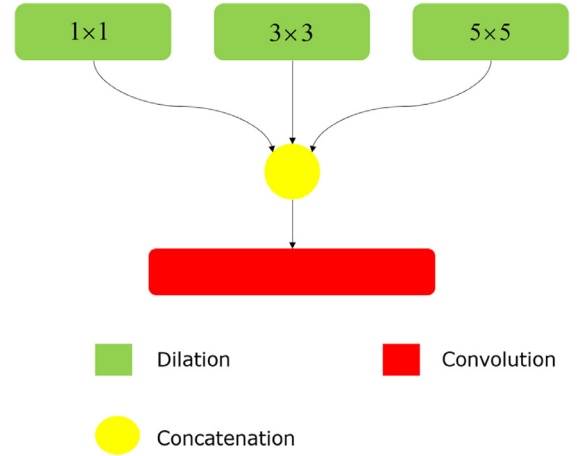


Fig. 5. Architecture for dilation block.

with 4 Transitions Up (TU). To capture multi-scale features, DL (see Fig. 5) is used with dilation rate of 1, 3 and 5 are concatenated for the convolution. The dense block (DB) is constructed by four layers. 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 pixels 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) and 2D cross Entropy as follows:

$$L = -\text{mean}(w_{10} * G * \log(O) + w_{11} * (1 - G) * \log(1 - O) + w_2 * \text{dice}(G)) \quad (5)$$

where w_{10} and w_{11} are the factors utilized to keep a balance between the positive and negative pixels, and w_2 is the factor utilized to control significance between dice and cross entropy loss. The values of w_{10} , w_{11} and w_2 were empirically set to 0.7, 0.3 and 0.4 respectively. The models were trained using Adam optimizer with default parameters, $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning 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 operating system.

5.1.4. SOONER (Yunzhi Wang et al.)

Wang et al. adopted U-Net (Ronneberger et al., 2015) architecture for solving retinal lesion segmentation problem. The network takes a 380×380 pixels fundus image patch as input and predicts the binary mask of the retinal lesion within 196×196 pixels central region of an input patch. They pre-processed fundus images by subtracting the local mean of each color channel and performed random flipping for data augmentation. Batch normalization 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 randomly extracted negative image patches from fundus images with

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 extracted 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) 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 preprocessing and augmentation. Subsequently, considering the correlation 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 segmentation of SE. While the model training of EX segmentation, a trade-off parameter (penalty) was assigned in loss function to control 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 $(1 - \frac{\text{epoch}-1}{\text{total epoch}})^{0.9}$ with the momentum of 0.9. The initial 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), 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 normalization, 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 image 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 pixels. 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 β_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 implementation of Mask R-CNN from Abdulla (2017) for solving the problem. This method could detect different objects while simultaneously 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 backbone. They implemented an algorithm in Keras with Tensorflow as backend and processed on 8 NVIDIA TITAN Xp GPUs. The experimental 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 follows:

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 augmentation. 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).

5.2.2. VRT (Jaemin Son et al.)

Son et al. used network (Son et al., 2018) for DR grading. Kaggle 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. Furthermore, 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, features with high importance were selected among numerous features in the initial training due to gradient boosting (for instance, XGBoost) was likely to overfit when provided with overly redundant 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 accuracy 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/isbi_2018_fundus_challenge.

5.2.3. Mammoth (Junyan Wu et al.)

Wu et al. proposed a unified framework that combines deep feature extractor and statistical feature blending to automatically predict the DR and DME severity scores. For DME, they used DenseNet (Iandola et al., 2014) to directly predict the severity score. Whereas for DR, Kaggle training dataset was used to pre-train the DenseNet model through a dynamic sampling mechanism to balance the training instances and later fine-tuned using IDRiD dataset. Initially, the background of all images was cropped and resized to 512×512 pixels. Later, morphological opening and closing are utilized to preserve bright and dark regions. For instance, the morphological opening can erase the EXs and highlight the MAs. Whereas, the closing operation can remove MAs and preserve EXs. These operations can be used to denoise specific levels of classifications, for example, the risk of DME only depends on the location of the EXs. Further, several standard data augmentation methods (as shown in Table 6) are also employed. Mean Squared Error (MSE) and cross-entropy with five classes were the loss functions employed to train the network and SGD for optimization. The initial learning rate was set to 0.0005 with a decrement of 0.1 after every 30 epochs. The initial training was done by 200 epochs and fine-tuning by 50 epochs. Afterward, the last layer was removed before the final prediction, and its statistical features were aggregated together into a boosting tree. Specifically, 50 pseudo-random augmentations were performed to get 50 outputs from last second FC layer (size of 4096), then the mean and standard deviation of 50 feature vectors for each image was computed, and both vectors were then concatenated together for training in LightGBM. The output from the second last layer of fine-tuning experiments was used to train a blending model, strategy adopted from team o.o's solution of Kaggle DR challenge. Finally, for the disease grading prediction, gradient boosting tree model was built on a combined second last layer from the pre-trained network and fine-tuned network.

5.2.4. Harangim1 (Balazs Harangi et al.)

Harangi et al. proposed an approach for the classification of retinal images via the fusion of two AlexNet (Krizhevsky et al., 2012) and GoogLeNet (Szegedy et al., 2015). For this aim, they removed FC and classification layers and interconnect them by inserting a joint FC layer followed by the classic softmax/ classification layers for the final prediction. In this way, single network architecture was created which allows to train the member CNN's simultaneously. For each $I^{(n)}$, let us denote the outputs of the final FC layers of the member CNN's by $\hat{O}_1^{(n)}, \hat{O}_2^{(n)}$. The FC layer of their ensemble aggregates them via

$$\hat{O}^{(n)} = A_1 \hat{O}_1^{(n)} + A_2 \hat{O}_2^{(n)} \quad (6)$$

where weight matrices A_1, A_2 were of size 5×5 and initialized as

$$A_1 = A_2 = \begin{bmatrix} 1/5 & 0 & 0 & 0 & 0 \\ 0 & 1/5 & 0 & 0 & 0 \\ 0 & 0 & 1/5 & 0 & 0 \\ 0 & 0 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 0 & 1/5 \end{bmatrix} \quad (7)$$

The last two layers of the ensemble were a Softmax and a classification one. Let $O_{SM}^{(n)}$ be an output of a former layer, the MSE was used for optimization as a loss function:

$$MSE = \frac{1}{2N} \sum_{n=1}^N (\hat{O}_{SM}^{(n)} - O^{(n)})^2 \quad (8)$$

During the training phase, back-propagation is applied to minimize the loss via adjusting all parameters of member CNNs and weight matrices A_1, A_2 .

For the grading of DME, the final layers of member CNNs consist of 3 neurons, and weight matrices A_1, A_2 were 3×3 , initialized as

$$A_1 = A_2 = \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \quad (9)$$

For training, they merged IDRiD and Kaggle training set. The parameters of architectures were found by SGD algorithm in 189 and 50 epochs respectively for DR and DME classification tasks. Learning rate was set to 0.0001. Training times required on the datasets for DR and DME were 96.6 (189 epochs) and 23.4 (50 epochs) hours respectively. Implementation of this work was done in MATLAB 2017b. The training was performed using an NVIDIA TITAN X GPU card with 7 TFlops of single-precision performance, 336.5 GB/s of memory bandwidth, 3072 CUDA cores, and 12 GB memory.

5.2.5. AVSASVA (Varghese Alex et al.)

Alex et al. used ensembles of pre-trained CNNs (on ImageNet dataset), namely, ResNets (He et al., 2016) and DenseNets (Iandola et al., 2014) for the task of disease grading. For DR grading, two ensembles of CNN's namely "primary" and "expert" classifiers were used. The primary classifier was trained to classify a fundus image as one of the 4 classes viz; Normal, Mild NPDR, Moderate NPDR or S-(N)-PDR, a class formed by clubbing Severe NPDR and PDR. The expert classifier was trained exclusively on Severe NPDR or PDR images and was utilized to demarcate the input image as one of the aforementioned classes. During inference, each fundus image was resized to a dimension of 256×256 pixels. For the task of grading of DR in fundus images, they used test time augmentation through the "Ten Crop" function defined in PyTorch. The images were first passed through the primary classifier and then through the expert classifier, only if the image was classified as S-(N)-PDR by the primary classifier. The final prediction was achieved by using a majority voting scheme.

For DME grading, two ensembles were trained in a one versus rest approach. Ensemble 1 was trained to classify the input as either "image with no apparent EXs" (Grade 0) or "presence of EXs in image" (Grade 1 & Grade 2), while the Ensemble 2 was trained to classify an image as "Grade 2" DME or not (Grade 0 & Grade 1). During inference, the resized images were fed to both ensembles, and the final prediction was obtained by combining the two predictions by utilizing a set of user-defined rules. Briefly, the user-defined rules were: an image was classified as Grade 0 DME if ensemble 1 and ensemble 2 predict the absence of EXs and the absence of grade 2 DME respectively. A scenario wherein ensemble 2 predicts the presence of grade 2 DME, images were classified under category "Grade 2 DME" irrespective of the prediction from ensemble 1. Lastly, images were classified as Grade 1 DME if none of the above conditions were satisfied.

Both models for DR and DME were initialized with the pre-trained weights and the parameters of networks were optimized by reducing cross-entropy loss with ADAM as an optimizer. The learning rate was initialized to 10^{-3} for DR and 10^{-4} for DME. For DR, the learning rate was reduced by a factor of 10% every instance when the validation loss failed to drop. Each network was trained for 30 epochs and the model parameters that yielded the lowest validation loss were used for inference. For DME, the learning rate was annealed step-wise with a step size of 10 and the multiplicative factor of learning rate decay value of 0.9.

5.2.6. Harangim2 (Balazs Harangi et al.)

Harangi et al. combined self-extracted, CNN-based features with traditional, handcrafted ones for disease classification. They modified AlexNet (Krizhevsky et al., 2012) to allow the embedding of

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 considered for DR (or DME) classification task. In this way, both final weightings 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). As a result, a 30-element feature vector was obtained, which reflects the intensity, geometry, and texture of structures contained in the image (Agurto et al., 2010). Whereas to extract features related to the lesions MA and EX, they employed two detector ensembles (Antal and Hajdu, 2012; Nagy et al., 2011), which consist of a set of <preprocessing method (PP), candidate extractor (CE)> pairs organized into a voting 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 ensembles 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. Training 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 encoder (with a backbone network of ResNet-50 (He et al., 2016)) to localize the OD and fovea centers as a whole by solving a regression 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 information 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 performance of multi-task learning is better than a single task. The predicted 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 Zisserman, 2014)) which only localizes the OD center or fovea center of their related bounding boxes. During the training stage, they designed an effective data augmentation scheme to solve the problem of insufficient training data. In particular, to build the training 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 coordinates. 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 regression loss for the central location was the Euclidean loss. The modified loss function for global and local encoders was $0.045(L_{OD} + L_{fovea})$ and $0.045(L_{OD}/L_{fovea})$ respectively. Where L_{OD} and L_{fovea} are losses 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) and CNN that takes a vessel image 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 standardized with its mean and standard deviation (SD). When calculating 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). Pixel values in a vessel image range from 0 to 1. It uses external datasets DRIONS-DB (Carmona et al., 2008) and DRIVE (van Ginneken et al., 2004) available with OD and vessel ground truths respectively. For augmentation, the fundus images were affine-transformed and additionally 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 network 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 $\beta_1 = 0.5$, $\beta_2 = 0.999$. 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 Bengio, 2010).

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 vessel branches were depth-concatenated to the main branches. After deriving an activation map that represents the probability of containing an anatomical landmark, a hard-coded matrix was multiplied 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 1st to 500th epochs and 10^{-4} from 501th to 1000th epochs. All implementation was done in Keras 2.0.8

with TensorFlow backend 1.4.0 using a server with 8 TITAN X (pascal). Source code is available at 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) to localize and segment OD and fovea simultaneously. It scans the image and generates 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 image into fixed dimensions as network input. A feature extractor (ResNet-50) with feature pyramid networks (FPN) generates feature 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 objects. Finally, a ROI head network (RHN) is employed to obtain the label, mask, and refined bounding box for each ROI. They also incorporated 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) (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 implemented in TensorFlow 1.3 and Python 3.4 (source code was modified from Abdulla (2017)). 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) for segmentation 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 brightness levels. However, in contrast to original SegNet, the final decoder 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 activations 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, preserving the aspect ratio. Additionally, Drishti-GS (Sivaswamy et al., 2014) dataset was used for data augmentation to account for the 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 optimizer with learning rate 10^{-3} and $\beta = 0.9$. Early stopping of the training based on the validation loss is adopted to prevent overfitting. It was observed that the validation loss started to increase after 200 epochs. One more softmax layer is introduced after the Sigmoid 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) for solving all tasks in this sub-challenge. Mask R-CNN could realize accurate target 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 TITAN 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 segmenting these structures, mainly based on color and vascular information. 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 segmentation were an improved version of two methods previously published (Mendonça et al., 2013; Dashtbozorg et al., 2015), the method proposed for fovea localization was completely new. Initially, 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) is applied. The positions of the support points which give rise to the maximum filter response were found and used for delineating the OD boundary. Since a relation between the fovea-OD distance and the OD diameter was known (Jonas et al., 2015), 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 images 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).

$$SN = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (10)$$

$$PPV = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (11)$$

The curve was obtained by thresholding the results at 33 equally spaced instances i.e. $[0, 8, 16, \dots, 256]$ in gray levels or $[0, 0.03125, 0.0625, \dots, 1]$ 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).

The AUPR measure is more realistic (Boyd et al., 2013; Saito and Rehmsmeier, 2015) for the lesion segmentation performance over 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 results, 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 decide the overall grade of disease. For instance, EXs presence in the macular region can affect the vision of patient to a greater extent 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 screening 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 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

```

1 for  $n = 1, 2, \dots, N$  do
2   Correct = 0;
3   if ( $DR_O(n) == DR_G(n)$ ) and ( $DME_O(n) == DME_G(n)$ ) then
4     Correct = Correct + 1;
5   end
6 end
7 Average Accuracy =  $\frac{\text{Correct}}{N}$ 

```

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 coordinates. 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 determining 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). It represents the proportion of overlapping area between the segmented OD (O) and the ground truth (G).

$$J = \frac{|O \cap G|}{|O \cup G|} \quad (12)$$

Higher J indicates better segmentation. For the segmented results, images in range $[0, 255]$, it was computed at 10 different equally spaced thresholds $[0, 0.1, \dots, 0.9]$ 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. 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 participated (a complete list is available on the challenge website) in the sub-challenge-1 whose results were evaluated and ranked using the AUPR values. Amongst them, 7 teams (see Table 5) having performance within top 4 positions in either of lesion segmentation 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 different lesion segmentation tasks, presented their work at ISBI.

Table 8 summarizes the individual performance (Off-site evaluation) of each solution listed in order of their final placement for each subtask. It also contains various approaches followed and external 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 segmentation results illustrated in Fig. 6 and their corresponding overall evaluation score from Table 8 gives a better idea of how the evaluation scores correlate with the quality of segmentation. Fig. 7 summarizes the performance of top-4 teams per lesion segmentation 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 segmentation 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 on the Set-B. This algorithm produced an average grading accuracy of joint DR and DME on all images. Table 9 summarizes 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 proposed by team LzyUNCC followed by team VRT and team Mammoth. Fig. 8 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 participating 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 segmentation results were evaluated and ranked using Jaccard similarity score as outlined in Section 6.3. Results from the on-site evaluations are reported in Table 10 and Table 11 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

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 segmentation task were from teams ZJU-BII-SGEX, VRT and IITKgp-KLIV. Some sample OD segmentation results from these teams are illustrated in Fig. 9.

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

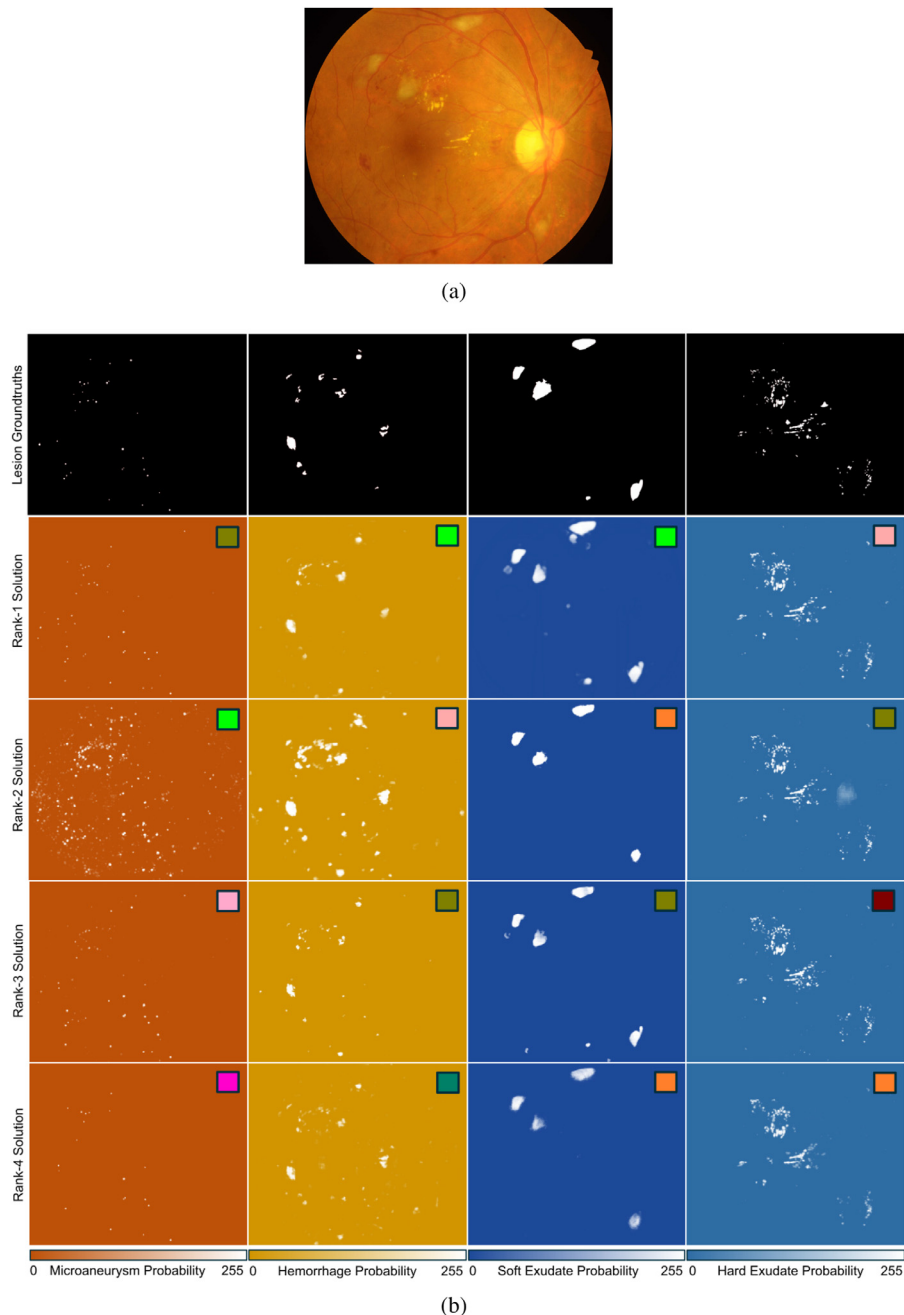


Fig. 6. Illustration of lesion segmentation results: (a) sample image and (b) segmentation outcome of top-4 teams (from left to right) (i) MAs, (ii) HEs, (iii) SEs, and (iv) EXs in retinal fundus images. Top row corresponds to ground truths, second row to entry from top performing team, similarly, third, 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	–

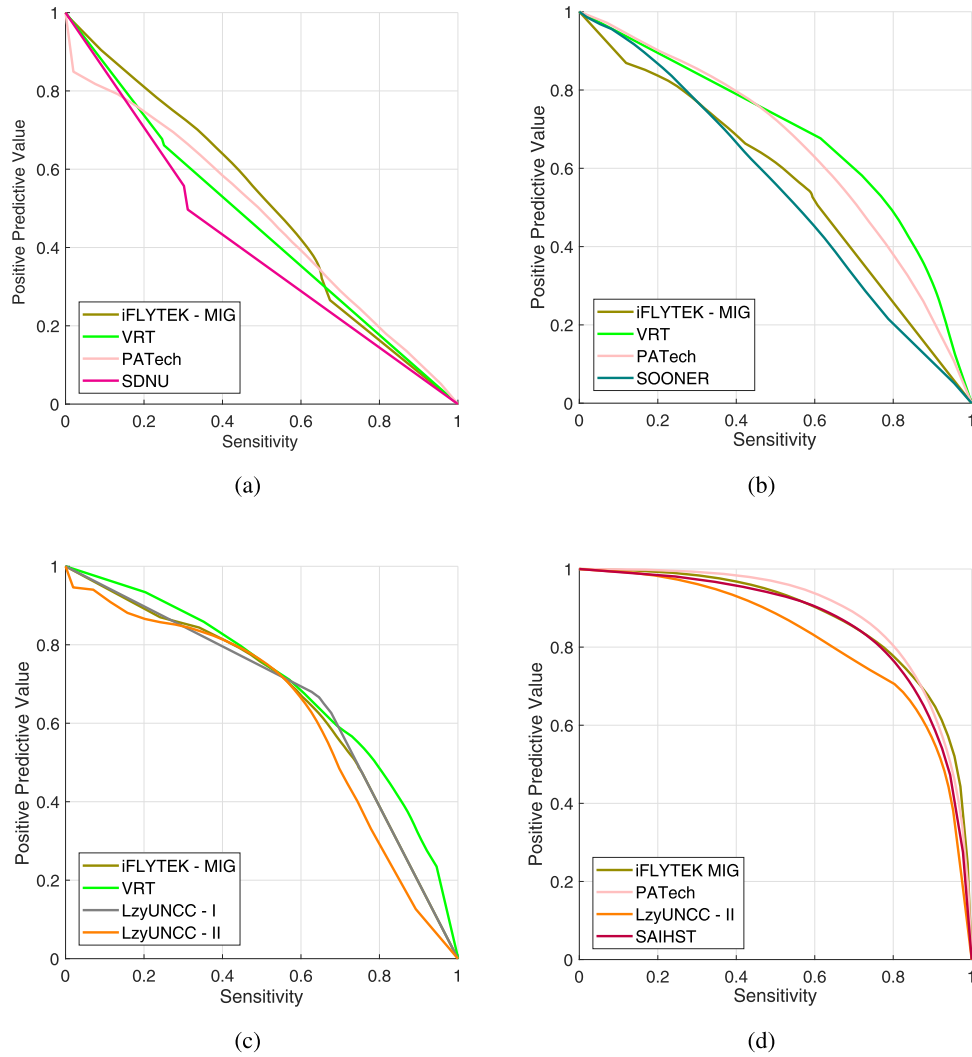


Fig. 7. The AUPR curves for the four top performing individual methods on the test dataset. These curves plot the sensitivity versus the positive predictive values for the different lesions, namely, (a) MAs, (b) HEs, (c) SEs, and (d) EXs.

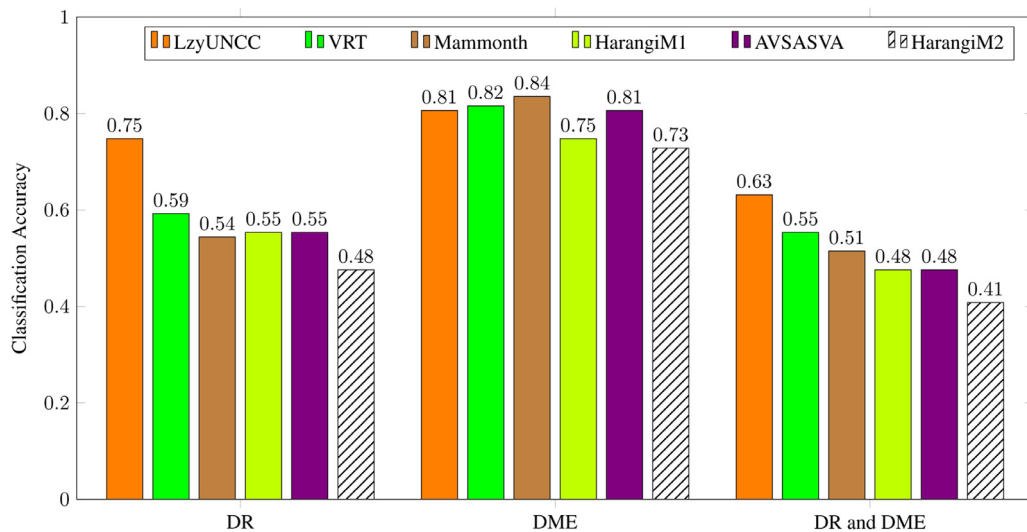


Fig. 8. Barplots showing separate and simultaneous classification accuracy of solutions developed by top-6 teams for grading of DR and DME.

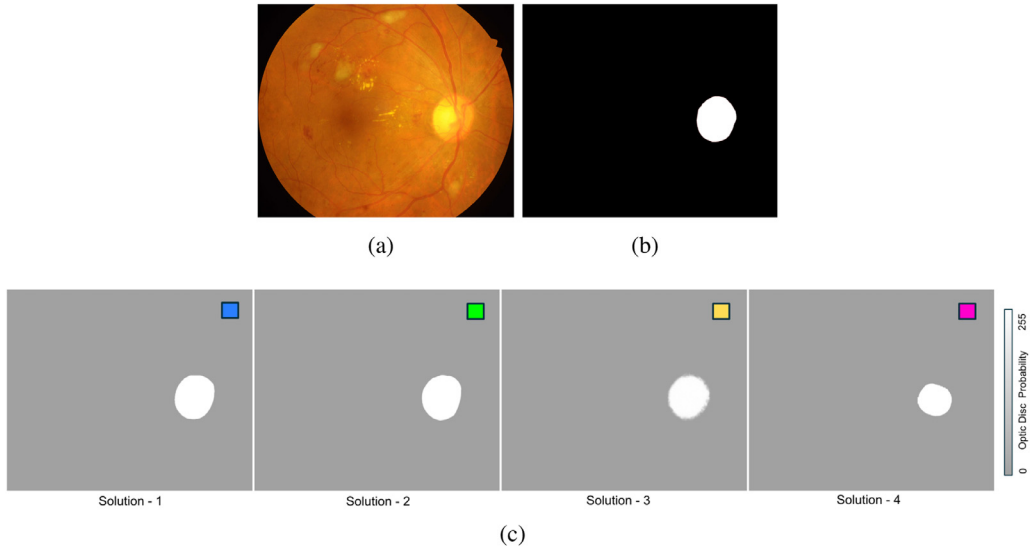


Fig. 9. Illustration of OD segmentation results: (a) sample image, (b) OD ground truth and (c) segmentation outcome of top-4 teams (from left to right).

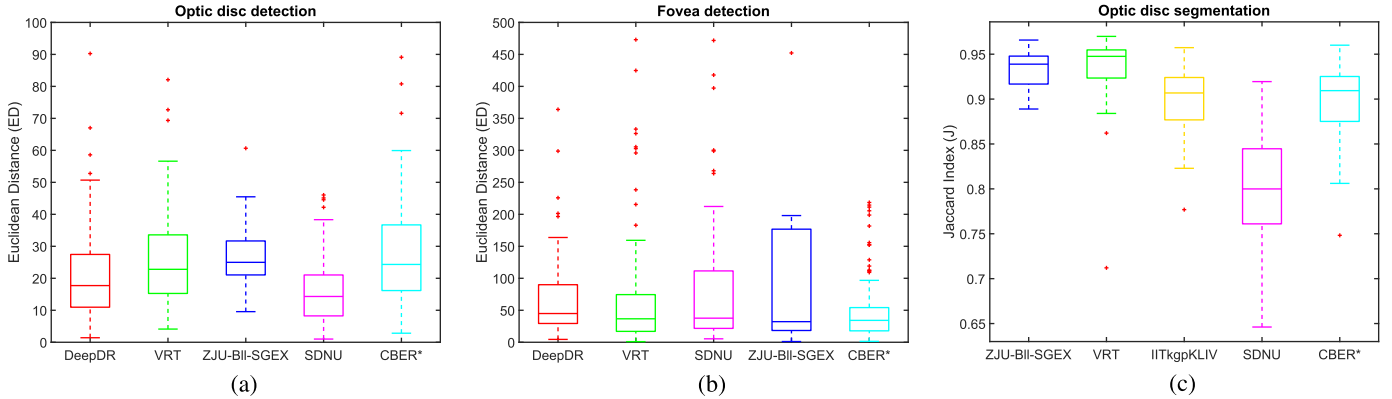


Fig. 10. Boxplots (a,b) showing dispersion of Euclidean distance for individual methods for OD and fovea and (c) showing the dispersion of Jaccard index for OD segmentation task. Boxplots show quartile ranges of the scores on the test dataset; plus sign indicate outliers (full range of data is not shown).

8. Discussion and conclusion

In this paper, we have presented the details of IDRid challenge including information about the data, evaluation metrics, an organization of the challenge, competing solutions and final results for all sub-tasks, i.e., lesion segmentation, disease grading and localization and segmentation of other normal retinal structures. Given the significant number of participating teams (37) and results obtained, we believe this challenge was a success. To the organizational end, efforts have been made in creating a relevant, stimulating and fair competition, capable of advancing collective knowledge 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 methods. 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 computing 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 highlights the performance of top solution for EX that performs significantly well in the presence of normal retinal structures and different challenging circumstances.

From the top-performing approaches, it is evident that solving the data imbalance problem improves the model performance significantly. Since background overwhelms foreground i.e. there are more background pixels than lesion pixels (see Fig. 6), the loss during 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 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 current 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

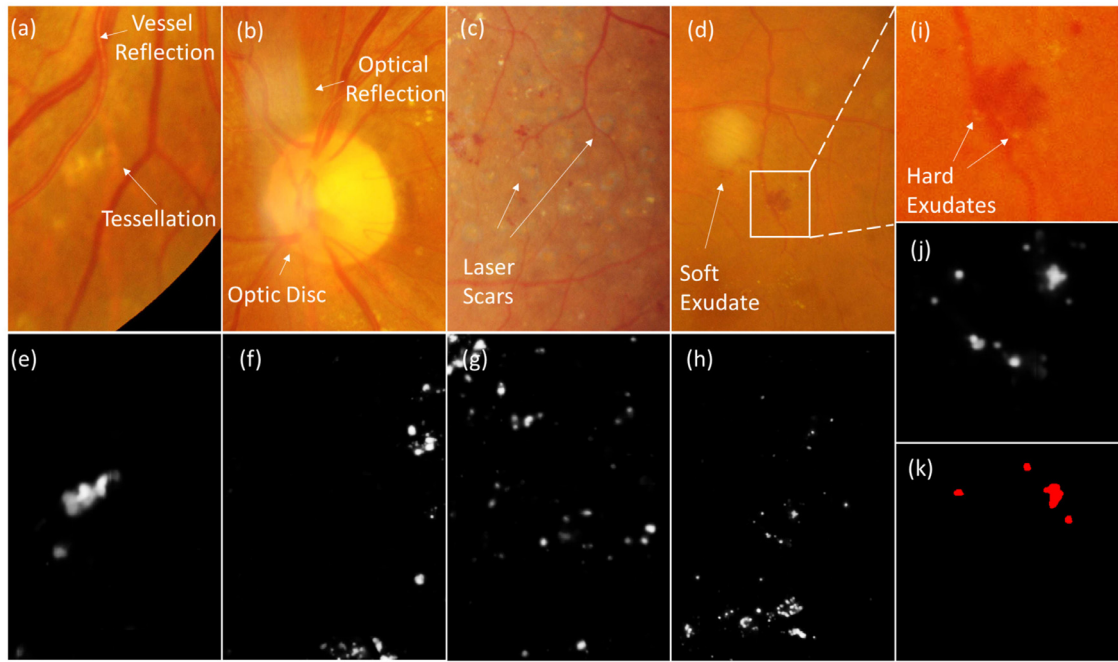


Fig. 11. Illustration of (a-d) different challenging circumstances for segmentation of EXs, (e-h) segmentation results (probability map) of the top-performing team for EXs, (i) enlarged part of Fig. (d), and (j) depicts its performance to be better than (k) the human annotator (The annotator tool had a limitation of the markup capability when there is an overlap of multiple types of lesion. In this case, EXs and HE).

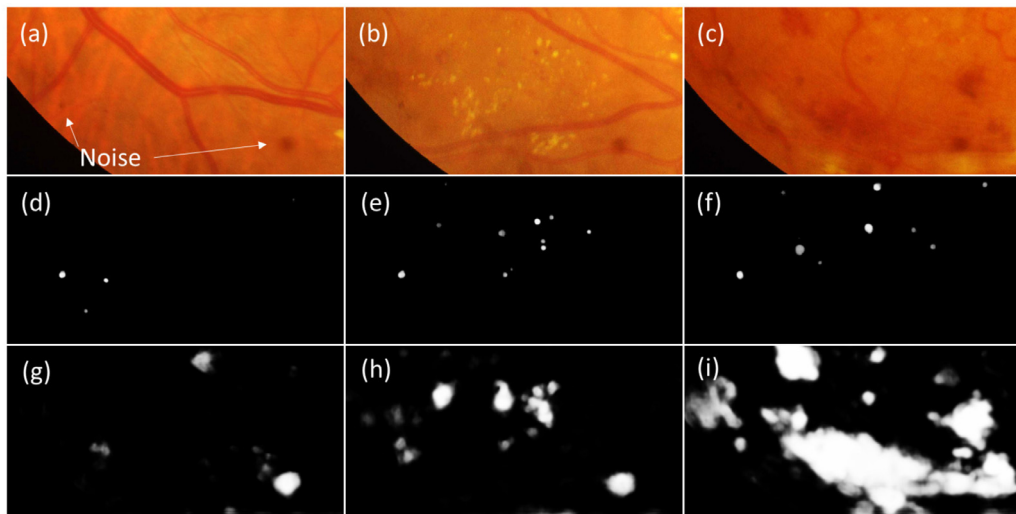


Fig. 12. Illustration of results by top performing solutions for (a-c) different images with noise causing most common false positives in the segmentation of (d-f) MAs, and (g-i) HEs respectively.

0.6311 as shown in Table 9. 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 object 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 training data is a crucial factor for a fair comparison of competing solutions. 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, along with the lesion information, it is essential to give attention towards characterization of intra-retinal microvascular abnormalities (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 segmentation. For the task of OD localization, the Euclidean distance varied between 21.072 and 36.22 (lower values indicate better performance). However, for Fovea localization task the same performance metric ranged between 64.492 and 570.133. This massive variation is due to outliers, e.g. team ZJU-BII-SGEX had 23 outliers

Table 12
Confusion matrix of retinal images predicted by top performing solution for DR (5 class) and DME (3 class).

		Predicted							Predicted			
Actual		0	1	2	3	4			0	1	2	
	0	30	0	2	1	1		0	40	2	3	
	1	3	1	1	0	0		1	5	2	3	
	2	3	2	22	4	1		2	5	2	41	
	3	2	0	1	13	3						
	4	1	0	1	0	11						

whose Euclidean distance exceeded 700. In the OD segmentation task, the average Jaccard similarity index score amongst the participants 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 significance 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 developed 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 models 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 reducing the original image size would prevent an accurate segmentation. Similarly, image resolution plays a very important role for disease classification task (see Table 9), the most likely reason is that presence of the disease is determined by the presence of lesions in the image, including the small ones that might be invisible at low resolution. This is corroborated by the confusion matrices in Table 12 which show misclassified instances in DR (particularly, grade 1 and 2) as well as DME (5 images each belonging to grade 1 and 2 are predicted as grade 0). For the localization 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 converted their predictions in the original image space. Comparative analysis indicates that the input image resolution has limited effect 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 approximate center location whereas the team SDNU utilized the input size of 1984 × 1318 pixels. Considering the OD average diameter of approximately 516 pixels, limited performance variation (10 to 15 pixels) is observed as compared to the top-performing solution for huge variation (multiple times) in input resolutions (see Table 10). 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., 2019), we hypothesized that algorithms developed using images with fine visibility and images having high resolution with ad-

judicated consensus grades yield better performance when compared 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 disagreement amongst the expert annotators. Even after these efforts to provide the best possible data, the annotation process is still inherently 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 allowed 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 solutions used deep convolutional networks. These approaches are comparably easier to implement than approaches based on feature engineering and do generalize well to multiple medical imaging domains, which in turn, dramatically reduces the need for specialized 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 superior performance. However, it seems there might be some impact 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 analyzing 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 assessment of participating solutions would have facilitated the submission procedure by providing real-time feedback to the teams performance. This would have enabled a maximum number of submissions 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 robust solutions for lesion segmentation, detection, and segmentation 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 human 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 problem 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.

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 FF450+	✓	TIFF
DIARETDB	130+89	1500 × 1152	50	Zeiss FF450+	✓	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 TRC – NW6	✓	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 TRV – 50	×	PPM
IDRiD	516 (81 with LA)	4288 × 2848	50	Kowa VX – 10α	✓	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 structures			Abnormalities				Multiple experts		DR grading	DME grading
	OD	VS	FA	MA	HE	EX	SE	Yes/No	#		
ARIA	✓	✓	✓	×	×	×	×	✓	2	×	×
DIARETDB1	×	×	×	✓	✓	✓	✓	✓	4	×	×
DRIVE	×	✓	×	×	×	×	×	✓	3	×	×
E-Ophtha	×	×	×	✓	×	✓	×	✓	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 Institute 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 Retinopathy 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 Sahasrabudhe (Government Medical College, Nanded) and Désiré Sidibé (Université de Bourgogne, 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 Promotion Agency \(NIPA-C0202-17-1045\)](#) in South Korea. *DeepDR*: This work was supported in part by the [National Natural Science Foundation of China](#) under Grant Grant 61872241, Grant 61572316, in part by the [National Key Research and Development Program of China](#) under Grant 2016YFC1300302 and Grant 2017YFE0104000, in part by the [Science and Technology Commission of Shanghai Municipality](#) under Grant 16DZ0501100 and Grant 17411952600. *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 Social Fund. *ZJU-BII-SGEX*: This work is supported by Beijing Shang-gong Medical Technology Co., Ltd., which provided ocular healthcare solutions in China. This research is partially supported by the A*STAR A1818g0022 grant of Singapore. Many thanks to the labeled 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 Development Fund through the Operational Programme for Competitiveness 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](#)) within project [CMUP-ERI/TIC/0028/2014](#). Teresa Araújo is funded by the FCT grant SFRH/BD/122365/2016. *SDNU*: This study was supported by the National Natural Science Foundation of China (Grant No. 61572300).

Appendix A. Comparison of Publicly Available Retinal Image Databases

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

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi: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.
- Abrahamoff, M.D., Garvin, M.K., Sonka, M., 2010. Retinal imaging and image analysis. *IEEE Rev. Biomed. Eng.* 3, 169–208.
- Abrahamoff, M.D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J.C., Niemeijer, M., 2016. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigat. Ophthalmol. Vis. Sci.* 57 (13), 5200–5206.
- Acharya, R., Chua, C.K., Ng, E., Yu, W., Chee, C., 2008. Application of higher order spectra for the identification of diabetes retinopathy stages. *J. Med. Syst.* 32 (6), 481–488.
- Acharya, U.R., Mookiah, M.R.K., Koh, J.E.W., Tan, J.H., Bhandary, S.V., Rao, A.K., Hagiwara, Y., Chua, C.K., Laude, A., 2017. Automated diabetic macular edema (DME) grading system using DWT, DCT features and maculopathy index. *Comput. Biol. Med.* 84, 59–68.
- Acharya, U.R., Ng, E.Y.-K., Tan, J.-H., Sree, S.V., Ng, K.-H., 2012. An integrated index for the identification of diabetic retinopathy stages using texture parameters. *J. Med. Syst.* 36 (3), 2011–2020.
- Adal, K.M., Sidibé, D., Ali, S., Chaum, E., Karnowski, T.P., Mériaudeau, F., 2014. Automated detection of microaneurysms using scale-adapted blob analysis and semi-supervised learning. *Comput. Method. Progr. Biomed.* 114 (1), 1–10.
- Agurto, C., Murray, V., Barriga, E., Murillo, S., Pattichis, M., Davis, H., Russell, S., Abrahamoff, M., Soliz, P., 2010. Multiscale AM-FM methods for diabetic retinopathy lesion detection. *IEEE Trans. Med. Imag.* 29 (2), 502–512.
- Almazroa, A., Alodhayb, S., Osman, E., Ramadan, E., Hummadi, M., Dlain, M., Alkatee, M., Raahemifar, K., Lakshminarayanan, V., 2018. Retinal fundus images for glaucoma analysis: the RIGA dataset. In: *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*, Vol. 10579. International Society for Optics and Photonics, p. 105790B.
- Antal, B., Hajdu, A., 2012. Improving microaneurysm detection using an optimally selected subset of candidate extractors and preprocessing methods. *Pattern Recognit.* 45 (1), 264–270.
- Antal, B., Hajdu, A., 2014. An ensemble-based system for automatic screening of diabetic retinopathy. *Knowl.-Based Syst.* 60, 20–27.
- 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 convolutional encoder-decoder architecture for image segmentation. *arXiv preprint arXiv:1511.00561*.
- Bai, J., Miri, M.S., Liu, Y., Saha, P., Garvin, M., Wu, X., 2014. Graph-based optimal multi-surface segmentation with a star-shaped prior: application to the segmentation of the optic disc and cup. In: *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, pp. 525–528.
- Bandello, F., Parodi, M.B., Lanzetta, P., Loewenstein, A., Massin, P., Menchini, F., Veritti, D., 2010. Diabetic macular edema. In: *Macular Edema*, Vol. 47. Karger Publishers, pp. 73–110.
- Biyani, R.S., Patre, B.M., 2018. Algorithms for red lesion detection in diabetic retinopathy: a review. *Biomed. Pharmacother.* 107, 681–688.
- Bourne, R.R.A., Stevens, G.A., White, R.A., Smith, J.L., Flaxman, S.R., Price, H., Jonas, J.B., Keeffe, J., Leasher, J., Naidoo, K., et al., 2013. Causes of vision loss worldwide, 1990–2010: a systematic analysis. *Lancet Global Health* 1 (6), e339–e349.
- Boyd, K., Eng, K.H., Page, C.D., 2013. Area under the precision-recall curve: point estimates and confidence intervals. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 451–466.
- Carin, L., Pencina, M.J., 2018. On deep learning for medical image analysis. *JAMA* 320 (11), 1192–1193.
- Carmona, E.J., Rincón, M., García-Feijó, J., Martínez-de-la Casa, J.M., 2008. Identification of the optic nerve head with genetic algorithms. *Artif. Intell. Med.* 43 (3), 243–259.
- Carson Lam, D.Y., Guo, M., Lindsey, T., 2018. Automated detection of diabetic retinopathy using deep learning. *AMIA Summit. Translat. Sci. Proc.* 2017, 147.
- Cheng, J., Yin, F., Wong, D.W.K., Tao, D., Liu, J., 2015. Sparse dissimilarity-constrained coding for glaucoma screening. *IEEE Trans. Biomed. Eng.* 62 (5), 1395–1403.
- Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.-M., Zietz, M., Hoffman, M.M., et al., 2018. Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* 15 (141), 20170387.
- Chudzik, P., Majumdar, S., Calivá, F., Al-Diri, B., Hunter, A., 2018. Microaneurysm detection using fully convolutional neural networks. *Comput. Method. Progr. Biomed.* 158, 185–192.
- Ciulla, T.A., Amador, A.G., Zinman, B., 2003. Diabetic retinopathy and diabetic macular edema: pathophysiology, screening, and novel therapies. *Diabetes Care* 26 (9), 2653–2664.
- Cuadros, J., Bresnick, G., 2009. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. *J. Diabetes Sci. Technol.* 3 (3), 509–516.
- Dai, L., Fang, R., Li, H., Hou, X., Sheng, B., Wu, Q., Jia, W., 2018. Clinical report guided retinal microaneurysm detection with multi-sieving deep learning. *IEEE Trans. Med. Imag.* 37 (5), 1149–1161.
- Das, V., Puhan, N.B., Panda, R., 2015. Entropy thresholding based microaneurysm detection in fundus images. In: *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*. IEEE, pp. 1–4.
- Dashtbozorg, B., Mendonça, A.M., Campilho, A., 2015. Optic disc segmentation using the sliding band filter. *Comput. Biol. Med.* 56, 1–12.
- Decenciére, E., Cazuguel, G., Zhang, X., Thibault, G., Klein, J.-C., Meyer, F., Marcotegui, B., Quéllec, G., Lamard, M., Danno, R., et al., 2013. Teleophtha: machine learning and image processing methods for teleophthalmology. *IRBM* 34 (2), 196–203.
- Decenciére, E., Zhang, X., Cazuguel, G., Laÿ, B., Cochener, B., Trone, C., Gain, P., Ordóñez Varela, J.-R., Massin, P., Erginay, A., et al., 2014. Feedback on a publicly distributed image database: the messidor database. *Image Anal. Stereol.* 33 (3), 231–234.
- Deepak, K.S., Sivaswamy, J., 2012. Automatic assessment of macular edema from color retinal images. *IEEE Trans. Med. Imag.* 31 (3), 766–776.
- Dhara, A.K., Mukhopadhyay, S., Bency, M.J., Rangayyan, R.M., Bansal, R., Gupta, A., 2015. Development of a screening tool for staging of diabetic retinopathy in fundus images. In: *Medical Imaging 2015: Computer-Aided Diagnosis*, Vol. 9414. International Society for Optics and Photonics, p. 94140H.
- Dobbin, K.K., Simon, R.M., 2011. Optimally splitting cases for training and testing high dimensional classifiers. *BMC Med. Genom.* 4 (1), 31.
- Esteves, T., Quelhas, P., Mendonça, A.M., Campilho, A., 2012. Gradient convergence filters and a phase congruency approach for in vivo cell nuclei detection. *Mach. Vis. Appl.* 23 (4), 623–638.
- Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The pascal visual object classes (voc) challenge. *Int. J. Comput. vis.* 88 (2), 303–338.
- Farnell, D.J.J., Hatfield, F.N., Knox, P., Reakes, M., Spencer, S., Parry, D., Harding, S.P., 2008. Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators. *J. Franklin Inst.* 345 (7), 748–765.
- Ferris, F.L., 1993. How effective are treatments for diabetic retinopathy? *JAMA* 269 (10), 1290–1291.
- Figueiredo, I.N., Kumar, S., Oliveira, C.M., Ramos, J.a.D., Engquist, B., 2015. Automated lesion detectors in retinal fundus images. *Comput. Biol. Med.* 66, 47–65.
- Fleming, A.D., Philip, S., Goatman, K.A., Olson, J.A., Sharp, P.F., 2006. Automated microaneurysm detection using local contrast normalization and local vessel detection. *IEEE Trans. Med. Imag.* 25 (9), 1223–1232.
- Fraz, M.M., Badar, M., Malik, A.W., Barman, S.A., 2018. Computational methods for exudates detection and macular edema estimation in retinal images: a survey. *Arch. Comput. Method. Eng.* 1–28.
- Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X., 2018. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *arXiv preprint arXiv:1801.00926*.
- Fu, H., Xu, Y., Wong, D.W.K., Liu, J., 2016. Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In: *Biomedical Imaging (ISBI)*, 2016 IEEE 13th International Symposium on. IEEE, pp. 698–701.
- García, G., Gallardo, J., Mauricio, A., López, J., Del Carpio, C., 2017. Detection of diabetic retinopathy based on a convolutional neural network using retinal fundus images. In: *International Conference on Artificial Neural Networks*. Springer, pp. 635–642.
- García, M., Sanchez, C.I., Poza, J., López, M.I., Hornero, R., 2009. Detection of hard exudates in retinal images using a radial basis function classifier. *Annal. Biomed. Eng.* 37 (7), 1448–1463.
- Gargeya, R., Leng, T., 2017. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 124 (7), 962–969.
- Gegundez-Arias, M.E., Marin, D., Bravo, J.M., Suero, A., 2013. Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques. *Comput. Med. Imag. Graph.* 37 (5–6), 386–393.
- Giachetti, A., Ballerini, L., Trucco, E., 2014. Accurate and reliable segmentation of the optic disc in digital fundus images. *J. Med. Imag.* 1 (2), 024001.
- Giancardo, L., Meriaudeau, F., Karnowski, T.P., Li, Y., Garg, S., Tobin, K.W., Chaum, E., 2012. Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Med. Image Anal.* 16 (1), 216–226.
- Giancardo, L., Meriaudeau, F., Karnowski, T.P., Li, Y., Tobin, K.W., Chaum, E., 2011. Microaneurysm detection with radon transform-based classification on retina images. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 5939–5942.
- Giancardo, L., Roberts, K., Zhao, Z., 2017. Representation learning for retinal vasculature embeddings. In: *Fetal, Infant and Ophthalmic Medical Image Analysis*. Springer, pp. 243–250.
- Glort, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pp. 249–256.
- Greenspan, H., Van Ginneken, B., Summers, R.M., 2016. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans. Med. Imag.* 35 (5), 1153–1159.
- van Grinsven, M.J., van Ginneken, B., Hoyng, C.B., Theelen, T., Sánchez, C.I., 2016. Fast convolutional neural network training using selective data sampling: application to hemorrhage detection in color fundus images. *IEEE Trans. Med. Imag.* 35 (5), 1273–1284.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahruday, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al., 2018. Recent advances in convolutional neural networks. *Pattern Recognit.* 77, 354–377.

- Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316 (22), 2402–2410.
- Harangi, B., Hajdu, A., 2014. Detection of exudates in fundus images using a markovian segmentation model. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 130–133.
- Hatanaka, Y., Nakagawa, T., Hayashi, Y., Kakogawa, M., Sawada, A., Kawase, K., Hara, T., Fujita, H., 2008. Improvement of automatic hemorrhage detection methods using brightness correction on fundus images. In: *Medical Imaging 2008: Computer-Aided Diagnosis*, Vol. 6915. International Society for Optics and Photonics, p. 69153E.
- Havlicek, J.P., 1996. *Am-Fm Image Models*. University of Texas at Austin.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn. In: *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, pp. 2980–2988.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
- Hinton, G., 2018. Deep learning technology with the potential to transform health care. *JAMA* 320 (11), 1101–1102.
- Hoo-Chang, S., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imag.* 35 (5), 1285.
- Hoover, A., 1975. *Stare Database*. Available. Available: <http://www.ces.clemson.edu/ahoover/stare>
- Howard, A.G., 2013. Some improvements on deep convolutional neural network based image classification. *arXiv preprint arXiv:1312.5402*.
- Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K., 2014. Densenet: implementing efficient convnet descriptor pyramids. *arXiv preprint arXiv:1404.1869*.
- ICO, 2017. Guidelines for diabetic eye care, 2nd edn. Int. Council Ophthalmol. (ICO).
- Jaccard, P., 1908. Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.* 44, 223–270.
- Javidi, M., Pourreza, H.-R., Harati, A., 2017. Vessel segmentation and microaneurysm detection using discriminative dictionary learning and sparse representation. *Comput. Method. Progr. Biomed.* 139, 93–108.
- Jelinek, H., Cree, M.J., 2009. Automated Image Detection of Retinal Pathology. *Crc Press*.
- Jonas, R.A., Wang, Y.X., Yang, H., Li, J.J., Xu, L., Panda-Jonas, S., Jonas, J.B., 2015. Optic disc-fovea distance, axial length and parapapillary zones. the Beijing eye study 2011. *PLoS One* 10 (9), e0138701.
- Jones, S., Edwards, R.T., 2010. Diabetic retinopathy screening: a systematic review of the economic evidence. *Diabet. Med.* 27 (3), 249–256.
- Jordan, K.C., Menolotto, M., Bolster, N.M., Livingstone, I.A.T., Giardini, M.E., 2017. A review of feature-based retinal image analysis. *Expert Rev. Ophthalmol.* 12 (3), 207–220.
- Joshi, S., Karule, P.T., 2019. Mathematical morphology for microaneurysm detection in fundus images. *Eur. J. Ophthalmol.* 1120672119843021
- Kamble, R., Kokare, M., Deshmukh, G., Hussin, F.A., Mériaudeau, F., 2017. Localization of optic disc and fovea in retinal images using intensity based line scanning analysis. *Comput. Biol. Med.* 87, 382–396.
- Kao, E.-F., Lin, P.-C., Chou, M.-C., Jaw, T.-S., Liu, G.-C., 2014. Automated detection of fovea in fundus images based on vessel-free zone and adaptive gaussian template. *Comput. Method. Progr. Biomed.* 117 (2), 92–103.
- Kauppi, T., Kamarainen, J.-K., Lensu, L., Kalesnykiene, V., Sorri, I., Uusitalo, H., Kälviäinen, H., 2012. A framework for constructing benchmark databases and protocols for retinopathy in medical image analysis. In: *International Conference on Intelligent Science and Intelligent Data Engineering*. Springer, pp. 832–843.
- Ker, J., Wang, L., Rao, J., Lim, T., 2018. Deep learning applications in medical image analysis. *IEEE Access* 6, 9375–9389.
- Khojasteh, P., Júnior, L.A.P., Carvalho, T., Rezende, E., Aliahmad, B., Papa, J.A.P., Kumar, D.K., 2018. Exudate detection in fundus images using deeply-learnable features. *Comput. Biol. Med.*
- Kim, J., Hong, J., Park, H., Kim, J., Hong, J., Park, H., 2018. Prospects of deep learning for medical imaging. *Precis. Future Med.* 2 (2), 37–52.
- Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kollias, A.N., Ulbig, M.W., 2010. Diabetic retinopathy: early diagnosis and effective treatment. *Deutsch. Arzteblatt Int.* 107 (5), 75.
- Krause, J., Gulshan, V., Rahimy, E., Karth, P., Widner, K., Corrado, G.S., Peng, L., Webster, D.R., 2018. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology* 125 (8), 1264–1272.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp. 1097–1105.
- Lam, C., Yu, C., Huang, L., Rubin, D., 2018. Retinal lesion detection with deep learning using image patches. *Invest. Ophthalmol. Vis. Sci.* 59 (1), 590–596.
- Li, H., Chutatape, O., 2004. Automated feature extraction in color retinal images by a model based approach. *IEEE Trans. Biomed. Eng.* 51 (2), 246–254.
- Lin, S., Ramulu, P., Lamoureux, E.L., Sabanayagam, C., 2016. Addressing risk factors, screening, and preventative treatment for diabetic retinopathy in developing countries: a review. *Clin. Exper. Ophthalmol.* 44 (4), 300–320.
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciampi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.
- Lynch, S.K., Shah, A., Folk, J.C., Wu, X., Abramoff, M.D., 2017. Catastrophic failure in image-based convolutional neural network algorithms for detecting diabetic retinopathy. *Investigat. Ophthalmol. Vis. Sci.* 58 (8), 3776–3776.
- Marin, D., Gegundez-Arias, M.E., Ponte, B., Alvarez, F., Garrido, J., Ortega, C., Vassallo, M.J., Bravo, J.M., 2018. An exudate detection method for diagnosis risk of diabetic macular edema in retinal images using feature-based and supervised classification. *Med. Biol. Eng. Comput.* 56 (8), 1379–1390.
- Marin, D., Gegundez-Arias, M.E., Suero, A., Bravo, J.M., 2015. Obtaining optic disc center and pixel region by automatic thresholding methods on morphologically processed fundus images. *Comput. Method. Progr. Biomed.* 118 (2), 173–185.
- Mary, M.C.V.S., Rajsingh, E.B., Jacob, J.K.K., Anandhi, D., Amato, U., Selvan, S.E., 2015. An empirical study on optic disc segmentation using an active contour model. *Biomed. Signal Process. Control* 18, 19–29.
- Medhi, J.P., Dandapat, S., 2014. Analysis of maculopathy in color fundus images. In: *2014 Annual IEEE India Conference (INDICON)*. IEEE, pp. 1–4.
- Mendonça, A.M., Sousa, A., Mendonça, L., Campilho, A., 2013. Automatic localization of the optic disc by combining vascular and intensity information. *Comput. Med. Imag. Graph.* 37 (5–6), 409–417.
- Mookiah, M.R.K., Acharya, U.R., Chandran, V., Martis, R.J., Tan, J.H., Koh, J.E.W., Chua, C.K., Tong, L., Laude, A., 2015. Application of higher-order spectra for automated grading of diabetic maculopathy. *Med. Biol. Eng. Comput.* 53 (12), 1319–1331.
- Mookiah, M.R.K., Acharya, U.R., Chua, C.K., Lim, C.M., Ng, E., Laude, A., 2013. Computer-aided diagnosis of diabetic retinopathy: a review. *Comput. Biol. Med.* 43 (12), 2136–2155.
- Mookiah, M.R.K., Acharya, U.R., Martis, R.J., Chua, C.K., Lim, C.M., Ng, E., Laude, A., 2013. Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: a hybrid feature extraction approach. *Knowl.-Based Syst.* 39, 9–22.
- Morales, S., Engan, K., Naranjo, V., Colomer, A., 2017. Retinal disease screening through local binary patterns. *IEEE J. Biomed. Health Informat.* 21 (1), 184–192.
- Morales, S., Naranjo, V., Angulo, J., Alcañiz, M., 2013. Automatic detection of optic disc based on PCA and mathematical morphology. *IEEE Trans. Med. Imag.* 32 (4), 786–796.
- Murphy, K.P., 2012. *Machine learning: a probabilistic perspective*. MIT press.
- Nagy, B., Harangi, B., Antal, B., Hajdu, A., 2011. Ensemble-based exudate detection in color fundus images. In: *Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on*. IEEE, pp. 700–703.
- Naqvi, S.A.G., Zafar, H.M.F., et al., 2018. Automated system for referral of cotton-wool spots. *Current Diabetes Rev.* 14 (2), 168–174.
- Niemeijer, M., Abramoff, M.D., Van Ginneken, B., 2009. Information fusion for diabetic retinopathy CAD in digital color fundus photographs. *IEEE Trans. Med. Imag.* 28 (5), 775–785.
- Niemeijer, M., Van Ginneken, B., Cree, M.J., Mizutani, A., Quéllec, G., Sánchez, C.I., Zhang, B., Hornero, R., Lamard, M., Muramatsu, C., et al., 2010. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE Trans. Med. Imag.* 29 (1), 185–195.
- Niemeijer, M., Van Ginneken, B., Staal, J., Suttrop-Schulten, M.S.A., Abramoff, M.D., 2005. Automatic detection of red lesions in digital color fundus photographs. *IEEE Trans. Med. Imag.* 24 (5), 584–592.
- Nørgaard, M.F., Grauslund, J., 2018. Automated screening for diabetic retinopathy—a systematic review. *Ophthalmic Res.*
- Orlando, J.L., Prokofyeva, E., del Fresno, M., Blaschko, M.B., 2018. An ensemble deep learning based approach for red lesion detection in fundus images. *Comput. Method. Progr. Biomed.* 153, 115–127.
- Osareh, A., Shadgar, B., Markham, R., 2009. A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images. *IEEE Trans. Inf. Technol. Biomed.* 13 (4), 535–545.
- Patton, N., Aslam, T.M., MacGillivray, T., Deary, I.J., Dhillon, B., Eikelboom, R.H., Yegesan, K., Constable, I.J., 2006. Retinal image analysis: concepts, applications and potential. *Progress Retinal Eye Res.* 25 (1), 99–127.
- Perdomo, O., Otalora, S., Rodríguez, F., Arevalo, J., González, F. A., 2016. A novel machine learning model based on exudate localization to detect diabetic macular edema.
- Pereira, C., Gonçalves, L., Ferreira, M., 2015. Exudate segmentation in fundus images using an ant colony optimization approach. *Inf. Sci.* 296, 14–24.
- Pereira, C.S., Mendonça, A.M., Campilho, A., 2007. Evaluation of contrast enhancement filters for lung nodule detection. In: *International Conference Image Analysis and Recognition*. Springer, pp. 878–888.
- Pires, R., Avila, S., Jelinek, H.F., Wainer, J., Valle, E., Rocha, A., 2017. Beyond lesion-based diabetic retinopathy: a direct approach for referral. *IEEE J. Biomed. Health Inform.* 21 (1), 193–200.
- Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabudde, V., Meriaudeau, F., 2018. Indian diabetic retinopathy image dataset (IDRID). *IEEE Dataport* doi:10.21227/H25W98.
- Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabudde, V., Meriaudeau, F., 2018. Indian diabetic retinopathy image dataset (IDRID): a database for diabetic retinopathy screening research. *Data* 3 (3/25). doi:10.3390/data3030025.

- Porwal, P., Pachade, S., Kokare, M., Giancardo, L., Mériaudeau, F., 2018. Retinal image analysis for disease screening through local tetra patterns. *Comput. Biol. Med.* 102, 200–210.
- Quelleg, G., Charrière, K., Boudi, Y., Cochener, B., Lamard, M., 2017. Deep image mining for diabetic retinopathy screening. *Med. Image Anal.* 39, 178–193.
- Quelleg, G., Lamard, M., Erginay, A., Chabouis, A., Massin, P., Cochener, B., Cazuguel, G., 2016. Automatic detection of referral patients due to retinal pathologies through data mining. *Med. Image Anal.* 29, 47–64.
- Quelleg, G., Lamard, M., Josselin, P.M., Cazuguel, G., Cochener, B., Roux, C., 2008. Optimal wavelet transform for the detection of microaneurysms in retina photographs. *IEEE Trans. Med. Imag.* 27 (9), 1230–1241.
- Raman, R., Gella, L., Srinivasan, S., Sharma, T., 2016. Diabetic retinopathy: an epidemic at home and around the world. *Indian J. Ophthalmol.* 64 (1), 69.
- Rangrej, S.B., Sivaswamy, J., 2017. Assistive lesion-emphasis system: an assistive system for fundus image readers. *J. Med. Imag.* 4 (2), 024503.
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., Yang, G.-Z., 2017. Deep learning for health informatics. *IEEE J. Biomed. Health Informat.* 21 (1), 4–21.
- Reichel, E., Salz, D., 2015. Diabetic retinopathy screening. In: *Managing Diabetic Eye Disease in Clinical Practice*. Springer, pp. 25–38.
- Rocha, A., Carvalho, T., Jelinek, H.F., Goldenstein, S., Wainer, J., 2012. Points of interest and visual dictionaries for automatic retinal lesion detection. *IEEE Transactions on biomedical engineering* 59 (8), 2244–2253.
- Romero-Oraá, R., Jiménez-García, J., García, M., López-Gálvez, M.I., Oraá-Pérez, J., Hornero, R., 2019. Entropy rate superpixel classification for automatic red lesion detection in fundus images. *Entropy* 21 (4), 417.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.
- Roychowdhury, S., Koozekanani, D.D., Parhi, K.K., 2014. Dream: diabetic retinopathy analysis using machine learning. *IEEE J. Biomed. Health Informat.* 18 (5), 1717–1728.
- Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 10 (3), e0118432.
- Sánchez, C.I., García, M., Mayo, A., López, M.I., Hornero, R., 2009. Retinal image analysis based on mixture models to detect hard exudates. *Med. Image Anal.* 13 (4), 650–658.
- Sánchez, C.I., Niemeijer, M., Išgum, I., Dumitrescu, A., Suttrop-Schulten, M.S.A., Abramoff, M.D., van Ginneken, B., 2012. Contextual computer-aided detection: improving bright lesion detection in retinal images and coronary calcification identification in CT scans. *Med. Image Anal.* 16 (1), 50–62.
- Seoud, L., Hurtut, T., Chelbi, J., Cheriet, F., Langlois, J.M.P., 2016. Red lesion detection using dynamic shape features for diabetic retinopathy screening. *IEEE Trans. Med. Imag.* 35 (4), 1116–1126.
- Shah, M.P., Merchant, S.N., Awate, S.P., 2018. Abnormality detection using deep neural networks with robust quasi-norm autoencoding and semi-supervised learning. In: *Biomedical Imaging (ISBI 2018)*, 2018 IEEE 15th International Symposium on. IEEE, pp. 568–572.
- Shen, D., Wu, G., Suk, H.-I., 2017. Deep learning in medical image analysis. *Annu. Rev. Eng.* 19, 221–248.
- Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874–1883.
- Shortliffe, E.H., Blois, M.S., 2006. The computer meets medicine and biology: emergence of a discipline. In: *Biomedical Informatics*. Springer, pp. 3–45.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sivaswamy, J., Krishnadas, S.R., Joshi, G.D., Jain, M., Tabish, A.U.S., 2014. Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation. In: *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, pp. 53–56.
- Son, J., Bae, W., Kim, S., Park, S.J., Jung, K.-H., 2018. Classification of findings with localized lesions in fundoscopic images using a regionally guided CNN. In: *Computational Pathology and Ophthalmic Medical Image Analysis*. Springer, pp. 176–184.
- Son, J., Park, S.J., Jung, K.-H., 2017. Retinal vessel segmentation in fundoscopic images with generative adversarial networks. *arXiv preprint arXiv:1706.09318*.
- Son, J., Shin, J.Y., Kim, H.D., Jung, K.-H., Park, K.H., Park, S.J., 2019. Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images. *Ophthalmology*.
- Sopaharak, A., Uyyanonvara, B., Barman, S., Williamson, T.H., 2008. Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Comput. Med. Imag. Graph.* 32 (8), 720–727.
- Sreng, S., Maneerat, N., Hamamoto, K., Panjaphongse, R., 2019. Cotton wool spots detection in diabetic retinopathy based on adaptive thresholding and ant colony optimization coupling support vector machine. *IEEE Trans. Electr. Electron. Eng.*
- Srivastava, R., Duan, L., Wong, D.W.K., Liu, J., Wong, T.Y., 2017. Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels. *Comput. Method. Progr. Biomed.* 138, 83–91.
- Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B., 2004. Ridge based vessel segmentation in color images of the retina. *IEEE Trans. Med. Imag.* 23 (4), 501–509.
- Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J., 2017. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, pp. 240–248.
- Suzuki, K., 2017. Overview of deep learning in medical imaging. *Radiol. Phys. Technol.* 10 (3), 257–273.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9.
- Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans. Med. Imag.* 35 (5), 1299–1312.
- Tan, J.H., Acharya, U.R., Bhandary, S.V., Chua, K.C., Sivaprasad, S., 2017. Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. *J. Comput. Sci.* 20, 70–79.
- Tan, J.H., Fujita, H., Sivaprasad, S., Bhandary, S.V., Rao, A.K., Chua, K.C., Acharya, U.R., 2017. Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. *Inf. Sci.* 420, 66–76.
- Tang, L., Niemeijer, M., Reinhardt, J.M., Garvin, M.K., Abramoff, M.D., 2013. Splat feature classification with application to retinal hemorrhage detection in fundus images. *IEEE Trans. Med. Imag.* 32 (2), 364–375.
- Thakur, N., Juneja, M., 2017. Clustering based approach for segmentation of optic cup and optic disc for detection of glaucoma. *Current Med. Imag. Rev.* 13 (1), 99–105.
- Ting, D.S.W., Cheung, G.C.M., Wong, T.Y., 2016. Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. *Clin. Exper. Ophthalmol.* 44 (4), 260–277.
- Trucco, E., Ruggeri, A., Karnowski, T., Giancardo, L., Chaum, E., Hubschman, J.P., Al-Diri, B., Cheung, C.Y., Wong, D., Abramoff, M., et al., 2013. Validating retinal fundus image analysis algorithms: issues and a proposal. *Invest. Ophthalmol. Visual Sci.* 54 (5), 3546–3559.
- Uribe-Valencia, L.J., Martínez-Carballido, J.F., 2019. Automated optic disc region location from fundus images: using local multi-level thresholding, best channel selection, and an intensity profile model. *Biomed. Signal Processing and Control* 51, 148–161.
- Voulodimos, A., Doulamis, N., Bebis, G., Stathaki, T., 2018. Recent Developments in deep learning for engineering applications. *Comput. Intell. Neurosci.* 2018.
- Walter, T., Klein, J.-C., Massin, P., Erginay, A., 2002. A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. *IEEE Trans. Med. Imag.* 21 (10), 1236–1243.
- Welfer, D., Scharcanski, J., Marinho, D.R., 2011. Fovea center detection based on the retina anatomy and mathematical morphology. *Comput. Method. Progr. Biomed.* 104 (3), 397–409.
- Winder, R.J., Morrow, P.J., McRitchie, I.N., Bailie, J.R., Hart, P.M., 2009. Algorithms for digital image processing in diabetic retinopathy. *Comput. Med. Imag. Graph.* 33 (8), 608–622.
- Wong, T.Y., Cheung, C.M.G., Larsen, M., Sharma, S., Simó, R., 2016. Diabetic retinopathy. *Nature Rev. Disease Primers*.
- Wu, B., Zhu, W., Shi, F., Zhu, S., Chen, X., 2017. Automatic detection of microaneurysms in retinal fundus images. *Comput. Med. Imag. Graph.* 55, 106–112.
- Wu, L., Fernandez-Loaiza, P., Sauma, J., Hernandez-Bogantes, E., Masís, M., 2013. Classification of diabetic retinopathy and diabetic macular edema. *World J. Diabetes* 4 (6), 290.
- Wu, X., Dai, B., Bu, W., 2016. Optic disc localization using directional models. *IEEE Trans. Image Process.* 25 (9), 4433–4442.
- Yu, F., Wang, D., Shelhamer, E., Darrell, T., 2017. Deep layer aggregation. *arXiv preprint arXiv:1707.06484*.
- Yu, H., Barriga, E.S., Agurto, C., Echegaray, S., Pattichis, M.S., Bauman, W., Soliz, P., 2012. Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets. *IEEE Trans. Inf. Technol. Biomed.* 16 (4), 644–657.
- Yun, W.L., Acharya, U.R., Venkatesh, Y.V., Chee, C., Min, L.C., Ng, E.Y.K., 2008. Identification of different stages of diabetic retinopathy using retinal optical images. *Inf. Sci.* 178 (1), 106–121.
- Zhang, B., Karray, F., Li, Q., Zhang, L., 2012. Sparse representation classifier for microaneurysm detection and retinal blood vessel extraction. *Inf. Sci.* 200, 78–90.
- Zhang, X., Thibault, G., Decencière, E., Marcotegui, B., Laÿ, B., Danno, R., Cazuguel, G., Quelleg, G., Lamard, M., Massin, P., et al., 2014. Exudate detection in color retinal images for mass screening of diabetic retinopathy. *Med. Image Anal.* 18 (7), 1026–1043.
- Zhou, W., Wu, C., Yi, Y., Du, W., 2017. Automatic detection of exudates in digital color fundus images using superpixel multi-feature classification. *IEEE Access* 5, 17077–17088.
- Zilly, J., Buhmann, J.M., Mahapatra, D., 2017. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. *Comput. Med. Imag. Graph.* 55, 28–41.