Medical Image Analysis 59 (2019) 101561

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

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 Li[i], Xiaodan Sui[j], Junyan Wu[l], Xiaolong Li[m], Ting Zhou[n], 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], Qinhao 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 Zhang[i], Aurélio Campilho[y,z], Bin Zheng[g], Dinggang Shen[k], Luca Giancardo[b,1], Gwenolé Quellec[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 Shanggong 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@sggs.ac.in (P. Porwal).

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

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

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.

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,

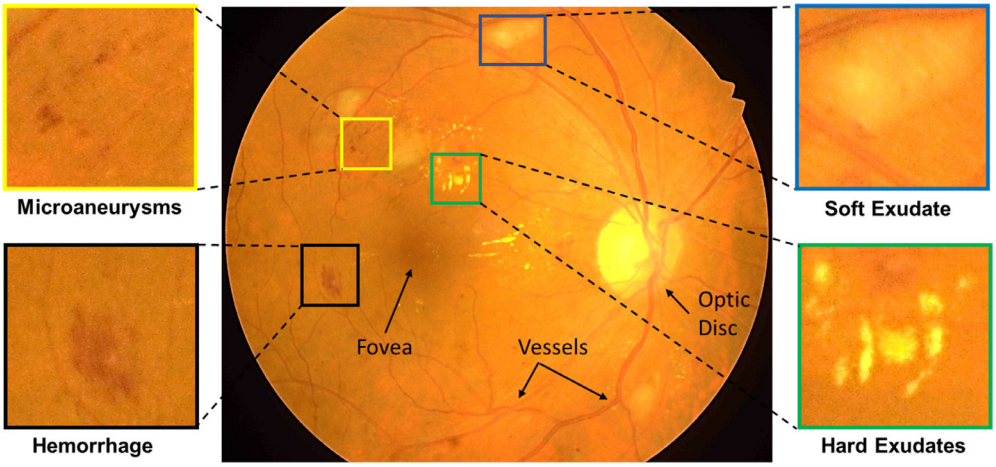

Fig. 1. Illustration of retinal image (in center) by highlighting normal structures (blood vessels, optic disc and fovea center) and abnormalities associated with DR: Enlarged regions (in left) MAs, and HEs and (in right) SEs, and EXs.

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 challenge[3]. 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 (Abràmoff 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 morphological 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; Ravıet 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.,