

Divide and Conquer Approach to Diabetic Retinopathy Classification

Sarthak Purohit¹, Ujjwal Kumar²

¹Student, Department of Computer Science and Engineering, SRM Institute of Science & Technology, Chennai, India

²Student, Department of Computer Science and Engineering, SRM Institute of Science & Technology, Chennai, India

Abstract - Medical image analysis using computer aided automation has been proven to be highly effective for classification of diseases like diabetic retinopathy. The present study reviews some of the recent AI driven solutions for diabetic retinopathy classification and gives a new solution to achieve state-of-the-art results on a retinal image dataset. A heuristic structure is used in the architecture of the proposed alternative that utilizes several binary classifiers to function as single multi-class classifier, thereby dividing the task amongst smaller groups. This new approach to diabetic retinopathy classification is described as 'Divide & Conquer Network' or 'D&C Network'. After evaluation, the D&C network exhibited promising results and achieved the area under the receiver operating characteristic curve of 0.906, with 0.965 specificity and 0.853 sensitivity.

Key Words: Medical image analysis, Computer-aided diagnosis, Convolutional neural networks, Transfer learning, Diabetic retinopathy classification, Deep learning,

1.INTRODUCTION

The complications that arise out of Diabetes Mellitus (DM) are unknown to none. World Health Organization reports that this chronic disease affects approximately 422 million people worldwide and is responsible for about 1.6 million deaths each year [1]. The most prevalent complication of DM is Diabetic Retinopathy (DR) [2] which is a vascular disease of the retina.

DR affects about 3.4% of the population [3] and in its advanced stages, it may even lead to complete vision loss. It is estimated that about 2.6% of worldwide blindness is caused by DR [4]. It is for these reasons that diabetic patients are prone to develop this disease and are in need for regular screening of their retina. In DR, the blood vessels that reach the retina swell up due to high blood sugar levels from DM and start leaking blood and other fluids. This results in the formation of various types of lesions that appear in the retinal fundus images. The various types of lesions can be categorized as microaneurysms (MA), haemorrhages (HM), hard exudates (HX) and soft exudates or cotton wool spots (SX). MA and HM are coined under the term red lesions while HX and SX are coined under bright lesions. MA appear as small red dots in the retina, and they eventually rupture to give rise to HM. HX are waxy yellow deposits while SX appear as white spots such as a cotton wool [3]. This can be seen in Figure 1. Based on the categorizations of different types of lesions, DR can be divided into five stages, namely:

No DR, Mild, Moderate, Severe and Proliferative. Proliferative DR here refers to neovascularization (abnormal growth of various blood vessels in the retina). DR without any abnormal growth vessels therefore fall under non-proliferative DR (Mild, Moderate and Severe). Table 1 gives a perspicuous overview of the different stages.

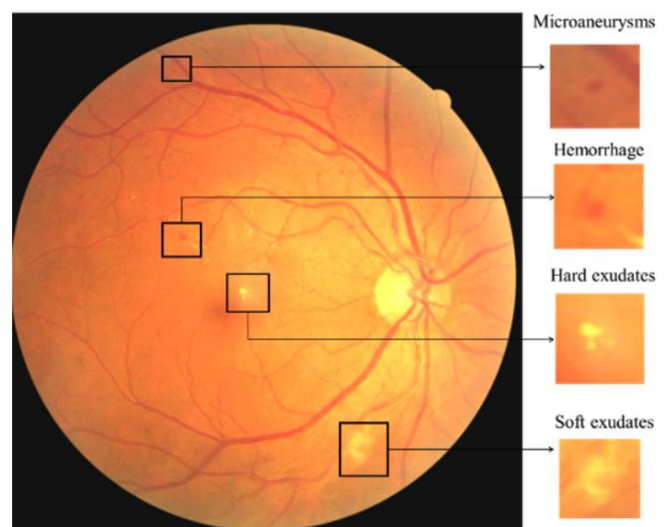


Fig -1: The various lesions of DR [5]

Despite knowing such intricacies of this disease, it is pivotal that DR be treated in its early stages and hence, its screening becomes a crucial task for ophthalmologists. Advancements in the field of computer aided diagnosis has made the job easier for several medical professionals in detecting various diseases including DR. Using automated techniques such as sophisticated pattern recognition algorithms for accurately predicting and classifying DR has numerous merits in terms of efficiency and speed. There should be no scope of error in the diagnostic process which makes manual screening of fundus images disputable as it demands considerable effort and is prone to misdiagnosis. Deep learning (DL) based methodologies have revolutionized several aspects of our life and medical image analysis is one of those. There has been considerable amount of work and research devoted for classifying DR from fundus images using DL as in [7] [8] [9] [10] [11]. DL outperforms other image processing techniques as it learns the features directly from the images in the training set and it does not require a lot of feature engineering, given that the model has several images to sample over [12]. These algorithms have shown high sensitivity and specificity for accurately classifying and

grading DR using retinal images. This paper describes the development of a new DL based architecture inherently using convolutional neural networks (CNN) to accurately predict various stages of DR. We are calling this new architecture as “D&C Network” (Divide and Conquer Network). The abstraction of using a CNN based model is because they have proven to be highly effective for medical image analysis [13] [14]. The D&C network is assisted through transfer learning [15] by fine-tuning the layers and parameters of ResNet [16], specifically for our task of DR grading. Our proposed architecture employs training and validating 4 separate CNN models, all sharing their outputs amongst themselves. Each CNN model in the D&C network serve as a binary classifier where one-vs-rest technique is used to convert a multi-class classification problem into several binary classification problems. This heuristic method involves developing the first model, which would predict from the input retinal image if it belonged to the DR class of {“No DR”} or either of {“Mild”, “Moderate”, “Severe”, “Proliferative”}. Similarly, the remaining 3 models would be designed with the last one being able to differentiate between the DR class of {“Severe”} or {“Proliferative”}. Here we are significantly dividing the task of DR grading by training a separate model for each class. This approach enhances each model to learn the visual representations of their corresponding class by ignoring to learn features of the other one, hence conquering the task of DR classification.

Table -1: Classification of various stages of DR [6]

Severity Level	Type of Lesion
No Diabetic Retinopathy	No abnormalities.
Mild Non-Proliferative Diabetic Retinopathy	Microaneurysms only.
Moderate Non-Proliferative Diabetic Retinopathy	More than microaneurysm but less than severe diabetic retinopathy.
Severe Non-Proliferative Diabetic Retinopathy	Any of the following: more than 20 intraretinal haemorrhages in each of 4 quadrants; definite venous beading in 2+ quadrants; prominent intraretinal microvascular abnormalities in 1+ quadrant and no signs of proliferative diabetic retinopathy.
Proliferative Diabetic Retinopathy	One or more of the following: neovascularization, vitreous/preretinal haemorrhage.

2. Related Works

Significant efforts have already been made in the domain of DR classification. This section is devoted to extensively review some of the research that already exists in order to gain a better perspective to where we currently stand. We have focused on reviewing works that essentially tried to detect DR by either using binary classifiers or multi-class classifiers.

2.1 Binary Classification

Here we attempt to list the work that have tried classifying retinal images into just two classes. The studies listed here does not include the five stages of DR. Esfahani et al. [17] used ResNet34 [18] to classify the retinal images into two classes of either normal images or images with DR. They used the Kaggle dataset [19] for training and testing their model. Gaussian filters and image normalization were some of the preprocessing techniques they used to enhance the image quality. Their work reported an accuracy of 85% with 86% sensitivity. K. Xu. et al. [20] used 1000 retinal images from the Kaggle dataset [19] to classify them into normal or DR images. They resized the images to 224*224 pixels with 3 color channels. Data augmentation techniques like shearing, translation, rotation and flipping were applied to increase the dataset size. A custom CNN architecture comprising of convolutional layers (CONV), pooling layers and fully connected layers were used in their study to achieve an accuracy of 94.5%. The work of Pires et al. [21] involved segmenting retinal images into referable (moderate DR or more) or non-referable (No DR or mild DR) images. They presented a novel CNN architecture with some resemblances to VGG-16 [22] and o_0 [23]. Their proposed architecture consisted of 16 weight layers with over 10 million parameters. The model training was done on the Kaggle dataset [19] and testing using the Messidor-2 [24] and DR2 datasets. They reported an area under the ROC curve of 98.2% when tested using Messidor-2.

2.2 Multi-Class Classification

In this section we analyze the studies that have attempted to use the retinal images to classify them into different stages of DR. Gulshan et al. [9] used the Messidor-2 [24] dataset and the EyePacs-1 dataset to detect DR and diabetic macular edema (DME). Their architecture included 10 pretrained Inception-v3 [25] networks. They reported 93% specificity on both the datasets and 96.1% and 97.5% sensitivity on the Messidor-2 and EyePacs-1 dataset respectively. The work of Alyoubi et al. [26] involved creating an automated system for detecting DR where they were able to classify retinal images into the five stages of DR and simultaneously localize lesions in the images. They used the APTOS 2019 Kaggle dataset [27] for classifying retinal images into different stages and the DDR dataset [28] for

lesion localization. They used three CNN architectures for the image classification part, of which two of them were designed from scratch: CNN512 and CNN229. The other model included fine-tuning the pretrained EfficientNetB0 [29] using transfer learning. To cater to the lesion localization problem, they used a modified YOLOv3 [30] model. At the end, they fused the CNN512 model and the modified YOLOv3 model to classify images into different stages of DR and localize the lesions in them too. Their proposed model obtained an accuracy of 89%, with 89% sensitivity and 97.3% specificity. Pratt et al. [31] used the Kaggle dataset [19] to classify the images into five stages of DR. Their approach involved building a custom CNN architecture made up of 10 CONV layers, eight max-pooling layers and three fully connected layers. They used preprocessing techniques like color normalization and image resizing for better results. To account for the generalization gap and overfitting, they used weight decay and dropout as regularization techniques. They reported an accuracy of 75%, with 95% specificity and 30% sensitivity. Shanthi et al. [32] used the Messidor dataset [24] to detect the five stages of DR. They used the pretrained AlexNet [33] architecture to achieve an accuracy of 96.35%.

3. Methods

In this section we give an overview of the dataset we have used for our model and the proposed architecture of the D&C network. Details about the preprocessing techniques and the hyperparameters used are also mentioned along this section.

3.1 Dataset: Preprocessing and Preparation

The open-source APTOS 2019 Kaggle dataset [27] was used for training and testing purpose of the D&C network. It consists of 3662 images in total which are split between training and testing sets. The dataset is divided between the five stages of DR, namely: 'No DR', 'Mild DR', 'Moderate DR', 'Severe DR' and 'Proliferative DR'. 1805 images belong to the first class of 'No DR', while the rest being 370, 999, 193 and 295 images respectively split amongst the 4 remaining classes in increasing levels of severity. It is an imbalanced dataset with unequal number of normal images. For the images in the dataset to be used as an input to the DL model, their quality needs to be enhanced so that feature extraction becomes easier for the model. Gaussian filter (a type of noise removal method) was applied to all the images in the dataset. Apart from this, data augmentation was used for significantly increasing the size of the training set. A set of operations like rotation, flipping, shearing and translation were applied to the images to achieve this. Color normalization was done on the images to normalize each channel of the RGB image. This balances the unnecessary variations in the pixel intensity of each image which makes it easier for the DL model to learn the visual representations.

For the preprocessed dataset to be used as an input to the D&C classification network, restructuring was needed. DR have 5 stages: Level 0 (No DR), Level 1 (Mild), Level 2 (Moderate), Level 3 (Severe) and Level 4 (Proliferative). A traditional multi-class classification expects the dataset images to be divided in the forementioned 5 separate directories. However, D&C classification expects the same dataset in a complex structure. Four directories are created:

- Directory 1 has two sub-directories. The first sub-directory has images of Level 0 DR, and the second sub-directory has images of Level 1, Level 2, Level 3 and Level 4 DR.
- Directory 2 has two sub-directories. The first sub-directory has images of Level 1 DR, and the second sub-directory has images of Level 2, Level 3 and Level 4 DR.
- Directory 3 has two sub-directories. The first sub-directory has images of Level 2 DR, and the second sub-directory has images of Level 3 and Level 4 DR.
- Directory 4 has two sub-directories. The first sub-directory has images of Level 3 DR, and the second sub-directory has images of Level 4 DR.

Figure 2 could be used as a reference to understand how images of different severity levels are split across directories.

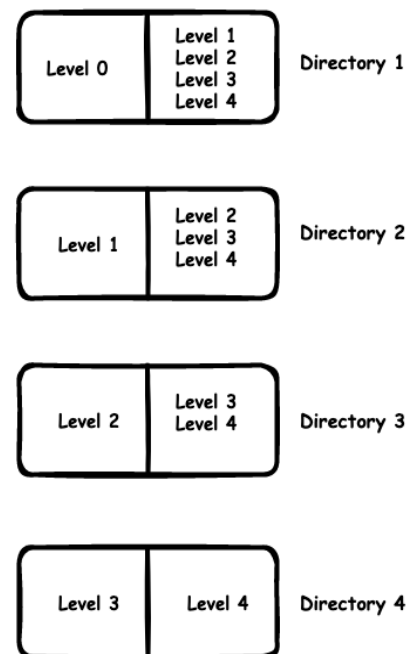


Fig -2: Organizing the dataset into different directories

3.2 Understanding the Model and the Architecture

The Divide and Conquer technique for DR classification is based on a heuristic method in which we utilize multiple binary classifiers together that function as a single multiclass classifier. The hypothesis being, dividing the classification task across several binary classifiers. In a traditional fully connected multiclass classification, it's sometimes evident that the information of inputs from few partial or biased classes (classes that have high correlation amongst themselves) disturb the classification process. It may seem miniature at first, but it could result in overall drop in performance. This could be avoided by extensively visualizing the data using heatmaps and checking for correlation. If we create separate classifiers for the indiscernible classes, we might make up for the lost performance. The dataset of DR is complex, few images of a particular severity level of DR resemble a lot with the images in the adjacent level. This resemblance in images may translate to erroneous gradients during back-propagation [34]. In multi-class classification, exploratory data analysis showed that a portion of images of a DR severity level were misclassified with the adjacent severity level as it was trained on the entire dataset. To solve this, separate classifiers are trained with a restructured dataset, such that each classifier have to binary classify a specific DR level. This somewhat decreases the erroneous optimization of parameters as each classifier is aimed to focus on just one severity level. Consequently, the task of classifying the inputs to n outputs could be sub-divided in up to n-1 binary classifiers, where each classifier handles only the specific set of classes it was trained on. Thus, putting some constraints on the confusing signals of input. Therefore, in the D&C network, a single input (retinal image) is passed sequentially through the binary classifiers. If a particular binary classifier recognizes the input with target label it was trained on, the input isn't further passed to other classifiers.

In this study, we have compared the performance of a traditional multiclass classification with our D&C network on DR images using transfer-learning. For a fair comparison, hyperparameters like loss function, optimizer, learning rate and batch size are kept the same. For the experiment using multi-class classification, a pre-trained ResNet152 is used for feature extraction. Its final fully connected linear layer is replaced with a custom linear layer having 5 output nodes (due to 5 output classes in DR). All the non-linear layers are disabled for training; thus, only the linear layer gets trained. In the D&C classification, a pre-trained ResNet152 network is used for same purpose of feature extraction with its final fully connected layer removed. A set of 4 multi-layer perceptrons (MLP) are added at the end of the ResNet152 model, each containing 2 output nodes, which completes the architecture of our 4 binary classifiers. All the 4 MLP networks are used in an if-else ladder formation. Each MLP network was trained for classifying a particular level of DR. In case if a MLP network returns false for classifying an

image with the target class of DR it was trained on, then the next MLP is used to check for the other target class and so on. In this way, the task of classifying the retinal images into different stages is divided across multiple models. Figure 3 represents this architecture in a comprehensible style. Both the models (multiclass and D&C network) are trained in the same traditional fashion: batches of images are computed by the network which compares true labels with the predicted labels.

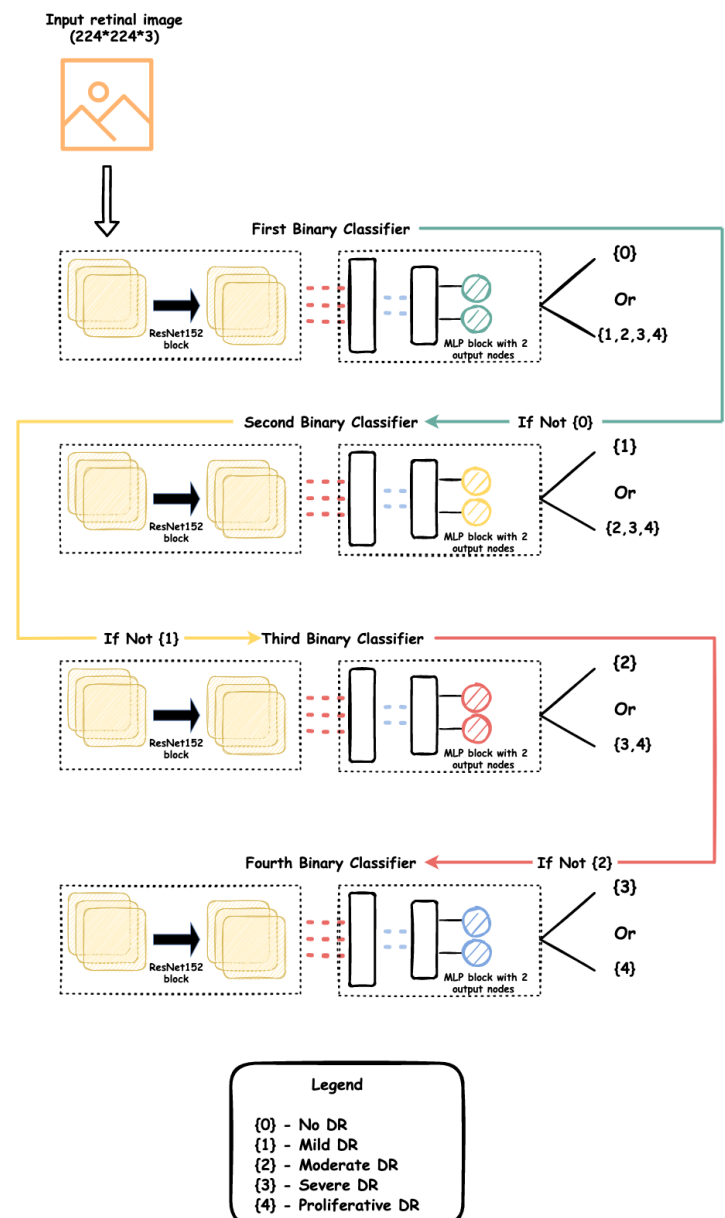


Fig -3: Proposed diagram of the D&C Classification Network

4. Results

In this section we will compare the performance of the D&C classification network with a traditional multi-class classifier in order to understand how our approach enhances the overall capabilities of the model. The quantitative performance of our experiment was based on some of the metrics that are customarily used for classification tasks. We have chosen to report the accuracy (ACC), precision (PR), recall/sensitivity (SEN), specificity (SP) and area under the receiver operating characteristic (ROC) curve (AUC) of both the D&C classification network and the multi-class classifier. Accuracy represents the percentage of images that were correctly classified by the model. Precision of the classifier denotes the accuracy of positive predictions that the model has made. Recall or sensitivity represents the true positive rate of the classifier or the ratio of positive images that are correctly detected by the model. Specificity represents the true negative rate. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1- specificity).

$$\text{Accuracy (ACC)} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}}$$

$$\text{Precision (PR)} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall or Sensitivity (SEN)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity (SP)} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

In the above equations, true positive (TP) refers to the number of DR images that are classified as DR by the model. True negative (TN) refers to the number of normal images that are classified as normal by the model. False positive (FP) refers to the number of normal images that are classified as DR by the model. False negative (FN) refers to the number of DR images that are classified as normal by the model. From our studies, it was found that the D&C classification network outperformed the traditional multi-class classifier by leaps and bounds. The D&C network recorded the highest accuracy of 0.854 on the APTOS 2019 dataset while the multi-class classifier reported a mere 0.7868. Figure 4 shows the variation of accuracy over the number of epochs of both the models. D&C network reported a precision of 0.8637, with 0.8536 sensitivity and 0.9656 specificity. For the D&C classification network, the area under the ROC curve was reported to be 0.9067 when compared to 0.8643 of the multi-class classifier. The confusion matrices of the two models used in this study are shown in figure 5 and figure 6. Table 2 shows the performance metrics of both the models against each other.

Table -2: Comparison of the performance metrics between the two models

Performance Metric	D&C Classifier	Multi-Class Classifier
Accuracy	0.8540	0.7868
Precision	0.8637	0.7844
Sensitivity	0.8536	0.7989
Specificity	0.9656	0.9135
Area Under the ROC Curve	0.9067	0.8643

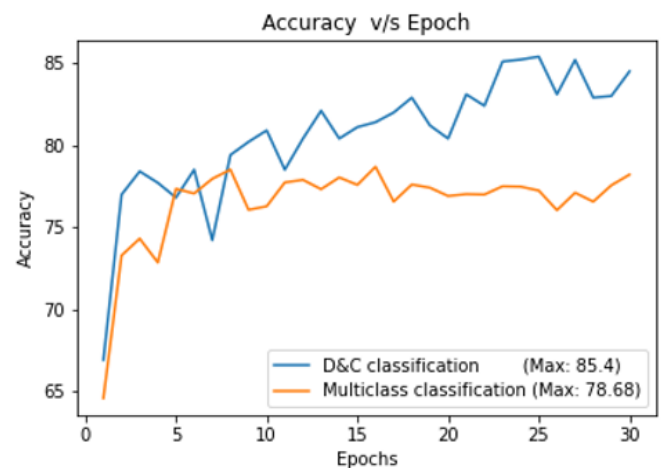


Fig -4: Plot of reported Accuracy v/s Epochs

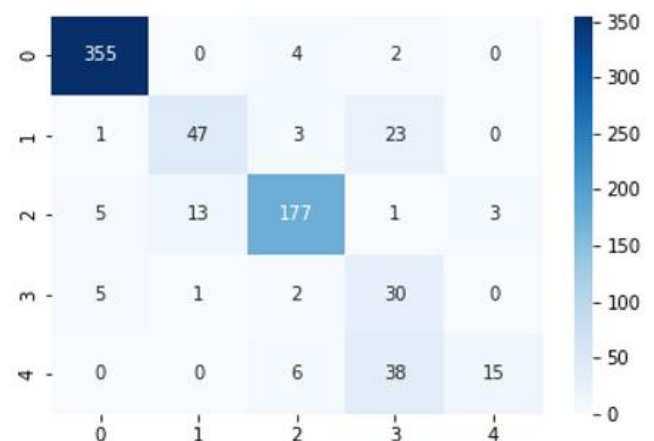


Fig -5: Confusion matrix of the D&C classification network as reported on the APTOS 2019 dataset

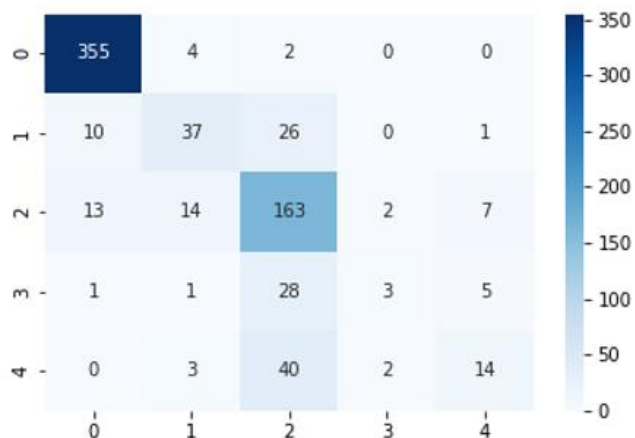


Fig -6: Confusion matrix of the multi-class classifier as reported on the APTOS 2019 dataset

5. CONCLUSION

Diabetic retinopathy is a prominent cause of blindness among adults suffering from diabetes mellitus. It progresses from a mild phase to moderate and then eventually to severe. This complication is a hazard to patient’s vision and should be diagnosed early and at the correct stage in order to decrease the threat of vision loss. Present automated systems excel at diagnosing diabetic retinopathy by making use of deep learning solutions. Advanced deep learning algorithms provide remarkable ability in diagnosing diseases accurately and the possibilities of improving architectures and algorithms always exists. This review has compared the performance of existing CNN architectures with our proposed alternative that resulted in an improved performance for diabetic retinopathy classification. We aim to use our proposed alternative in the future on other medical imaging tasks as well as look out for the possibility of an improved deep learning algorithm.

REFERENCES

[1] World Health Organization [Online]. Available at: <https://www.who.int/health-topics/diabetes>

[2] Marshall SM, Flyvbjerg A. Prevention and early detection of vascular complications of diabetes. *BMJ*. 2006 Sep 2;333(7566):475-80.

[3] Vislisel J, Oetting T. Diabetic Retinopathy: From One Medical Student to Another. University of Iowa Health Care, Ophthalmology and Visual Sciences. September 1, 2010

[4] Bourne, R.R.; 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. Causes of vision loss worldwide, 1990-2010: A systematic analysis. *Lancet Glob. Health* 2013, 1, 339-349.

[5] Alyoubi, W.L.; Abulkhair, M.F.; Shalash, W.M. Diabetic Retinopathy Fundus Image Classification and Lesions

Localization System Using Deep Learning. *Sensors* 2021, 21, 3704.

[6] Wilkinson, C.P.; Ferris, F.L.; Klein, R.E.; Lee, P.P.; Agardh, C.D.; Davis, M.; Dills, D.; Kampik, A.; Pararajasegaram, R.; Verdager, J.T.; Lum, F. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Am. Acad. Ophthalmol.* 2003, 110, 1677-1682.

[7] Abràmoff, M. et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest. Ophthalmol. Vis. Sci.* 57, 5200-5206 (2016).

[8] Gargeya, R. & Leng, T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology* 124, 962-969 (2017).

[9] Gulshan, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 2402-2410 (2016).

[10] Ting, D. S. W. et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 318, 2211-2223 (2017).

[11] van der Heijden, A. A. et al. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmol.* 96, 63-68 (2018).

[12] Deng, L.; Yu, D. Deep learning: Methods and applications. *Found. Trends Signal Process.* 2014, 7, 197-387.

[13] Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* 2017, 42, 60-88.

[14] Lu, L.; Zheng, Y.; Carneiro, G.; Yang, L. *Deep Learning and Convolutional Neural Networks for Medical Image Computing*; Springer: Berlin/Heidelberg, Germany, 2017.

[15] Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How transferable are features in deep neural networks? In *International Conference on Neural Information Processing Systems* 3320-3328 (NIPS, 2014).

[16] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 26 June-1 July 2016; pp. 770-778.

[17] Esfahani MT, Ghaderi M, Kafiyeh R. Classification of diabetic and normal fundus images using new deep learning method. *Leonardo Electron J Pract Technol* 2018; 17(32):233-48.

[18] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: *The thirty-first AAAI conference on artificial intelligence*; 2016. p. 4278-84.

[19] Kaggle dataset [Online]. Available at: <https://www.kaggle.com/c/diabetic-retinopathy-detection>

[20] Xu K, Feng D, Mi H. Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image. *Molecules* 2017;22(12): 2054.

- [21] Pires R, Avila S, Wainer J, Valle E, Abramoff MD, Rocha A. A data-driven approach to referable diabetic retinopathy detection. *Artif Intell Med.* 2019 May; 96:93-106.
- [22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations*; 2015.
- [23] Team o-O solution [Online]. Available at: <https://www.kaggle.com/c/diabetic-retinopathydetection/discussion/15617>
- [24] Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., Gain, P., Ordonez, R., Massin, P., Erginay, A., Charton, B., & Klein, J. (2014). FEEDBACK ON A PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE. *Image Analysis & Stereology*, 33(3), 231-234.
- [25] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *IEEE Conference on computer Vision and pattern recognition*; 2016. p. 2818–26.
- [26] Alyoubi WL, Abulkhair MF, Shalash WM. Diabetic Retinopathy Fundus Image Classification and Lesions Localization System Using Deep Learning. *Sensors.* 2021; 21(11):3704.
- [27] APTOS 2019 Blindness Detection [Online]. Available at: <https://www.kaggle.com/c/aptos2019-blindness-detection/overview/evaluation>
- [28] Li, T.; Gao, Y.; Wang, K.; Guo, S.; Liu, H.; Kang, H. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Inf. Sci.* 2019, 501, 511–522.
- [29] Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Beach, CA, USA, 10–15 June 2019*; pp. 6105–6114.
- [30] Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* 2018, arXiv:1804.02767.
- [31] Pratt H, Coenen F, Broadbent DM, Harding SP, Zheng Y. Convolutional neural networks for diabetic retinopathy. *Procedia Comput Sci* 2016; 90:200–5.
- [32] Shanthy T, Sabeenian RS. Modified Alexnet architecture for classification of diabetic retinopathy images. *Comput Electr Eng* 2019; 76:56–64.
- [33] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017;60(6).
- [34] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation", *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations.* MIT Press, pp. 318-362, 1986.