A Lightweight Deep Learning Approach for Diabetic Retinopathy Classification

Ruchika Bala[1](✉), Arun Sharma[1], and Nidhi Goel[2]

[1] Department of Information and Technology, Indira Gandhi Delhi Technical University for Women, Delhi, India
[2] Department of Electronics and Communication, Indira Gandhi Delhi Technical University for Women, Delhi, India

Abstract. In the present time, chances of suffering from diabetes have drastically increased due to the genetic probability, lack of physical activities, high blood pressure and modern lifestyle related problems. Diabetic Retinopathy (DR) is an intense problem which affects blood vessels in the eye retina. Early detection of DR can avoid severe eye damage. Several machine-learning and deep-learning based techniques have been used for DR detection and classification. However, these techniques are complex, time consuming, and take millions of parameters in training and deploying the DR classifier. In this paper, a lightweight dual-branch based CNN architecture is proposed for DR classification. The proposed architecture involves 84,645 (0.084 M) parameters for training and deploying the model. APTOS dataset has been used for analysis.

Keywords: Diabetic retinopathy · Classification · Grading · Fundus images · Deep learning · Transfer learning · CNN · Convolutional Neural Network

1 Introduction

Diabetic Retinopathy (DR) is caused due to excess of sugar levels in people across the world. Early screening or detection helps in avoiding severe eye problems. This is especially important in rural areas where people are not aware about early diagnosis. Diabetes affects our eyes and causes various lesions. Figure 1 represents the lesions including microaneurysms (MA), heamorrhages (HR), hard exudates (yellow, white spots), soft exudates (cotton wool spots), etc. In accordance with the presence of these symptoms in the eye retina, severity level of DR is measured in different grades. DR is primarily categorized into two types as non-proliferative DR (NPDR) and proliferative DR (PDR). Based on the outcome of Early Treatment of Diabetic Retinopathy Study (ETDRS), DR is labelled as no DR, mild DR, moderate DR, severe DR, and PDR [1].

As per the WHO global report on diabetes (2016), by the age of 70, there have been 43% of deaths happened due to high blood sugar level [2]. Since, the number of patients is considerably more than the number of eye specialists, it is highly required to automate the process of screening.

A. Dev et al. (Eds.): AIST 2021, CCIS 1546, pp. 277–287, 2022.
https://doi.org/10.1007/978-3-030-95711-7_25

Fig. 1. Diabetic retinopathy lesions

Ophthalmologists perform scanning of various types of images like photographic retinal fundus images, Fluorescein Angiograms (FA), and Optical Coherence Tomography (OCT) to identify pathologies and their severity [3]. Analysis of FA and OCT for classification is tedious and not preferred for clinical investigation. Analysis of blood vessel patterns in the photographic retinal fundus images is the most recommended means in identifying DR pathologies and their severity level. Initially, various mathematical operations are performed to enhance contrast, smooth or sharpen edges, remove noise, resize, crop, etc. [4]. Separation of foreground pixels from the background pixels i.e. segmentation techniques are applied to extract features. The extracted features include shape, width, size, edges, color, tortuosity, ridges, bifurcation, etc. [5]. The extracted features are input to the model to categorize fundus images into DR severity scales.

Convolutional Neural Network (CNN) has been widely utilized for DR classification. Murugan et al. [6] has proposed CNN based model for feature extraction and classification to identify abnormal fundus images. Labhsetwar et al. [7] has implemented CLAHE with VGG [8] and ResNet [9] architectures on different datasets for DR classification. There is a trade-off between computational cost and the performance metrics. Either, these models give very high accuracy or involve very high computational cost.

This paper proposes a lightweight CNN model for DR classification. The proposed architecture uses two interconnected branches with each convolution layer employing 32 filter units having 3×3 kernel size. At the end, two dense layers are incorporated with the model to perform binary DR classification. Data augmentation and dropout are used to avoid overfitting and produce generalize results. Experiments have been performed on Kaggle APTOS dataset including 3662 images. Precise comparative analysis indicates that the computational cost of the proposed framework is very low in terms of a smaller number of parameters for training and deploying the network.

The next section discusses the state-of-the-art work in the diabetic retinopathy domain. Section 3 and 4 discusses the proposed architecture and experimental analysis respectively. The last section discloses the conclusion.

2 Related Work

Currently, widely used approaches for classification are based on machine-learning and deep-learning techniques [10–13]. Machine learning approaches are based on extracting features and feeding to the algorithms like SVM, KNN, ANN, fuzzy C means, decision tree, random forest, adaboost, etc. for DR classification [14].

Deep learning is gaining great interest in pattern recognition and computer vision. With the wide use of deep learning in industries, its use is highly recommended in medical imaging such as DR classification, cancer detection, MRI scanning, etc. In transfer-learning, a pre-trained model is used for other similar visual recognition problem domains to share knowledge.

Gulshan et al. [15] has proposed has proposed a deep CNN framework based on Inception-V3 model for detecting referable DR and DME. Two sets have been formed for validation based on EyePACS-1 and Messidor-2. These sets have been validated by 7 US based ophthalmologists and achieved 99% accuracy. Referral recommendations have been provided for future clinical testing. However, the framework used 22 million parameters for binary classification and need to be tested in real-time clinical testing.

Further, Agarwal et al. [16] proposed DR classification based on Inception-v3 architecture with combination of preprocessing operations. To increase the size of dataset, augmentation techniques are utilized. With pretrained weights of Inception-v3 architecture, accuracy of 69% and specificity of 94% have been obtained. Sahlsten et al. [17] has also proposed an architecture based on Inception-v3 having combination of five classifiers for DR classification. Initially, images have been pre-processed by cropping to a square shape having circular fundus area. Then, resize operation is performed on each of the cropped images. Further, the effect of resolution has been studied in terms of performance measures of the model. Dekhil et al. [18] implemented VGG model on APTOS dataset and obtained 77% accuracy.

Tymchenko et al. [19] has implemented ensemble of pre-trained model with regression, classification, and ordinal regression together. Ensemble of two EfficientNet models B4 and B5 with SE-ResNeXt50 has been developed for DR classification of fundus images. The method obtained 99.1% in both recall and specificity. Orlando et al. [20] proposed extraction of intensity and shape based hand-crafted features for DR classification based on random forest ensemble technique and LeNet convolution network. Sridhar et al. [21] has implemented binary and multi-class classification using ResNet modules based on adaboost and ANN through ensemble techniques.

Instead of directly using the pre-trained model, Kassani has modified the Xception model by incorporating auxiliary features extracted from intermediate convolutional layers [22]. Modified Xception has obtained better performance than the Xception network. Further, Gong et al. [23] developed DR classification by extracting feature distractor interference from CNN learning and used them in backpropagation to cover the loss. Four deep CNNs (ResNet50 [9], VGGNet [8], InceptionNet V4 [24], and EfficientNet-B0 [25]) are considered as base models. Shaban et al. [26] developed deep convolution neural network for DR classification and obtained 0.92 kappa score. Most of these models are based on millions of pretrained parameters, which make them computationally expensive.

The state-of-the-art techniques involve trade-off between time complexity and model performance by using complex architectures with millions of model parameters. To overcome these issues, the proposed work presents a lightweight architecture which makes it simple and provides good evaluation measures.

3 Material and Methods

3.1 Dataset Description

The Experimental analysis is performed on APTOS (Asia Pacific Tele-Ophthalmology Society) dataset obtained from Kaggle [27]. APTOS dataset includes 3662 fundus images. All images are guassian filtered and resized to a standard size of 224 × 224. These images are arranged in respective folders based on the severity level of DR. The distribution of these images as per the severity level is shown in Fig. 2. There are 1805 images with no symptoms of DR, 370 images with mild DR, 999 images with moderate DR, 193 images with severe DR, and 295 images with PDR.

Augmentation procedures are employed to remove the data imbalancing and increase the size of dataset. During data augmentation step, all pixels are rescaled to (0, 1) to avoid differences in pixel intensities and boost the training process. Rescaled images have undergone several geometrical transformations such as rotation, width-shift, height-shift, shear, zoom, etc. to bring diversity in the dataset.

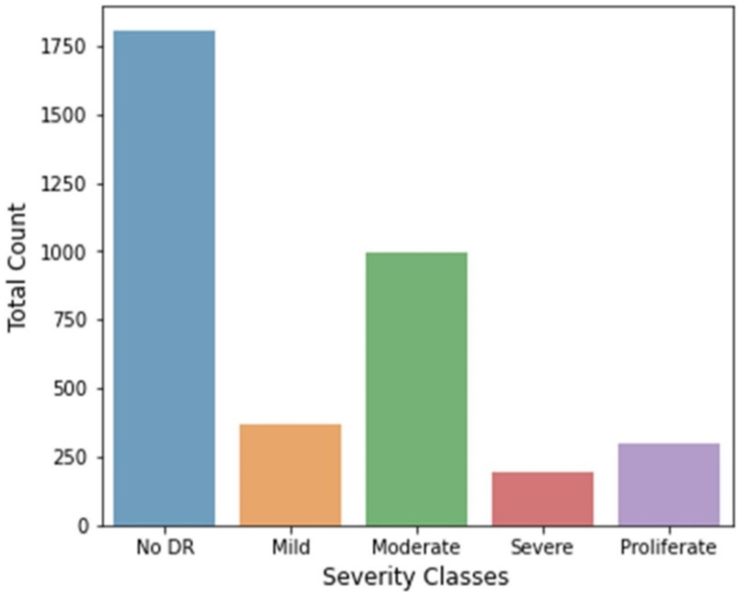

Fig. 2. Distribution of images

3.2 Proposed Architecture

DR classification is primarily based on identifying both lesions and their severity level. The proposed framework comprised of on-the-fly data augmentation followed by implementing a light-weight CNN model architecture to obtain pre-processed fundus images whereas the second step includes a light-weight CNN model architecture to output one of the multi-level DR classifications by extracting intrinsic features from the fundus images. Our model is significantly simpler than other techniques in term of number of parameters, and still achieves better results than those techniques. The broad-level block diagram of the proposed framework is shown in Fig. 3.

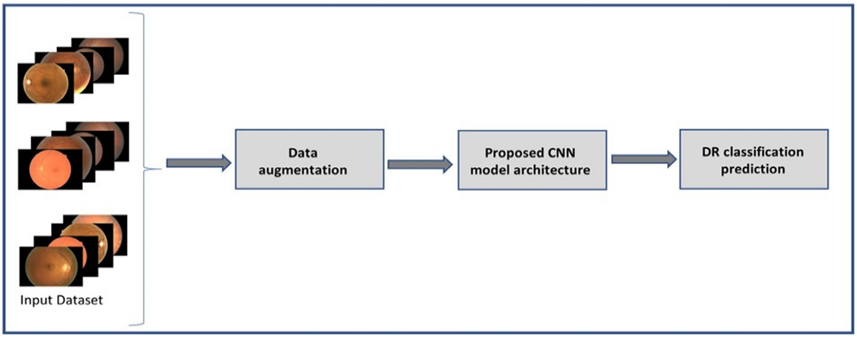

Fig. 3. Block diagram of proposed framework

The proposed Architecture performs DR screening of retinal fundus images. The proposed CNN architecture is a lightweight CNN-framework having two branches. The first branch includes a convolutional layer C11 followed by three blocks B11, B12, and B13, whereas the second branch consists of three blocks B21, B22, and B23. Each block in both branches consists of a convolutional layer followed by down-sampling using a max-pooling layer of size 2×2. Each convolutional layer uses 3×3 kernel size, and ReLU activation with 32 units. The structure of a block B is illustrated in Fig. 4.

Initially, pre-processed input images are given as input to the C11 layer of the first branch. Output of the C11 layer is fed to the block B11, output of the block B11 is fed to the block B12, and output of the block B12 is fed to the block B13. Further, concatenation of the output of the C11 layer and the pre-processed input images is given as input to the first block B21 of the second branch. Similarly, concatenation of output of the block B21 and the block B11 is fed to the block B22, and concatenation of output of the block B22 and the block B12 is fed to the block B23. Then output of the block B13 and the block B23 are merged together and flattened by feeding to global max pooling layer. At end, the output of convolution layer is fed to a dense layer of size 128 units, which is connected to an output dense layer of size 1 unit in the network to yield binary classification of DR. In DR screening, ‘0’ label indicates no signs of DR in the fundus images and ‘1’ indicates DR affected images. Both the branches are interconnected to extract different positional features of the fundus images, which is depicted in Fig. 5.

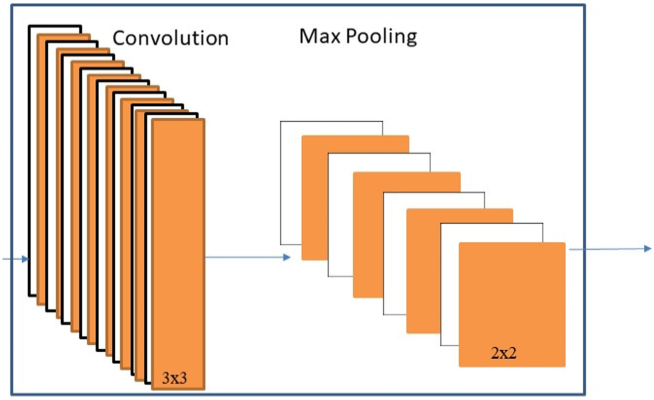

Fig. 4. Structure of blocks of CNN architecture

Dropout layer is used for regularization of the network [28]. Our proposed model uses 0.3 dropout, which means one in 3 neurons is randomly dropped or not used in training the network, and updating the weights in back-propagation to avoid overfitting.

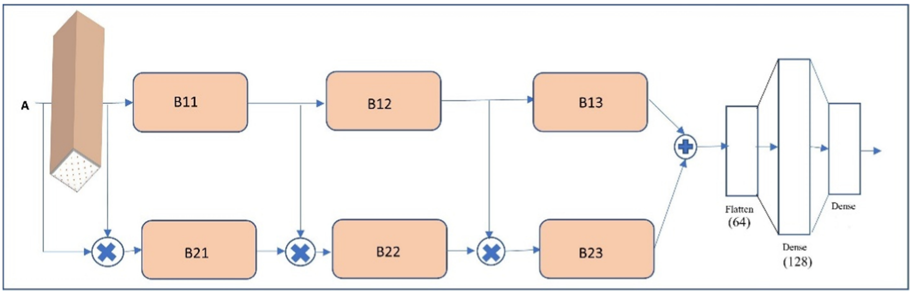

Fig. 5. Lightweight CNN model architecture.

3.3 Evaluation Metrics

In diabetic retinopathy, several retinal lesions are formed in the eye based on the severity of the DR stage. Since the lesions are of very small size, so small size kernels are consistently used to capture their presence. During the evaluation stage, the effectiveness of the proposed framework is measured by analyzing accuracy, sensitivity, specificity, recall, precision, and F1 score. The representation of these metrics is provided below.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Fscore = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

where TP denotes true positive that is predicted output is positive when the actual label is positive. FP denotes false positive that is predicted output is positive when the actual label is negative. TN denotes true negative that is predicted output is negative when the actual label is negative. FN denotes false negative that is predicted output is negative when the actual label is true.

4 Experimental Analysis

During experimentation, dataset is randomly divided into 3 sets in the ratio of 80:10:10 for training, validation and testing respectively before commencing the training. During the training stage, different parameters are set like 5 as a batch-size, and 0.0001 as a learning rate (lr).

Table 1. Performance metrics of the proposed frameworks

Metric (in %)	Proposed work
Training accuracy	92.50
Training loss	0.1655
Validation accuracy	98.75
Validation loss	0.072
Testing accuracy	92.50
Testing loss	0.1639
Precision	64
Recall	80
F-1 Score	71

The performance metric values are obtained on APTOS dataset images of size 224 × 224 [19]. The proposed framework is analyzed in terms of training- validation accuracy and loss measures. Figure 6 shows accuracy plots of training and validation datasets for DR screening. The training-validation accuracy plot shows training accuracy score values with best accuracy score of 0.9250, and validation-accuracy score of 0.9875 for binary classification. The loss plot shows loss score value with least training loss value of 0.1655 and least validation-loss score of 0.072.

Table 1 summarizes the obtained performance metrics of the proposed framework on test dataset. The obtained values for binary classification indicate very good training and testing accuracy of 0.9250. The recall, precision, and F1-score are 0.80, 0.64 and 0.71 respectively.

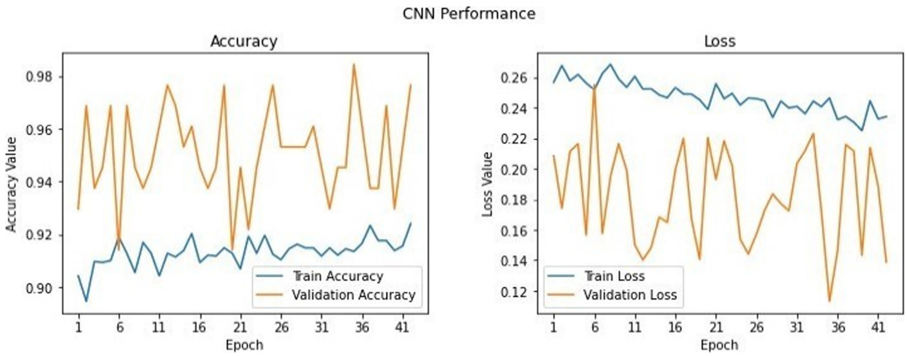

Fig. 6. Binary classification - Accuracy loss plots

5 Discussion

In this section, comparison of the proposed CNN architecture with some of the prior arts in the field of DR classification is presented. With respect to DR classification, the literature covering APTOS dataset for experimentation are selected for comparison in Table 2.

Agarwal et al. [16] proposed DR classification based on Inception-v3 architecture. At first, image segmentation is performed via combination of preprocessing operations. Augmentation techniques are utilized and Inception-v3 architecture is applied on the images to predict DR classification and obtains accuracy of 69% and specificity of 94%. The method of Agarwal et al. involves lot of preprocessing which makes their method more complex and tedious. Even though Inception-v3 uses a smaller number of parameters than other deep CNN architectures, but yet requires much more parameters than the proposed model. The present proposed model obtains accuracy of 92.50% with simple architecture and less computational complexity, which are the main attributes of our model.

Deep CNN architectures for binary and multi-class DR classification based on skip connections is implemented using ResNet model by Sridhar et al. [9, 21]. For binary classification, four ResNet models are built considering images having no signs of DR with each set of images selected from mild, moderate, severe and proliferate DR images. Output of these sub-models are combined together by Adaboost ensemble technique to predict final output. For multi-class classification, multiple ResNet models are built with all images together. The outputs are fed to the ANN for predicting the final grade. The entire architecture is based on ensemble techniques to boost the accuracy of the algorithm. Different sizes of filters such as 8×8, 16×16, and 32×32 and variable filter units such as 16, 64, 128, and 256 are used in different stages of the model. Based on these hyper-parameters, the overall number of parameters used in the model is roughly calculated as 0.13M. Binary classification architecture gives better results than multi-class classification. Use of large number of big filters, ResNet architecture, and application of ensemble techniques make the overall architecture complex and involves lot of parameters in training. However, the proposed method is a simple architecture, which uses 32 filters each of size 3x3 across the network and involves 0.084M parameters in total, which makes our approach simple and efficient.

Shaban et al. [26] proposed DR classification using deep convolution network having 18 CNN layers and 3 dense layers. They performed augmentation to remove data imbalancing. During training, 5-fold and 10-fold cross validation is performed with pre-trained weights of VGG. The method obtained 89% in validation accuracy and sensitivity, and 95% in specificity. The use of VGG pretrained weights contributes to 138 M parameters which makes the model heavy and complex.

Dekhil et al. [18] proposed a DR classifier based on transfer learning using VGG with preprocessing and obtained 77% accuracy score.

Overall, the precise comparison of various evaluation metrics of the proposed framework and prior arts is mentioned in Table 2.

Table 2. Comparative analysis.

Metric	Agarwal et al. [16]	Sridhar et al. [21]	Shaban et al. [26]	Dekhil et al. [18]	Proposed work
Accuracy	69	78.89	89	77	97.50
Precision	–	86	–	–	64
Recall	–	79	89	–	80
Specificity	94	–	95	–	–
F-1 score	–	79	–	–	71
Number of parameters	24M	0.013	138M	–	*0.084M*

6 Conclusion

Existing approaches of implementing DR classification are complex, time-consuming and require millions of parameters in training and deploying the model. In this paper, the proposed architecture presents a lightweight dual branch CNN framework for DR classification. The architecture involves two branches with each convolutional layer having 32 filters and 3x3 kernel. Both the branches are interconnected to extract different positional features. The proposed architecture is a lightweight model including 0.084 million parameters in training and deploying the network, which makes it simple and efficient. In future, research is required to confirm whether the proposed work can be applied in clinical settings. Research is also planned to train the proposed architecture further to improve the DR classification performance metrics.

References

1. Wilkinson, C., et al.: Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology **110**(5), 1677–1682 (2003)

2. Kamm, J.: Global report on diabetes, Technical Report, World Health Organization 330 (2016)
3. Wu, L., Fernandez-Loaiza, P., Sauma, J.: Classification of diabetic retinopathy and diabetic macular edema. World J. Diabetes **4**(6), 290–294 (2013)
4. Memari, N., Ramli, A.R., Saripan, M.I.B., Syamsiah, M., Moghbel, M.: Retinal blood vessel segmentation by using matched filtering and fuzzy c-means clustering with integrated level set method for diabetic retinopathy335assessment. J. Med. Biol. Eng. **39**(5), 713–731 (2019)
5. Panchal, P., Bhojani, R., Panchal, T.: An algorithm for retinal feature ex-traction using hybrid approach. In: 7th International Conference on Communication, Computing and Virtualization (ICCCV), vol. 99, pp. 61–68, Mumbai, India, 26–27 February 2016
6. Murugan, R., Roy, P., Singh, U.: An abnormality detection of retinal fundus images by deep convolutional neural networks. Multimedia Tools Appl. **79**(33–34), 24949–24967 (2020). https://doi.org/10.1007/s11042-020-09217-6
7. Labhsetwar, S.R., Salvi, R.S., Kolte, P.A., Venkatesh, V.S., Baretto, A.M.: Predictive analysis of diabetic retinopathy with transfer learning, arXiv arXiv:2011.04052.35016 (2020)
8. Simonyan, K., Zisserman, A.: Very deep convolutional networks for largescale image recognition. In: 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
10. Kaur, S., Goel, N.: A dilated convolutional approach for inflammatory lesion detection using multi-scale input feature fusion. In: 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM), pp. 386–393. IEEE (2020)
11. Kumari, R., Dev, A., Kumar, A.: An efficient adaptive artificial neural network based text to speech synthesizer for Hindi language. Multimedia Tools Appl. **80**(16), 24669–24695 (2021). https://doi.org/10.1007/s11042-021-10771-w
12. Jaiswal, G., Sharma, A., Yadav, S.K.: Critical insights into modern hyperspectral image applications through deep learning, Wiley Interdisc. Rev. Data Min. Knowl. Discovery **11**(6), e1426 (2021)
13. Carrera, E.V., Gonzalez, A., Carrera, R.: Automated detection of diabetic retinopathy using SVM. In: IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), Cuzco, Peru, 15–18 August 2017
14. Yu, S., Xiao, D., Kanagasingam, Y.: Machine learning based automatic neo- vascularization detection on optic disc region. IEEE J. Biomed Health In-365form **22**(3), 886–894 (2018). https://doi.org/10.1109/JBHI.2017.2710201
15. 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)
16. Agarwal, R., Mahamuni, A., Gautam, N., Awachar, P., Sagar, P.: Detection of diabetic retinopathy using convolutional neural network. Int. J. Recent Technol. Eng. (IJRTE) **8**(4), 1957–1960 (2019)
17. Sahlsten, J., et al.: Deep learning fundus image analysis for diabetic retinopathy and macular edema grading. Tech. Rep. Sci. Rep. **9**(10750), 405 (2019). https://doi.org/10.1038/s41598-019-47181-w
18. Dekhil, O., Naglah, A., Shaban, M., Ghazal, M., Taher, F., Elbaz, A.: Deep learning based method for computer aided diagnosis of diabetic retinopathy. In: 2019 IEEE International Conference on Imaging Systems and Techniques (IST), pp. 1–4 (2019). https://doi.org/10.1109/IST48021.2019.9010333
19. Tymchenko, B., Marchenko, P., Spodarets, D.: Deep learning approach to diabetic retinopathy detection, arXiv arXiv:2003.02261 (2020)

20. Orlando, J.I., Prokofyeva, E., Del Fresno, M., Blaschko, M.B.: An ensemble deep learning based approach for red lesion detection in fundus images. Comput. Methods Program. Biomed. **153**, 115–127 (2018). https://doi.org/10.1016/j.cmpb.2017.10.017. Epub 2017 October 14 PMID: 29157445
21. Sridhar, S., Sanagavarapu, S.: Detection and prognosis evaluation of diabetic retinopathy using ensemble deep convolutional neural networks. In: IEEE International Electronics Symposium (IES), Surabaya, Indonesia, pp. 78–85, 29–30 September 2020. https://doi.org/10.1109/IES50839.2020.9231789.420
22. Kassani, S.H., Kassani, P.H., Khazaeinezhad, R., Wesolowski, M.: Diabetic retinopathy classification using a modified Xception architecture. In: 9th IEEE International Symposium on Signal Processing and Information Technology, Ajman, United Arab Emirates, pp. 1–6, 10–12 December 2019. https://doi.org/10.1109/ISSPIT47144.2019.9001846.425
23. Gong, L., Ma, K., Zheng, Y.: Distractor-aware neuron intrinsic learning for generic 2D medical image classifications, In: Martel, A.L. et al. (ed.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, Lecture Notes in Computational Science and Engineering, vol.12262, pp. 591–601. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-957.430
24. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: 2016 Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA (2016)
25. Mingxing, T., Quoc, V. L.: Efficientnet: rethinking model scaling for convolutional neural networks. In: 2019 International Conference on Machine Learning, pp. 6105–6114 (2019)
26. Shaban, M., Ogur, Z., Mahmoud, A., Switala, A., Shalaby, A., Abu Khalifeh, H., et al.: A convolutional neural network for the screening and staging of diabetic retinopathy. PLoS ONE **15**(6), e0233514 (2020). https://doi.org/10.1371/journal.pone.0233514
27. Kaggle, APTOS dataset. https://www.kaggle.com/sovitrath/diabetic-retinopathy-224x224-gaussian-filtered
28. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdi-nov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**, 1929–1958 (2014)