The Scientific Temper (2023) Vol. 14 (2): 351-357

E-ISSN: 2231-6396, ISSN: 0976-8653

Doi: 10.58414/SCIENTIFICTEMPER.2023.14.2.17

https://scientifictemper.com/

RESEARCH ARTICLE

Assessment of transfer learning models for grading of diabetic retinopathy

Sowmiya M[1], Banu Rekha B[2] and Malar E[3]

Abstract

Diabetic retinopathy is a potentially mortal diabetic complication. The severity level of DR must be identified earlier to reduce the medical complications. Effective automated ways for identifying DR and classifying its severity stage are necessary to reduce the burden on ophthalmologists. Transfer learning methods are utilized to automatically grade the severity of diabetic retinopathy in this study. The stages of DR are diagnosed using pretrained VGG16, Inception v3, and ResNet50 models on pre-processed retinal images of DDR dataset. Out of three implemented models, Inception v3 achieved higher validation accuracy of 76.47% and testing accuracy of 67% compared to VGG16 and ResNet50 models. This research contributes to the analysis of deep learning architectures for the creation of automated diabetic retinopathy stage diagnosis and grading.

Keywords: Transfer learning, retinal image, diabetic retinopathy, VGG16, Inception v3, ResNet50.

Introduction

Diabetic retinopathy (DR), which primarily affects working people, is the leading cause of vision impairment and blindness. Diabetes causes diabetic retinopathy, which is caused by elevated blood sugar levels. This medical disorder causes vascular damage and aberrant blood vessels (Duh *et al.*, 2017) . The International Diabetes Federation predicts there will be 643 million diabetic people globally in 2030 and 783 million by 2045 (Ogurtsova *et al.*, 2017). Nonproliferative Diabetic Retinopathy (NPDR) and Proliferative Diabetic Retinopathy (PDR) are two stages of DR. NPDR manifests itself in the early phases, which are further divided into Mild, Moderate, and Severe. Microaneurysms (MA), dot and flame shaped haemorrhages, exudates, and cotton wool patches are all common NPDR abnormalities (Duh *et al.*, 2017). PDR is a more highly developed form of DR in which aberrant vessels begin to form and lead to neovascularization. In addition to lesions, vascular changes include vessel calibres, tortuosity, and branching angles (Habib *et al.*, 2014). The mild stage might progress to advanced PDR if an early diagnosis is not made. Early screening and treatment are required to limit retinal complications and avoid vision loss. Because manual grading needs skill and time, Computer Assisted Diagnosis (CAD) techniques have been demonstrated to be helpful in diagnosing the DR at an early phase. Table 1 shows the abnormal lesions appear in various stages of DR.

Several deep learning-based techniques for diagnosing DR using fundus images have been proposed. CNN models were created specifically to rate the severity degree (Alyoubi *et al.*, 2020). However, the DL model necessitates a vast amount of data. Because the current database has a restricted number of photographs. Transfer learning strategies have been highlighted in studies as a way to overcome restricted resources (Oltu *et al.*, 2021). The pre-trained CNN models were utilized specifically for assessing the severity level of DR. To diagnose the severity of DR, pretrained architectures Alexnet, VGG16, VGG19, Inceptionv3, ResNet were used (Oltu *et al.*, 2021) . The Kaggle EyePACS, Messidor, DIARETDB, and e-optha databases were used in the majority of the studies (Alyoubi *et al.*, 2020). The EyePACS dataset comprises

[1]Electronics and Communication Engineering PSG Institute of Technology and Applied Research Coimbatore, Tamil Nadu, India

[2]Biomedical Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India

[3]Electrical and Electronics Engineering, PSG Institute of Technology and, Applied Research, Coimbatore, Tamil Nadu, India

***Corresponding Author:** Sowmiya M, Electronics and Communication Engineering PSG Institute of Technology and Applied Research Coimbatore, Tamil Nadu, India, E-Mail: swm.muruganantham@gmail.com

How to cite this article: Sowmiya, M., Banu, R.B., Malar, E. (2023). Assessment of transfer learning models for grading of diabetic retinopathy. The Scientific Temper, **14**(2):351-357.

Doi: 10.58414/SCIENTIFICTEMPER.2023.14.2.17

Source of support: Nil

Conflict of interest: None.

Received: 01/04/2023 **Accepted:** 15/05/2023 **Published** : 30/06/2023

Table 1: Abnormal Lesions Present in Stages of Dr [1]

Stages of DR	*Retinal anomalies*
No DR	No anomalous signs
Mild NPDR	Few Microaneurysms
Moderate NPDR	Microaneurysms, hemorrhages, Exudates
Severe NPDR	venous beading, intraretinal microvascular abnormality
Proliferative DR	neovascularization, vitreous and preretinal hemorrhages

around 88,000 images, which is enough data for DL models. However, it has the drawback of including ungraded and low-quality photos (Alyoubi *et al.*, 2020) . In this study, a second large diabetic retinopathy dataset DDR is used, and the performance of transfer learning models is examined. The sample images for stages of DR taken from the DDR dataset are shown in Figure 1.

Employed DDR datasets, hence this work focuses on constructing transfer learning models for DDR datasets.

Pre-processing, selecting a pretrained architecture, and fine-tuning the architecture are the primary processes necessary for detecting DR using a transfer learning approach. Riaz *et al.* classified DR into five groups using the CNN architecture. The model was created with the help of the kaggle and messidor datasets. Using the EyePACS and Messidor-2 datasets, Gulshan *et al.* developed an Inception-v3 model to predict diabetic retinopathy**.** The model had a sensitivity of 90.3% for the EyePACS dataset and 87% for the Messidor-2 dataset (Riaz *et al.*, 2020). Data augmentation was used in conjunction with CNN to diagnose five categories of DR severity levels, with a sensitivity of 95% (Chlap *et al.*, 2021). On the Kaggle dataset, Kanungo *et al.* built the Inceptionv3 architecture for automated detection of various of diabetic retinopathy. For a batch size of 128, they were able to attain an accuracy of 88% (Kanungo *et al.*, 2017). Masood *et al.* used CNN transfer learning with Inceptionv3, which had been pretrained on ImageNet. They were able to attain 48.2% accuracy on the EyePACS five class dataset Masood *et al.*, 2017). For the diagnosis of diabetic retinopathy, Mansour *et al.* created AlexNet architecture with various optimization approaches (Mansour, 2018). On a dataset containing 30,000 photos, Gosh *et al.* used a CNN-based model to achieve an accuracy of 95 and 85% for two and five category diabetic retinopathy disorders (Ghosh *et al.*, 2017). Chandrakumar *et al.* devised the stagewise classification model, which includes dropout to reduce overfitting. The model was developed using the Kaggle dataset and tested using the DRIVE and STARE datasets (Chandrakumar & Kathirvel, 2016). Using the EyePACS dataset, Wang *et al.* created AlexNet, VGG16, and Inception-v3 models. The accuracy rates were 37.43, 50.3, and 63.23%, respectively (Wang *et al.*, 2018). Wan *et al.* applied pretrained VGG, AlexNet, ResNet, and GoogLeNet models to a dataset of 35,126 images from the

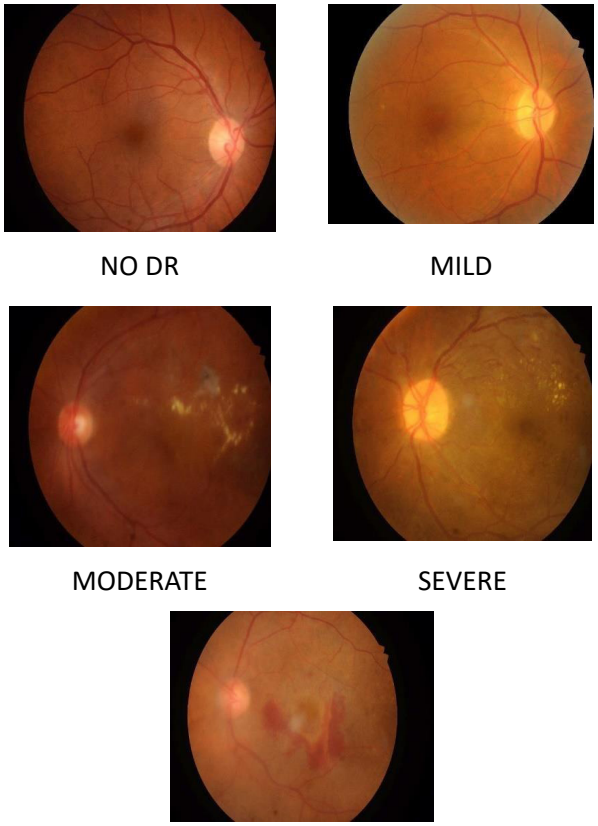

Figure 1: Stages of DR from DDR dataset

EyePACS and obtained accuracy of 89.75, 93.17, 93.73, 93.36, and 90.40%, respectively (Wan*et al.*, 2018).

According to the literature, transfer learning is extensively utilized to diagnose the DR severity level for EyePACS Kaggle and MESSIDOR datasets. Only a few research have

The contributions of this work are,

- This work develops the transfer learning pretrained models VGG16, Inception v3, ResNet50 for DDR dataset. From the literature, it is inferred that few studies have used this dataset.
- For grade five stages of DR, the DDR database will be used for training and validation. Performance criteria such as accuracy, sensitivity, precision, and F1 score will be examined for each architecture.
- This study evaluates the performance of three models for DDR dataset.

The paper is organized as follows; Section 2 presents the description of dataset and section 3 presents the methods and transfer learning models. Experimental results and discussions are presented in section 4 followed by a conclusion.

Dataset

The DDR dataset contains 13,637 fundus pictures showing evidence of DR with a resolution of 2124 x 2056 and a FOV

Table 2: Shows the Dr grading distribution

Class	*DR stage*	*Total*	*Training*	*Validation*	*Testing*
0	No DR	6266	3133	1253	1880
1	Mild	630	315	126	189
2	Moderate	4477	2238	895	1344
3	Severe	236	118	47	71
4	Proliferative	913	456	182	275
5	Ungradable	1151	575	230	346
	Total	13,673	6835	2733	4105

of 45 degrees. Images are collected between 2016 to 2018 in china from 147 hospitals. The level of DR severity includes No DR, mild NPDR, moderate NPDR, severe NPDR, and PDR. Images are graded by multiple graders. The dataset contains training, validation, and testing sets separately (Li *et al.*, 2019).

Methodology

The steps in this work are depicted in Figure 2. Pre-processing, training, and classification with performance metrics evaluation are the three stages of the transfer learning method given in this paper. The images were first resized and normalized as part of the pre-processing procedure as shown in Figure 3. The data augmentation was used to reduce overfitting and improve the model's generalization ability (Tariq *et al.*, 2021). Three CNN models were employed in the training stage: VGG16, Inception v3, and ResNet50, all of which were pre-trained using the ImageNet database. The last layers of each model were fine-tuned, and a dense layer was added for grading.

Transfer Learning Models

In this study, pre-trained CNN architectures, VGG16, Inception v3, and ResNet50, were trained and validated on the DDR database. For DR grading, fine-tune the architecture's final fully connected layers. The following is a model training procedure:

- Resize, normalize and augment the data in the training set as part of the pre-processing.
- Load the models that have been pre-trained.
- Configure the trained model's parameters, such as batch size, epochs, optimizer, and metrics.
- Save the model and run it.
- Evaluate the model's loss and other performance measures.

VGG16

VGG16 is a CNN model developed for the 2014 ImageNet Challenge by the Visual Geometry Group (Simonyan & Zisserman, 2014). There are 13 convolution layers, 5 pooling layers, and 3 fully connected layers in the VGG16 network. The image is resized to a size of 224x224. The model is pretrained using the ImageNet database and weights are

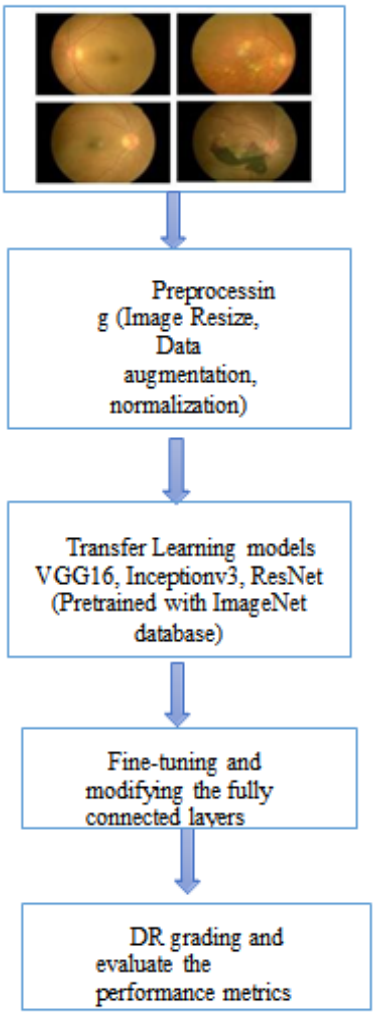

Figure 2: Framework for grading the DR severity level

utilized for DR classification. The layer weights are locked during the training phase, and one fully connected layer, and a dense layer with 6 neurons are added for grading DR. SoftMax function is utilized in dense layer with 6 neurons. The model s validated and tested on the respective DDR datasets. Figure 4 shows the pretrained VGG16 model with modified layers for classification. Adam optimizer was used to train for 25 epochs with a batch size of 32. The dropout was included to help prevent overfitting.

Pre-processing

DDR database contains fundus images of size 2124 x 2056 from five DR classes. First pre-process the images to a specific dimension based on the transfer learning models. The DDR database is class imbalanced, with 6266 samples present in no DR and 236 samples present in severe DR, which leads to biasing of the model. To reduce the model from overfitting, data augmentation was implemented. Rotation, flipping, brightness, shearing was included for creating augmented images. Image was normalized to make convergence faster.

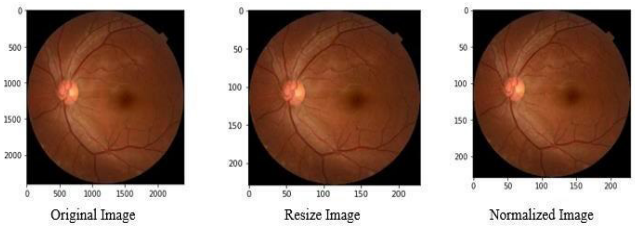

Figure 3: Sample images from pre-processing stage

Inception v3

The Inception v3 is an enhanced version of the Inception v1, which was first released as GoogLeNet in 2014. convolutional layers, pooling layers with an inception block, and fully connected layers are among the 48 layers in Inception v3. The Inception modules combine the characteristics of preceding levels to speed up calculation. Batch normalization and dropout of 0.5 were employed to the design to boost performance (Szegedy*et al.*, 2015). The model allows images with a resolution of 299x299. In this work, pretrained layers with fully connected layer and SoftMax layer are employed for classification.

ResNet50

ResNet50, also known as Residual Network, is a 50-layer residual network with skip connections that reduces the gradient problem (He *et al.*, 2015). The architecture begins with a 7x7 convolution layer, which is followed by a maxpooling layer. Four stages with three residual blocks are layered, each with three convolution layers of kernel size 1x1, 3x3, 1x1. Finally, DR is graded using a 0.5 dropout, fully connected layer, and SoftMax layer. Figures 5 and 6 illustrate the Inception v3 and ResNet50 architecture.

RESULTS AND DISCUSSIONS

Performance Metrics

Accuracy, sensitivity, precision, and F1 score are utilized as performance metrics to quantitatively evaluate the model presented in Table 4.

where, TP – True Positive; TN – True Negative; FP – False Positive; FN – False Negative.

The percentage of correctly categorized samples is called accuracy. Precision is the fraction of accurately classified samples among the classified positive samples. The fraction of samples accurately categorized from all positive samples is measured by recall.

Table 3: Hyperparameters used in the training phase

Metrics	*Formula*
Accuracy	$\frac{TN + TP}{TP + TN + FP + FN}$
Sensitivity (Recall)	$\frac{TP}{TP + FN}$
Precision	$\frac{TP}{TP + FP}$
F1 score	$\frac{Precision \times Recall}{^{2\times}Precision + Recall}$

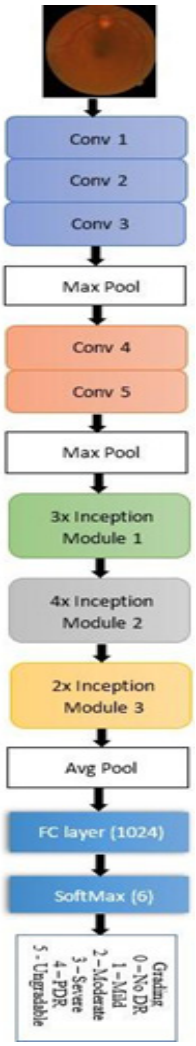

Figure 4: VGG16 architecture for DR Grading

Table 4: Performance evaluation metrics for dr grading

Parameters	*VGG16*	*Inception v3*	*ResNet50*
Input Image Size	224 x 224	299 x 299	224 x 224
Optimizer	Adam	Adam	Adam
Learning Rate	0.01	0.01	0.01
No. of Epoch	25	25	25
Batch size	32	32	32
Total number of Parameters	14,714,688	21,802,784	23,587,712

Results of Transfer Learning Models

This section contains the setup and outcomes for all of the pretrained models, as well as the loss gained during training and each model's class-wise performance. VGG16, Inception v3, and ResNet50 are the transfer learning architectures employed. For DR grading, each model employed pretrained weights with modified last layers. Six

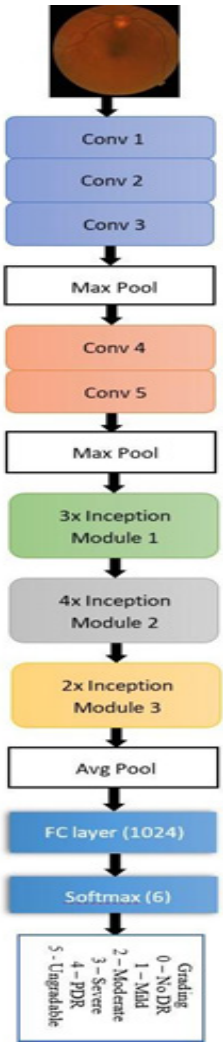

Figure 5: Inception v3 architecture for DR Grading

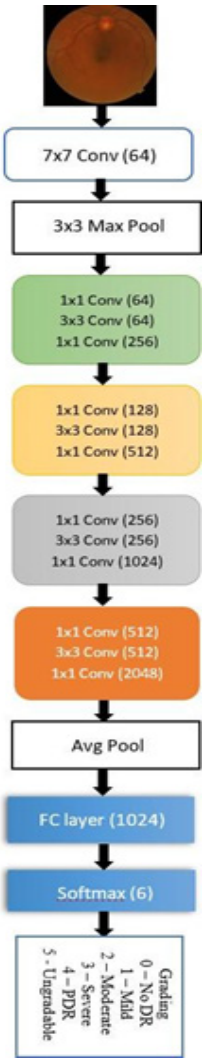

Figure 6: Res Net50 architecture for DR Grading

Table 5: Accuracy of transfer learning architectures

Model	*Training (%)*	*Validation (%)*	*Testing (%)*
VGG16	55.92	65.85	57
Inceptionv3	68.01	76.47	67
ResNet50	42	51.9	53.6

Table 6: Performance metrics of Vgg16 mode1

Class	*Precision*	*Recall*	*F1 score*
No DR (0)	0.56	0.97	0.71
Mild DR (1)	0.11	0.06	0.08
Moderate DR (2)	0.36	0.16	0.22
Severe DR (3)	0.1	0.02	0.01
Proliferative DR (4)	0.52	0.20	0.29
Ungradable DR (5)	0.80	0.38	0.52

class targets are added to the SoftMax layer. The pretrained models are fine-tuned using a 6835-image training dataset, validated using 2733 photos, and tested with 4105 images. The model's hyperparameters are optimized to maximize the results displayed in Table 3. The Adam optimizer with sparse categorical cross entropy loss is used for all of the architectures. Table 5 presents the accuracy of all the models. Training, validation, and testing accuracy have all been presented.

In comparison to the other two models, the inception v3 model gave better outcomes, as seen in Table 5. Furthermore, the training and validation accuracy have not been varied with a considerable variation, indicating that the overfitting problem has been reduced. Validation accuracy of 76.47% and testing accuracy of 67% were attained by Inception v3.

When trained on the training and validation datasets, the accuracy and loss performance of the three models is shown in Figure 7. Because the models were resistant to overfitting, they performed well during training and validation. Over each epoch, the losses in training and validation have decreased.

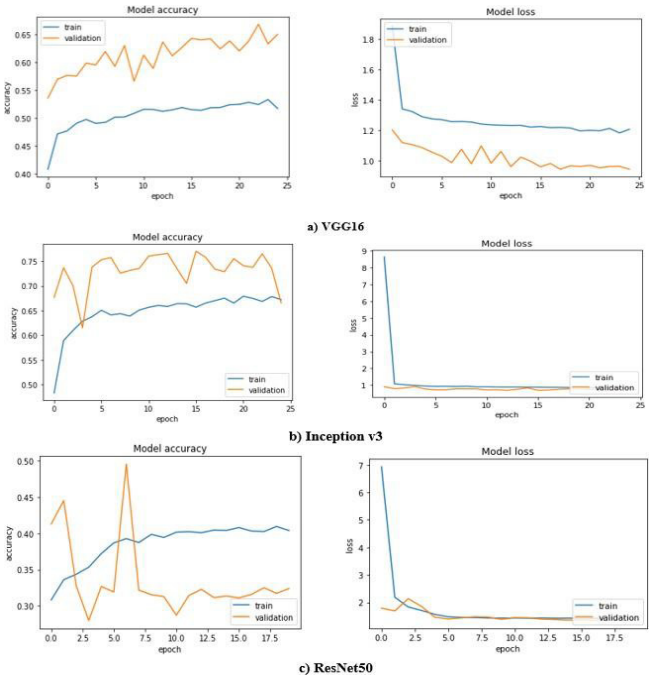

Figure 7: Training and validation loss

Table 7: Performance metrics of inception V3 mode1

Class	*Precision*	*Recall*	*F1 score*
No DR (0)	0.72	0.83	0.78
Mild DR (1)	0.1	0.01	0.02
Moderate DR (2)	0.54	0.57	0.55
Severe DR (3)	0.62	0.07	0.13
Proliferative DR (4)	0.74	0.31	0.43
Ungradable DR (5)	0.78	0.90	0.84

Table 8: Performance metrics of resnet50 mode1

Class	*Precision*	*Recall*	*F1 score*
No DR (0)	0.65	0.81	0.72
Mild DR (1)	0.1	0.00	0.00
Moderate DR (2)	0.39	0.47	0.43
Severe DR (3)	0.02	0.03	0.14
Proliferative DR (4)	0.01	0.01	0.01
Ungradable DR (5)	0.34	0.16	0.22

The metrics collected for each class for the DDR database are presented in Tables 6-8. Because of the enormous sample size, models perform better in the no DR and ungradable classes. It demonstrates that the models were successful in categorizing these two classes. However, all the models fail to account for mild DR; specifically, VGG16 have not identified the severe DR class. The achieved results shows that the models were identified each class with better performance. Of all three models, Inception v3 produced the highest validation accuracy of 76.47% and testing accuracy of 67%, respectively. Inception blocks produce better results for DR classification than VGG16 and ResNet50 architectures.

Conclusion

Transfer learning-based diagnosis models for grading diabetic retinopathy are presented in this research. Three pre-trained CNN models were used, and the final layers for multi-class grading were fine-tuned. Because the bulk of the class belongs to the no DR, there may be a lot of bias. Dropout and data augmentation were used to reduce bias and overfitting. The performance of VGG16, Inception v3, ResNet50 models have been analyzed for DDR dataset. From the results, the Inceptionv3 model achieved better performance with an accuracy of validation accuracy of 76.47% and testing accuracy of 67%. VGG16 produced a validation accuracy of 65.87% and testing accuracy of 57%. ResNet50 produced a validation accuracy of 51.9% and a testing accuracy of 53.6%. The research findings could be used to larger datasets in the future to increase the performance. Additionally, models will be trained from scratch and analyze its performance for clinical implementation.

References

Alyoubi, W. L., Shalash, W. M., & Abulkhair, M. F. (2020). Diabetic retinopathy detection through deep learning techniques: A review. Informatics in Medicine Unlocked, 20. https://doi.org/10.1016/j.imu.2020.100377

Chandrakumar, T., & Kathirvel, R. (June 2016). Classifying diabetic retinopathy using deep learning architecture. International Journal of Engineering Research and Technology, 5, 19–24.

Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. Journal of Medical Imaging and Radiation Oncology. Wiley, 65(5), 545–563. https://doi.org/10.1111/1754-9485.13261

Duh, E. J., Sun, J. K., & Stitt, A. W. (2017). Diabetic retinopathy: Current understanding, mechanisms, and treatment strategies. JCI Insight, 2(14). https://doi.org/10.1172/jci.insight.93751

Ghosh, R., Ghosh, K., Maitra, S., Maitra, S. (2017). Automatic Detection and Classification of Diabetic Retinopathy stages using CNN. In Proceedings of the 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)," Noida, 550–554. https://doi.org/10.1109/SPIN.2017.8050011

Habib, M. S., Al-Diri, B., Hunter, A., & Steel, D. H. (2014). The association between retinal vascular geometry changes and diabetic retinopathy and their role in prediction of progression – An exploratory study. BMC Ophthalmology, 14, 89. https://doi.org/10.1186/1471-2415-14-89

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. Cornell University. https://doi.org/10.1109/CVPR.2016.90

Kanungo, Y. S., Srinivasan, B., Choudhary, S. (2017). Detecting diabetic retinopathy using deep learning. In Proceedings of the 2nd IEEE International Conference on Recent Trends in Electronics, Information y Communication Technology," Bangalore, May 2017 (pp. 801–804). https://doi.org/10.1109/RTEICT.2017.8256708

Li, T., Gao, Y., Wang, K., Guo, S., Liu, H, & Kang, H. (2019). Diagnostic assessment of deep learning algorithms for diabetic

retinopathy screening. Information Sciences. Elsevier, 501, 511–522. https://doi.org/10.1016/j.ins.2019.06.011

Mansour, R. F. (2018). Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy. Biomedical Engineering Letters, 8(1), 41–57. https://doi.org/10.1007/s13534-017-0047-y

Masood, S., Luthra, T., Sundriyal, H., & Ahmed, M. (2017). Identification of diabetic retinopathy in eye images using transfer learning. In Proceedings of the 2017 International Conference on Computing, Communication and Automation (ICCCA)," Greater Noida, 1183–1187. https://doi.org/10.1109/CCAA.2017.8229977

Ogurtsova, K., da Rocha Fernandes, J. D., Huang, Y., Linnenkamp, U., Guariguata, L., Cho, N. H., Cavan, D., Shaw, J. E., & Makaroff, L. E. (2017). IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Research and Clinical Practice, 128, 40–50. https://doi.org/10.1016/j.diabres.2017.03.024

Oltu, B., Karaca, B. K., Erdem, H., & Ozgur, A. (2021). A systematic review of transfer learning based approaches for diabetic retinopathy detection. Cornell University.

Riaz, H., Park, Jisu, Choi, H., Kim, H., & Kim, J. (2020). Deep and densely connected networks for classification of diabetic retinopathy. Diagnostics, 10(1), 24. https://doi.org/10.3390/diagnostics10010024

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the inception architecture for computer vision. Cornell University. https://doi.org/10.1109/CVPR.2016.308

Tariq, H., Rashid, M., Javed, A., Zafar, E., Alotaibi, S. S., & Zia, M. Y. I. (2021). Performance analysis of deep-neural-network-based automatic diagnosis of diabetic retinopathy. Sensors, 22(1). https://doi.org/10.3390/s22010205

Wan, S., Liang, Y., & Zhang, Y. (2018). Deep convolutional neural networks for diabetic retinopathy detection by image classification. Computers and Electrical Engineering, 72, 274–282. https://doi.org/10.1016/j.compeleceng.2018.07.042

Wang, X., Lu, Y., Wang, Y., & Chen, W. B. (2018). Diabetic retinopathy stage classification using convolutional neural networks. In Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI)," Salt Lake City, 465–471. United States. https://doi.org/10.1109/IRI.2018.00074