Diabetic Retinopathy Diagnosis with InceptionResNetV2, Xception and EfficientNetB3

Mukkesh Ganesh[1], Sanjana Dulam[2], Pattabiraman Venkatasubbu[3]

[1, 2 & 3]Vellore Institute of Technology, Chennai, Tamil Nadu, India
[1]g.mukkesh2017@vitstudent.ac.in, [2]sanjana.dulam2017@vitstudent.ac.in,
[3]pattabiraman.v@vit.ac.in

Abstract. Diabetic retinopathy (DR) is a leading cause of visual impairments in adults. Early diagnosis of this disease is crucial for a chance at fast recovery. The advent of advanced deep learning techniques for computer-aided diagnosis has led to widespread adoption of the same in the medical field. Deeper and more powerful neural network architectures have emerged in the past few years consistently outperforming the previous generation models or offering similar levels of performance with an exponentially fewer number of parameters resulting in more efficient neural networks. Transfer learning has also taken the center stage since it introduced the concept of model reusability which led to faster convergence rates and made the training of deeper neural networks possible with limited data. This paper explores the feasibility of transfer learning with three specific architectures, namely Xception, InceptionResNetV2, and EfficientNetB3 pre-trained on the Imagenet problem, for the task of DR diagnosis. The Xception model achieved the highest accuracy of 84% on the binary classification of IDRiD, while the InceptionResNetV2 gave us the best result of 64% for 5-ary classification. EfficientNetB3, with significantly fewer parameters, was able to provide comparable results achieving an accuracy of 81% and 62% for 2-ary and 5-ary classifications respectively.

Keywords: EfficientNet, Xception, InceptionResNetV2, Deep Learning, Diabetic Retinopath.

1 Introduction

Diabetic Retinopathy (DR) is a diabetic condition that affects the eyes. Worldwide, one-third of the estimated diabetic population show signs of DR. Elevated sugar levels in the blood can lead to the blockage of blood vessels in the retina. This condition is termed as non-proliferative DR (NPDR) which could worsen to proliferative DR (PDR). If left untreated, scar tissues stimulated by the growth of new blood vessels may cause the retina to detach from the back of your eye which can cause complete blindness. Hence, early diagnosis is critical to the mitigation of this medical condition. Over the past decade, Deep Learning (DL) assisted diagnostic systems have risen in number and have outperformed the traditional image processing based systems. From detecting cancerous tumors in lungs and breast scans to the diagnosis of COVID-19 from CT scans, this technology has gained wide acceptance within the medical field. Rapid innovation in the Deep Learning based computer vision has giv-

en rise to powerful neural network architectures which have greatly enhanced the performance of these models. In much more recent years, researchers have started to use the power of transfer learning to make models converge faster and better for tasks that previously had limited training resources.

In this paper, we will be utilizing transfer learning to explore and compare the performance of state-of-the-art neural network architectures for diagnosing the severity of DR from retinal fundus images. For training and validating these models, we make use of the Kaggle dataset and test the performance of the same on the Messidor-2 dataset and IDRiD (Indian Diabetic Retinopathy Image Dataset).

2 Related Work

In [1], the authors used Convolutional Neural Networks (CNN) and achieved a validation sensitivity of 95% on the Messidor dataset. They also explored the use of transfer learning with GoogLeNet and AlexNet and were able to achieve accuracies of 74.5% on 2-ary, 68.8% on 3-ary and 57.2% on 4-ary classifications. In [2] Inception V3 architecture was pre-trained in various ways with a subset of the CHCF dataset and neural networks were compared and a few were able to achieve an accuracy of 88% test set. Using the Kaggle dataset, [3] were able to achieve an accuracy of 75% and sensitivity of 95% on 5000 validation images. The study in [4] compared the performance of a novel DL algorithm for the detection of DR with previous results on the benchmark Iowa Detection Program (IDP) dataset. By replacing IDP's feature-based lesion detectors with CNN based lesion detectors, the hybrid model achieved a sensitivity of 96.8% on the Messidor-2 dataset. The retrospective study of [5] analyzed posterior pole photographs of patients with diabetes. A randomly initialized GoogLeNet was trained on 95% of the photographs using manual modified Davis grading of three additional adjacent ones achieving an accuracy of 81%. AlexNet has been applied in [6] to enable an optimal DR computer-aided diagnosis (CAD) solution. The proposed system on the standard Kaggle fundus dataset, with LDA feature selection, reaches a classification accuracy of 97.93% and 95.26% with PCA. The Inception V4 model architecture in [7] was trained on a large dataset of more than 1.6 million retinal fundus images, and was then tested on 2000 images; it showed a 5-class accuracy of 88.4%. An accuracy of 72.5% with VGGnet was achieved in [8] and a weighted Fuzzy C-means algorithm was used to diagnose the severity of the disease. In [9], residual U-net architecture consistently outperformed the traditional non-residual U-net models in segmentation tasks. [10] extracts features such as blood vessels and hard exudates that were used to train a support vector machine (SVM), which achieved a maximum sensitivity of 94.6% on the Messidor dataset.

3 Dataset Description

The standard Kaggle dataset published with the consent of EyePACs in 2015 consists of 35126 images which we made use of for training the models. The dataset of high-

resolution retina images taken under varying conditions is labelled based on the severity of DR in each image on a scale ranging from 0 to 4, where 0 - No DR, 1 - Mild, 2 - Moderate, 3 – Severe, and 4 -Proliferative DR, being the most severe. The Messidor-2 dataset and IDRiD [11], which we used for evaluating the performance of our trained models, are significantly different from the Kaggle dataset. These variations help give us a realistic estimation of the models' performance since more often than not the images do not belong to the same distribution as the data on which the models were trained.

The Messidor-2 dataset is an aggregation of several DR eye evaluations, consisting of 1748 macula-centered eye fundus images. The 512 fundus images in IDRiD (Indian Diabetic Retinopathy Image Dataset) were captured by a retinal ophthalmologist in India. The Kaggle training data was split into 90% training and 10% validation, while the two remaining datasets were used to evaluate the models.

(a) No DR - Kaggle (b) Moderate DR - Messidor-2 (c) Proliferative DR - IDRiD

Fig. 1. Sample images with varying severity of DR from different datasets

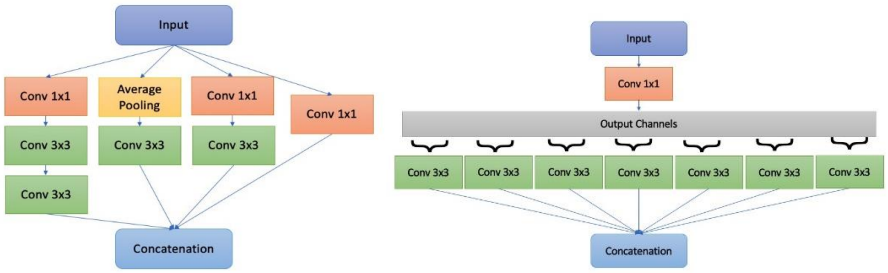
4 Model Architecture

CAD systems have started to increasingly utilize deep learning methodologies over traditional image processing workflows. The recent architectures developed by computer vision and deep learning researchers offer better performance while using lesser computing resources.

In this study, we focus on three major architectures, namely, Xception, Inception-ResnetV2 and EfficientNetB3, and compare their performance on the task of DR diagnosis. In trained models, the initial convolution kernels have been shown to detect basic patterns such as edges. By initializing the weights of these network layers to values learnt from other tasks such as ImageNet and CoCo image classification challenges, the models can converge faster or sometimes even perform better with having limited data at hand. This process known as transfer learning is leveraged in this paper in an attempt to achieve better results while training these architectures for this specific task.

4.1 Xception

Xception (extreme inception) is an architecture proposed by Google as an improvement over its Inception V3 architecture. The original Inception architecture used depthwise convolution followed by a 1x1 convolution to modify output dimension. Depthwise convolution involved channel-wise NxN spatial convolution. Depthwise separable convolution on the other hand made use of 1x1 convolution before performing the NxN channel-wise spatial convolutions. This was shown to yield better results than the vanilla Inception model.



(a) InceptionV3 (b) Xception

Fig. 2. Differences in Inception and Xception architectures

4.2 InceptionResNetV2

Deeper neural networks often take longer to train and may fail to converge due to vanishing gradients. To mitigate these issues, Microsoft introduced the Residual Neural Network architecture, which had skip connections between convolution blocks, which solves the vanishing gradient problem caused by the deeper architectures while also greatly speeding up the training process. InceptionResnetV2 takes this concept of skip/residual connections from ResNets and applies it to the Inception architecture thereby enhancing the performance of the model.

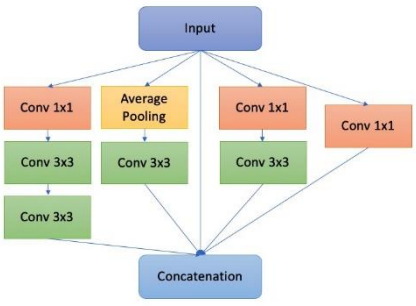


Fig. 3. InceptionResNetV2 Architecture

4.3 EfficientNetB3

The traditional practice for model scaling is to randomly increase the model depth or width or to use greater resolutions of input images for training and evaluation. This results in tedious manual fine-tuning and longer training times. Introduced by Google, EfficientNets are scaled in a more principled manner. The resulting architecture makes use of mobile inverted bottleneck convolution. Each stem of the 8 Efficient-Nets contains 7 blocks, each further consisting of an increasing number of subblocks from EfficientNet B0 to B7.

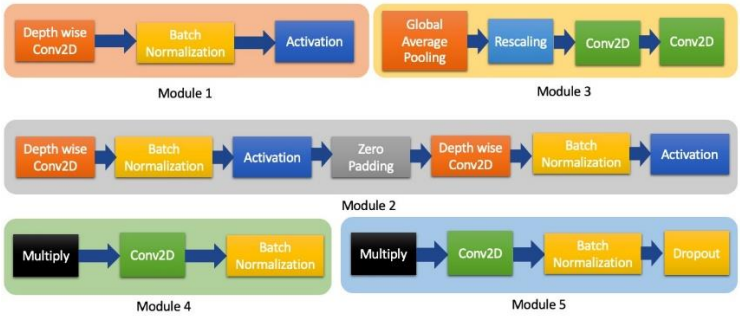


Fig. 4. Modules in EfficientNets

5 Implementation

Our study was conducted using the Kaggle environment. The Tensorflow framework was chosen to develop these deep learning models. The preprocessing stage consisted of resizing the images to 299x299. Tensorflow's image data generator class was used to augment our dataset by using a few transformations such as horizontal flip, zoom, shear, and brightness modifiers. For these experiments, we used ImageNet pre-trained models as our baseline architecture. The output layers were removed and a two-dimensional Global Average Pooling layer was added to reduce the dimensions to the number of channels. We added another layer of 256 neurons, after which a Dropout layer with a dropping probability of 0.4 was used for better generalization. The output layer was decided based on the classification problem at hand and used softmax activation. Nesterov implemented Adam was chosen as the optimizer, while categorical cross-entropy loss was chosen as the loss function for our models.

Previous studies have performed both binary and 5-ary classification on this dataset. For the task of binary classification, classes 0-2 (No DR, Mild DR and Moderate DR) and classes 3 and 4 (Severe DR and Proliferative DR) were combined respectively. Hence, two versions of these models were developed for both binary and 5-label classification.

6 Results and Discussion

The variations that arise from different scanning techniques and imaging procedures invariantly hamper the real-world application of these DL systems as these models which are trained to detect patterns in a particular distribution of images might not be able to replicate similar results under varying conditions. 3511 images from the Kaggle dataset were used for validation while the real-world performance of the model was tested on images from both the Messidor-2 and IDRiD. 2-ary and 5-ary classifications were conducted and the accuracies are compared in Fig. 5.

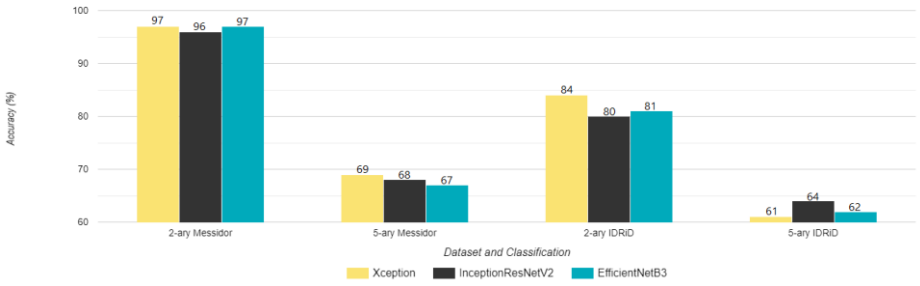


Fig. 5. Comparison of model accuracies

The inherent difficulty of 5-ary classification is amplified by the complexity of IDRiD, which was shown in the IDRiD grand challenge [12] held in 2018. A baseline accuracy of less than 60% was achieved by most of the participating teams. The accuracies achieved by the models discussed in this paper are comparable to those of the top teams from the grand challenge. Our best 5-ary classification model, InceptionResNetV2, was able to achieve a higher accuracy of 64% on this dataset while providing similar performance on class-wise sensitivity and specificity.

Table 1. Xception model performance

	5-ary				2-ary			
Dataset	Class	Specificity	Sensitivity	Accuracy	Class	Specificity	Sensitivity	Accuracy
Messidor-2	0	74	90	69	0	97	99	97
	1	0	0					
	2	54	66		1	81	59	
	3	78	41					
	4	53	46					
IDRiD	0	74	89	61	0	94	83	84
	1	0	0					
	2	62	54		1	68	86	
	3	49	29					
	4	42	72					

The Xception model was able to reach the highest accuracy of 69% on 5-ary classification of the Messidor-2 dataset and is tied with EfficientNetB3 on 2-ary classification with an accuracy of 97%. It also achieves the highest accuracy of 84% on the 2-

ary classification of IDRiD, while also offering significantly higher sensitivity and specificity.

Table 2. InceptionResNetV2 model performance

	5-ary				2-ary			
Dataset	Class	Specificity	Sensitivity	Accuracy	Class	Specificity	Sensitivity	Accuracy
Messidor-2	0	72	89	68	0	97	99	96
	1	0	0					
	2	57	62		1	75	58	
	3	65	77					
	4	58	31					
IDRiD	0	76	90	64	0	92	79	80
	1	0	0					
	2	66	57		1	62	83	
	3	46	62					
	4	52	41					

Apart from achieving the highest accuracy of 64% on the 5-ary classification on IDRiD, the InceptionResNetV2 model also offers high sensitivities on most of the classes. In the task of 2-ary classification, it achieves an exceptionally high sensitivity of 99% on Messidor-2.

Table 3. EfficientNetB3 model performance

	5-ary				2-ary			
Dataset	Class	Specificity	Sensitivity	Accuracy	Class	Specificity	Sensitivity	Accuracy
Messidor-2	0	69	98	67	0	97	99	97
	1	0	0					
	2	63	46		1	83	56	
	3	77	13					
	4	50	49					
IDRiD	0	75	88	62	0	92	81	81
	1	0	0					
	2	59	57		1	64	72	
	3	64	15					
	4	37	85					

EfficientNet B3, while having only 12 million parameters when compared to the 22 million and 55.8 million parameters of Xception and InceptionResNetV2 respectively, can achieve comparable results with faster inferences, training, and convergence rates making it a viable solution for automated diagnostic systems. It achieved an accuracy of 62% and 81% on the 5-ary and 2-ary classification of IDRiD. The general trend of achieving higher sensitivity (85% on class 4 IDRiD) at the cost of a lower specificity (37% on class 4 IDRiD) is observed here too.

7 Conclusion and Future Enhancements

Diagnosis of DR at its earlier stages greatly increases the probability of successful recovery. Given this, the need for efficient diagnosis systems is highly essential for which various state of the art deep learning techniques are increasingly being adopted.

The time taken to train, validate, and test the performance of these models is critical due to its direct impact on the research throughput in this field. Our study shows that EfficientNets can be used as an alternative to the previous state of the art model architectures as it converges faster, takes less time to train, and produces comparable results while utilising lesser computing resources. Transfer learning with this model also greatly speeds up the convergence rate.

Although binary classification produces better results, 5-ary classification should be given more attention since diagnosing DR at early stages would help greatly reduce the risk of it developing into more severe stages. This multi-label classification suffers from a larger loss in accuracy due to the lack of sufficient images in the mild to moderate categories of DR. In the future, we plan to extend our work by evaluating the performance of these models with augmented datasets and also make use of undersampling to see if we can mitigate this particular drawback. We also plan to explore image processing techniques such as denoising to make these models more robust.

References

1. Lam, Carson, et al. "Automated detection of diabetic retinopathy using deep learning." AMIA summits on translational science proceedings 2018 (2018): 147.
2. Kanungo, Yashal Shakti, Bhargav Srinivasan, and Savita Choudhary. "Detecting diabetic retinopathy using deep learning." 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). IEEE, 2017.
3. Pratt, Harry, et al. "Convolutional neural networks for diabetic retinopathy." Procedia Computer Science 90 (2016): 200-205.
4. Abràmoff, Michael David, et al. "Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning." Investigative ophthalmology & visual science 57.13 (2016): 5200-5206.
5. Takahashi, Hidenori, et al. "Applying artificial intelligence to disease staging: Deep learning for improved staging of diabetic retinopathy." PloS one 12.6 (2017): e0179790.
6. Mansour, Romany F. "Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy." Biomedical engineering letters 8.1 (2018): 41-57.
7. Sayres, Rory, et al. "Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy." Ophthalmology 126.4 (2019): 552-564.
8. Dutta, Suvajit, et al. "Classification of diabetic retinopathy images by using deep learning models." International Journal of Grid and Distributed Computing 11.1 (2018): 89-106.
9. Alom, Md Zahangir, et al. "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation." arXiv preprint arXiv:1802.06955 (2018)
10. Carrera, Enrique V., Andrés González, and Ricardo Carrera. "Automated detection of diabetic retinopathy using SVM." 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON). IEEE, 2017.
11. Porwal, Prasanna, et al. "Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research." Data 3.3 (2018): 25.
12. Porwal, Prasanna, et al. "IDRiD: Diabetic Retinopathy–Segmentation and Grading Challenge." Medical image analysis 59 (2020): 101561.