EfficientNet for Blindness Detection using Retina Images: A Diabetic Retinopathy Case

Amit Kumar Jaiswal[a], Prayag Tiwari[b], Vipin Kumar Rathi[c], Sachin Kumar[d], Deepak Gupta[e], Sagar Uprety[f], Victor Hugo C. de Albuquerque[g]

[a]*Institute for Research in Applicable Computing, University of Bedfordshire, UK*
[b]*Department of Information Engineering, University of Padova, Padova, Italy*
[c]*School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India*
[d]*Department of System Programming, South Ural State University, Chelyabinsk, Russia*
[e]*Maharaja Agrasen Institute of Technology, New Delhi, India*
[f]*The Open University, Milton Keynes, United Kingdom*
[g]*Graduate Program in Applied Informatics, University of Fortaleza, Fortaleza/CE, Brazil*

Abstract

Diabetic eye disease is the primary cause of blindness between working aged peoples. The major populated Asian countries such as India and China presently account for millions of people and at the verge of an eruption of diabetic inhabitants. These growing number of diabetic patients posed a major challenge among trained doctors to provide medical screening and diagnosis. Our goal is to leverage the deep learning techniques to automate the detection of blind spot in an eye and identify how severe the stage may be. In this paper, we propose a optimized technique on top of recently released pre-trained EfficientNet models for blindness identification in retinal images along with a comparative analysis among various other neural network models. Our fine-tuned EfficientNet-B5 based model evaluation follows the benchmark dataset of retina images captured using fundus photography during varied imaging stages and outperforms CNN and ResNet50 models.

Keywords: Diabetic Retinopathy, Medical Diagnosis, Convolutional Neural Network, Retina Images

Email addresses: amitkumar.jaiswal@beds.ac.uk (Amit Kumar Jaiswal), prayag.tiwari@unipd.it (Prayag Tiwari), vipin68_scs@jnu.ac.in (Vipin Kumar Rathi), sachinagnihotri16@gmail.com (Sachin Kumar), deepakgupta@mait.ac.in (Deepak Gupta), sagar.uprety@open.ac.uk (Sagar Uprety), victor.albuquerque@unifor.br (Victor Hugo C. de Albuquerque)

2010 MSC: 00-01, 99-00

1. Introduction

Diabetes is one of the chronic disease around the world that occurs when our body is not able to produce insulin hormone or not able to use it effectively. World Health Organization (WHO) reported that diabetes caused more than 1.6 million deaths in 2016 [1]. Diabetic patients are tend to produce high glucose level in the blood which may cause body organ's damage and failure. According to International Diabetes Federation (IDF), 1 out of 10 people is diabetic, which is a serious concern. Possible complications of diabetes may result in heart stroke, kidney failure, serious problems with eye vision etc.

One of the serious complication of diabetes is Diabetic Retinopathy (DR). DR may cause complete blindness and a lot of people around the world are victims of DR [2, 3]. Around 25% people out of total diabetic patients are affected from DR only, which make DR a complicated disease [4]. A long term diabetes may results in DR, which is a progressive disease and may cause partial or permanent vision impairment. Majority of people affected from DR are working age group which are the main work force of growing economy of any country [5]). IDF also reported that only India alone shares a large portion of diabetic people around the world and this share is growing rapidly every year [5].

It is difficult to detect the symptoms of DR in early stage, which is a challenging and important issue in medical science. The initial symptoms of DR are very few and patients are usually unaware until it results in irreversible damage to the retina or it is diagnosed through a medical test. Therefore, it is highly desirable that DR must be detected as early as possible. However, it is possible to detect DR but it required highly efficient and trained person who can evaluate the digital color fundus photographs of retina. The rear part of the human eye is known as fundus. These fundus images are evaluated by locating the lesions associated with vascular abnormalities that arises due to diabetes. Therefore, the solution is available and effective but its a time consuming process and also

requires a highly efficient and trained practitioner which may not be available everywhere.

Deep learning nowadays is very popular approach in medical image analysis and also providing solutions in computer assisted intervention problems [6]. Convolutional Neural Network (ConvNets) is a deep learning approach which is very popular for image data analysis in several domains [7, 8]. The accuracy of ConvNets can be improved by scaling up different parameters subject to the availability of resources. These scaled versions of ConvNets are very efficient for medical domain where accuracy is highly desirable. Therefore, in this study, EfficientNet architecture is used to analyze the retina images in order to detect the DR.

The rest of the article is organized as follows: Section 2 provides state of art literature review. In Section 3, methodology used is discussed in detail followed by Section 4 provides the details about experiments followed by the evaluation results in Section 5, and conclusion of the study in Section 6.

2. Related Work

Diabetic Retinopathy is a very common issue found in diabetic patients that can lead to eventual blindness. It's early detection hence is extremely important. Considering the state of the art technology, the exact detection of DR requires continuous observation of patients by a skilled physician. The process is not only time consuming but also requires immense resources. In [9], they highlighted the fact that only 10-15 physicians manually diagnose DR over 2 million retinal images per year while sharing their experience in the Aravind Eye Hospital, the largest eye care facility in the world. They have reinstated the extreme effort in terms of infrastructure and time required in the task when actually a large number of cases turn out to be normal. Automating this process can prove to be highly beneficial for the patients and can be an assistance for the doctors. Considering this aspect in mind, several researchers have worked in this area and utilized the machine learning artifact to devise a model that can predict

the presence of DR in a patient. In the following paragraphs, we review some of these key researches in this field.

In [9], they presented an account of the initial efforts they put in to develop an automated system using computer vision technique to operate patients in early stage of DR using retinal color fundus images. The project was a continuation of their previous work on WiLDNet at Aravind eye hospital. Their main focus was to create a model based on different retinopathy types using SIFT descriptors. The characteristics features include hemorrhages and exudates. The intention was to develop a robust and long term model. Besides that Support Vector Machine classifiers are trained to label each image patch; post which these images are aggregated and a decision over the entire image is identified based on patch-level prediction. The authors' reported an equal error rate of 87% using 1000 images.

In [10], they developed a solution for early-detection of DR using deep convolutional neural network technique to detect micro-aneurysms (MAs) in diabetic patients. They also perform multi-label classification by assigning labels to retinal fundal images on five categories. Earlier they proposed multi-layer convolutional neural networks (CNNs) by considering two fully connected layer and a single output layer to efficiently detect DR. Catering to the issue of oversampled classes, they use a small capacity network with L2 regularization and dropout, 4X4 kernel and pre-processing and augmentation methods. The performance of the scheme was evaluated on Kaggle diabetic retinopathy dataset[1]. The approach gave a kappa score of more than 0.85 in severity grading. They claim the approach to be also computationally time and space efficient. However, a mean squared error is used as objective function on regression problem.

A model built using convolutional neural networks such as AlexNet, VggNet, GoogleNet and ResNet is proposed in [11] to identify Diabetic Retinopathy in diabetic patients. The fundus image of DR are classified into five classes based on which the model identifies the stage of DR for the patient. Moreover, the use

[1] https://www.kaggle.com/c/diabetic-retinopathy-detection/data

of transfer learning and hyper-parameter tuning will leverage the CNNs model for better accuracy which was not possible before using non-transferring learning for noisy data. Normalization techniques and data augmentation have been deployed for preprocessing of images and non-local means denoising (NLMD) to eject noise. The performance of the system was tested on datasets available on Kaggle and a classification accuracy of 95.68% was reported attributed to the application of CNN and transfer learning.

An EyeWeS model was proposed to achieve high performance and greater efficiency by converting the pre-trained convolutional neural network architecture for DR detection into a weekly supervised model by eliminating the data automation lesion-wise required for pixel-level training [12]. Moreover they not only focused on identification of diabetic retinopathy but also directs to the region of eye that got effected. Also this model consists of bag labels to train instances (i.e, image patches) to identify the pooling function f to encode relation among instances and bag labels. The objective of pooling function is to train instance classifier $P(y_i|x_i, \theta)$. They have checked the performance of their scheme on the Messidor dataset. On E-optha MA, micro-aneurysms of size 3 pixels in diameter could be detected with a significant improvement in the results has been obtained from Inception V3 with a higher AUC of 95.8% using only 5% of the Inception V3s number of parameters. A remarkable 97.1% AUC has been reported for a cross-dataset experiment.

Past work [13] employed a deep learning methodology to detect the diabetic retinopathy (DR). They gave attention to regression activation Maps (RAM) model to capture the discriminative area of input retina image. The network architecture was proposed by authors' considered that the global averaging pooling (GAP) layer to identify the total contribution in final prediction mentioned by neuron. The RAM model is depicted by $G(i, j) = \sum_{k=1}^{K} g_k(i, j)w_k$. In addition, the comparison of their own neural network with conventional neural network such as AlexNet, GoogLeNet is highlighted in their work. The authors proposition claims better performance due to the addition of RAM that distinguishes the exact area of severity in terms of vascular abnormalities.

3. Methodology

The ConvNets can be scaled based on the requirements to achieve better accuracy subject to the availability of resources. There are several ways to achieve this scalability. For example, we can increase the number of layers in ResNet [14] to scale it up from ResNet-18 to ResNet-200. The popular approaches to scale up a ConvNets model is to increase the width or depth of the model or a high resolution image can be used for model training and testing. However, these approaches do not use a well defined criteria to select the width, depth or resolution of the input image. In this study, EfficientNet [15], a balanced and more accurate deep ConvNets is used to detect DR in retina image data set which uses a well defined criteria to scale up on all dimensions and provides better accuracy.

3.1. Model

This section details the adapted recent convolutional neural network-based EfficientNet model [15] for Diabetic Retinopathy – a disease caused by damage to the retina. If not treated, it can progress to blindness. Existing method of diagnosis is defined by classes 0 to 4, where 0 represents no presence of diseases and 4 represents severe progression. Our proposed model is based on the EfficientNet-B5 [15] with ImageNet as pre-trained weights, and also with different input image sizes of 224x224 and 256x256 respectively. The EfficientNet uses a compound dimension scaling approach [15]. This approach uses a coefficient parameter ϕ to apply uniform scaling on width, depth and resolution as given in Eq.1 as follows:

$$width : w = \beta^{\phi}, \qquad depth : d = \alpha^{\phi}, \qquad resolution : r = \gamma^{\phi}, \tag{1}$$

$$\text{such that } \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$
$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

ϕ is a user defined coefficient and acts as a control parameters to measure the availability of resources for model scaling. α, β and γ are constants to control

the number of resources to be assigned to scale width, depth and resolution. In the main architecture of EfficientNet-B5, we replaced the top head with GlobalAveragePooling2D and a dense layer as linear output. Our models uses mean squared error as a loss function and the model output were linear activation function.

Our aim is to optimize the quadratic weighted kappa score [16, 17] which we consider as a regression problem. In order to be flexible in the optimization and we can yield higher scores than solely optimizing for accuracy. We optimize a pre-trained EfficientNet-B5 with a few added layers. The metric that we optimize is the mean squared error which is the squared mean differences between the predictions and labels, as reported in the formula below. By optimizing the given metric we are also optimizing for quadratic weighted kappa in case if we round the predictions afterwards.

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{2}$$

where n is the number of data points, Y_i denotes observed values, and $\hat{Y}_i$ denotes predicted values.

Since we do not have much training data (3662 images), we augment the data to increase the robustness of our proposed model. We also rotate the data on any angle. Also, we flip the data both horizontally and vertically. Finally, we then divide the data by 128 for normalization.

Also, we found one of the earlier work [18] which developed a technique to deal with imbalanced image data (in our study, its' retinal images) by sampling method, but it does not improved our prediction due to augmentation step as we remove black background, without augmentation it improves a little which our model seems to overfit due to continuous loss at the same rate with no learning even on using cyclic learning rate scheduler,

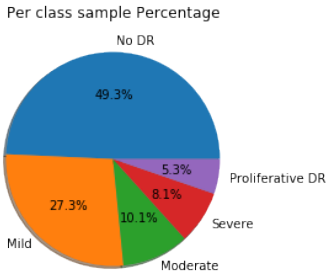

Figure 1: Class Distribution

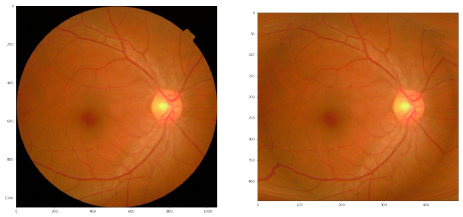

Figure 2: Typical Augmentation - ResizeAndCrop

4. Experimental Evaluation

In this section, we will detail the employed dataset followed by the preprocessing and data augmentation techniques as well as the training steps including inference of trained models.

4.1. Dataset

In this work, we employ a real-world eye disease image dataset which has a large collection of high resolution retina images captured using fundus photography[2] underlying varied circumstances of imaging situations. A trained clinician has assessed the existence of diabetic retinopathy in each image on standard ICDR (International Clinical Diabetic Retinopathy) severity scale of 0 to 4. The rating scores of each DR causes are $\{0 - No\ DR\}$, $\{1 - Mild\}$, $\{2 - Moderate\}$, $\{3 - Severe\}$, $\{4 - Proliferative\ DR\}$. The collection consists of 5,590 images labeled with a subject for each eye image out of which 3,662 and 1928 images are splitted as train and test set with corresponding classes of 2 and 1, respectively.

The size of images are quite large which is 2896**x**1944. According to Figure 1, it shows that the training data is highly unbalanced due to larger shift in "No DR" class as compared to "Severe" is very less. The dataset is noisy in terms of images and labels such as some of the image artifacts are out of focus,

[2]https://en.wikipedia.org/wiki/Fundus_photography

overexposed or underexposed which is due to collection from various clinics using different camera over a long period of time.

4.2. Preprocessing

This section details the data processing step which we performed before actually modeling this regression task so that training can be made faster on such unbalanced dataset. First and foremost, we transformed each retinal images in RGB channels followed by circular crop on each images based on image center to separate black regions from real-squared images by keeping the original aspect ratio so the images appearance natural. Also, we employ Ben Graham's method[3] to improve the illumination condition of images so that we can enrich the insights from eye images. Then, we resize the original images to 224x224 and 256x256 (different sizes to check the correctness /feasibility of our modeling approach), and we trained these collection of reduced image sizes with a large batch size of 32. After preprocessing the dataset, we found that images when

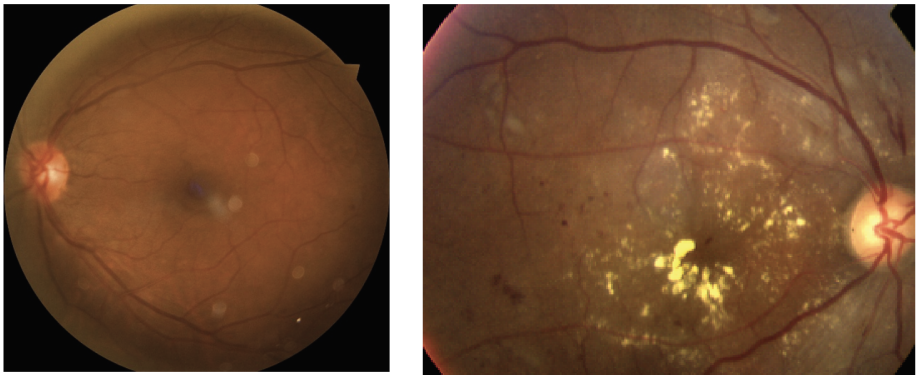

Figure 3: Preprocessed Sample Image: Left image is zoomed on the right

zoomed to center are exhibiting close relation to the light spots on left. To solve this zoom issue as reported in Figure 3, we performed augmentation on these processed images to make sure it generalises better and does not overfits.

[3]https://github.com/btgraham/SparseConvNet/tree/kaggle_Diabetic_Retinopathy_competition

4.3. Data Augmentation

For better generalisation of the processed dataset, we look into this problem by executing some augmentation steps for the images. We employ the Albumentations library[4] to perform augmentation where we flipped images horizontally/vertically, image rotation to 360°, zoom images to 1.3x and contrasting for lightning. We report these augmentation steps in Figure 4.

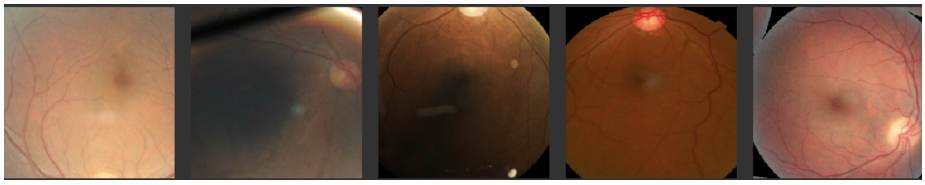

Figure 4: Sample Images after Augmentation

We adapted a new augmentation step i.e. *Polar unrolling* which allows to better leverage pixel space, separate "rotation" from an augmentations list and acquire uniformly scaled eye images (for the instances of no or partial cropping of the fundus image with preservation of radius).

Originally we have images with noticeable black regions. In order to whittle away such regions we at the beginning apply an autocrop. After autocrop, we have the circle's radius (depicted by the broadest side of an image). We extracted circle using polar unrolling. By unrolling we changed coordinate space which do not require rotation augmentation. As rotation becomes just plain shift by x axis (which does not matter for convolutional neural networks). It is more than absence of this type of augmentation. This made all possible rotations for our model considered (except some borders, which can be tackled by the single 50% shift by x axis).

Our augmentation technique improves the existing contourlet transform method on retinal images [2] in terms of processing the imbalanced data and training technique on the state-of-the-art EfficientNet models.

[4]https://github.com/albu/albumentations

4.4. Training and Inference

We trained our proposed EfficientNet-B5 based model with a larger batch size of 32 for 5 epochs as in the warm-up stage where we freeze all layers except the last two given the learning rate of *4e-3*, and we use Adam optimizer and cosine with learning rate (LR) scheduler. In the complete model training, we trained over 30 epoch under the fine-tuning stage provided all layers are unfreeze. Also, we use Early stopping monitoring validation loss for 5 epoch. The LR scheduler at different steps of training is schematically shown in Figure 5 followed by the training and validation loss in Figure 6

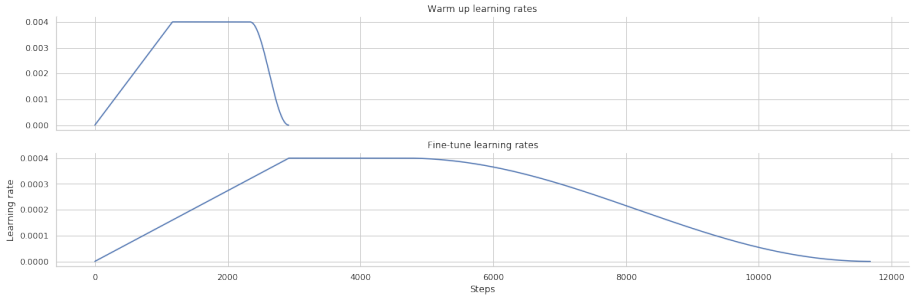

Figure 5: Learning rate Policy Scheduler

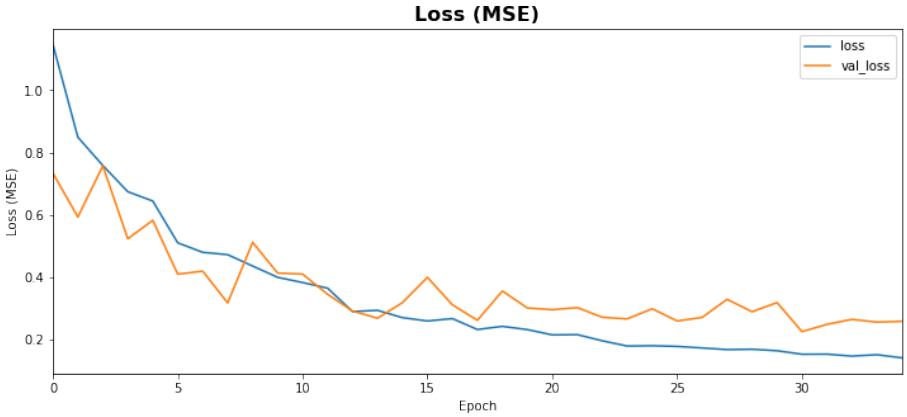

Figure 6: Training Loss

As we found that the training loss becomes smoother around $\sim$ 0.4 and

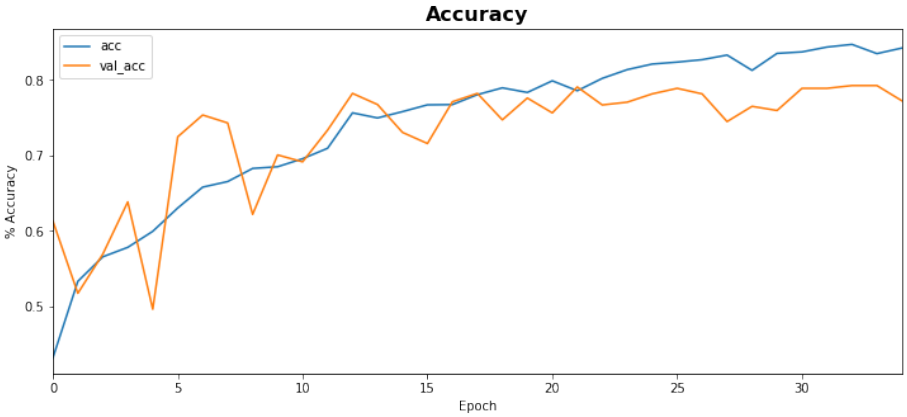

Figure 7: Accuracy

steeply improved to less than 0.2 at 30th epoch. This shows that TTA keeps the retinal feature during validation while training and could be improved if trained for long on higher batch size.

We cross-validated our EfficientNet-B5 model 5 folds and then inference these trained models by doing test-time augmentation (TTA) [19] 10 times and taken average of 5 models.

We employ 1XP100 and 1xT4 NVIDIA Tesla GPU for training models.

4.5. Evaluation Measures

To evaluate the detection model, we use *quadratic weighted kappa* (or Cohen' s Kappa), which measures the consensus between expert ratings and submitted ratings. The quadratic weights triggers an optimization factor in the rounding operation. This metric ranges from 0 (random agreement among raters) to 1 (complete agreement among raters). A perfect score of 1.0 is allowed when both the actuals and predictions are same, otherwise the least possible score is -1 which is provided when the predictions are furthest away from actuals. In this work, we treat all actuals as 0's and all predictions as 4's. This will give rise

Training Set	Validation Set
0.96754	**0.92321**

Table 1: Fine-tuned EfficientNet-B5 Regression - Cohen Kappa Score

a quadratic weighted kappa score of -1[5]. The weighted kappa is given below:

$$\kappa = 1 - \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} x_{ij}}{\sum_{i=1}^{k} \sum_{j=1}^{k} w_{ij} m_{ij}} \tag{3}$$

where we are going to optimize mean squared error (MSE) and by optimizing MSE, we will also optimize quadratic weighted kappa as the problem is expressed as regression.

5. Experimental Results

This section reports the evaluation result which is the predicted labels of DR for blind spot. For evaluation we predict values from the generator and round it to the nearest integer to get valid predictions. After that we compute the quadratic weighted kappa score on the training set and the validation set which is reported in Table 1.

We perform Grid search in order to optimize the validation score over a range of thresholds and the quadratic weighted kappa score is 0.92 with threshold range of (0.5, 1.5, 2.5, 3.5).

We perform test time augmentation (same augmentation steps as mentioned in above section) 10 times and averaged all 5 models with their TTA predictions. The prediction result is reported in Figure 8 and Figure 9.

We also employed different composition of EfficientNet-Bx (where x=0, 1, 2, 3, 4 and 5), ResNet50 and other CNN models which is reported below in Table 2.

We adopted several convolutional neural network based models reported in Table 2 and found that DenseNet169

[5]Generally, the score near to 1 is best

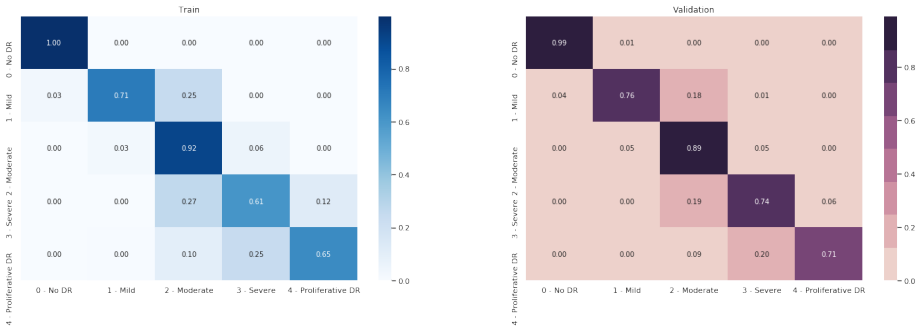

Figure 8: Confusion Matrix - Inference of 5 Folds

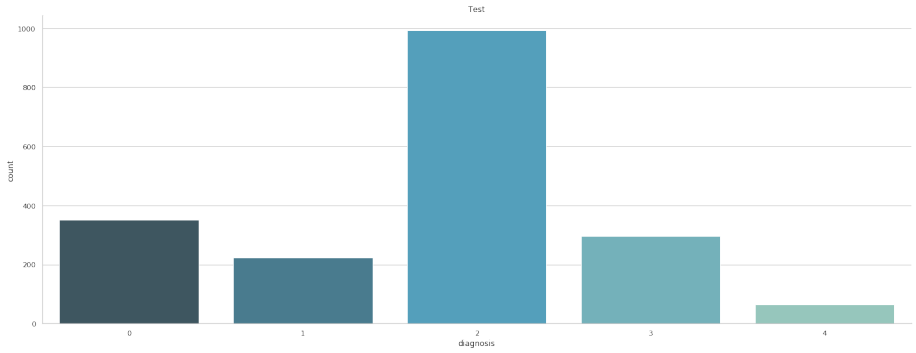

Figure 9: Label Distribution - Predictions

The prediction scores of convolutional neural network models reported in Table 2 treats the identification problem as regression which we also considered for our proposed model on top of EfficientNet-B5 model. The models are trained in general without any fine-tuning steps and EfficientNet-B5 performs better than any other CNN models with a significant improvement of 0.10%.

6. Conclusion and Future Work

In this work, we introduce the state-of-the-art EfficientNet-B5 based model for identification of blindness in eye disease (diabetic retinopathy) evaluated on a retinal image dataset. Our fine-tuned Efficient-B5 based model outperforms CNN and ResNet50 models with 92.32% validation accuracy which predicts diagnosis of diabetic retinopathy severity (eye blindness) on the five-point scale

Model	QWK Score
EfficientNet-B0	0.883
EfficientNet-B1	0.892
EfficientNet-B2	0.893
EfficientNet-B3	0.900
EfficientNet-B4	0.891
EfficientNet-B5	**0.902**
ResNet50	0.858
InceptionNetResNetV2	0.884
DenseNet169	0.881

Table 2: Comparison of CNN models with EfficientNet weights

from retinal images. Our baseline model Efficient-B5 (with fine-tuning) model training on the average doctor opinion, a tactics that output state-of-the-art results on identifying blindness by 90.20% of validation accuracy. The freezing and unfreezing for fine-tuned EfficientNet-B5 significantly improved the prediction with 92.32% validation accuracy. We found that there are more number of 0's and 2's i.e., no diabetic retinopathy symptom and moderate retinopathy shown in around 89% of the images. We intend to adopt other CNN architectures such as Unet with ResNet models and EfficientNet weights with Unet for such imbalanced data for this prediction task, and pseudo labeling the imbalanced dataset may potentially improve the prediction for given classes including consideration of this identification task as binary classification by individually labeling [20] the data could be an added advantage. The limitation in doing so is processing power which greatly increases if using EfficientNet-B6 or B7 weights.

References

[1] W. H. Organization, Death among diabetic patients.
URL http://www.who.int/diabetes/global-report/en

[2] P. Feng, Y. Pan, B. Wei, W. Jin, D. Mi, Enhancing retinal image by the contourlet transform, Pattern Recognition Letters 28 (4) (2007) 516–522.

[3] Z. Liu, C. Opas, S. M. Krishnan, Automatic image analysis of fundus photograph, in: Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering'(Cat. No. 97CH36136), Vol. 2, IEEE, 1997, pp. 524–525.

[4] R. L. H. T. P. Z. L.M Rutta, D.J Magliano, J. Shaw, Prevalence of diabetic retinopathy in type 2 diabetes in developing and developed countries, Diabetic medicine 30 (2013) 387–398.

[5] I. D. F. Team, Latest diabetes figures paint grim global picture.
URL `http://www.idf.org/press_releases`

[6] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, Medical image analysis using convolutional neural networks: A review, Journal of Medical Systems 42 (2018) 226.

[7] A. K. Jaiswal, P. Tiwari, S. Kumar, D. Gupta, A. Khanna, J. R. Joel, Identifying pneumonia in chest x-rays: A deep learning approach, Measurement 145 (2019) 511–518.

[8] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, J. Liang, Convolutional neural networks for medical image analysis: Full training or fine tuning?, IEEE Transactions on Medical Imaging 35 (2016) 1299–1312.

[9] N. Silberman, K. Ahrlich, R. Fergus, L. Subramanian, Case for automated detection of diabetic retinopathy, in: 2010 AAAI Spring Symposium Series, 2010.

[10] S. M. S. Islam, M. M. Hasan, S. Abdullah, Deep learning based early detection and grading of diabetic retinopathy using retinal fundus images, arXiv preprint arXiv:1812.10595.

[11] S. Wan, Y. Liang, Y. Zhang, Deep convolutional neural networks for diabetic retinopathy detection by image classification, Computers & Electrical Engineering 72 (2018) 274–282.

[12] P. Costa, T. Araújo, G. Aresta, A. Galdran, A. M. Mendonça, A. Smailagic, A. Campilho, Eyewes: weakly supervised pre-trained convolutional neural networks for diabetic retinopathy detection, in: 2019 16th International Conference on Machine Vision Applications (MVA), IEEE, 2019, pp. 1–6.

[13] Z. Wang, J. Yang, Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation, arXiv preprint arXiv:1703.10757.

[14] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[15] M. Tan, Q. V. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, arXiv preprint arXiv:1905.11946.

[16] J. Cohen, Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit, Psychological Bulletin 70 (1968) 213–220.

[17] M. J. Warrens, Some paradoxical results for the quadratically weighted kappa, Psychometrika 77 (2016) 315–323.

[18] A. DAddabbo, R. Maglietta, Parallel selective sampling method for imbalanced and large data classification, Pattern Recognition Letters 62 (2015) 61–67.

[19] M. S. Ayhan, P. Berens, Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks.

[20] M. Y. Guan, V. Gulshan, A. M. Dai, G. E. Hinton, Who said what: Modeling individual labelers improves classification, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.