Diabetic Retinopathy Detection Using Deep Learning

Eren Tekin[1] and Salih Yalcin[2]

[1,2]Department Of Computer Engineering, Abdullah Gul University, Kayseri, Turkey

{eren.tekin, salih.yalcin}@agu.edu.tr

Keywords: Deep learning, diabetic retinopathy, convolutional neural network, classification, APTOS, Kaggle, transfer learning

Abstract: Diabetic retinopathy is one of the most threatening complications of diabetes that leads to permanent blindness if left untreated. One of the essential challenges is early detection, which is very important for treatment success. Unfortunately, the exact identification of the diabetic retinopathy stage is notoriously tricky and requires expert human interpretation of fundus images. Simplification of the detection step is crucial and can help millions of people. Convolutional neural networks (CNN) have been successfully applied in many adjacent subjects, and for diagnosis of diabetic retinopathy itself. However, the high cost of big labeled datasets, as well as inconsistency between different doctors, impede the performance of these methods. In this paper, we propose an automatic deep-learning-based method for stage detection of diabetic retinopathy by single photography of the human fundus. Additionally, we propose the multistage approach to transfer learning, which makes use of similar datasets with different labeling.

1. INTRODUCTION

Diabetic retinopathy (DR) is one of the most threatening complications of diabetes in which damage occurs to the retina and causes blindness. It damages the blood vessels within the retinal tissue, causing them to leak fluid and distort vision. Along with diseases leading to blindness, such as cataracts and glaucoma, DR is one of the most frequent ailments, according to the US, UK, and Singapore statistics (NCHS, 2019; NCBI, 2018; SNEC, 2019). DR progresses with four stages:

• Mild non-proliferative retinopathy, the earliest stage, where only microaneurysms can occur;

• Moderate non-proliferative retinopathy, a stage which can be described by losing the blood vessels’ ability of blood transportation due to their distortion and swelling with the progress of the disease;

• Severe non-proliferative retinopathy results in deprived blood supply to the retina due to the increased blockage of more blood vessels, hence signaling the retina for the growing of fresh blood vessels;

• Proliferative diabetic retinopathy is the advanced stage, where the growth features secreted by the retina activate proliferation of the new blood vessels, growing along inside covering of retina in some vitreous gel, filling the eye.

Each stage has its characteristics and particular properties, so doctors possibly could not take some of them into account, and thus make an incorrect diagnosis. So, this leads to the idea of creation of an automatic solution for DR detection.

In 2019, APTOS (Asia Pacific Teleophthalmology Society) and competition ML platform Kaggle challenged ML and DL researchers to develop a five-class DR automatic diagnosing solution (APTOS 2019 Blindness Detection Dataset) [1]. In this paper, we propose the transfer learning approach and an automatic method for detection of the stage of diabetic retinopathy by single photography of the human fundus.

2. RELATED WORK

Many research efforts have been devoted to the problem of early diabetic retinopathy detection. First of all, researchers were trying to use classical methods of computer vision and machine learning to provide a suitable solution to this problem. For instance, Priya et al. (Priya and Aruna, 2012) proposed a computer-vision-based approach for the detection of diabetic retinopathy stages using color fundus images. They tried to extract features from the raw image, using the image processing techniques, and fed them to the SVM for binary classification and achieved a sensitivity of 98%, specificity 96%, and accuracy of 97.6% on a testing set of 250 images. Also, other researchers tried to fit other models for multiclass classification, e.g., applying PCA to images and fitting decision trees, naive Bayes, or k-NN (Conde et al., 2012) with best results 73.4% of accuracy, and 68.4% for F-measure while using a dataset of 151 images with different resolutions.

With the growing popularity of deep learningbased approaches, several methods that apply CNNs to this problem appeared. Pratt et al. (Harry Pratt, 2016) developed a network with CNN architecture and data augmentation, which can identify the intricate features involved in the classification task such as micro-aneurysms, exudate, and hemorrhages in the retina and consequently provide a diagnosis automatically and without user input. They achieved a sensitivity of 95% and an accuracy of 75% on 5,000 validation images. Also, there are other works on CNNs from other researchers (Carson Lam and Lindsey, 2018; Yung-Hui Li and Chung, 2019). It is useful to note that Asiri et al. reviewed a significant amount of methods and datasets available, highlighting their pros and cons (Asiri et al., 2018). Besides, they pointed out the challenges to be addressed in designing and learning about efficient and robust deep learning algorithms for various problems in DR diagnosis and drew attention to directions for future research.

Other researchers also tried to make transfer learning with CNN architectures. Hagos et al. (Hagos and Kant, 2019) tried to train InceptionNet V3 for 5- class classification with pretrain on ImageNet dataset and achieved accuracy of 90.9%. Sarki et al. (Rubina Sarki, 2019) tried to train ResNet50, Xception Nets, DenseNets and VGG with ImageNet pretrain and achieved best accuracy of 81.3%. Both teams of researchers used datasets, which were provided by APTOS and Kaggle.

3. PROBLEM STATEMENT

3.1. Dataset

We used on Kaggle APTOS 2019 Blindness Detection (APTOS2019) dataset. The full dataset consists of 18590 fundus photographs, which are divided into 3662 training, 1928 validation, and 13000 testing images by organizers of Kaggle competition. All datasets have similar distributions of classes; distribution for APTOS2019 is shown in Figure 1.

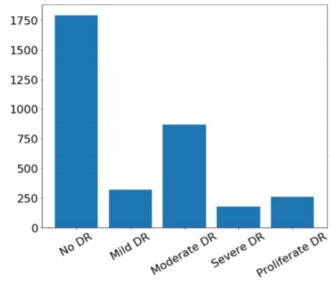

Figure 1: Classes distributions of dataset

3.2.Evaluation Metrics

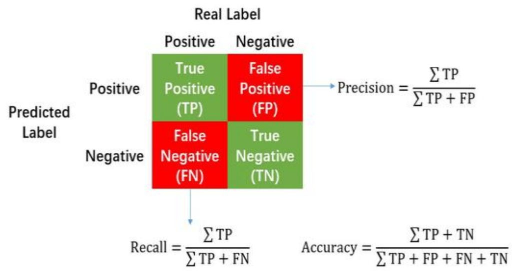

Figure 2: Calculation of Precision, recall and accuracy in the confusion matrix

Accuracy - Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same. Therefore, you have to look at other parameters to evaluate the performance of your model.
Accuracy = TP+TN/TP+FP+FN+TN

Precision (Spesificity) - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labeled as survived, how many actually survived? High precision relates to the low false positive rate.
Precision = TP/TP+FP

Recall (Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes. The question recall answers is: Of all the passengers that truly survived, how many did we label?
Recall = TP/TP+FN

F1 score - F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.
F1 Score = 2*(Recall * Precision) / (Recall + Precision)

4. METHOD

In this study, we used Python programming language and PyTorch (Facebook, USA).

4.1.Models

In this study, transfer learning was carried out through the following models.

- EfficientNet-B2, B3 and B4 Noisy Student-Weights (arXiv:1905.11946v5)
- ResNet101D and ResNet200D (arXiv:1512.03385)
- NfNet-F0 (arXiv:2102.06171)

- Vit-Base_Patch16 (arXiv:2010.11929)

4.2.Data Augmentation

In this study we used Data Augmentation (DA). Data Augmentation is a technique that can be used to artificially expand the size of a training set by creating modified data from the existing one. It is a good practice to use DA if you want to prevent overfitting, or the initial dataset is too small to train on, or even if you want to squeeze better performance from your model.

Data Augmentation is not only used to prevent overfitting. In general, having a large dataset is crucial for the performance of both ML and Deep Learning (DL) models. However, we can improve the performance of the model by augmenting the data we already have. It means that Data Augmentation is also good for enhancing the model's performance.

4.3.Training

4.3.1. Pretraining

- Before training, we determined 5 classes for each stage of DR.

  ({No DR, Mild, Moderate, Severe, Proliferative} {0, 1, 2, 3, 4})

- The following configurations have been made.

  "batch_size": 8,
  "learning_rate": 1e-3,
  "weight_decay": 0,
  "img_size": 448,
  "early_stopping": 10

4.3.2. Main Training

We initialized feature extractor with weights from ImageNet-pretrained CNN (Transfer Learning). We trained our models with 30 epochs.

4.3.3. Post Training

Confusion Matrix is a performance measurement for machine learning classification problem where output can be two or more classes. It is a table with 4 different combinations of predicted and actual values.

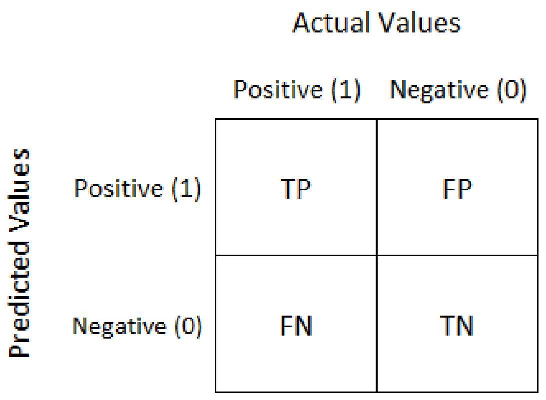

Figure 3: Example confusion matrix for binary problem

Figure 4: An example confusion matrix from our NFNET model (Y Axis: True Labels, X Axis: Predicted Labels)

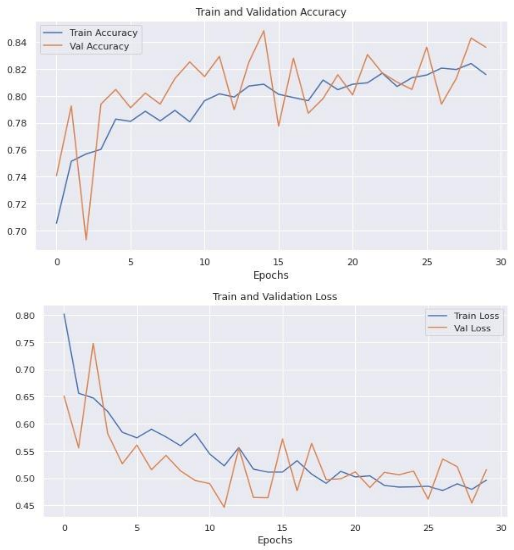

Figure 5: An example accuracy and loss graphs from our EfficientNet-B3 model

We examined Confusion Matrix and Accuracy/Loss graphs to analyze our models' reliability.

5. RESULTS

	Val. Acc.	Precision	Recall	F1-Score
Resnet50-Baseline	0.81	0.81	0.81	0.80
EffNet-B2	0.85	0.85	0.85	0.84
EffNet-B3	0.85	0.85	0.85	0.83
EffNet-B4	0.85	0.84	0.85	0.84
Resnet101D	0.87	0.87	0.87	0.86
Resnet200D	0.85	0.85	0.85	0.84
Vision Transformer	0.72	0.63	0.72	0.66
NFNet-F0	0.83	0.83	0.83	0.81
MLP-Mixer-B16(IMG SIZE 224)	0.72	0.62	0.72	0.68

6. CONCLUSION

In this paper, we tried newest models such as NfNet, EfficientNet(s) and the others on APTOS Diabetic Retinopathy Dataset and examined results in Section5. We reached best results with ResNet101D model. However, models may improve using different optimizers, scheduler, weight decay.

7. REFERENCES

[1] https://www.kaggle.com/c/aptos2019-blindness-detection/data (Accessed: 7 June 2021)

[2]https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9 (Accessed: 7 June 2021)

[3] https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy (Accessed: 7 June 2021)

[4] Muthu Rama Krishnan Mookiah, U. Rajendra Acharya, Chua Kuang Chua, Choo Min Lim, E.Y.K. Ng, Augustinus Laude, Computer-aided diagnosis of diabetic retinopathy: A review, Computers in Biology and Medicine, Volume 43, Issue 12, 2013, Pages 2136-2155

[5] Wejdan L. Alyoubi, Wafaa M. Shalash, Maysoon F. Abulkhair, Diabetic retinopathy detection through deep learning techniques: A review, Informatics in Medicine Unlocked, Volume 20, 2020, 100377

[6] Borys Tymchenko, Philip Marchenko, Dmitry Spodarets, Deep Learning Approach to Diabetic Retinopathy Detection, arXiv:2003.02261

[7] Lam, Carson et al. “Automated Detection of Diabetic Retinopathy using Deep Learning.” AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science vol. 2017 147-155. 18 May. 2018

[8] Sahlsten, J., Jaskari, J., Kivinen, J. et al. Deep Learning Fundus Image Analysis for Diabetic Retinopathy and Macular Edema Grading. Sci Rep 9, 10750 (2019).

[9] Revathy R, Nithya B, Reshma J , Ragendhu S, Sumithra M D, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-018, Vol. 9 Issue 06, June-2020

[10] R. Sarki, K. Ahmed, H. Wang and Y. Zhang, "Automatic Detection of Diabetic Eye Disease Through Deep Learning Using Fundus Images: A Survey," in *IEEE Access*, vol. 8, pp. 151133-151149, 2020, doi: 10.1109/ACCESS.2020.3015258.

[11] Abràmoff MD, Reinhardt JM, Russell SR, Folk JC, Mahajan VB, Niemeijer M, Quellec G. Automated early detection of diabetic retinopathy. Ophthalmology. 2010 Jun;117(6):1147-54.

[12] Harry Pratt, Frans Coenen, Deborah M. Broadbent, Simon P. Harding, and Yalin Zheng. "Convolutional Neural Networks for Diabetic Retinopathy" Procedia Computer Science, vol. 90, 2016.

[13] Goh, James Kang Hao et al. “Retinal Imaging Techniques for Diabetic Retinopathy Screening.” Journal of diabetes science and technology vol. 10,2 282-94. 1 Feb. 2016,

[14] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 2016 Dec 13;316(22):2402-2410.

[15] Ratanapakorn, Tanapat et al. “Digital image processing software for diagnosing diabetic retinopathy from fundus photograph.” Clinical ophthalmology (Auckland, N.Z.) vol. 13 641-648. 17 Apr. 2019