■ RESEARCH ARTICLE

AITA
ARTIFICIAL INTELLIGENCE
THEORY AND APPLICATIONS

Classification of Eye Disease from Fundus Images Using EfficientNet

Batuhan Bulut [a], **Volkan Kalın** [a], **Burcu Bektaş Güneş** [b†], **Rim Khazhin** [a]

[a] Ural Telekom, Antalya, Turkey

[b] Department of Computer Engineering, Naval Academy, National Defence University, İstanbul, Turkey

[†] bbektas@dho.edu.tr, corresponding author.

RECEIVED JANUARY 4, 2022
ACCEPTED APRIL 3, 2022

CITATION Bulut, B., Kalın, V., Berktaş Güneş, B., & Khazin, R. (2022). Classification of Eye Disease from Fundus Images Using EfficientNet. *Artificial Intelligence Theory and Applications, 2*(1), 1-7.

Abstract

Studies show that at least 2.2 billion people in the world have some kind of visual impairment or blindness. The prevalence of conditions progressing into preventable blindness is quite high. As more and more public data sets are available, the training of deep learning in the medical field is a possible choice, but the practical application of deep learning in clinical practice is still an open issue. We work for solving this problem and continue developing clinical data sets and models to create a practically usable model that will identify “referrable” retinal disorders that can be treated or are at the stage sufficiently progressed to start treatment as opposed to the “non-referrable” disorders with too early stage that doesn’t require treatment, or disorders having no known treatment methods. Important difference between the two is: diagnosing a “non-referrable” disorder will result in unnecessary visit to a retina specialist, while missing the “referrable” disorders might result in permanent blindness or vision loss.

In this study, we explored the use of deep convolutional neural network methodology for the automatic classification of eye diseases using color fundus images. More than 10 retinal disorders have been effectively classified using the proposed model. The proposed method is tested using the public datasets and the EyeCheckup dataset we created. Our deep learning model achieved sensitivity of 0.9439, specificity of 0.8604, and an Accuracy of 0.86 with the test data set.

Keywords: eye diseases, fundus data set, deep learning, EfficientNet, CNN

1. Introduction

According to World Health Organization (WHO), it has been reported that at least 2.2 billion people around the world have a vision impairment, of whom at least 1 billion have a vision impairment that could have been prevented or is yet to be addressed [1]. The major causes of these disorders are eye diseases. Visual impairment will continue to increase with the aging population growth, highest prevalence being diabetic retinopathy, glaucoma, age-related macular degeneration, and cataract. Visual impairment and blindness have profound human and socioeconomic consequences in all societies. People with vision loss experience reduced quality of life, greater difficulty with daily living and social dependence, higher rates of depression, and increased risk of falls and related hip fractures. In most cases vision loss can be prevented with early diagnosis, surveillance, and timely treatment. Deep learning, a state-of-art machine learning (ML) technique, has shown promising diagnostic performance. It has been widely adopted in

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN : 978-605-69730-2-4 © 2022 University of Bakırçay

many domains including social media, cybersecurity, and medicine. Retinal photographs, widely used since the 1990s, can scan large amounts of data in minutes with low cost and show high diagnostic performance via artificial intelligence (AI) systems [2].

There is a number of research on the classification of eye diseases in the literature [3–13]. Studies are generally about diabetic retinopathy, glaucoma, diabetic macular edema, and abnormality detection. Our goal is to develop a system that will detect more than 20 retinal disorders. This comprehensive study, that we describe has not been found in the literature.

In this study, we explored use of deep convolutional neural network methodology for the automatic classification of eye diseases using color fundus images. For this purpose in addition to the public data sets, 21.842 color fundus images obtained from Department of Ophthalmology in Akdeniz University (AU) Hospital. Images were analyzed by EfficientNet model. The model has been optimized with various hyperparameter tuning techniques and 50 different parameters were trained. A model with the best performance results was selected.

2. Material and Method

2.1. Data Set

This research was carried out using color fundus images depicted in Figure 1. In addition to 80.000 images from public data sets (Kaggle 2015 DR Competition-EyePACS [14, 15], DIARETDB0 [16], IDRID [17], MESSIDOR [18], MESSIDOR2 [18], Kaggle 2019 APTOS [19]), 21.842 color fundus images have been obtained from Department of Ophthalmology of Akdeniz University Hospital, Turkey (AU) after Ethical committee approval. Thus, the current data set was enhanced from 80.000 to 101.842.

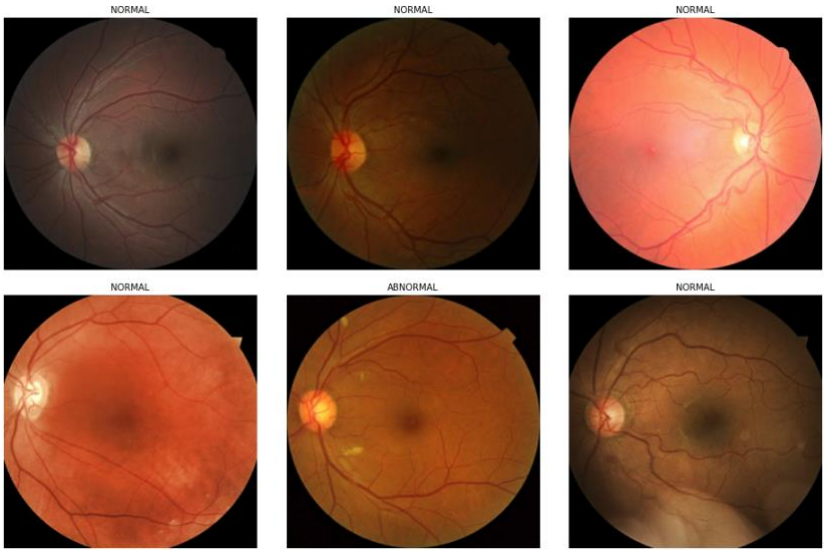

Figure 1. Sample images from the training data set.

For each fundus image of the AU, a Retina specialist has graded Diabetic Retinopathy (DR) severity score (values between 0 and 4; 0 being non-DR, 1 - Mild DR, 2 – Moderate DR, 3- Severe Non-proliferative DR, 4 – Proliferative DR), identified whether there was any anomaly other than DR (0 or 1), whether there was Diabetic Macular Edema (DME)

(0 or 1) and whether laser treatment (photo coagulation) was applied to the eye (values 0 and 1). Ungradable images were marked as -1. This indicates that the images are not of sufficient quality for analysis. In this research, images were rearranged for binary classification, normal and abnormal, and images marked with -1 (n = 3135) were excluded from the study.

Building a quality deep learning or machine learning model depends largely on the quality of existing data, its correct labeling, the presence of sufficient variation in the data, and its lack of bias. Therefore, a model have been designed to evaluate and confirm the accuracy of image labels for all images. Labelling was finalized both by the model and by submitting to the approval of the Retina specialist physician.

101.842 retinal images were included in the analysis process where 80% of the data set was allocated for training and 20% was reserved for validation. Test data set different from these images was created. Details of all test data set are described in Table 1. The test data set did not overlap with the image data used in training.

Table 1. Distribution of Test set across classes

Class	Disease	Number of data
Normal	normal	1576
Normal	normal (AU)	7
Normal	healthy-no microaneurysms	108
Abnormal	vein occlusion (AU)	135
Abnormal	vein occlusion	18
Abnormal	DME	149
Abnormal	DR	12
Abnormal	moderate non-p dr	642
Abnormal	severe non-p dr	77
Abnormal	severe prolif dr	5
Abnormal	dr level 2	56
Abnormal	dr level 3	92
Abnormal	dr level 4	106
Abnormal	wet ARMD (Age-related macular degeneration)	39
Abnormal	exudate	46
Abnormal	soft exudates	40
Abnormal	hard exudates	81

2.2. Preparing the Data

Images were resized to 528x528 resolution to prepare for the modeling phase and saved as tfrecord (a Tensorflow optimized data format). In order to read data efficiently, serializing them and storing in a series of files (100-200MB each) which can be read linearly, is important to get better performance in large data sets [20]. In one tfrecord file, more than one image and information such as label, file path, resolution, dataset source of those images were stored. In order to prevent the tfrecord files from being a bottleneck for TPUs as required by the documentation, we made sharding to keep the sizes between 100mb and 200mb, so we divided the dataset into parts so that images can be contained in more than one tfrecord file.

Some pre-processing operations and data enhancements were applied to the images in order to bring them into suitable conditions for the classification model. Some operations applied are: random horizontal flip, random crop, random adjust hue, random adjust contrast, random adjust saturation, random adjust brightness, random absolute pad image.

2.3. Proposed System

In this study, EfficientNet-B6 algorithm with different hyperparameters was used to classify the images. 101.842 retinal images were included in the analysis process and 80% of data set was allocated for training and 20% was reserved for validation. The test for different disease types (test data set) was performed with the created model. Analysis was performed using the Google Colab [21] environment using Python language.

While analyzing the performances of various pre-trained models under validation, we observed that the ideal classification performances was achieved using EfficientNet-B6. When designing the model, the first choices we make for hyperparameters do not usually lead us to correct results [12]. Therefore, in this study, the performance of the model was observed by iteratively changing the hyperparameters. The most suitable hyperparameter group was chosen for the data set. Parameter selection process was carried out with the help of libraries such as scikit-optimized [22] and tensorflow [23].

Adjusting the learning coefficient in optimization algorithms plays a critical role in the training of the model. Rectified Adam (radam) was used as the learning algorithm. This is a new variant of the classic Adam optimizer that provides an automatic, dynamic adjustment to the adaptive learning rate regarding the effects of variance and momentum during training [12, 24].

The activation function is used to control the output value and to decide whether a neuron will be active or not. Some layers were added in the hyperparameter optimization. Swish function [25] seen in Equation (1) and prelu function [26] seen in Equation (2) was used as the activation function.

$$y = x \cdot \text{sigmoid}(x) \qquad (1)$$

$$y = x.\ (1 / (1+e^{-x})) = x / (1+e^{-x})$$

$$f(y_i) = \begin{cases} y_i, & \text{if } y_i > 0 \\ a_i y_i, & \text{if } y_i \leq 0 \end{cases} \qquad (2)$$

3. Experimental Results

The dataset contains fundus images with the hemisphere of the retina surrounded by black margins. These black regions are cropped and the images are resized to 528×528 pixels (input size of EfficientNet-B6). All the images are normalized and pre-processed. Radam was used as the gradient descent algorithm and Cross-Entropy as the loss function. Batch size was selected as 128 and learning rate 0.001. To reduce overfitting and improve generalization of deep neural networks, early stopping has been applied due to validation loss.

The search space consisting of values defined for hyperparameters was scanned and EfficientNet-B6 training consisting of approximately 50 different parameter combinations was carried out. Two different hyperparameter combination features that give the best performance result at the end of the training are shown in Table 2. The "Percentage of Trainable Layers" tab in this table refers to the inclusion of the specified number of trailing EfficientNet-B6 layer blocks in the training to learn the attributes specific to the data set. It contains the information on how many layers are included in training as a percentage. During training, the first layers can be frozen (weights are fixed) and the rest can be included in the re-training process depending on the need. The "Global Pooling Layer"

tab informs which type of pooling layer is used or not. On the "Optional Layer Properties" tab presents the information regarding the added layer number.

Table 2. Hyperparameter choices

Model Name	Percentage of Trainable Layers	Global Pooling Layer	Optional Layer Properties		
			Number of Added Layers	First Layer P AF D	Second Layer P AF D
1	0.3831	max	2	128 swish 0.867104279593865	128 swish 0.527870247076625
2	0.9169	average	2	64 prelu 0.396098832448106	128 swish 0.527870247076625

P=Perceptrons of Optional Layer, AF=Activation Function of Optional Layer, D= Dropout of Optional Layer

The performance of the models is assessed using Accuracy (ACC), Loss, Sensitivity (SN) and Specificity (SP) metrics. The results are shown in Table 3. Our deep learning model achieved a sensitivity of 0.9439, a specificity of 0.8604 and an ACC of 0.86 on test set.

Table 3. ACC, Loss, Sensitivity and Specificity of our deep learning model

Model Name	Validation Set Binary ACC	Validation Set Loss	Test Set Binary ACC	Test Set Loss	Test Set Sensitivity	Test Set Specificity
1	0.80	0.437	0.82	0.404	93.52%	82.73%
2	0.85	0.356	0.86	0.328	94.39%	86.04%

Table 4 shows the ACC values of each disease for the test data set. While calculating the ACC score, each image containing a disorder was considered "Abnormal" against "Normal" images and a binary classification was made. Images obtained from Akdeniz University Hospital (AU) and graded by trusted Retina specialists as “Normal” were regarded as “Trusted/AU”. Additionally Retinal Vein Occlusion (RVO) dataset from the Akdeniz University Hospital provided by their Retina specialists (AU Vein Occlusion dataset) was used as “Vein Occlusion (trusted)”. The abnormality ratios of the test data set for all models are presented. During the labeling of the training data set from Akdeniz University Hospital (AU), it was stated whether there was an anomaly in the image, but the type of anomalies were not specified by the physicians. The test data set was used to explore how models predict 17 different disorders.

Table 4. Performance comparison for test set

Disease	Model 1 ACC	Model 2 ACC
normal	0.8179	0.8591
normal (trusted/AU)	1.0000	0.8571
healthy-no microaneurysms	0.9537	0.8796
vein occlusion (trusted)	0.9778	0.9926
vein occlusion	0.9444	0.8333
diabetic macular edema	0.9732	1.0000
dr	1.0000	1.0000
moderate non-p dr	0.8769	0.9003
severe non-p dr	1.0000	1.0000

severe prolif dr	1.0000	1.0000
dr level 2	0.8929	0.9643
dr level 3	0.9891	1.0000
dr level 4	1.0000	1.0000
wet ARMD	0.9487	0.6410
exudate	0.9783	1.0000
soft exudates	1.0000	1.0000
hard exudates	1.0000	1.0000

4. Conclusion and Evaluation

In this work, we developed a deep convolutional neural network (CNN) based on the recently developed EfficientNET-B6 CNN model. The model was fine-tuned and trained for the detection of several eye diseases. As more and more public data sets are available, the training of deep learning in the medical field is a possible choice, but the practical application of deep learning in clinical practice is still an open issue. We work for solving this problem and continue developing clinical data sets and models to create a practically usable model that will identify “referrable” retinal disorders that can be treated or are at the stage sufficiently progressed to start treatment as opposed to the “non-referrable” disorders with too early stage that doesn’t require treatment, or disorders having no known treatment methods. Important difference between the two is: diagnosing a “non-referrable” disorder will result in unnecessary visit to a retina specialist, while missing the “referrable” disorders might result in permanent blindness or vision loss.

The clinical big fundus data set (EyeCheckup) will be created and will soon be available globally. In this work, in addition to the public data sets, 21.842 color fundus images have been used from Department of Ophthalmology in Akdeniz University(AU) Hospital. More than 10 retinal disorders have been classified effectively with the proposed model. The proposed method is tested using the public datasets and the EyeCheckup dataset we created. Our deep learning model achieved sensitivity of 0.9439, specificity of 0.8604, and an ACC of 0.86 with the test data set. In the future work, possible improvent will be detection of the disease biomarkers in the fundus images using object recognition techniques.

References

[1] World Health Organization, “World Report on Vision,” Oct. 2019. Accessed: May 05, 2020. [Online]. Available: https://www.who.int/publications/i/item/world-report-on-vision.

[2] A. Grzybowski *et al.*, “Artificial intelligence for diabetic retinopathy screening: a review,” *Eye*, vol. 34, no. 3, pp. 451–460, Mar. 2020, doi: 10.1038/s41433-019-0566-0.

[3] N. Yalçin, S. Alver, and N. Uluhatun, “Classification of retinal images with deep learning for early detection of diabetic retinopathy disease,” in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, May 2018, pp. 1–4, doi: 10.1109/SIU.2018.8404369.

[4] E. Şatir, F. Azboy, A. Aydin, H. Arslan, and Ş. Haciefendioğlu, “Veri İndirgeme ve Sınıflandırma Teknikleri ile Glokom Hastalığı Teşhisi,” *El-Cezeri J. Sci. Eng.*, vol. 3, no. 3, Art. no. 3, Sep. 2016, doi: 10.31202/ecjse.258576.

[5] B. Dizdaroğlu and B. Çorbacioğlu, “Deep Diagnosis of Non-Proliferative Diabetic Retinopathy in a Mobile System,” in *2019 Medical Technologies Congress (TIPTEKNO)*, Oct. 2019, pp. 1–4, doi: 10.1109/TIPTEKNO.2019.8894946.

[6] Ö. Deperlıoğlu and U. Köse, “Diagnosis of Diabetic Retinopathy by Using Image Processing and Convolutional Neural Network,” in *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Oct. 2018, pp. 1–5, doi: 10.1109/ISMSIT.2018.8567055.

[7] Y. Yu *et al.*, “Detecting abnormal fundus images by employing deep transfer learning,” In Review, preprint, Feb. 2020. doi: 10.21203/rs.2.24133/v1.

[8] S. Gayathri, A. K. Krishna, V. P. Gopi, and P. Palanisamy, “Automated Binary and Multiclass Classification of Diabetic Retinopathy Using Haralick and Multiresolution Features,” *IEEE Access*, vol. 8, pp. 57497–57504, 2020, doi: 10.1109/ACCESS.2020.2979753.

[9] V. Sathananthavathi, G. Indumathi, and R. Rajalakshmi, “Abnormalities detection in retinal fundus images,” in *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Mar. 2017, pp. 89–93, doi: 10.1109/ICICCT.2017.7975165.

[10] A. Pak, A. Ziyaden, K. Tukeshev, A. Jaxylykova, and D. Abdullina, “Comparative analysis of deep learning methods of detection of diabetic retinopathy,” *Cogent Eng.*, vol. 7, no. 1, p. 1805144, Jan. 2020, doi: 10.1080/23311916.2020.1805144.

[11] J. Wang, L. Yang, Z. Huo, W. He, and J. Luo, “Multi-Label Classification of Fundus Images With EfficientNet,” *IEEE Access*, vol. 8, pp. 212499–212508, 2020, doi: 10.1109/ACCESS.2020.3040275.

[12] B. Bulut, V. Kalın, B. B. Güneş, and R. Khazhin, “Deep Learning Approach For Detection Of Retinal Abnormalities Based On Color Fundus Images,” in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Oct. 2020, pp. 1–6, doi: 10.1109/ASYU50717.2020.9259870.

[13] M. Chetoui and M. A. Akhloufi, “Explainable Diabetic Retinopathy using EfficientNET*,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, Jul. 2020, pp. 1966–1969, doi: 10.1109/EMBC44109.2020.9175664.

[14] EyePACS, “Diabetic Retinopathy Detection.” https://kaggle.com/c/diabetic-retinopathy-detection (accessed May 06, 2021).

[15] “Resized version of the Diabetic Retinopathy Kaggle competition dataset.” https://kaggle.com/tanlikesmath/diabetic-retinopathy-resized (accessed May 06, 2021).

[16] T. Kauppi *et al.*, “DIARETDB0: Evaluation Database and Methodology for Diabetic Retinopathy Algorithms,” p. 17.

[17] P. Porwal, “Indian Diabetic Retinopathy Image Dataset (IDRiD).” IEEE, Apr. 24, 2018, Accessed: May 06, 2021. [Online]. Available: https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid.

[18] E. Decencière *et al.*, “Feedback on a publicly distributed database: the Messidor database,” *Image Anal. Stereol.*, vol. 33, no. 3, pp. 231–234, Aug. 2014, doi: 10.5566/ias.1155.

[19] “APTOS 2019 Blindness Detection.” https://kaggle.com/c/aptos2019-blindness-detection (accessed May 06, 2021).

[20] “TFRecord ve tf.train.Example | TensorFlow Core,” *TensorFlow*. https://www.tensorflow.org/tutorials/load_data/tfrecord?hl=tr (accessed May 06, 2021).

[21] “Google Colab.” https://colab.research.google.com/notebooks/intro.ipynb#recent=true (accessed Jul. 19, 2020).

[22] “scikit-optimize: sequential model-based optimization in Python — scikit-optimize 0.7.4 documentation.” https://scikit-optimize.github.io/stable/ (accessed Jul. 19, 2020).

[23] “TensorFlow,” *TensorFlow*. https://www.tensorflow.org/ (accessed Jul. 19, 2020).

[24] L. Liu *et al.*, “On the Variance of the Adaptive Learning Rate and Beyond,” *ArXiv190803265 Cs Stat*, Apr. 2020, Accessed: Jul. 19, 2020. [Online]. Available: http://arxiv.org/abs/1908.03265.

[25] P. Ramachandran, B. Zoph, and Q. V. Le, “Swish: a Self-Gated Activation Function,” p. 12.

[26] W. QingJie and W. WenBin, “Research on image retrieval using deep convolutional neural network combining L1 regularization and PRelu activation function,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 69, p. 012156, Jun. 2017, doi: 10.1088/1755-1315/69/1/012156.