The International Society of Data Scientists
Journal of Data Science and Artificial Intelligence
Vol. 1, No. 1. Online ISSN: 2831-4794
Available at https://www.isods.org/publication/

A Deep Learning Approach using EfficientNet-based Transfer Learning for Diabetic Retinopathy Diagnosis

Cuong Do*[a], Lan Vu**[b]
[a]Department of Mathematics, University of Connecticut, 341 Mansfield Road U1009, Storrs, Connecticut 06269-1009.
[b]Department of Epidemiology and Biostatistics, Hanoi University of Public Health, 1A Duc-Thang road, North Tu-Liem District, Hanoi, Vietnam

ABSTRACT

Detecting and treating diabetic retinopathy early is crucial to preventing blindness among the working-age population worldwide. The aim of the present study is to evaluate the efficacy of deep learning models in diagnosing diabetic retinopathy. The EfficientNets, which are claimed to be more efficient than many current well-performing network architectures, were tested to see how transfer learning performed with color fundus photography as input.

Keywords: Deep Learning, Convolutional Neural Networks, classification, Transfer Learning

***First Authors**, E-mail: cuong.do@engineer.uconn.edu, vhl@huph.edu.vn
*Cuong Do and Lan Vu contributed equally to this manuscript

1. INTRODUCTION

Globally diabetes mellitus (DM) is one of the main health concerns. Diabetes patients can lose their vision and become blind due to diabetic retinopathy (DR). Approximately 2.6 million people worldwide are visually impaired due to DR [1]. Because of the increasing trends of DM in developed countries and the poor management of DM in developing countries, the burden of visual impairment and blindness caused by DR is expected to rise worldwide. More than 30% of people with DM have never had their eyes checked [2].

The prevention of vision impairment is highly dependent on early detection of DR. DR is currently diagnosed by examining digital color fundus photographs. It is estimated that in a population of 1 million, there are between 30,000 and 50,000 people with diabetes, of whom 5000 to 10,000 suffer from diabetic retinopathy, with 200 to 1,000 of them requiring laser photocoagulation [3]. This manual screening process requires time, expertise and equipment which are often lacking in low resource areas [4], where the rate of diabetes in local populations is high and DR detection is most needed. As the number of individuals with diabetes continues growing, the infrastructure needed to prevent blindness due to DR will become even more insufficient.

There has been research that suggests using color fundus photography as input to automated models to screen for diabetic retinopathy. Several studies have shown that automated computer-aided diagnostic tools powered by deep learning algorithms using convolutional neural networks (CNN) are an effective means of analyzing chest color fundus photographs. Deep Learning is applied to learn the feature representations from medical images to identify typical disease manifestations and localize suspicious regions of interest. A study by Gulsan et al found that deep learning can accurately detect diabetic retinopathy in images of the retinal fundus and macular edema in images of the retina [5]. Ting et al. found that Deep Learning can detect referable diabetic retinopathy, vision-threatening diabetic retinopathy, possible glaucoma, and age-related macular degeneration with high sensitivity and specificity that was superior to professional graders [6]. According to Gargeya et al., a fully data-driven artificial intelligence-based grading algorithm can accurately identify which diabetic patients should be referred to an ophthalmologist by analyzing fundus photographs obtained from diabetics and then refer individual cases to an ophthalmologist for further evaluation and treatment [7].

The purpose of this study is to investigate a method for detecting DR in deep learning. Implementations of the model could be developed and hosted on the cloud to serve the needs, especially at low-resource locations.

2. NETWORK ARCHITECTURES

Transfer learning refers to the reuse of learned weights of a model for another task. The lower layers learn low-level structures of the images, which can be considered the basis for constructing the upper layers. The weights of the base model are frozen, allowing only the new deep layers to be trained. In the proposed architectures, all the dense layers are removed, and new deep layers are built on the top [8].

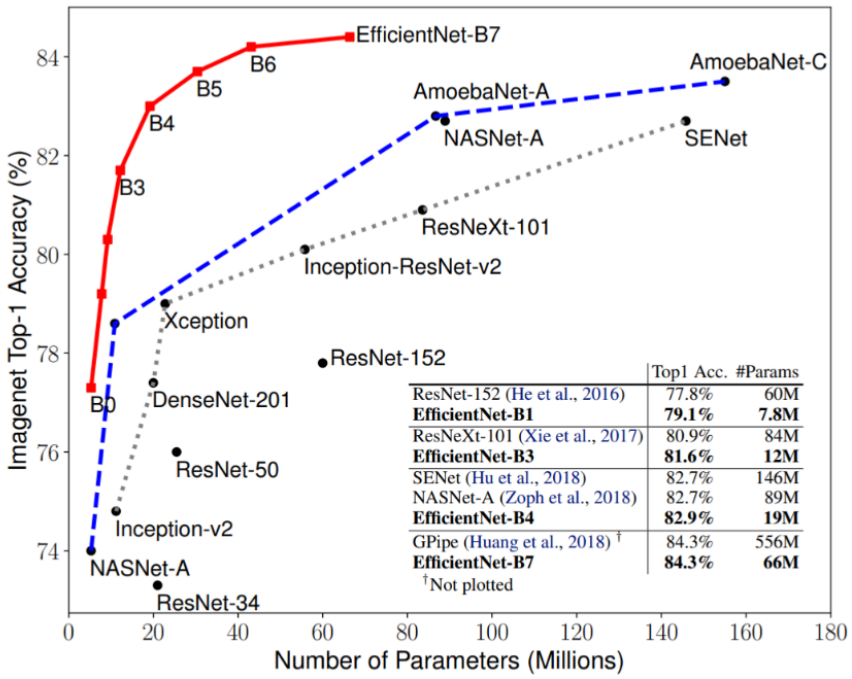


Figure 1. Model Size vs. Accuracy Comparison. EfficientNet-B0 is the baseline network developed by AutoML MNAS, while Efficient-B1 to B7 are obtained by scaling up the baseline network.

Several popular architectures of CNNs have recently made a significant contribution to the success of various research efforts involving transfer learning, including VGG, Inception, Xception, ResNet, and DenseNet. In 2014 Simonyan and Zisserman presented the VGG architecture, which consists of 3x3 convolutional layers stacked on top of each other along the depth of the architecture [9]. Introducing the Inception architecture in 2014, Szegedy et al. made use of dense modules/blocks to increase resource utilization [10]. The updated VGG and Inception variants are designed to improve performance primarily by increasing number of layers in the architecture. ResNet solved the issue of degraded accuracy of deep networks by using skip connections in building deeper models [11]. Based on Inception V3 and using some ideas borrowed from Microsoft ResNet documents, Inception-ResNet-v2 is a variant of the Inception V3 concept [12]. The Xception architecture is an extension of Inception, which introduces CNNs based on depth-wise programmable convolutional layers [13].

Huang et al., in 2018, introduced densely linked convolutional networks (DenseNets). The technique can reduce the vanishing gradient problem, enhance the propagation of features, encourage the reuse of features, and significantly reduce the number of parameters. In comparison with ResNet or other architectures, DenseNet achieves greater accuracy with fewer parameters. DenseNet-201 is a 201-layer convolutional neural network[14].

The EfficientNet model scaling method was recently released by Google AI, which allows CNNs to be scaled up in a more structured way. Instead of arbitrarily scaling network dimensions based on width, depth and resolution, the method uniformly scales each dimension based on a fixed set of scaling coefficients. EfficientNet became the new state-of-the-art

accuracy model with increased efficiency by developing a family of EfficientNet models. EfficientNet-B0 is the baseline network, which is scaled up to obtain other variants from EfficientNet-B1 up to B7 (See Figure 1) [15].

3. MODELING AND RESULT

This dataset is provided to the APTOS 2019 Blindness Detection Competition by the Aravind Eye Hospital in India. Each instance has been rated according to its severity, including 0 - No DR, 1 – Mild, 2 – Moderate, 3 – Severe, 4 - Proliferative DR (See Figure 2). A total of 3662 instances are included in the training dataset for each of the five classes (See Figure 3). The dataset is highly imbalanced.

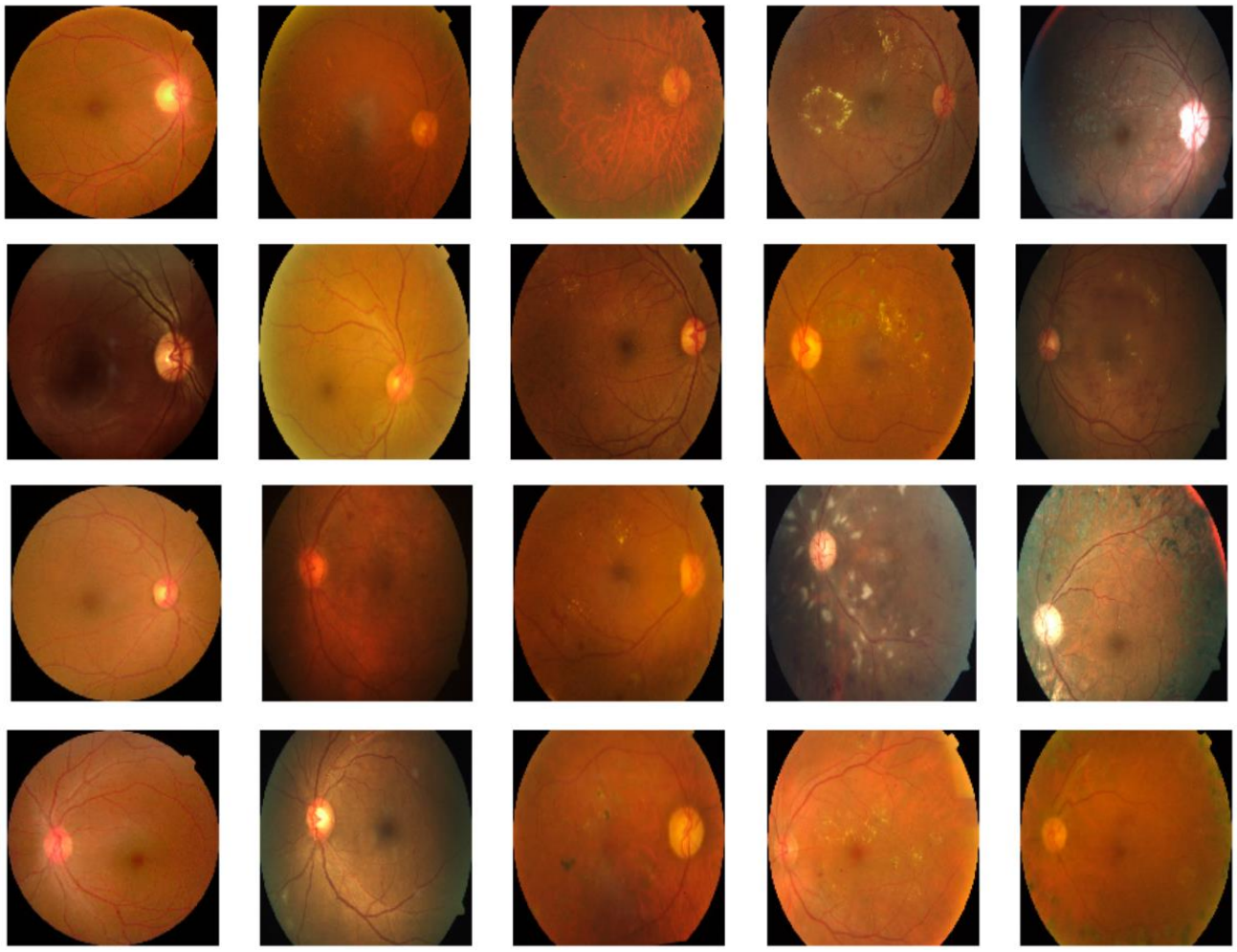


Figure 2. Sample images for class 0, 1, 2, 3, and 4, in columns 1, 2, 3, 4, and 5 respectively

The dataset is divided such that 10% is used for testing the model performance, the rest is divided further such that 10% is used for validation and 90% for training. We resize the images to make them fit the input format of the selected network architecture, which for EfficientNetB5 is 456x456x3. The training dataset is augmented by rotating clockwise a given number of degrees from 0 to 360, in order to add more data and overcome overfitting.

The authors of the EfficientNets architectures claimed that it is superior than the state-of-the-art accuracy on the ImageNet image database, with up to 10x better efficiency. Figure 1 shows that from the smallest to the largest EfficientNets architecture, accuracy increases while the model size remains relatively small. The accuracy of several large EfficientNets is much higher than that of the other architectures described in section 2. Our focus in this section is on these architectures.

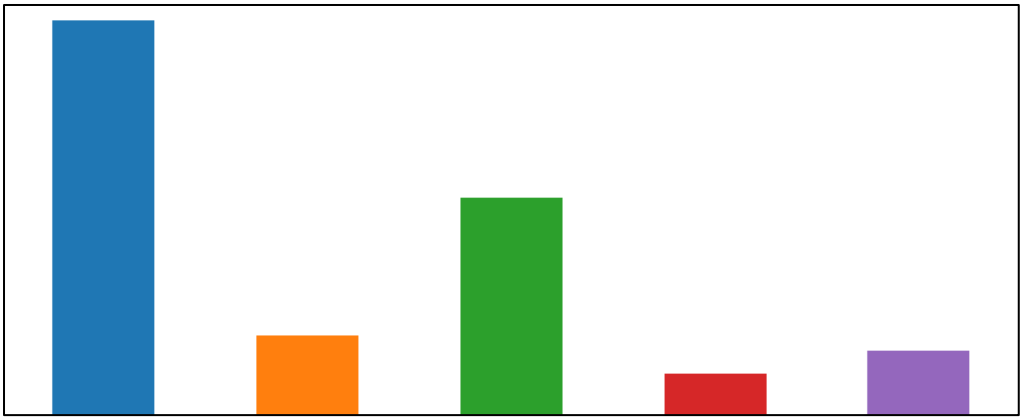


Figure 3. Distribution of class labels

A model based on EfficientNetB5 is chosen for transfer learning, in which the top is removed, and we build an average pooling layer, a dropout layer, and two dense layers on the base top. The problem is modeled as a regression model with mean square error as the loss function. Predictions are rounded to calculate Quadratic Weighted Cohen's kappa. During training at the end of each epoch, the model is saved if its Kappa score improves. An early stopping condition ensures that the model will not overfit when the validation loss degrades while the training loss continues to decline. As learning stagnates, validation loss is monitored to reduce the learning rate. The best Cohen's kappa of 93.96% is obtained for validation data, and 92.34% is obtained for test data, which is in the top range of Kappa value interpretation by Landis & Koch [16].

4. CONCLUSION

This paper describes the proposed approach based on transfer learning using EfficientNetB5, a member of the new state-of-the-art EfficientNets family. As for the base for new trainable layers, we borrowed weights from the pretrained model and froze them before adding new layers on top of it. EfficientNets architectures require smaller number of parameters. Additionally, as a result of transfer learning, much fewer parameters need to be trained in comparison to training a whole network from scratch. We dealt with overfitting by dropping nodes at random during training, augmenting training data, and using cross validation. It was concluded that the method could detect and classify diabetic retinopathy at different stages automatically.

REFERENCES

[1] Lee, R., Wong, T. Y. and Sabanayagam, C., "Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss," Eye Vis (Lond) **2** (2015).

[2] Elshafei, M., Gamra, H., Khandekar, R., Al Hashimi, M., Pai, A. and Ahmed, M. F., "Prevalence and determinants of diabetic retinopathy among persons ≥ 40 years of age with diabetes in Qatar: a community-based survey," Eur J Ophthalmol **21**(1), 39–47 (2011).

[3] Dandona, L., Dandona, R., Naduvilath, T. J., McCarty, C. A. and Rao, G. N., "Population based assessment of diabetic retinopathy in an urban population in southern India," Br J Ophthalmol **83**(8), 937–940 (1999).

[4] Vashist, P., Singh, S., Gupta, N. and Saxena, R., "Role of Early Screening for Diabetic Retinopathy in Patients with Diabetes Mellitus: An Overview," Indian J Community Med **36**(4), 247–252 (2011).

[5] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L. and Webster, D. R., "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," JAMA **316**(22), 2402–2410 (2016).

[6] Ting, D. S. W., Cheung, C. Y.-L., Lim, G., Tan, G. S. W., Quang, N. D., Gan, A., Hamzah, H., Garcia-Franco, R., San Yeo, I. Y., Lee, S. Y., Wong, E. Y. M., Sabanayagam, C., Baskaran, M., Ibrahim, F., Tan, N. C., Finkelstein,

E. A., Lamoureux, E. L., Wong, I. Y., Bressler, N. M., et al., "Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes," JAMA **318**(22), 2211–2223 (2017).

[7] Gargeya, R. and Leng, T., "Automated Identification of Diabetic Retinopathy Using Deep Learning," Ophthalmology **124**(7), 962–969 (2017).

[8] Yosinski, J., Clune, J., Bengio, Y. and Lipson, H., "How transferable are features in deep neural networks?," arXiv:1411.1792 [cs] (2014).

[9] Simonyan, K. and Zisserman, A., "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556 [cs] (2015).

[10] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., "Rethinking the Inception Architecture for Computer Vision," arXiv:1512.00567 [cs] (2015).

[11] He, K., Zhang, X., Ren, S. and Sun, J., "Deep Residual Learning for Image Recognition," arXiv:1512.03385 [cs] (2015).

[12] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A., "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," arXiv:1602.07261 [cs] (2016).

[13] Chollet, F., "Xception: Deep Learning with Depthwise Separable Convolutions," arXiv:1610.02357 [cs] (2017).

[14] Huang, G., Liu, Z., van der Maaten, L. and Weinberger, K. Q., "Densely Connected Convolutional Networks," arXiv:1608.06993 [cs] (2018).

[15] Tan, M. and Le, Q. V., "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," arXiv:1905.11946 [cs, stat] (2020).

[16] Landis, J. Richard, and Gary G. Koch. "The Measurement of Observer Agreement for Categorical Data." Biometrics, vol. 33, no. 1, 1977, pp. 159–174. JSTOR, www.jstor.org/stable/2529310.