Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 216 (2023) 57–66

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Computer Science and Computational Intelligence 2022

Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning

Mira Hayati[a], Kahlil Muchtar[a,b,*], Roslidar[a], Novi Maulina[c], Irfan Syamsuddin[d], Gregorius Natanael Elwirehardja[e], Bens Pardamean[e,f]

[a]Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia
[b]Telematics Research Center (TRC) Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia
[c]Faculty of Medicine, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia
[d]Department of Computer and Networking Engineering, Politeknik Negeri Ujung Pandang, Makassar, 90245, Indonesia
[e]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, 11480, Indonesia
[f]Computer Science Department, BINUS Graduate Program – Master of Computer Science Program, Bina Nusantara University, Jakarta, 11480, Indonesia

Abstract

Diabetic retinopathy (DR) is a type of diabetes mellitus that attacks the retina of the eye. DR will cause patients to experience blindness slowly. Generally, DR can be detected by using a special instrument called an ophthalmoscope to view the inside of the eyeball. However, in conditions where there is a very small difference between the normal image and the DR image, computer-based assistance is needed for maximizing image reading value. In this research, a method of image quality improvement will be carried out which will then be integrated with a classification algorithm based on deep learning. The results of image improvement using Contrast Limited Adaptive Histogram Equalization (CLAHE) shows that the average accuracy of the method on several models is very competitive, 91% for the VGG16 model, 95% for InceptionV3, and 97% for EfficientNet compared to the results original image which only has an accuracy of 87% for VGG16 model, 90% for InceptionV3 model, and 95% for EfficientNet. However, in ResNet34 better accuracy is obtained in the original image with an accuracy of 95% while in the CLAHE image the accuracy value is only 84%. The results of this comprehensive evaluation and recommendation of famous backbone networks can be useful in the computer-aided diagnosis of diabetic retinopathy.

Keywords: Diabetic retinopathy; Deep Learning;CNN;Image Enhancement; CLAHE

* Corresponding author. Tel.: +62 651-755 3205
E-mail address: kahlil@unsyiah.ac.id

10.1016/j.procs.2022.12.111

1. Introduction

Among diabetic patients, the Diabetic Retinopathy (DR) has been known as one of the leading causes of vision loss. The main complication in people diagnosed with diabetes is elevated glucose levels, which can affect various organs, such as the liver, eyes, and brain. Diabetes in particular may impair the retina, a phenomenon known as DR [1]. It is a microvascular complication of diabetes mellitus that targets the blood vessels in the retina of the eye. To identify this disease, manual identifications have to be performed by doctors, an approach that is known to be extensive. Therefore, a method is needed to facilitate and assist ophthalmologists in diagnosing DR by using retinal images to assist doctors in performing faster examinations and making more accurate diagnosis.

Epidemiology research in Europe, America, Asia, and Australia reports that the number of people with diabetic retinopathy will increase from 100.8 million in 2010 to 154.9 million in 2030, with 30% of them at risk of blindness. Diabetic retinopathy disease begins with the destruction or weakening of the small capillaries in the retinal area, leaking blood, which can then cause thickening of the retinal tissue, swelling, and more extensive bleeding. So that over time will cause blurry vision, reduced or distorted. If it is not detected properly and correctly, vision will get worse, and eventually, blindness occurs [2].

Despite the worrisome statistics, studies have shown that more than 90% of new DR cases could be reduced under the condition that proper care, alertness, and monitoring of the eyes were put in place. DR is diagnosed in five stages: mild, moderate, severe, proliferative, or no disease. The characteristics of DR include microaneurysms, abnormal growth of new blood vessels, leaking blood vessels, retinal swelling, and also damaged nerve tissue [3]. As such, Computer-Aided Diagnosis (CAD) systems will be significantly beneficial to enable early detection.

The Artificial Intelligence (AI) technology, especially deep learning, have been widely deployed for medical image classification tasks aimed to assist early disease detection. The Convolutional Neural Network (CNN) in particular have obtained high accuracy in a wide variety of such tasks. In 2018, Pardamean et al. used the CheXNet model to classify breast cancer from mammograms and obtained 90.38% accuracy [4]. In a similar task, Muljo et al. used the DenseNet121 CNN model to classify lung diseases [5]. Izzaty et al. had also used several CNN) architectures and compared them to classify histology images and detect colorectal cancer cases [6]. In all of these studies, the CNN had proven to be both reliable and accurate.

A previous work proposed by Adriman et al. [7] proposed a deep learning model for DR classification task with two main parts, namely LBP-based feature extraction and CNN. The CNN architectures used are ResNet, DenseNet, and DetNet. Based on the experiments conducted, the system has shown a good ability to detect DR. Experiments carried out in this study showed that the proposed method achieved an average accuracy of 96,35% (ResNet), 84,05% (DenseNet), and 93,99% (DetNet). Apart from deep learning evaluations, it is also compared against SVM for training and testing purposes where the average value obtained is 83,33%.

Queentinela and Triyani [8] classified DR disease by applying a system based on Deep Learning. Three CNN architectures were evaluated in the research, which are AlexNet [9], GoogleNet [10], and SqueezeNet [11]. The STARE (Structured Analysis of the Retina) dataset [12] was used in training the models. This system detects and classifies diabetic retinopathy based on the characteristics of the appearance of microaneurysms, hard exudates, soft exudates, and bleeding in the form of dots, lines, and spots on the retina. The accuracy values obtained from this study are 64.4%, 66.7%, and 75.6%, for the AlexNet [9], GoogleNet [10], and SqueezeNet [11] models, respectively. All in all, the research had proven that these CNN architectures are suitable to be deployed on this task albeit further improvements are necessary. Rizal, et al. [13] had also conducted research by designing a system that can detect Diabetic Retinopathy by using Convolutional Neural Network (CNN). The EfficientNet architecture was trained on datasets that have been pre-processed previously. The proposed model obtained an accuracy of 79.8% in classifying 5 levels of Diabetic Retinopathy.

Based on the problems above, a new diagnosis system for DR disease with higher accuracy level is urgently needed. Image classification is the process of grouping images into certain classes. In performing the classification, the quality of the images must first be improved using image enhancement, such as CLAHE and its variant due to the low contrast of the input images [14,15]. Hence, in this study, a computer vision-based model that utilizes CLAHE pre-processing is proposed to classify DR through deep learning. The models evaluated in this study are ResNet34 [16], VGG16 [17], InceptionV3 [18], and EfficientNetB4 [19] architectures. It is noteworthy that the aforementioned networks have been widely used for comparative evaluation due to its performance and popularity as the backbone networks [20–22].

2. Proposed Approach and Dataset

Several CNN architectures were used in classifying DR and CLAHE pre-processing was implemented to improve the quality of the images. CLAHE is utilized to produce images with better quality, which may assist the CNN models to distinguish the images. The CLAHE pre-processed image is used as input for deep learning to be trained with the CNN architecture. As shown in the literature, compared with other approaches, CLAHE and its variant are commonly used in order to improve the low contrast medical images that frequently affect the clinical diagnosis step [14,15,23].

2.1. Dataset

The dataset used is the APTOS 2019 dataset, which was obtained from Kaggle [24] with a total of 3,288 images contained. It consisted of 1,799 negative images and 1,489 positive images for DR classification task.

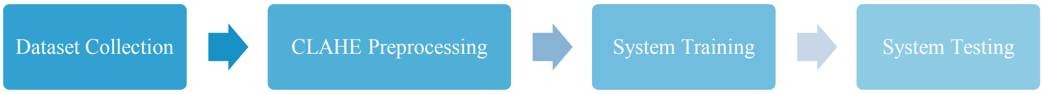


Fig. 1. The proposed flowchart.

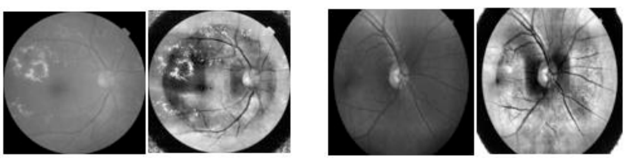
2.2. Implementation of CLAHE

CLAHE is an improvement method from the previous method, namely AHE. Because AHE has a contrast level that can be excessive, CLAHE is made to provide a contrast limit so that the expected results are achieved. The problem in the form of excessive contrast enhancement in the AHE method can be overcome by using Contrast Limited Adaptive Histogram Equalization (CLAHE), which gives a limit value to the histogram [25]. The CLAHE process aims to obtain an image with better contrast without compromising the quality of the image itself.

The CLAHE method operates on a small area of the image which is usually called a tile. The contrast contained in each tile is corrected so that the histogram generated from that area matches the specified histogram shape. Adjacent tiles are connected using bilinear interpolation. This method is done so that the results of the merging of the tiles look smooth. The CLAHE method is defined as follows:

β represents the limit value (clip limit), variable M represents the area size, N represents the grey-level value (256), α represents clip factor expresses the addition of a histogram limit with a value of 1 to 100 [26]. S_{max} is the maximum allowable slope.

$$\beta = \frac{M}{N}\left(1 + \frac{\alpha}{100}\left(S_{max} - 1\right)\right) \tag{1}$$



(a) (b)

Fig. 2. Different Between improved image through CLAHE and Original image (a) Positive Image (b) Negative Image.

2.3. Model Training

The dataset was split into 3 subsets, which are the training set used in training the DR classification models, the validation used in testing the models during training, and an independent test set to further assess the models an unseen data. To be more specific, as recommended in previous works [27,28], each class of the dataset was divided with a ratio of 7:2:1. 70% of the data was used in training the models, 20% for validation, and the other 10% was used as the independent test set. The distribution of samples in each subset are listed in Table 1.

Table 1. Details of image distributions

Types	Infected	Uninfected
Training Data	1.040	1.259
Validation Data	297	362
Testing Data	149	178
Total	1.489	1.799

The models were trained using the SGD (Stochastic Gradient Descent) optimizer. SGD computes the gradient parameters using only a single or a few training examples. This led to a faster training process. It should be noted that two types of experiment were conducted. On the first one, the models were trained on the images after CLAHE was applied, and the other one used the original images. Detailed configurations of the hyperparameters used are available in Table 2. The configurations used is set by taking the large number of trained datasets and the limitations of the devices used into account. These hyperparameters were fine-tuned through random search.

Table 2. Hyperparameter configurations for the experiment

Hyperparameter	Values
Optimizer	SGD
Epoch	16
Batch Size	4
Learning rate	0.01
Learning rate decay	0.5
Learning rate step size	3
Weight decay	1×10^{-5}
Number of workers	2
Image size	224

2.4. Model Evaluation

At this stage, the four models were evaluated. The metric used in assessing the models is the accuracy score, calculated using the equation below. The model is said to be good if the accuracy value obtained is high enough with at least 90% accuracy.

$$Accuracy = \frac{Number\ of\ true\ positives}{Number\ of\ data} \quad (2)$$

3. Results

3.1. Model Training Results

Table 3. Training results for the experiment using CLAHE

Model	VGG16	ResNet34	InceptionV3	EfficientNetB4
Train Time	25,28 minutes	13,8 minutes	3 hour 54 minutes	4 hour 30 minutes
Epoch	16	16	16	16
Train acc	0.9957	0.9974	0.9987	**1.0000**
Train loss	0,0044145	0.0023345	0.00097	**0.00044**
Val acc	0.9803	0.9772	0.9757	**0.9833**
Val loss	**0.02021**	0.04385	0.02867	0.11897

Table 4. Training results for the experiment using the original images

Model	VGG16	ResNet34	InceptionV3	EfficientNetB4
Train Time	1 hour 12 minutes	30 minutes	3 hour 51 minutes	4 hour 15 minutes
Epoch	16	16	16	16
Train acc	0.9861	0.9987	**1.0000**	0.9983
Train loss	0.0142	0.0013	**0.0005**	0.0116
Val acc	**0.9803**	0.9757	0.9742	0.9742
Val loss	**0.0159**	0.0331	0.0273	0.0678

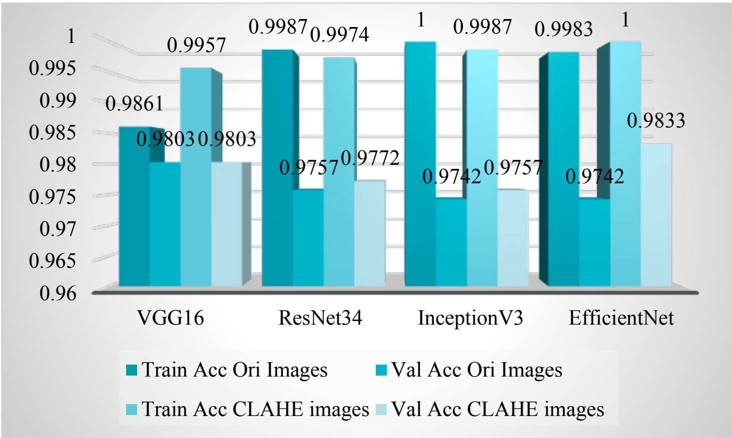


Fig 3. Comparison of Training Accuracy and Validation Accuracy of Models Trained Using the Original Images and CLAHE Images.

The results of the training phase are summarized in Table 3 and Table 4. Both the ResNet34 and InceptionV3 architecture has a higher training accuracy when trained on the original images compared to the CLAHE images. However, the validation accuracy is higher when the models are trained using CLAHE images. Similar results on the validation set can also be observed on the VGG16 and EfficientNetB4. These results indicate that CLAHE allows more distinguishing features to be observed by the CNNs, hence the validation accuracy becomes higher. For the VGG16 and EfficientNetB4 architectures, it is clear that using CLAHE resulted in better training accuracy compared to the models trained on the original images. Thus, it can be inferred that performing CLAHE on the DR images can yield significant improvement on the models, especially on lighter models, as EfficientNetB4 possess the least number of parameters among the four models despite being the deepest model. The validation loss of this model also got reduced by more than 0.05. To further validate this finding, the evaluation results on the independent test set are further analyzed in section 3.2.

3.2. Model Evaluation Results

Table 5. Average accuracy of the models on the test set

	VGG16	ResNet34	InceptionV3	EfficientNetB4
Original	0.8740	**0.9596**	0.9080	0.9557
CLAHE	**0.9187**	0.8443	**0.9520**	**0.9783**

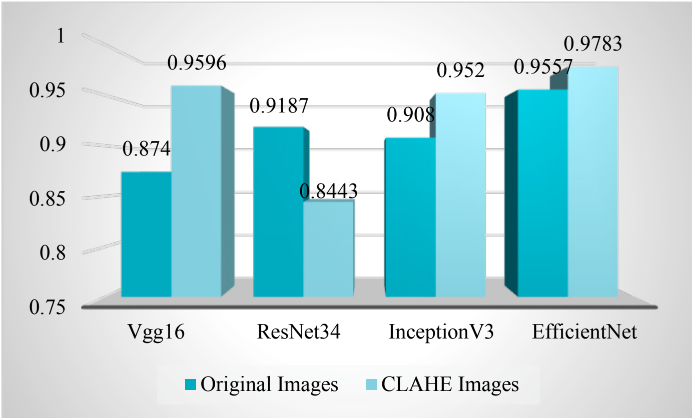
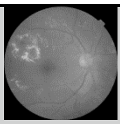
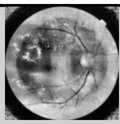
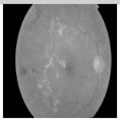
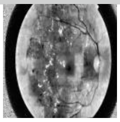
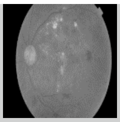
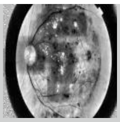
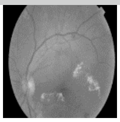
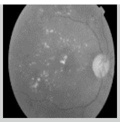
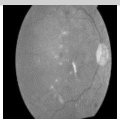
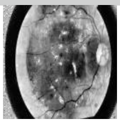


Fig 4. Comparison the models' performance on the test set

Table 6. Illustration of EfficientNetB4's prediction results for positive images

Original Images	Softmax Confidence Score	CLAHE Images	Softmax Confidence Score
	0.9995		0.9928
	0.9999		0.9999
	0.9999		0.9998
	0.9999		0.9995
	0.9999		0.9987
	0.9999		0.9998

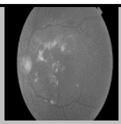
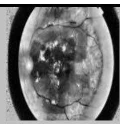
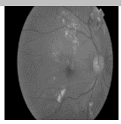
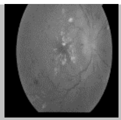
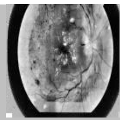
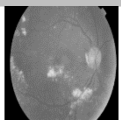
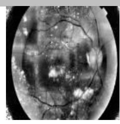
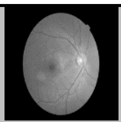
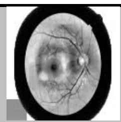
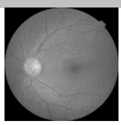
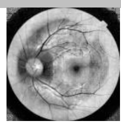
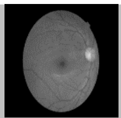
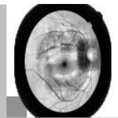
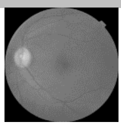
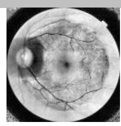
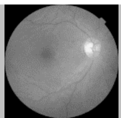
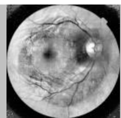
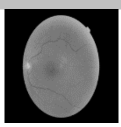
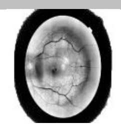
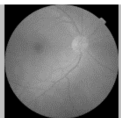
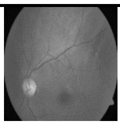
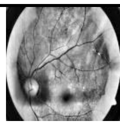
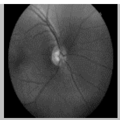
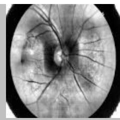
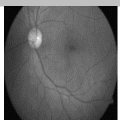
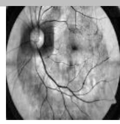
	0.9999		0.9995
	0.9999		0.9998
	0.9999		0.9994
	0.9999		0.9994

Table 7. Illustration of EfficientNetB4's prediction results for positive images

Original Images	Softmax Confidence Score	CLAHE Image	Softmax Confidence Score
	0.9985		0.9993
	0.9975		0.9974
	0.9998		0.9995
	0.9987		0.9920
	0.9965		0.9964
	0.9999		0.9991
	0.9365		0.9895

	0.9374		0.9536
	0.9998		0.9997
	0.9743		0.9685

After accumulating the accuracy values, it can be seen that for most of the models, training using the CLAHE images brought about better results compared to the models trained on original images. The VGG16, InceptionV3, and EfficientNetB4 models obtained 5.11%, 4.85%, and 2.36% improvement in accuracy, respectively. However, for ResNet34 the accuracy got reduced by 12.02% when trained on the CLAHE images. As no fine-tuning had been performed in this research, such results are logical as each model may require different hyperparameter values to achieve optimal results. The highest accuracy value was obtained by using CLAHE on the dataset and deploying the EfficientNetB4 model, which attained an astounding 97.83% accuracy.

When compared to the models' performance on the validation set, it is possible that the other models overfit except the EfficientNetB4. The VGG16 and ResNet34 in particular experienced the largest drop in accuracy on both versions of the images. However, the accuracy drop is larger on the models trained on the original images for VGG-16 and InceptionV3. As such, it can be inferred that CLAHE allowed the models to extract more general features, allowing them to maintain their accuracy on the test set. The EfficientNetB4 had proven to be the best model in this task with 98.33% and 97.83% accuracy on the validation and test sets, respectively. All in all, CLAHE allowed most of the models to achieve better accuracy compared to using the original images.

3.3. Discussion

Although CLAHE allowed most of the models to obtain better accuracy, it can be seen that not all of the models improved. However, such results may have been affected by the hyperparameter settings used as utilizations of hyperparameter grid-search can generally allow CNNs to perform better with personalized hyperparameter settings. As ResNet-34 is neither distinctively different from the other models in terms of depth and number of parameters aside from the back-end mechanisms, specifically the residual attention, finding the optimal hyperparameter values may be the key to enhance its accuracy, which is a limitation of the current study. Overall, the models generally obtained better results after being trained on the enhanced images. The VGG16 and InceptionV3 backbones in particular obtained the most improvement, with 4.47% and 4.4% higher accuracy, respectively. Such results imply that by enhancing the contrast, the models were able to extract better distinguishing features. However, the EfficientNet architecture once again proved how compound scaling allows the model to outperform other architectures. In the future, deeper analysis such as by using Class Activation Maps (CAM) to further analyze how CLAHE affects the feature extraction process can be studied.

4. Conclusions

The purpose of this study is to evaluate the impact of image enhancement on the accuracy of DR classification. On the four tested models, which are ResNet34, VGG16, InceptionV3, and EfficientNetB4, the method produced generally better results. The best accuracy was achieved by EfficientNetB4, with 98.33% and 97.83% accuracy on the validation and test sets, respectively when using the CLAHE images. On the original images, its accuracy got reduced slightly, a phenomenon that is also observed on the VGG16 and InceptionV3 models. As such, it was further proven that image enhancements for medical images may lead to better results. This means that other enhancement methods may be further explored to improve the performance of AI models in Computer-Aided

Diagnosis (CAD). In the future, deeper analysis on the impact of image quality enhancement for DR classification can be conducted to further validate the findings of this study. Based on our findings, it can be concluded in general that by combining image quality improvement with deep learning, DR classification could be enhanced.

Acknowledgements

We thank Nona Zarima for assistance in conducting the research.

References

[1] Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In: Proceedings of 12th International Conference on Pattern Recognition. 1994. p. 582–5 vol.1.

[2] Sembiring K. Penerapan teknik support vector machine untuk pendeteksian intrusi pada jaringan. Institut Teknologi Bandung. 2007;

[3] Elveny M, Anjulina T, Siregar B, Syah R. Identification of Diabetic Retinopathy with Retinal Fundus Imagery Using Probabilistic Neural Network. J Phys Conf Ser. 2020 Nov;1641(1):12055.

[4] Pardamean B, Cenggoro TW, Rahutomo R, Budiarto A, Karuppiah EK. Transfer Learning from Chest X-Ray Pre-trained Convolutional Neural Network for Learning Mammogram Data. Procedia Comput Sci. 2018;135:400–7.

[5] Muljo HH, Pardamean B, Purwandari K, Cenggoro TW. Improving lung disease detection by joint learning with COVID-19 radiography database. Commun Math Biol Neurosci. 2022;2022.

[6] Izzaty AMK, Cenggoro TW, Elwirehardja GN, Pardamean B, others. Multiclass classification of histology on colorectal cancer using deep learning. Commun Math Biol Neurosci. 2022;2022.

[7] Adriman R, Muchtar K, Maulina N. Performance Evaluation of Binary Classification of Diabetic Retinopathy through Deep Learning Techniques using Texture Feature. Procedia Comput Sci. 2021;179:88–94.

[8] Queentinela VA, Triyani Y. Klasifikasi Penyakit Diabetic Retinopathy pada Citra Fundus Berbasis Deep Learning. In: 9th Applied Business and Engineering Conference. 2021.

[9] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira F, Burges CJ, Bottou L, Weinberger KQ, editors. Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2012.

[10] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going Deeper With Convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.

[11] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. arXiv; 2016.

[12] Hoover AD, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans Med Imaging. 2000;19(3):203–10.

[13] Rizal S, Ibrahim N, Pratiwi NKC, Saidah S, Fu'adah RYN. Deep Learning untuk Klasifikasi Diabetic Retinopathy menggunakan Model EfficientNet. ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika. 2020;8(3).

[14] Dabass J, Arora S, Vig R, Hanmandlu M. Mammogram Image Enhancement Using Entropy and CLAHE Based Intuitionistic Fuzzy Method. In: 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN). 2019. p. 24–9.

[15] Salem N, Malik H, Shams A. Medical image enhancement based on histogram algorithms. Procedia Comput Sci. 2019;163:300–11.

[16] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

[17] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv; 2014.

[18] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.

[19] Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Chaudhuri K, Salakhutdinov R, editors. Proceedings of the 36th International Conference on Machine Learning. PMLR; 2019. p. 6105–14. (Proceedings of Machine Learning Research; vol. 97).

[20] Saleem MH, Potgieter J, Arif KM. Plant Disease Classification: A Comparative Evaluation of Convolutional Neural Networks and Deep Learning Optimizers. Vol. 9, Plants. 2020.

[21] Afif M, Ayachi R, Said Y, Pissaloux E, Atri M. An Evaluation of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation. Neural Process Lett. 2020;51(3):2265–79.

[22] Wang G, Wu M, Wei X, Song H. Water Identification from High-Resolution Remote Sensing Images Based on Multidimensional Densely Connected Convolutional Neural Networks. Vol. 12, Remote Sensing. 2020.

[23] Li X, Li T, Zhao H, Dou Y, Pang C. Medical image enhancement in F-shift transformation domain. Health Inf Sci Syst. 2019;7(1):13.

[24] Asia Pacific Tele-Ophthalmology Society (APTOS). APTOS 2019 Blindness Detection Detect diabetic retinopathy to stop blindness before it's too late [Internet]. 2019. Available from: https://www.kaggle.com/c/aptos2019-blindness-detection/

[25] kumar Rai R, Gour P, Singh B. Underwater image segmentation using clahe enhancement and thresholding. International Journal of Emerging Technology and Advanced Engineering. 2012;2(1):118–23.

[26] Reza AM. Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. J VLSI Signal Process Syst Signal Image Video Technol. 2004;38(1):35–44.

[27] Hu S, Luan X, Wu H, Wang X, Yan C, Wang J, Liu G, He W, Bai HX, Wang R, Xiong Z, Hsieh B, Chang K, Halsey K, Tran TML, Choi JW, Wang DC, Shi LB, Mei J, Jiang XL, Pan I, Zeng QH, Hu PF, Li YH, Fu FX, Huang RY, Sebro R, Yu QZ, Atalay MK, Liao WH. ACCV: automatic classification algorithm of cataract video based on deep learning. Biomed Eng Online. 2021 Sep;20(1):78.

[28] Bai HX, Wang R, Xiong Z, Hsieh B, Chang K, Halsey K, Tran TML, Choi JW, Wang DC, Shi LB, Mei J, Jiang XL, Pan I, Zeng QH, Hu PF, Li YH, Fu FX, Huang RY, Sebro R, Yu QZ, Atalay MK, Liao WH. Artificial Intelligence Augmentation of Radiologist Performance in Distinguishing COVID-19 from Pneumonia of Other Origin at Chest CT. Radiology. 2020 Sep;296(3):E156–65.