MobileNetV2-based Deep Learning for Retinal Disease Classification on a Mobile Application

Pattarakorn Intaraprasit
Department of Biomedical Engineering,
School of Engineering,
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
63011245@kmitl.ac.th

Toan Huy Bui
College of Science and Technology,
Tokai University,
Kanagawa 259-1292, Japan.
9LTAD006@cc.u-tokai.ac.jp

May Phu Paing*
Department of Biomedical Engineering,
School of Engineering,
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
may.pa@kmitl.ac.th
*Corresponding author

2023 15th Biomedical Engineering International Conference (BMEiCON) | 979-8-3503-4524-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/BMEiCON60347.2023.10322079

***Abstract*— (1) Background: Cataract, glaucoma, and diabetic retinopathy are prevalent retinal diseases that can have significant adverse effects on vision and eye health. Cataracts result in the clouding of the lens, leading to blurred vision and reduced visual acuity. Glaucoma damages the optic nerve, causing peripheral vision loss and the potential for blindness. Diabetic retinopathy, which arises as a complication of diabetes, impacts the blood vessels in the retina and can result in vision impairment and blindness. Timely detection, appropriate treatment, and regular monitoring play a crucial role in effectively managing these conditions and minimizing their impact on vision. (2) Methods: Using the MobileNetV2 deep learning network, an automated classification technique for retinal disease detection from optical coherence tomography (OCT) images was created. The OCT scans, with a sample size of 224 × 224 pixels, are divided into four classes— cataract (CAT), glaucoma (GLC), diabetic retinopathy (DR), and normal retina. Using MobileNetV2 makes it simple to export models into Android mobile apps for point-of-care diagnosis. Using validation samples, the model was adjusted, and its accuracy and sensitivity were assessed. Evaluation metrics such as accuracy, precision, recall, and F1-score were used to assess the performance of the model. (3) Results: According to the experimental findings on an extensive dataset, the proposed MobileNet V2 can deliver an accuracy of 0.9983 for training, 0.9952 for validating, and 1.00 for testing. The precision, recall, and F1-score values were also obtained and demonstrated strong performance in the classification of retinal diseases.**

Keywords*— *deep learning, medical image classification, retinal diseases, optical coherence tomography

I. Introduction

When the eye's lens becomes cloudy, it is known as a cataract. If untreated, this condition can result in blurry vision and eventually vision loss [1]. The optic nerve is typically damaged in glaucoma, a group of eye disorders that is frequently accompanied by elevated intraocular pressure. Glaucoma can cause irreparable vision loss if it is not treated appropriately [2]. The blood vessels in the retina are harmed by a diabetes complication called diabetic retinopathy. In the worst cases, it may result in blindness and vision issues [3].

Manually reviewing medical photographs is one of the more time-consuming and unreliable traditional methods for identifying eye illnesses. Recent advancements in deep learning and AI have made it possible to construct computer-aided diagnostic (CAD) systems that can increase the accuracy and speed of diagnosing retinal illnesses [4]. In this research, we will concentrate on developing an automated classification approach for identifying retinal diseases from optical coherence tomography (OCT) images. We will also investigate whether creating Android mobile apps for point-of-care diagnosis using the suggested approach is feasible. Concisely, the ultimate goal of this study is to incorporate a deep learning model into an Android mobile application to quickly and accurately diagnose retinal diseases using retinal images. The application uses the MobileNetV2 deep learning network to classify eye conditions from retinal images with the goal of improving the speed and accuracy of eye condition diagnosis, which eventually leads to better patient outcomes.

II. Literature Review

Before conducting this research, a comprehensive review of the relevant literature was conducted. Various approaches were employed to gather the literature, including exploration of prior research utilizing the MobileNetV2 methodology, analysis of simulation results, and examination of outcomes published in reputable sources such as the Institute of Electrical and Electronics Engineers (IEEE) and the International Journal of Social Science & Economic Research (IJSSER).

A study by Taufiqurrahman et al. (2020) [5] introduced the MobileNe-tV2-SVM model, which demonstrated impre-ssive performance on a spe-cific dataset. This model displayed high accuracy, a quadratically weighted kappa, and areas under the receiving operating characteristic (AUROC) curve for different severity classes of diabetic retinopathy (DR). Moreover, the authors compared their proposed model with previous approaches and established its comparable or superior performance while considering various factors such as data, complexity, and scale.

On another hand, Patel and Chaware (2020) [6] observed significant enhancements in accuracy by fine-tuning the network. Their approach led to a notable incre-ase in both training and validation accuracy, suggesting a well-adjusted model. To assess the performance of their technique, the authors utilized a dataset comprising retinal fundus images from the Kaggle diabetic retinopathy dataset. They employed an extensive- number of training, validation, and testing images.

Moreover, Yildirim et al. (2022) [7] introduced a model based on the MobileNetV2 architecture. This model incorporates nested patch division and neighborhood component analysis (NCA) feature selection to detect diabetic retinal disorders in fundus angiography images. They applied the k-nearest neighbors (kNN) classification function and validated the model using 10-fold cross-validation. The proposed approach achieved an impre-ssive classification accuracy of 87.40% on the collected dataset.

III. Materials and Methods

A. Materials

Retinal images applied in our study were collected from [13], which comes from various public databases, including the High-Resolution Fundus (HRF) Image Database, the Indian Diabetic Retinopathy Image Dataset (IDRiD), and Ocular Recognition Databases. The image dataset was divided into three categories, with 70% of the images used for model training, 20% for model validation, and the remaining 10% for model testing. The images were categorized into four classes: cataract (CAT) with 1,038 images (24.6%), glaucoma (GLC) with 1,007 images (23.9%), diabetic retinopathy (DR) with 1,098 images (26%), and normal retinal images with 1,074 images (25.5%). As we can see in Fig. 1, the dataset is balanced. Moreover, examples of OCT images in our dataset can be found in Fig 2, showing cataract (Fig. 1a), glaucoma (Fig. 1b), diabetic retinopathy (Fig. 1c), and a normal retinal image (Fig. 1d).

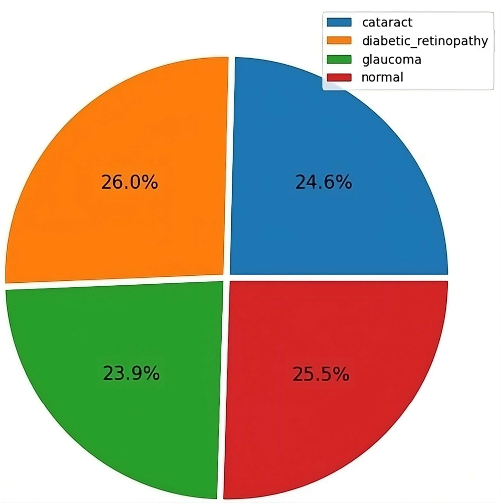
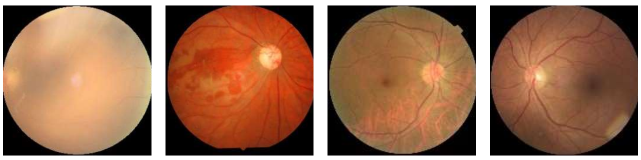


Fig. 1. The percentage of the images in each class:
(a) DR (orange), (b) CAT (blue), (c) GLC (green), and (d) normal (red).



(a) Cataract (b) Glaucoma (c) Diabetic Retinopathy (d) Normal retinal image

Fig. 2. Example of the retinal OCT images in the dataset [13].

For implementation, we utilized Jupyter Notebook as our Integrated Development Environment (IDE) for developing and executing the code. We used the Keras library (version 2.10.0) in combination with TensorFlow (version 2.10.0) within the Jupyter Notebook (version 7.0.2) environment. The experiments were conducted on a computer running Python 3.8.0, equipped with a 12th Gen Intel(R) Core(TM) i3-12100F 3.30 GHz CPU, 16GB of RAM, and a NVIDIA GeForce GTX 1050Ti GPU. For the mobile application part, we used HUAWEI P20 mobile phone (Android version 10) and Android Studio (version 2022.1.1 Patch 2) to build the app.

B. Methodology

Utilizing neural networks' ability for image processing, a convolutional neural network (CNN) was created for the categorization of retinal diseases. By selecting pertinent features from retinal images during training, CNN was created to mimic human vision [8]. The images were examined to determine their characteristics, which CNN used as input to categorize the images according to the different types of retinal diseases.

MobileNetV2 [9], known for its effectiveness in terms of model size and computational complexity, is a deep learning architecture created for picture categorization tasks. With further optimizations to make it more suitable for deployment on mobile devices with constrained computational resources, MobileNetV2 was launched by Google in 2018 [9] as an enhancement to the original MobileNet architecture. Using depthwise separable convolutions, MobileNetV2 can perform image classification tasks with a high degree of accuracy while using fewer parameters and operations. MobileNetV2 contains 53 convolution layers and 1 global average pooling [12]. The depthwise separable convolution filters in MobileNetV2 consist of pointwise convolution filters that linearly combine the output of the depthwise convolution filters with a 1×1 convolution and depthwise convolution filters that perform a single convolution on each of the input channels. For MobileNetV2 to effectively trade off latency and accuracy, two parameters—a width multiplier and a resolution multiplier—were included. The input width of a layer is controlled by the width parameter, and the resolution of the input image is controlled by the resolution parameter. The architecture of MobileNetV2 is shown in Fig. 3 [14]. The advantages of parameter efficiency, computational efficiency, better model efficiency, and improved generalization are offered by depthwise separable convolutions.

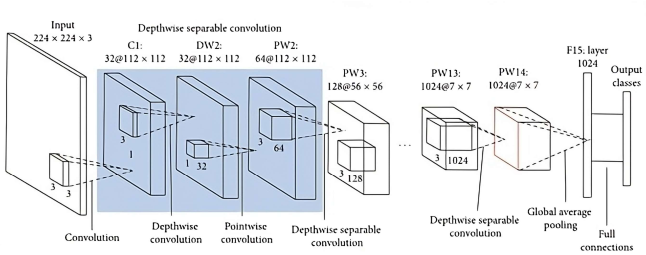


Fig. 3. MobileNetV2 Architecture [14].

(i) Evaluation Criteria Equations

The evaluation of the model's predictions in this study involved using a formula that considered the ratio of accuracy on the training data to the probability obtained from the test data. To measure the performance of the categorization task, standard metrics like accuracy, precision, recall, and F1-score were utilized. These metrics offer valuable insights into various aspects of the model's performance, including overall correctness, its ability to minimize false positives, its capacity to capture true positives, and the trade-off between precision and recall. These widely used metrics in deep learning evaluation provide a quantitative assessment of the model's predictive accuracy and validate its performance against the expected outcomes or ground truth. The accuracy, precision, recall, and F1-score are evaluated in Equations (1-4) [10]:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

$$F1 - score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (4)$$

where TP stands for true positive, TN for true negative, FP for false positive (Type I error), and FN for false negative (Type II error).

(ii) K-Fold Cross Validation

Using the 5-fold cross-validation method, we assessed how well our deep learning model performed. This method separates the dataset into a number of subsets to evaluate the generalizability of the model. The dataset was divided into five equal folds as shown in Fig. 4 to provide a balanced representation of various classes in our investigation.

The procedure involves repeatedly iteratively training and assessing the model. Four folds were utilized to train the model in each iteration, with the final fold acting as the validation set for performance evaluation. Each fold served as the validation set exactly once throughout the five times this technique was carried out. This strategy reduces the risk of overfitting by allowing us to use the given data efficiently for both training and evaluation.

We were able to acquire a more robust and assessment of our model's performance by using 5-fold cross-validation. Each iteration's findings were then averaged to provide an overall assessment of the model's accuracy, precision, recall, and F1 - score. This strategy mitigates the impact of data variability and gives a more realistic assessment of the model's capacity to generalize unknown data.

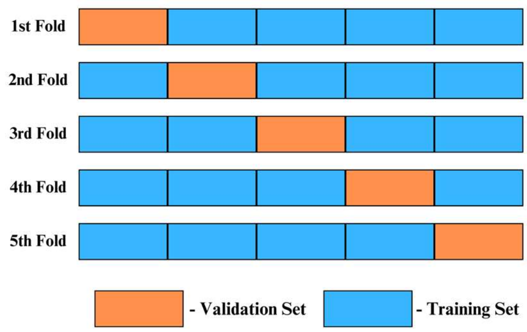


Fig. 4. Five-Fold Cross Validation.

IV. Experimental Results and Discussion

A. Experiments

Initially, an image directory is used to construct a dataset that will serve as training data. Scaling, normalization, standardization, and data separation are all used in picture preprocessing to improve the model's performance. After training the model with the preprocessed images, the training output is obtained. A testing dataset is used to assess accuracy, precision, recall, and F1-score to see how well the model works. Simultaneously, a smartphone app is being created that will allow users to interact with the trained model. An easy-to-use mobile phone application has been developed and is being used to categorize images. The classification results and the prediction score of the trained model using the testing images are then analyzed and discussed in the study paper, showcasing how successful the approach is at correctly classifying photos on a mobile platform. These processes are shown in Fig. 5.

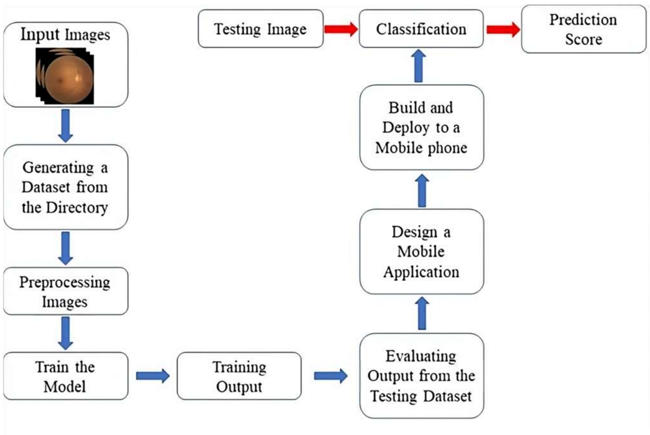


Fig. 5. Proposed Methodology.

This study used the MobileNetV2 architecture for training the picture classification model. For the training dataset, 224x224 pixel images were used as the input size. In order to control how many samples were processed during each training iteration, a batch size of 64 was used. For weight optimization, the Adamax optimizer was used, and the sparse categorical cross-entropy loss function—which is frequently employed for multi-class classification problems [11]—was selected. Softmax, a function that is appropriate for multi-class classification, was chosen as the activation function for the output layer. The step size for updating the weights of the model during optimization was set by the training procedure, which was carried out across 100 epochs with a learning rate of 0.001. These parameters are listed in TABLE I.

TABLE I. Parameter Settings.

Parameters	Settings
Input size	224
Batch size	64
Activation	Softmax
Optimizer	Adamax
Loss	Sparse categorical cross-entropy
Epoch	100
Learning rate	0.001

The best epoch was determined to be epoch 58 after 78 epochs of training a model with early stopping enabled. 99.83% accuracy and 99.52% validation accuracy were attained by the model. The validation loss was somewhat lower at 0.0727 than the training loss, which was 0.0752, as shown in TABLE II. According to these outcomes, the model performed effectively, obtaining high accuracy on both the training and validation sets. Based on the validation accuracy, the best epoch was identified, which may have signaled the beginning of a performance plateau for the model. The model was able to generalize well to previously unreported data since its accuracy and validation accuracy were consistently high. The great performance of the model was facilitated by the use of early stopping, which also helped prevent overfitting. Training and Validation Accuracy vs. Epoch and Training and Validation Loss vs. Epoch of MobileNetV2 are shown in Fig. 6 and Fig. 7 respectively.

TABLE II. COMPARISON OF DIFFERENT MODELS.

Architecture	*accuracy*	*val_accuracy*	*loss*	*val_loss*
VGG16	0.9415	0.9026	0.3696	0.4469
InceptionV3	0.9934	0.9567	0.1688	0.2558
ResNet50	0.8860	0.2644	1.4973	8.6319
MobileNet	0.9987	0.9531	0.1090	0.2226
MobileNetV2	0.9983	0.9952	0.0752	0.0727

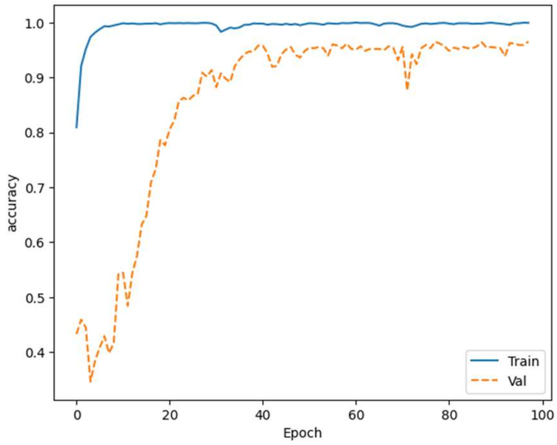


Fig. 6. Training and Validation Accuracy vs. Epoch of MobileNetV2.

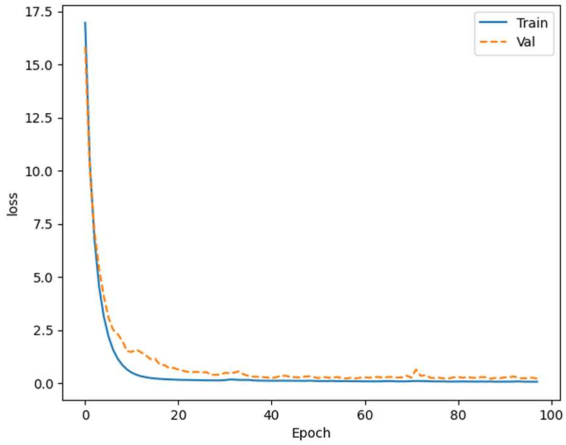


Fig. 7. Training and Validation Loss vs. Epoch of MobileNetV2.

After getting acceptable training and validation accuracy, we tested our trained model on the separated test set, which is 10% of the original dataset as mentioned in III.A Materials. Fig. 8 shows the confusion matrix, and TABLE III. shows the evaluation measurements, which are calculated on testing data. From these results, we can prove that the proposed model achieved perfect prediction results, showing 100% in all assessment measurements.

Class	*CAT*	*DR*	*GLC*	*Normal*
CAT	96	0	0	0
DR	0	79	0	0
GLC	0	0	113	0
Normal	0	0	0	89

Fig. 8. Confusion Matrix of MobileNetV2 refers to Cataract, Diabetic Retinopathy, Glaucoma and Normal, respectively.

TABLE III. CLASSIFICATION REPORT ON THE TESTING DATASET.

Class	*Precision*	*Recall*	*F1 - score*
Cataract	1.00	1.00	1.00
Diabetic Retinopathy	1.00	1.00	1.00
Glaucoma	1.00	1.00	1.00
Normal	1.00	1.00	1.00

Finally, an Android application was created using Android Studio and then released on a mobile device. Users can import test images into the application, which subsequently classifies them using the trained model. A high prediction score was attained by the model's accurate prediction of the proper class for the test image. This effective deployment of the trained model into the Android application shows the model's capability for real-world deployment in a practical scenario and highlights its practical applicability and usability for real-time image categorization tasks on mobile devices. The prediction of the DR class is shown in Fig. 9.

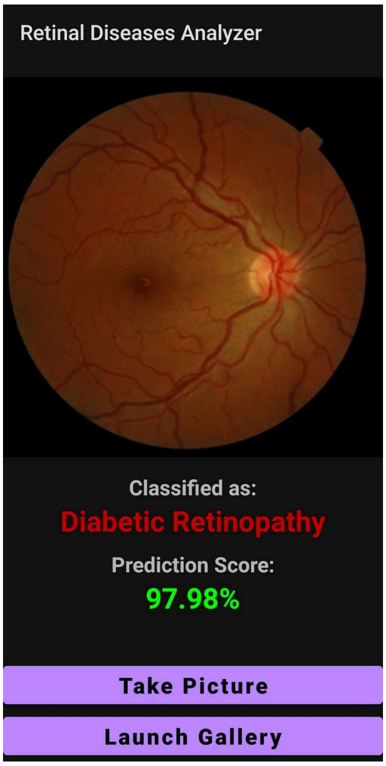


Fig. 9. Classification of Diabetic Retinopathy Testing Image and a Prediction Score of 97.98%

B. Discussion

TABLE IV. COMPARISON BETWEEN THE LITERATURE REVIEW PAPER AND THE PRESENT STUDY.

Ref (Year)	Dataset	Number of classes:	Based Deep-learning Model	Performance (Validation Accuracy)
[5] (2020)	APTOS 2019	5	MobileNetV2-SVM	85%
[6] (2020)	APTOS 2019	5	MobileNetV2	91%
[7] (2022)	Firat University Hospital	5	MobileNetV2	87.40%
Our Method (2023)	IDRiD, HRF	4	MobileNetV2	99.52%

We compared our proposed method with some related research in the literature [5], [6], and [7], and the comparative results are described in TABLE IV. From this table, we can see that our method has achieved superior performance, showing 99.52% validation accuracy on the IDRiD and HRF datasets with just four classes. Even though all the methods in TABLE IV. used the same model architecture, MobileNetV2, we achieved better outcomes for three main reasons: (i) we used MobileNetV2 in a transfer learning style using ImageNet pretrained weights; (ii) we made our dataset more balanced to prevent bias and enhance the generalization capability of the model; and (iii) we used five-fold cross-validation to ensure robustness and minimizes overfitting. However, this comparison has some limitations, as the aforementioned methods worked on different datasets, so it cannot be said to be a fair comparison. Due to the unavailability of the dataset, we couldn't test our method on their dataset. Moreover, another limitation is that the proposed method can classify four classes of retinal diseases.

V. CONCLUSION AND FUTURE SCOPE

In conclusion, this paper presents a promising method for diagnosing retinal diseases using deep learning and retinal images, with a training accuracy of 99.83%, a validation accuracy of 99.52%, and a testing accuracy (as shown in Fig. 8 and TABLE III. of 100% achieved using the MobileNetV2 model. The high accuracy of our approach in identifying various retinal diseases holds potential for improving patient outcomes through early and accurate detection. Nevertheless, further investigation and testing in clinical settings are needed to confirm its efficiency. And we believe that the findings of this paper contribute to the growing knowledge of research on deep learning in retinal disease diagnosis and further related research. In future work, the performance of the classifier proposed in this paper can be improved using additional images or classes and trying different models.

ACKNOWLEDGMENT

The authors would like to thank the owners of the Eye Disease Retinal Images dataset, Guna Venkat Doddi, a student at the GMR Institute of Technology in Visakhapatnam, Andhra Pradesh, India, and also the IDRiD, HRF, and Ocular Recognition Databases for freely releasing the retinal image datasets used in this study. Their efforts in gathering and sharing these datasets have been essential in advancing retinal imaging research. This study would not have been possible without their dataset.

FUNDING

This work was financially supported by School of engineering, King Mongkut's Institute of Technology Ladkrabang (KMITL) under Grant Number 2566-02-01-055.

REFERENCES

[1] Padalia, D., Mazumdar, A., & Singh, B. (2022). A CNN-LSTM Combination Network for Cataract Detection using Eye Fundus Images. https://doi.org/10.48550/ARXIV.2210.16093

[2] Casson, R. J., Chidlow, G., Wood, J. P., Crowston, J. G., & Goldberg, I. (2012). Definition of glaucoma: Clinical and experimental concepts: Definition of glaucoma. Clinical & Experimental Ophthalmology, 40(4), 341–349. https://doi.org/10.1111/j.1442-9071.2012.02773.x

[3] Morello, C. M. (2007). Etiology and natural history of diabetic retinopathy: An overview. American Journal of Health-System Pharmacy, 64(17_Supplement_12), S3–S7. https://doi.org/10.2146/ajhp070330

[4] Asiri, N., Hussain, M., Al Adel, F., & Alzaidi, N. (2019). Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. Artificial Intelligence in Medicine, 99, 101701. https://doi.org/10.1016/j.artmed.2019.07.009

[5] CA: University Science, 1989 Taufiqurrahman, S., Handayani, A., Hermanto, B. R., & Mengko, T. L. E. R. (2020). Diabetic Retinopathy Classification Using A Hybrid and Efficient MobileNetV2-SVM Model. 2020 IEEE REGION 10 CONFERENCE (TENCON), 235–240. https://doi.org/10.1109/TENCON50793.2020.9293739

[6] Patel, R., & Chaware, A. (2020). Transfer Learning with Fine-Tuned MobileNetV2 for Diabetic Retinopathy. 2020 International Conference for Emerging Technology (INCET), 1–4. https://doi.org/10.1109/INCET49848.2020.9154014

[7] Yildirim, H., ÇeliKer, Ü., Güngör Kobat, S., Dogan, S., Bayğin, M., Yaman, O., Tuncer, T., & Erdağ, M. (2022). An automated diabetic retinopathy disorders detection model based on pretrained MobileNetv2 and nested patch division using fundus images. Journal of Health Sciences and Medicine, 5(6), 1741–1746. https://doi.org/10.32322/jhsm.1184981

[8] S. Mutasa, S. Sun, and R. Ha, "Understanding artificial intelligence-based radiology studies: CNN architecture," Clin. Imaging, vol. 80, pp. 72–76, Dec. 2021, doi: 10.1016/j.clinimag.2021.06.033.

[9] Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.

[10] Sonarra, W., Vongmanee, N., Wanluk, N., Pintavirooj, C., & Visitsattapongse, S. (2022). Detection and Classification of COVID-19 Chest X-rays by the Deep Learning Technique. In 2022 14th Biomedical Engineering International Conference (BMEiCON).IEEE. https://doi.org/10.1109/bmeicon56653.2022.10012094

[11] Alwaqfi, Y., Mohamad, M., & Al-Taani, A. (2022). Generative Adversarial Network for an Improved Arabic Handwritten Characters Recognition. International Journal of Advances in Soft Computing and Its Applications, 14(1), 177–195. https://doi.org/10.15849/IJASCA.220328.12

[12] Khan, Misha Urooj & Saeed, Zubair & Raza, Ali & Abbasi, Zeeshan & Ali, Syeda & Khan, Hareem. (2022). Deep Learning-based Decision Support System for classification of COVID-19 and Pneumonia patients. JAREE (Journal on Advanced Research in Electrical Engineering). 6. 19-28. 10.12962/jaree.v6i1.229.

[13] Doddi, G. V. (2022). Eye_diseases_classification [dataset]. https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification

[14] Aji, Sani & Kumam, Poom & Siricharoen, Punnarai & Bukar, Ali & Adamu, Mohammed Sani. (2021). Deep Transfer Learning for Automated Artillery Crater Classification. Thai Journal of Mathematics. 19. 1068-1081.