East African Scholars Journal of Engineering and Computer Sciences
Abbreviated Key Title: East African Scholars J Eng Comput Sci
ISSN: 2617-4480 (Print) & ISSN: 2663-0346 (Online)
Published By East African Scholars Publisher, Kenya

OPEN ACCESS

Volume-7 | Issue-8 | Nov-2024 | DOI: https://doi.org/10.36349/easjecs.2024.v07i08.002

Original Research Article

Detection of Diabetic Retinopathy Using VGG19 and ResNet 50 Models

Odigie, M. O[1*], George, G. O[1], Igodan, E. C[2], Ukaoha, K. C[3]

[1]Department of Optometry, University of Benin, Benin City, Nigeria
[2]Department of Computer Science, University of Benin, Benin City, Nigeria
[3]School of Science and Computing, Wigwe University Isiokpo, Rivers State, Nigeria

Article History
Received: 06.09.2024
Accepted: 13.10.2024
Published: 02.11.2024

Journal homepage:
https://www.easpublisher.com

Quick Response Code

Abstract: Diabetic Retinopathy (DR) is an ocular condition that can manifest in individuals living with Diabetes Mellitus (DM). Retinal fundus examinations must be conducted on DM individuals as early identification and treatment of DR can reduce the risk of impaired vision or blindness. The manual diagnosis of DR conducted by eye-care professionals can be tedious and time-consuming, especially during mass screenings. Deep learning (DL) techniques are being used to provide automated diagnosis of DR. This study adopted two CNN (VGG 19 and ResNet50) models for the binary classification of DR (Non-referable DR and Referable DR). Both models were trained and validated with retinal fundus images from publicly available datasets. After training with Kaggle dataset, VGG 19 and ResNet50 models achieved accuracies of 94.3% and 96.9% respectively. For external validation, varying levels of accuracy, sensitivity and specificity were obtained for the two models on different datasets. The sensitivity of the VGG 19 model for the Messidor 2 dataset was 78.8% while the sensitivity of the ResNet50 model for the Indian Diabetic Retinopathy Image Dataset (IDRiD) was 85.7%. Findings in this study have shown that DR detection with deep learning techniques can serve as an assistive tool for eye-care professionals in the future.

Keywords: Deep Learning, Diabetic Retinopathy, Retinal fundus Images, VGG19 model, ResNet 50 model.

INTRODUCTION

Diabetes Mellitus (DM) is a condition in which a person has high blood sugar either as a result of the body not being able to produce enough insulin, or because the cells do not respond to the insulin that is produced. Diabetes mellitus is the second biggest negative total effect on reducing global health adjusted life expectancy worldwide (Chen *et al.*, 2019). The retinal microvasculature may be affected in individuals with DM leading to progressive damage of the eyes. This condition known as diabetic retinopathy (DR) could lead to symptoms such as blurred vision, dark spots in the field of view and even blindness. Clinical Signs of DR include microaneurysms which appear as small red round dots due to the weakness of the vessel's walls and haemorrhages that appear as larger red spots in the retina. Other signs include exudate which appears as bright yellow spots on the retina and are caused by leakage of plasma and soft exudates also known as cotton wool spots that appear as white spots on the retina caused by the swelling of the nerve fiber. DR is broadly classified into non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). NPDR is further classified into mild, moderate and severe DR. PDR, a more severe form of DR, occurs as a result of production of new fragile blood vessels that can leak leading to retinal detachment and blindness. The different classes of DR is depicted in Figure 1.

Globally, 600 million people will have diabetes mellitus by 2040, with a third expected to have diabetic retinopathy (Yau *et al.*, 2012). The rate of DR progression is approximately five times higher among the African population compared with the European population (Burgess *et al.*, 2017). Through regular eye examinations and adequate DM management, the diabetes-related vision loss can be prevented in 98% of cases (Rohan *et al.*, 1989, and Ferris 1993). Fundus photography is a rapid, non-invasive, well-tolerated and widely available imaging technique (Kwan and Fawzi, 2019). However, the computerized screening tools has been enabled to deal and document many retinal diseases with little or no intervention of clinical experts (Quereshi *et al.*, 2019). Diagnosis of DR in retinal fundus images is

***Corresponding Author:** Odigie, M. O
Department of Optometry, University of Benin, Benin City, Nigeria

necessary because of its non-invasiveness, reliability and better sensitivity (Hutchinson *et al.,* 2000). There could be challenges in manual inspection of morphological changes in retinal images as experienced clinicians may not be able to keep up with the demand for screening services (Goh *et al.,* 2016, Qureshi, *et al.,* 2019). The limiting factors for screening large numbers of people with diabetes include lack of adequate number of ophthalmologists and optometrists (Rema *et al.,* 2007, Namperumalsamy *et al.,* 2003, and Mohan *et al.,* 2014). Radical measures are required to identify and reduce blindness due to diabetes to achieve the Sustainable Development Goals by 2030 (Bellemo *et al.*, 2019).

Figure 1: Different classes of DR (Porwal *et al.*, 2018)

The need for regular and improved eye screening for diabetic individuals has led to the application of artificial intelligence (AI) for detection of DR. DL is a form of machine learning, a branch of AI that trains a neural network to carry out a task such as image classification prediction (LeCun *et al.,* 2015, Alshareel *et al*., 2022, and Tirumala and Narayanan, 2018). In order to train a neural network, a large number of images are needed in which the severity of a disease such as DR is already known (Gulshan *et al.,* 2016). Convolutional neural network (CNN) as depicted in Figure 2, is a type of neural network in deep learning that is designed for analyzing mainly two-dimensional images (Valueva *et al.*, 2020, Brownlee, 2019). CNN has become dominant in different computer vision tasks and is attracting a lot of attention across various disciplines (Yamashita *et al.*, 2018). Pre-trained CNN architectures provide a simpler and faster way of training using randomly initialized weights (Deniz *et al.*, 2018). CNNs are highly effective due to their capability to perform parallel computations with Graphic processing units (GPUs) (Kiran *et al.,* 2018). Convolutional neural network is composed of multiple building blocks, such as convolution layers, pooling layers, and fully connected layers, and is designed to automatically and adaptively learn spatial hierarchies of features through a backpropagation algorithm (Yamashita *et al.,* 2018).

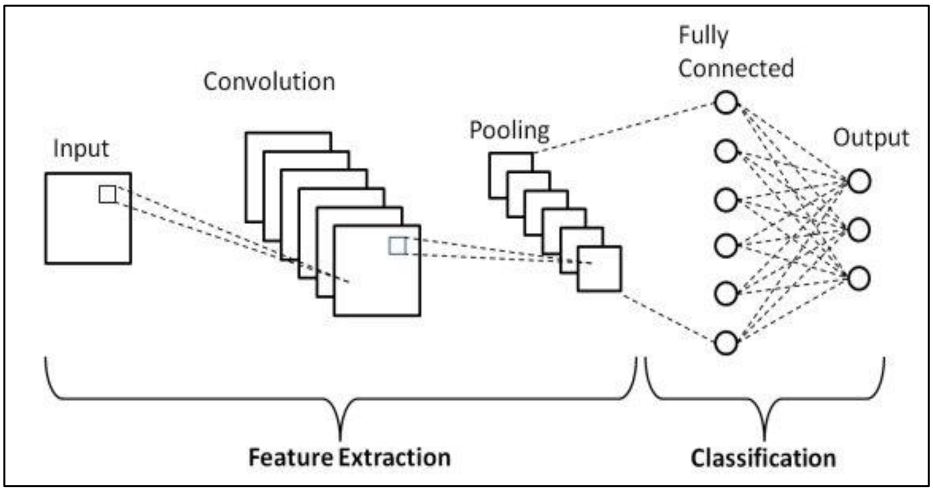

Figure 2: Convolutional Neural Architecture
(Phung and Rhee, 2018)

2. LITERATURE REVIEW

The detection of DR has been extensively studied using computer-aided diagnosis (CAD) systems for a fast and accurate diagnosis that integrate image processing, machine learning, and deep learning techniques in the literature with varying degrees of classification performance. Pratta *et al.,* (2016) developed an algorithm using CNN architecture to diagnose DR from retinal images. The network was trained on Kaggle dataset and achieved an accuracy and sensitivity of 75% and 95% respectively. The network has no issue learning to detect an image of a healthy eye, however, the network struggled to learn deep enough features to detect some of the more intricate aspects of DR due to the low sensitivity, mainly from the mild and moderate classes. Gulshan *et al.,* (2016) trained a neural network known as inception-V3 architecture to detect referable diabetic retinopathy (moderate non-proliferative diabetic retinopathy or worse) using retinal images from EyePACS and three eye hospitals located in United States and India respectively. Two datasets EyePACS 1 and Messidor 2 were used for validation. The sensitivity and specificity of the EyePACS 1 dataset was 90.3% and 98.1% respectively. The sensitivity and specificity of the Messidor 2 dataset was 87.0 % and 98.5 % respectively. Dutta *et al.*, 2018 proposed a DL model which was trained using backpropagation Neural Network (NN), DNN and VGG-16, for DR detection. Two thousand images from Kaggle dataset in the ratio of 7:3 were used for training and testing the models. During the training and the testing phase, the DNN model has achieved an accuracy of 89.6% and 86.3%, respectively while backpropagation NN achieving an accuracy of 62.7% and testing accuracy of 42% while VGG-16 which has achieved a training accuracy of 76.4% and testing accuracy of 78.3%, for image classification respectively. Li *et al.,* (2018) developed a DL algorithm for detecting vision threatening referable DR using retinal photographs acquired from a web-based, crowdsourcing platform (http://www.labelme.org). In the internal validation data set, sensitivity, and specificity of the DL algorithm for vision-threatening referable DR were 97.0%, and 91.4%, respectively. The algorithm was also validated on independent retinal images from population-based cohorts of Malay, Caucasian Australians, and Indigenous Australians and achieved a sensitivity, and specificity of 92.5%, and 98.5%, respectively. Sarki *et al.,* (2019) conducted experiments with 13 CNN architectures, pre-trained on large-scale ImageNet database in order to detect mild DR. Several performance improvement techniques such as fine-tuning, data augmentation, and volume increase were experienced. The maximum accuracy of 86% on No DR/Mild DR classification task was obtained for ResNet50 model after extensive experimentation. Sahlsten, *et al.,* (2019) trained an Inception-V3 architecture to detect referable DR using dataset of graded DR retinal images provided by Digifundus Ltd. Multiple resolutions were done for the purposes of analyzing the effect of the input image resolution on the classification performance. The study showed that increasing the input image resolution from 256×256 to 512×512 clearly improved the results, and even better results were obtained as the resolution was further increased. In the NRDR/RDR classification on the primary validation set, the algorithm achieved the sensitivity of 89.6% and specificity 97.4%. Gadekallu *et al.,* (2020) used principal component analysis (PCA) and firefly algorithm on Diabetic Retinopathy Debrecen dataset. The dataset was fed into Deep Neural Network Model for classification. The accuracy, sensitivity and specificity of this method were 97.0%, 92.0% and 95% respectively. Deep residual learning has been proposed by Rahman *et al.,* (2020) due to the challenge of time and space complexity while efficiently detecting DR. For each image in the training dataset, a 224x224 image was processed in the training dataset. About 66.66% of Kaggle dataset was used to train Resnet50 architecture while the remaining dataset was used for validation. An accuracy of 93.2% and sensitivity of 95.6% was achieved with this model. Shaban *et al.,* (2020) proposed a deep Convolutional Neural Network (CNN) with 18 convolutional layers and 3 fully connected layers to analyze fundus images and automatically distinguish between controls (i.e. no DR), moderate DR (i.e. a combination of mild and moderate NPDR and severe DR (i.e. a group of severe NPDR, and PDR with different ranges of validation accuracy of 88% to 89%, sensitivity of 87% to 89%, specificity of 94% to 95% were obtained in their results. Khanusiya and Savani 2021 used two different CNN architectures for DR detection. Image processing involved the application histogram balance on images from Kaggle dataset. An accuracy of 75.50% and 77.33% was obtained for VGG 16 and AlexNet respectively. Lam *et al.,* (2018) trained and tested CNN architectures (VGG16 and GoogLeNet) using the Kaggle dataset with 5 class labels and Messidor 1 dataset with 4-class labels. Contrast adjustment was performed using the contrast limited adaptive histogram equalization (CLAHE) filtering algorithm. The 4-ary classifier encounters a problem of simply not having enough images to effectively train a deep CNN such as GoogLeNet. The multi-class model was unable to distinguish between different classes and behaves as a majority classifier, attempting to classify all images into a single class. Deep CNNs (Alexnet, VGG16, and InceptionNet V3) techniques were employed by Wang *et al.,* (2018) for classification of DR. The dataset had only 166 images. The accuracy of Alexnet, VGG16, InceptionNet V3 was 37.43%, 50.03%, and 63.23% respectively for a 5 classification task. However, it was observed that they trained the networks with a limited number of images which affected the CNN learning capability. Nguyen *et al.,* (2020) presented an automated classification system, in which fundus images were analyzed with varying illumination and fields of view and generated a severity grade for DR using machine learning models such as CNN, VGG-16 and VGG-19. This system achieved 80% sensitivity, 82% accuracy and 82% specificity for classifying images into 5 categories ranging from 0 to 4, where 0 is no DR and 4 is proliferative DR. Sharma *et al.,* (2019) applied CNN for the detection of DR. An accuracy of 74.04% was achieved with a 5 class classification task. They stated that accuracy can be further improved by increasing the size of the dataset as only a subset of the data set was considered for implementation of the model as a result of hardware constraints. CNN was utilized on 88,700 retinal fundus images, from EyePACS dataset, and achieved 81.12%, 89.16%, and 84.16% for sensitivity, specificity, and accuracy, respectively for classifying DR into five stages in the works of Khaled *et al.,* (2021). Deshpande and Pardhi, (2021) applied pre-trained VGG-16 to detect the severity of Diabetic Retinopathy. The Asia-Pacific Tele-Ophthalmology Society (A.P.T.O.S) 2019 Blindness Detection dataset containing 3668 retinal images was used for training. The model achieved 74.58% accuracy when tested on 1728 images. In the works of Sudha and Gareshbadu (2020), VGG-19 was trained and tested with images from the Kaggle dataset to achieve a sensitivity of 82% for classifying DR into different stages. The 70% accuracy score obtained by ResNet50 was almost three times that of VGG-16 with an accuracy of 25% using the Kaggle dataset in Aatila *et al.,* (2021). Ayala *et al.,* 2021 implemented a DenseNet 121 to process fundus images to determine the severity of diabetic retinopathy. Their proposal achieved a suboptimal performance under both unbalanced datasets with 81% and 64% for APTOS and Messidor datasets, respectively. It was observed that the model trained over APTOS learned more useful features for most of the classes than the model trained over the Messidor dataset. DL algorithms might serve both as a promising solution to reduce human grading workload and as a cost-effective screening alternative for both high- and low-resource countries (Nguyen *et al.,* 2016). Although reliable results are obtained from current studies in the literature, this field is still an active research topic. Therefore, this study seeks a robust approach for the detection of DR form fundus images to detect either the non–referable (No DR and Mild DR) or Referable (Moderate DR and above) using the Visual Geometry Group 19 (VVG19) and Residual Network with 50 layers (ResNET50) convolution neural networks as classifiers. One of the primary factors in choosing these models is that the VVG19 architecture, with its 19 layers (thus the name), is deeper than previous models such as VGG16 and requires less computing power than deeper architectures like ResNet or Inception. It also offers a balance between model complexity and efficiency. ResNet50 is computationally efficient in comparison to previous networks of comparable depth, despite its depth. ResNet50 is feasible for real-world applications even with constrained computer resources thanks to the addition of residual connections, which lowers the computational cost of training deeper networks.

3. MATERIALS AND METHODS

3.1. Dataset for training and internal validation

The Kaggle Diabetic Retinopathy (DR) dataset containing 35,108 images was used to train the CNN architectures. The images were captured under various conditions by various fundus cameras with different resolutions at multiple primary care sites in the United States of America. Clinicians graded the images into five stages. 0 - No DR, 1-Mild, 2-Moderate, 3-Severe, 4-Proliferative DR. Classes 0 and 1 were regrouped together as non-referable DR while classes 2, 3 and 4 were grouped as referable DR. Eighty per cent of the images was used for training while twenty per cent for used for internal validation to evaluate the performance of the models from the same dataset. The classification of DR images using CNN architectures is initiated by data collection and by employing the necessary preprocessing to advance and boost the images (Shekar *et al.*, 2021). Image pre-processing, a process of cleaning raw data, helps to enhance the features and consistency of images which is relevant for subsequent processing and analysis (Dutta *et al.*, 2018). The images were first resized and centre cropped to minimize memory usage, the resulting images after being pre-processed are depicted in Figure 3.

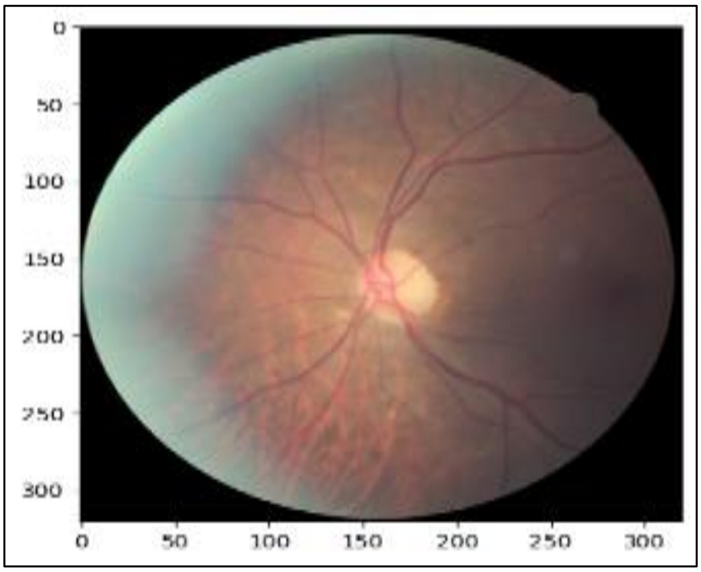

Figure 3: Preprocessed Retinal fundus Image

The images were resized to 320pixel by 320pixel. Gaussian blurring, a type of filter that takes surrounding pixels and returns a single number calculated with a weighted average was adopted in the study. This technique was applied to remove noise from the images thereby adjusting the transparency and improving the visibility of blood vessels. A representative image from the resulting image preprocessing using Gaussian blurring filtering method is shown in Figure 4.

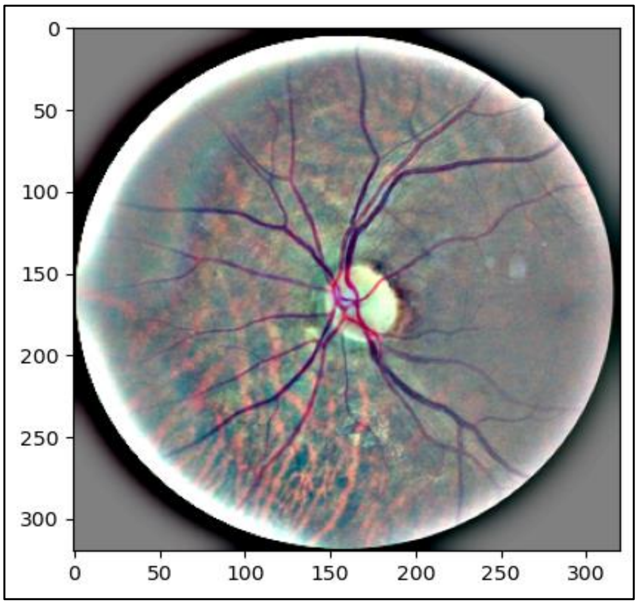

Figure 4: Pre-processed retinal fundus image

Image data generator processing was applied to make training results invariant to image orientation. These include (i) Image flip: this is the rotation of an image along a horizontal or vertical axis. The images were flipped randomly so that the CNN architectures can learn how to identify flipped images and classify them accurately when it sees them. The flipping techniques include: horizontal flip - this is reversing an image's full row and column pixels horizontally (left to right) in which the flipping occurs on the vertical axis. (ii) Image zoom: zooming is expanding an image such that its features become more visible and distinct. Zooming an image allows the empty areas to be excluded and allow the model to emphasize more on the areas which are truly needed.

3.2. Convolutional Neural Network Architectures

Visual Geometry Group (VGG 19) is designated for the visual geometry community and is the oldest of all the architectures tested. The neural network was ranked first in the ImageNet competition in 2014 (Simonyan and Zisserman, 2014). It was created to decide how the depth of a network influences its accuracy. It has various layers including 16 convolutional layers, 5 max pools, 3 fully connected, and 1 softmax layer. Residual Network (ResNet 50) is a deep convolutional neural network which was developed by He *et al.*, 2016. It comprises 48 convolutional layers with 64 different kernels, 1 max pool layer with a stride of size 2. These layers were replicated 3 times to give a total of 9 layers. The next layer has different kernels and repeated 4 times to give a total of 12 layers. The following layers consist of other variants of kernels which were repeated many times to form a total of 49 layers. Consequently, an average pool is obtained and a softmax function, which produced the last layer of this architecture. The ResNet architecture popularized the idea of using deeper networks as compared to VGG19 layers. Furthermore, the ResNet 50 architecture introduced skip connections, also known as residual connections to avoid information loss during training of deep network. Skip connection technique enables to train very deep networks and can boost the performance of the model.

A. Training of the CNN:

Two pre-trained CNN (VGG 19, and ResNet 50) models were adopted and implemented on Kaggle image dataset using Python programming language and Graphical Processing Unit (GPU) in cloud for the training of the CNN models. The performance of GPU is faster than central processing unit (CPU) and can carry out multiple calculations across data streams at the same time (Alhadi *et al.*, 2019, Bustamam *et al.*, 2009), being the reason for its usage in this work. The two CNN models were trained by passing the network with 16 batches of labelled images (Non referable DR and Referable DR) from the Kaggle dataset, thereby exposing the network to the key features of the images that is associated with each class of DR. The CNN models gradually adjusted their weight parameters to differentiate between the two classes through backpropagation (backward pass that occurs in order to adjust the CNN models parameters). Although the CNN model does not explicitly detect lesions (Drusen, Exudates, Microaneurysms, Hemorrhages, or Cotton-wool Spots), it likely learns to recognize them using the local features. The loss and accuracy during the training and validation at the end of each epoch were recorded. Two callback functions were utilized in the training process, Early Stopping and ReduceLROnPlateau. The Early Stopping function was assigned to monitor the validation loss. The EarlyStopping function stopped the training process once the monitored validation loss stop improving for 10 epochs. The weights that gave the best validation loss were recorded during the training, and the weights were restored to the model when the training terminates. The ReduceLROnPlateau function also monitored the validation loss. The ReduceLROnPlateau function reduced the current learning rate by 0.5 when the validation loss ceases to decrease for 3 epochs.

B. Hyper-parameters Settings:

These are variables that are set before training the CNN models. The following the hyper-parameters used in this study include: adaptive moment estimation (Adam) optimizer - aids in the adjustment of the parameters of a neural network in real time in order to enhance its accuracy and speed. Learning rate - this controls the rate at which the model learns and has a small positive value between the range of 0.0 and 1.0. The learning rate set for this experiment was 0.0001 (Gulshan *et al.*, 2016). Epochs - the number of complete forward and backward passes through the CNN model. The Epochs was set to a maximum number of 50.

After training and internal validation with the Kaggle dataset, the CNN models were tested on two external datasets.

C. Datasets for external validation:

External validation refers to the examination of an existing model's performance using dataset completely different from that used for development of the model (Riley *et al*., 2021). The following two datasets were adopted for the validation of our models: Messidor 2 dataset consisting of 1748 images captured in the Ophthalmology department of Brest University Hospital (France) with different resolutions was used for testing. The images were captured using a Topcon TRC NW6 non-mydriatic fundus camera with a 45degree field of view (Decencièsre *et al.*, 2014). Indian Diabetic Retinopathy Image Dataset (IDRiD), is a representative of an Indian population. The 516 fundus images in IDRiD were acquired from an Eye Clinic located in Nanded, India. Images were acquired using a Kowa VX-10α digital fundus camera with 50° field of view (Porwal, *et al.*, 2018).

3.3. Performance Metrics

The proposed models were evaluated using the following performance metrics. Confusion Matrix: It can be defined as a table that visualizes and describes the performance of the classification task on a test dataset samples are correctly classified (Gumaei *et al.,* 2021 and Sheikh, 2020). Figure 5 shows the confusion matrix of binary classification. Metrics calculated from the confusion among others are:

	Predicted Positive	**Predicted Negative**
Actual Positive	Number of True Positives (TP)	Number of False Positives (FP)
Actual Negative	Number of False Negatives (FN)	Number of True Negatives (TN)

Figure 5: Confusion matrix of binary classification

i. Accuracy: is the number of correct predictions made by the model.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

ii. Sensitivity: is the rate of actual positives overall predicted values that are positive.

$$Sensitivity = TP/(TP + FN) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

iii. Specificity: Specificity measures the True negatives over the sum of true negatives and false positives.

$$Specificity = TN/(TN + FP) \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

4. RESULTS

The performance metrics for binary classification task (non-referable DR and referable DR) were compared between the two CNN models (VGG19 and ResNet 50). The datasets used for external validation (Messidor-2 and IDRiD) were also compared. The results are presented graphically in Figures 6-13 and Table 1 respectively. Figure 6 indicates the training and validation accuracy for VGG19 model with Kaggle dataset obtaining 94.3% and 91.3% accuracy respectively after 23 epochs. While Figure 7 shows accuracy of 97.5% and 90.1% respectively after 16 epochs both for the training and validation using the ResNet 50 model with Kaggle dataset. Figure 8 and 9 shows the minimum validation loss which occurred at the 13th epoch for VGG19 model and at the 6th epoch for ResNet50 model respectively.

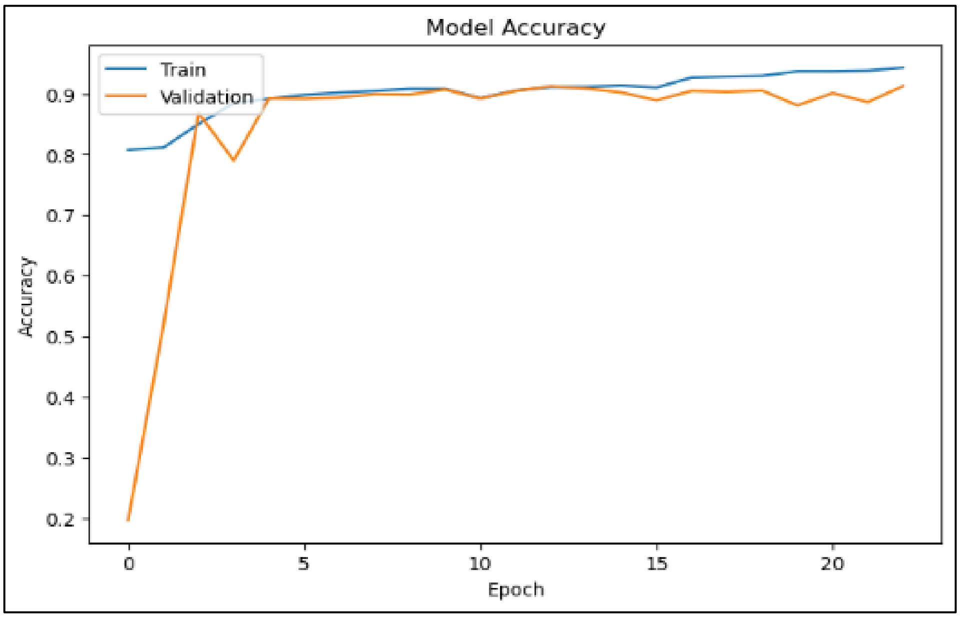

Figure 6: Training and Validation accuracy for VGG19 Model

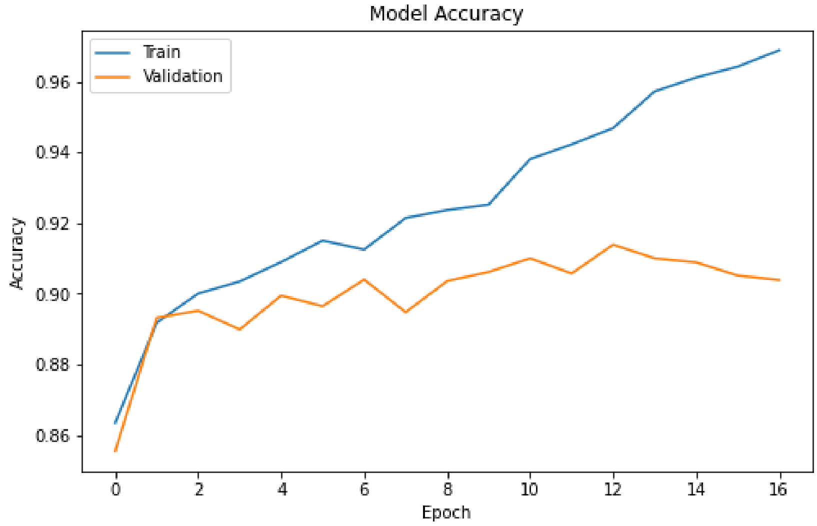

Figure 7: Training and Validation Accuracy for ResNet 50 Model

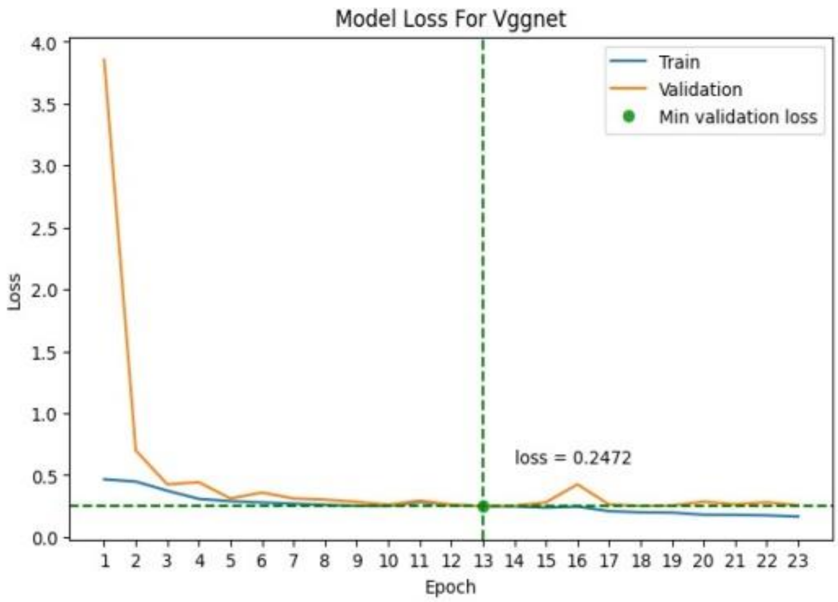

Figure 8: Training and Validation loss for VGG19 Model

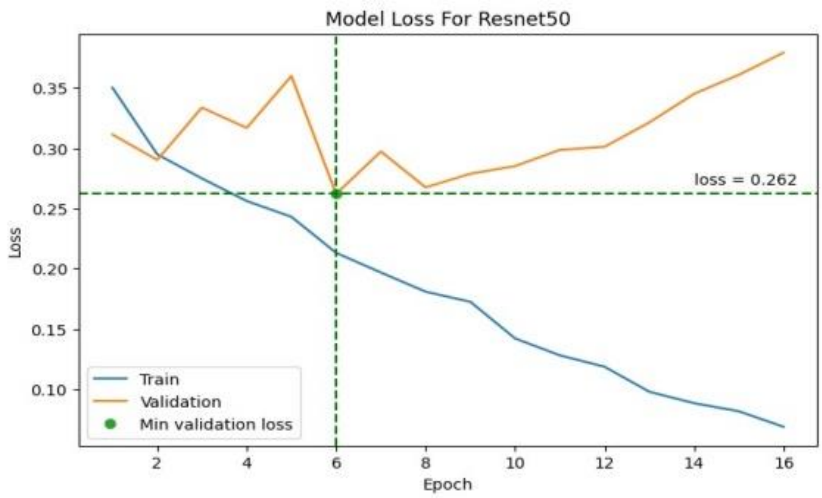

Figure 9: Training and Validation loss for ResNet50 Model

The confusion matrix in Figure 10 shows that 1134 images classified as non-referable DR were correctly classified, 360 images were also correctly classified as referable DR, while 153 and 97 were incorrectly classified as referable DR and non-referable DR respectively. The confusion matrix in Figure 11 depicts that 129 and 247 images were both classified as non-referable and referable DRs images correctly. While 25 and 12 images were wrongly classified as referable and non-referable DRs respectively. The confusion matrix in Figure 12, shows that 1142 images were truly classified as Non Referable DR while 316 images were truly classified as Referable DR. The number of falsely classified Non Referable DR and Referable DR were 141 and 145 images respectively. Lastly, the confusion matrix in Figure 13 depicts 148 images were truly classified as non-referable DR while 222 images were truly classified as referable DR. The number of falsely classified non-referable DR and referable DR were 37 and 6 images respectively.

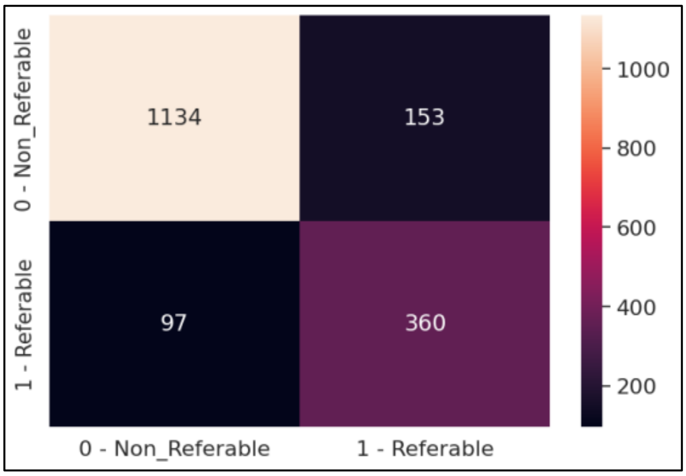

Figure 10: Performance metrics of VGG 19 model on Messidor 2 dataset

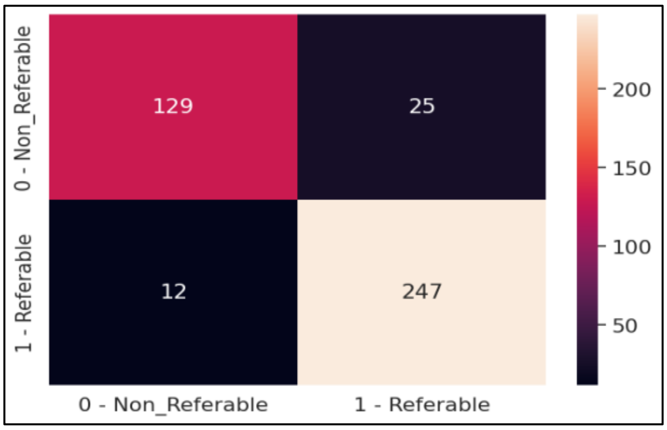

Figure 11: Performance metrics of VGG 19 model on Indian Diabetic Image Retinopathy dataset

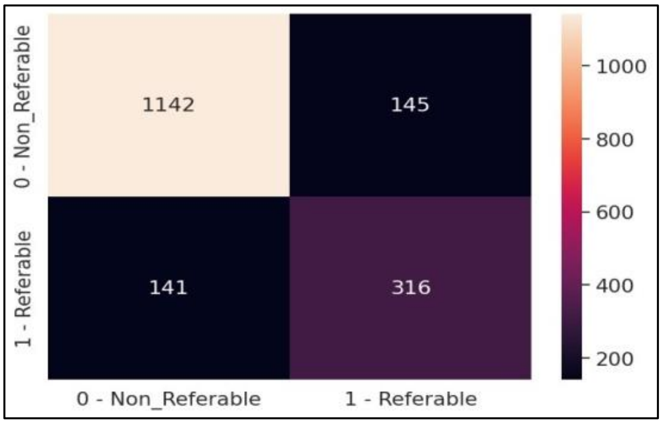

Figure 12: Performance metrics of ResNet 50 model on Messidor 2 dataset

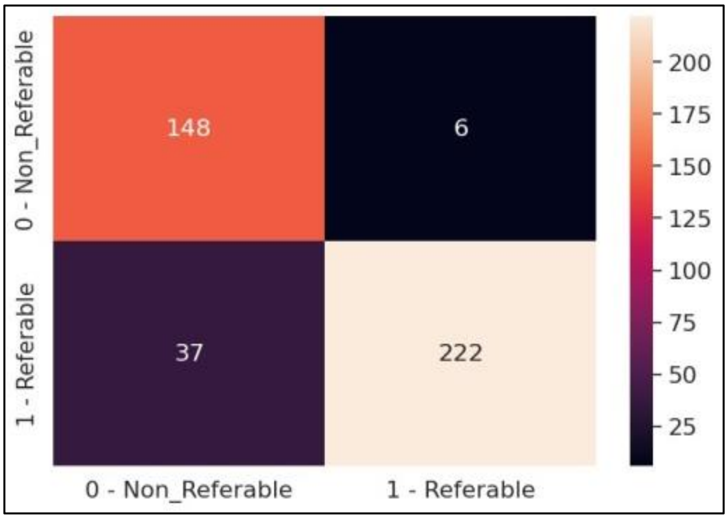

Figure 13: Performance metrics of ResNet 50 model on Indian Diabetic Image Retinopathy dataset

The VGG19 model had a higher accuracy and sensitivity values than ResNet 50 model in both datasets. However, ResNet 50 had a higher specificity values than VGG 19 in both datasets

Table 1: Performance metrics of Models

Dataset	Performance (%)	VGG19	ResNet 50	Validation (%)
Messidor 2	Accuracy	85.7	83.6	91.3
	Sensitivity	78.8	69.1	
	Specificity	88.1	88.7	
Indian Diabetic Retinopathy Image	Accuracy	91.0	89.6	90.1
	Sensitivity	95.4	85.7	
	Specificity	83.8	96.1	

5. DISCUSSION

In this study, two pre-trained CNN models (VGG19 and ResNet 50) were used to classify DR into two stages. These models have been previously trained on large datasets and as such they have been employed to recognize the features of DR without the need to train from scratch (Alzubaidi *et al.*, 2021). Fewer number of epochs (17) were used for the training of ResNet 50 model in order to achieve its best accuracy compared to the number of epochs (23) used for the training of VGG19 model. A lower accuracy was obtained for internal validation 91.3%, and 90.1% for VGG19 and ResNet 50 model respectively compared to the training accuracy. The models were tested on a proportion of Kaggle dataset (20%) that was not used for training which resulted in minimal differences in the accuracy values. Two external datasets (Messidor-2 and IDRiD) were used for further evaluation of the models. This was necessary in order to access the model's reproducibility and generalizability to fundus images of new and different patients (Debray *et al*., 2015, Ramspek *et al*., 2020). The accuracy obtained for training and internal validation for both models were higher than the accuracy obtained for external validation. The performance of models has been found to be lower for new patients than in the population used for development of the models hence they should not be recommended for clinical use until external validity is established (Moons *et al.*, 2015). Varying levels of accuracy, sensitivity and specificity values were obtained with the two models on two external datasets. This clearly shows that the depth of the models had no effect on improving the performance metrics as a shallower model such as VGG 19 obtained a higher accuracy and sensitivity compared the deeper model (ResNet50). It was observed that the VGG19 model had a higher ability to detect Referable DR compared to ResNet 50 model due to its higher sensitivity in this study. However, a higher specificity was obtained with ResNet 50 model in both datasets indicating that the model has a higher ability to detect Non Referable DR compared to VGG19 model. In order to accelerate trust in, and the adoption of, CNN models they should be developed in environments where retinal images are captured under different conditions (Muhammad *et al.,* 2022). The application of CNN models for the analysis of retinal images have the potential to provide alternative solution for DR screening in the future as the networks and the datasets continue to improve (Pratt *et al.,* 2016, Bellemo *et al.,* 2019). This can offer several advantages such as the consistency of interpretation and the near instantaneous reporting of results.

6. CONCLUSION AND FUTURE WORKS

Through the use of deep learning techniques, this research effort established an automated system for detecting two classes of DR: referable and non-referring DR. The features of DR were extracted from fundus pictures using two pre-trained CNN models. The study's outcomes demonstrated a high degree of accuracy of 91%. But in order to boost the models' confidence, accuracy can be increased in the future. Subsequent investigations should employ various machine learning methodologies that may result in enhanced performance measures derived from this investigation. These methods include using various pre-processing techniques to the fundus images and augmenting data to address the issue of class imbalance linked to the publically accessible DR dataset. Additionally, a fresh collection of fundus photos from eye clinics and hospitals will be used to assess the models.

REFERENCES

- Ayala, A., Ortiz Figueroa, T., Fernandes, B., & Cruz, F. (2021). Diabetic retinopathy improved detection using deep learning. *Applied Sciences*, *11*(24), 11970.
- Chen, H., Chen, G., Zheng, X., & Guo, Y. (2019). Contribution of specific diseases and injuries to changes in health adjusted life expectancy in 187 countries from 1990 to 2013: retrospective observational study. *British Medical Journal*, 364, l969.
- Yau, J. W., Rogers, S. L., Kawasaki, R., Lamoureux, E. L., Kowalski, J. W., Bek, T., ... & Meta-Analysis for Eye Disease (META-EYE) Study Group. (2012). Global prevalence and major risk factors of diabetic retinopathy. *Diabetes care*, *35*(3), 556-564.
- Burgess, P. I., Harding, S. P., García-Fiñana, M., Beare, N. A., Glover, S., Cohen, D. B., ... & Allain, T. J. (2017). Incidence and progression of diabetic retinopathy in Sub-Saharan Africa: A five year cohort study. *PloS one*, *12*(8), e0181359.
- Rohan, T. E., Frost, C. D., & Wald, N. J. (1989). Prevention of blindness by screening for diabetic retinopathy: a quantitative assessment. *British Medical Journal*, *299*(6709), 1198-1201.
- Ferris, F. L. III. (1993). How effective are treatments for diabetic retinopathy? *JAMA*, *269*(10), 1290-1291.
- Qureshi, I., Ma, J., & Abbas, Q. (2019). Recent Development on Detection Methods for the Diagnosis of Diabetic Retinopathy. *Symmetry*, 11, 749.
- Hutchinson, A., McIntosh, A., Peters, J., O'Keeffe, C., Khunti, K., Baker, R., & Booth, A. (2000). Effectiveness of screening and monitoring tests for diabetic retinopathy – a systematic review. *Diabetic Medicine*, *17*(7), 495-506.
- Goh, J. K. H., Cheung, C. Y., Sim, S. S., Tan, P. C., Tan, G. S. W., & Wong, T. Y. (2016). Retinal imaging techniques for diabetic retinopathy screening. *Journal of Diabetes Science and Technology*, *10*(2), 282-294.
- Rema, M., & Pradeepa, R. (2007). Diabetic retinopathy: An Indian perspective. *Indian Journal of Medical Research*, 125, 297-310.
- Namperumalsamy, P., Nirmalan, P. K., & Ramasamy, K. (2003). Developing a Screening Program to Detect Sight Threatening Diabetic Retinopathy in South India. *Diabetes Care*, 26, 1831-1835.
- Mohan, V., Prathiba, V., & Pradeepa, R. (2014). Tele-diabetology to Screen for Diabetes and Associated Complications in Rural India: The Chunampet Rural Diabetes Prevention Project Model. *Journal of Diabetes Science and Technology*, 8, 256–261.
- Bellemo, V., Lim, Z. W., Lim, G., Nguyen, Q. D., Xie, Y., Yip, M. Y., ... & Ting, D. S. (2019). Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. *The Lancet Digital Health*, *1*(1), e35-e44.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521, 436-444.
- Alshareef, A. M., Alsini, R., Alsieni, M., Alrowais, F., Marzouk, R., Abunadi, I., & Namri, N. (2022). Optimal Deep Learning Enabled Prostate Cancer Detection Using Microarray Gene Expression. *Journal of Healthcare Engineering*. Pp. 1-12. https://doi.org/10.1155/2022/7364704.
- Tirumala, S. S., & Narayanan, A. (2018). Classification and diagnostic Prediction of Prostate Cancer Using Gene Expression and Artificial Neural Network. Neural Computing and Application. https://doi.org/10.1007/s00521-018-3589-8
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *jama*, *316*(22), 2402-2410.
- Valueva, M. V., Nagornov, N. N., Lyakhov, P. A., Valuev, G. V., & Chervyakov, N. I. (2020). Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. Mathematics and Computers in Simulation. *Elsevier BV*, 177, 232-243.
- Brownlee, J. (2019). How Do Convolutional Layers Work in Deep Learning Neural Networks? Available from URL: https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks. Accessed 28th March 2021.
- Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, 9, 611-629.
- Deniz, E.¸Sengür, A., Kadiroglu, Z., Guo, Y. Bajaj, V., & Budak, U. (2018). Transfer Learning based histopathologic image classification for breast

cancer detection. *Health Information Science System, 6*(1), 18.

- Kiran, B. R., Thomas, D. M., & Parakkal, R. (2018). An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. *Journal of Imaging, 4*(2), 36.
- Pratta, H., Coenenb, F., Broadbentc, D. M. Hardinga, S. P., & Zhenga, Y. (2016). Convolutional Neural Networks for Diabetic Retinopathy. *Procedia Computer Science*, 90, 200-205.
- Dutta, S., Manideep, B. C., Basha, S. M., Caytiles, R. D., & Iyengar, N. C. S. N. (2018). Classification of diabetic retinopathy images by using deep learning models. *International Journal of Grid and Distributed Computing*, *11*(1), 89-106.
- Sarki, R., Michalska, S., Ahmed, K., Wang, H., & Zhang, Y. (2019). Convolutional neural networks for mild diabetic retinopathy detection: an experimental study. Available from URL: https://doi.org/10.1101/763136.
- Sahlsten, J., Jaskari, J., Kivinen, J., Turunen, L., Jaanio, E., Hietala, K., & Kaski, K. (2019). Deep learning fundus image analysis for diabetic retinopathy and macular edema grading. *Scientific reports*, *9*(1), 10750.
- Gadekallu, T. R., Khare, N., Bhattacharya, S., Singh, S., Maddikunta, P. K. R., Ra, I. H., & Alazab, M. (2020). Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics*, *9*(2), 274.
- Rahman, M. A., Rahman, M. A., & Noshin, J. A. (2020). Automated Detection of Diabetic Retinopathy using Deep Residual Learning. *International Journal of Computer Applications, 177*(42), 25-32.
- Shaban, M., Ogur, Z., Mahmoud, A., Switala, A., Shalaby, A., Abu Khalifeh, H., ... & El-Baz, A. S. (2020). A convolutional neural network for the screening and staging of diabetic retinopathy. *Plos one*, *15*(6), e0233514.
- Khanusiya, A., & Savani, V. (2021). Diabetes Recognition using Convolution Neural Network. *International Journal of Scientific & Engineering Research, 12*(40), 496-500.
- Lam, C., Yi, D., Guo, M., & Lindsey, T. (2018). Automated detection of diabetic retinopathy using deep learning. *AMIA summits on translational science proceedings*, *2018*, 147-155.
- Wang, X., Lu, Y., Wang, Y., & Chen, W. B. (2018, July). Diabetic retinopathy stage classification using convolutional neural networks. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)* (pp. 465-471). IEEE.
- Nguyen, Q, H., Muthuraman, R., Singh, L., Sen, G., Tran, A. C., & Nguyen, B. P. (2020). Diabetic Retinopathy Detection using Deep Learning. Proceedings of the 4th International Conference on Machine Learning and Soft Computing, January 17 -19, 2020. Haiphong city Vietnam. Association of Computing Machinery, New York, United States……
- Khaled, O., ElSahhar, M., Alaa El-Dine, M., Talaat, Y., Hassan, Y. M. I., & Hamdy, A. (2021). Automatic Classification of Preliminary Diabetic Retinopathy Stages using CNN. *International Journal of Advanced Computer Science and Applications, 12*(2), 713-721.
- Deshpande, A., & Pardhi, J. (2021). Automated detection of Diabetic Retinopathy using VGG-16 architecture. *International Research Journal of Engineering and Technology (IRJET), 8*(3), 2936-2940.
- Sudha, V., & Ganeshbabu. T. R. A. (2021). Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning. *Computers, Materials & Continua*, *66*(1), 827-841.
- Aatila, M., Lachgar, M., Hrimech, H., & Kartit, A. (2021). Diabetic retinopathy classification using ResNet50 and VGG-16 pretrained networks. *International Journal of Computer Engineering and Data Science (IJCEDS)*, *1*(1), 1-7.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Alhadi, B., Ardaneswari, G., Tasman, H., & Lestari, D. (2014). Performance evaluation of fast smith-waterman algorithm for sequence database searches using CUDA GPU-based parallel computing. *Journal of Next Generation Information Technology*, *5*(2), 38-46.
- Bustamam, M., Sehgal, S., Hamilton, N. A., Wong, S., Ragan, M. A., & Burrage, K. (2009). An Efficient parallel Evaluation of Convolutional Neural Network Variants implementation of Markov clustering algorithm for large-scale protein-protein interaction networks that uses MPI, in 5th IMT-GT International Conference on Mathematics, Statistics and Their Applications, 2009.
- Riley, R. D., Debray, T. P., Collins, G. S., Archer, L., Ensor, J., van Smeden, M., & Snell, K. I. (2021). Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Statistics in medicine*, *40*(19), 4230-4251. doi: 10.1186/s40537-021-00444-8
- Decencière, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., Trone, C., ... & Klein, J. C. (2014). Feedback on a publicly distributed image database: the Messidor database. *Image Analysis & Stereology*, 33(3), 231-234.
- Porwal, P., Pachade, S., Kamble, R., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., & Meriaudeau, F. (2018). Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. *Data*, *3*(3), 25. [Online] Avaliable: https://ieee-datport.org/open-

access/indian-diabetic retinopathy-image-dataset-idrid.

- Diabetic Retinopathy Detection. Available online: https://www.kaggle.com/c/diabetic-retinopathy-detection.
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, *8*(1), 1-74.
- Debray, T. P., Vergouwe, Y., Koffijberg, H., Nieboer, D., Steyerberg, E. W., & Moons, K. G. (2015). A new framework to enhance the interpretation of external validation studies of clinical prediction models. *Journal of clinical epidemiology*, *68*(3), 279-289.
- Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., ... & Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine*, *162*(1), W1-W73.
- Nadeem, M. W., Goh, H. G., Hussain, M., Liew, S. Y., Andonovic, I., & Khan, M. A. (2022). Deep learning for diabetic retinopathy analysis: a review, research challenges, and future directions. *Sensors*, *22*(18), 6780.
- Sharma, H. S., Singh, A., Chandel, A. S., Singh, P., & Sapkal, P. (2019, May). Detection of diabetic retinopathy using convolutional neural network. In *Proceedings of International Conference on Communication and Information Processing (ICCIP)*. Available at SSRN: https://ssrn.com/abstract=3419210 or http://dx.doi.org/10.2139/ssrn.3419210
- Phung, V. H., & Rhee, E. J. (2018). A deep learning approach for classification of cloud image patches on small datasets. *Journal of information and communication convergence engineering*, *16*(3), 173-178. doi:10.6109/jicce.2018.16.3.173.
- Nguyen, H. V., Tan, G. S. W., Tapp, R. J., Mital, S., Ting, D. S. W., Wong, H. T., ... & Lamoureux, E. L. (2016). Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in Singapore. *Ophthalmology*, *123*(12), 2571-2580.
- Gumaei, A., Sammouda, R., Al-Rakhami, M., AlSalman, H., & El-Zaart, A. (2021). Feature selection with ensemble learning for prostate cancer diagnosis from microarray gene expression. *Health Informatics Journal*, *27*(1), 1460458221989402. doi:10.1177/1460458221989402.
- Ramspek, C. L., Jager, K. J., Dekker, F. W., Zoccali, C., & van Diepen, M. (2021). External validation of prognostic models: what, why, how, when and where?. *Clinical Kidney Journal*, *14*(1), 49-58. doi: 10.1093/ckj/sfaa188. PMID: 33564405; PMCID: PMC7857818.

Cite This Article: Odigie M. O, George G. O, Igodan E. C, Ukaoha K. C (2024). Detection of Diabetic Retinopathy Using VGG19 and ResNet 50 Models. *East African Scholars J Eng Comput Sci*, *7*(8), 84-96.