Classification of Retinal OCT Images Using Deep Learning

Malliga Subramanian
Professor, Department of Computer Science and Engineering,
Kongu Engineering College, Erode, Tamil Nadu, India

Kogilavani Shanmugavadivel
Associate Professor, Department of Computer Science and Engineering,
Kongu Engineering College, Erode, Tamil Nadu, India

Obuli Sai Naren, K Premkumar, K Rankish
Students, Department of Computer Science and Engineering,
Kongu Engineering College, Erode, Tamil Nadu, India

2022 International Conference on Computer Communication and Informatics (ICCCI) | 978-1-6654-8035-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCCI54379.2022.9740985

***Abstract*— In recent years, retinal disorders have become a serious public health concern. Retinal disorders develop slowly and without obvious signs. Every year, millions of individuals all around the world are diagnosed with retinal illness. Retinal illnesses manifest themselves in a variety of ways, but the majority of them result in visual problems. Retinal illnesses may damage any portion of your retina, causing visual problems, and some can even lead to blindness. Various retinal illnesses include Diabetic retinopathy, Macular pucker, Glaucoma, Macular hole, Age-related macular degeneration, Drusen, Central serous retinopathy, Macular edema, Vitreous traction, and Optic nerve abnormalities. Millions of light-sensitive cells (rods and cones) and other nerve cells make up the retina, which receives and organizes visual information. To avoid vision deterioration, early identification and treatment are critical. Optical Coherence Tomography (OCT) is a high-resolution diagnostic technology that can analyze and determine the quantitative distinction in diseased retinal layers. OCT - Optical coherence tomography which uses light waves is a non-invasive imaging procedure that takes cross-section photographs of your retina. It obtains a large number of detailed images of the retina, which are valuable for diagnosing and tracking changes in the retina and optic nerve over time. It is crucial in both diagnosing and selecting the appropriate treatment options. The accuracy of traditional approaches for classifying retinal disorders has ranged from 80% to 91%. As a result, a deep learning image identification system based on convolutional neural networks is presented to classify retinal illnesses more precisely and ideally in their early stages. The OCT pictures of the retina are classified into "AMD, CNV, DRUSEN, DMR, DR, MH, CSR, and Normal Eye" using a lightweight Deep neural network. On the Retinal OCT Images dataset, the accuracy achieved in this study with the aid of VGG16 is around 97%. When compared to other methodologies in the literature, it has a high level of accuracy in categorizing the illness.**

Keywords*— *Retinal Diseases, Convolutional Neural Network (CNN), Retinal OCT Images Dataset, VGG16, VGG19, DenseNet, Inception

I. INTRODUCTION

The sensitive membrane that lines the inner surface of the back of the eyeball is known as the retina. It is made up of multiple layers, one of which comprises photoreceptors, which are specialized cells. Rods and cones are the two kinds of photoreceptor cells in the human eye. Rod photoreceptors detect motion, give black-and-white vision, and are well-suited to low-light conditions. Cones are in charge of center vision and colour vision, and they work best in medium to strong light. Rods can be found all across the retina. Cones are concentrated in the macula, a tiny center portion of the retina. The fovea is a tiny depression in the center of the macula. The fovea is a spot in the retina that includes solely cone photoreceptors and is responsible for maximal visual acuity and colour perception. In recent years, retinal disorders have become a serious public health concern. Retinal disorders develop slowly and without obvious signs. Every year, millions of individuals all around the world are diagnosed with retinal illness.

Retinal illnesses manifest themselves in a variety of ways, but the majority of them result in visual problems. Retinal illnesses may damage any portion of your retina, causing visual problems, and some can even lead to blindness. Various retinal illnesses include Diabetic retinopathy, Macular pucker, Glaucoma, Macular hole, Age-related macular degeneration, Drusen, Central serous retinopathy, Macular edema, Vitreous traction, and Optic nerve abnormalities. These illnesses lead to (i) loss of vision (ii) Floaters and cobwebs (which are more visible when staring at a white backdrop) (iv) Flashing lights (v) Objects seeming smaller or larger than they are (vi) Decrease in peripheral vision or presence of shadows (vii) Distortion of straight lines The term "prevention of retinal disease" refers to measures taken in advance to reduce the chance of vision loss, as well as the degree and effect of vision loss. In roughly 80% of cases, blindness and visual impairment may be avoided. It is about four times as common in impoverished and uneducated persons living in disadvantaged and rural regions as it is in wealthier places.

978-1-6654-8035-2/22/$31.00 ©2022 IEEE

A little preventative step can have a significant impact. Ophthalmologists often diagnose and treat retinal diseases. An ophthalmologist is a medical specialist who specializes in the diagnosis and treatment (surgical and nonsurgical) of all sorts of eye illnesses. An ophthalmologist does a complete eye examination and searches for abnormalities anywhere in the eye. The Amsler grid test, Optical Coherence Tomography (OCT), Indocyanine Green Angiography, Ultrasound, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) are just a few of the techniques used to determine the location and severity of an illness.

The categorization of seven categories of retinal illnesses is presented in this research. Fig. 1 displays the grayscale OCT images utilised in the proposed work for the CNV, DME, DRUSEN, AMD, DR, MH, CSR, and NORMAL instances.

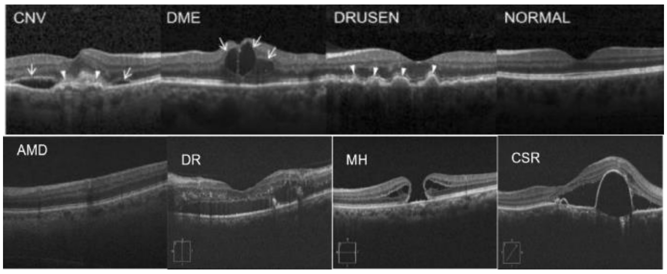


Fig. 1. Types of Retinal Diseases

OCT stands for optical coherence tomography, which is a non-invasive way to look at things. OCT takes cross-section photographs of your retina using light waves. Your ophthalmologist can use OCT to observe each of the retina's different layers. Your ophthalmologist will be able to map and assess their thickness as a result of this. These metrics aid in the diagnosing process. They also offer medical advice for glaucoma and retinal problems. OCT may be used to diagnose a variety of eye disorders. OCT scans can be used to identify all of the eye illnesses discussed in this research. Various models, including VGG16, VGG19, Densenet201, and InceptionV3, have been suggested and compared.

The rest of the work is orchestrated as follows: Section 2 briefly summarizes prior attempts to detect fire using CNNs. The materials and procedures utilized in this project are described in Section 3. In Section 4, the recommended models are explained. Section 5 presents the results of the experiments and examines the performance of the proposed models. Finally, Section 5 brings this article to a close.

II. Related Work

Several Deep Learning models based on well-known DL architectures have been utilized to diagnose retinal disorders. In addition, various researchers have released revised versions to improve the accuracy of retinal illness diagnosis in a range of situations. In recent years, deep CNN has been effectively applied to the detection and diagnosis of retinal diseases, with higher performance on detection tasks.

L. Huang et a. came up with a layer-guided convolutional neural network (LGCNN) that can find the difference between a normal retina and three common macular diseases such as edoema, drusen, and choroidal neovascularization. Retinal layer segmentation maps, in particular, are built using an effective segmentation network that can distinguish between two lesion-related retinal layers associated to relevant retinal lesions. The information from two lesion-related layers is then combined using two well-designed subnetworks in LGCNN. Thus, LGCNN is able to focus on the key lesion-related layer areas in order to improve OCT classification. The accuracy was estimated to be approximately 88%.

Sunija A.P., Sunija A.P., Sunija A. et al. presented a deep CNN architecture for accurately diagnosing and categorizing patients into Normal, DMD, and DME. To reduce intrinsic speckle noise, the Kuan filter is used to despeckle the input OCT pictures first. Furthermore, hyperparameter optimization techniques are used to tune the CNN network. In addition, K-fold validation is done to guarantee that the dataset is fully used, and the proposed model is assessed using many performance metrics utilizing the Mendeley database of labelled OCT pictures. The suggested model's classification accuracy is 95.7% as a consequence.

According to Feng Li et al., a classification strategy based on an ensemble of four classification model instances based on improved residual neural networks was proposed to classify retinal OCT images (ResNet50). This work utilized a 10-fold cross-validation method on the retinal OCT dataset, which was being developed. It was found that the proposed approach had classification accuracy of 0.973 (95 percent confidence interval [CI], 0.971–0.975), sensitivity of 0.960–0.966), and specificity of 0.985-0.987, which was comparable to or better than that of ophthalmologists with extensive clinical experience.

Deep learning architectures are increasingly being employed in the diagnosis of retinal illness from photographs, as indicated by the research described above. However, various gaps in the usage of deep learning architectures must be solved, such as quicker training time, fewer parameters, and so on.

III. Materials and Methods

A. Datasets

Data collection is a vital part of the initial stages of the project. The dataset collected from the various sources has been mentioned below.

The OCT images for the retinal disease are collected from Kaggle contain the categories AMD, CNV, DME, DRUSEN, and NORMAL.

An open-source OCT Image Database featuring a variety of retinal OCT pictures associated with various clinical diseases. CSR, DR, and MH categories were gathered from this source.

The image collected from various sources such as Kaggle and Open-ICPSR is now pre-processed. Both datasets contain different classes and different numbers of images. To use this to give as input to the model, the datasets have to be equalized.

The quantity of photos available typically enhances the performance of deep learning neural networks. Image augmentation is a technique for creating new training images from current training data artificially. Cropping, padding, and horizontal flipping are examples of image augmentation techniques that have been used to assist lessen the model's over-fitting. The dataset has roughly 24,000 photos after augmentation.

After equalizing the data collected from both datasets, a new dataset is created and hosted as a public dataset in Kaggle named “Retinal OCT – C8”.

The dataset was randomized and then, the Testing dataset and Training dataset split is performed by allocating 15% of the dataset for testing purposes, 15% for validation purposes and 75% of the dataset for training the classifier. This is done to ensure maximum availability of data for training which in turn would yield an accurate model.

B. *Convolutional Neural Networks*

The most extensively used and well-established deep learning model is the Convolutional Neural Network (CNN). CNN are a sort of deep learning algorithm that has been utilized in a range of picture categorization tasks and is gaining prominence in industries like health and music. A CNN has numerous layers, including convolutional, pooling, and dense layers, which are used to automatically learn data hierarchies using the Backpropagation method.

1) Convolutional layer : The convolutional layer is CNN's first and most important layer. A convolution is just the application of a filter to an input, which reduces the image size while also bringing all of the information in the field together into a single pixel. Fig. 2 depicts an image convolution that detects the edges of the animal in the image.

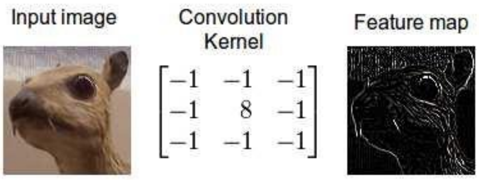
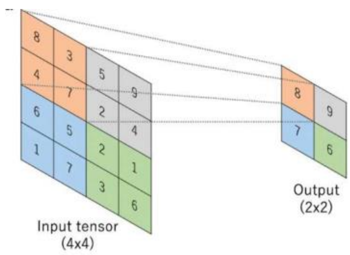


Fig. 2. Proposed workflow

2) Pooling Layer : Pooling layers are used to reduce the size or dimensions of the images or feature maps that are fed into the convolutional layer. Because of this, the number of features to be learned and the amount of computation required in the network are lowered. Maximum pooling layers, average pooling layers, and global pooling layers are the three types of pooling layers. The max pooling is used in this work. Fig. 3



shows how the max pooling layer reduces the image's dimension.

Adam is a learning rate optimization approach that was designed specifically for training Neural network topologies. To update deep neural network weights more efficiently, the Adam optimizer can be used instead of SGD (Stochastic Gradient Descent). This optimizer performs well when dealing with complex issues containing a large number of parameters or data. By adjusting the neural network's parameters, such as weights and learning rates, the Adam optimizer is used to reduce losses.

Fig. 3. Max Pooling operation Adam Optimizer

3) Activation Function : An activation function is a non-linear adjustment that may be done on the input before it is passed to the next layer of neurons, and it decides whether or not a neuron should be activated. In this work, the activation function is the Rectified Linear Unit (ReLU).

IV. PROPOSED MODELS

Retinal diseases are life-threatening and must be identified as soon as possible. Although retinal disease detection technologies have received a lot of attention in recent years, they still fall short in terms of comprehensiveness, speed, and accuracy. Using Retinal OCT pictures as input, the proposed approach compares several of the deep learning models. Retinal images are equalized, improved, and augmented as part of the data pre-processing processes. Finally, the pre-processed data is input into these models with changing parameters in order to choose the best appropriate model for assistance in the detection of illnesses before they cause blindness. In this work, CNN architectures such as VGG16, VGG19, DenseNet201, and InceptionV3 are examined and compared. Fig. 4 represents the workflow of the retinal disease detection system.

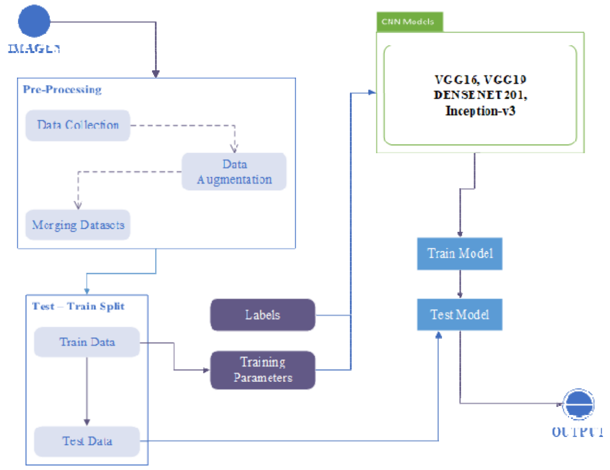


Fig. 4. Proposed workflow

A. VGGI6

The VGG16 was launched in 2014 by Simonyan and Ziserman of Oxford University's Visual Geometry Group Lab. It utilises a 16-layer network that is significantly deeper than AlexNet but delivers a simpler network due to the replacement of big kernel-size filters with several 33 kernel-size filters. This model achieved a top-five accuracy of 92.7 percent in ImageNet, a dataset of over 14 million images classified into 1000 classes. Fig.5. shows the architecture of VGG16.

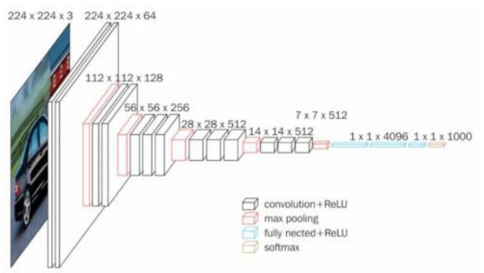


Fig. 5. VGGI6 Architecture

B. VGG19

The VGG19 model is a variant of the VGG model with a total of 19 layers. VGG11, VGG16, and others are additional VGG variants. VGG19 contains 19.6 billion FLOPs. VGG19 is a sophisticated CNN with pre-trained layers and an in-depth understanding of how an image's form, colour, and structure are created. VGG19 is a CNN that was trained on millions of images with difficult classification issues. The RGB image 224*224 is used as the input to the VGG-based convNet. When an RGB image with pixel values ranging from 0–255 is used as the input for the preprocessing layer, it is subtracted from the mean image values generated for the whole ImageNet training set. The VGG19 model's general design is seen in Figure 6.

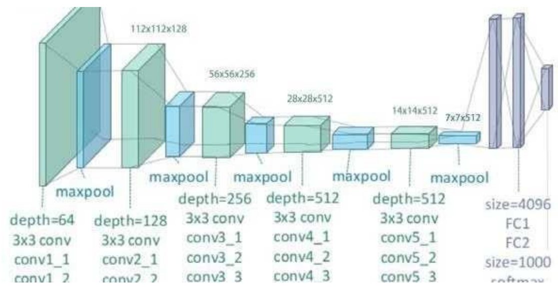


Fig. 6. Architecture of VGG19

C. DenseNet201

A DenseNet201 is a sort of convolutional neural network that uses Dense Blocks to create dense connections between layers, where all levels are directly linked (with matching feature-map sizes). In a feed-forward approach, each layer is directly linked to every other layer (within each dense block). When a layer is created, the feature maps of all preceding layers are handled as distinct inputs for each layer, whereas the feature maps of the current layer are passed on as inputs to all following layers. CIFAR10/100 (with or without data augmentation) and SVHN accuracy are achieved using this connection configuration at the cutting edge of technology. With much fewer parameters and nearly half the FLOPs of ResNet on the large-scale ILSVRC 2012 (ImageNet) dataset, DenseNet achieves the same degree of accuracy as ResNet on the same dataset. Fig.7 shows the overall architecture of the DenseNet201 model.

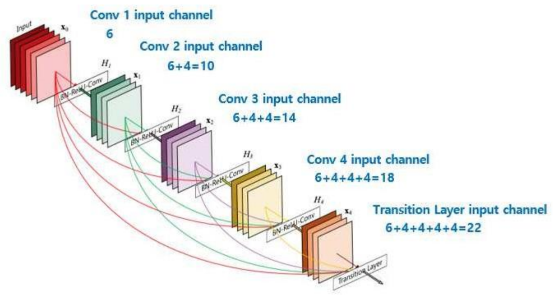


Fig. 7. Architecture of DenseNet201

D. Inception-V3

Inception-v3, a convolutional neural network design from the Inception family, includes label smoothing, factorized 7 x 7 convolutions, and an auxiliary classifier The general architecture of the Inception model is seen in Figure 8.

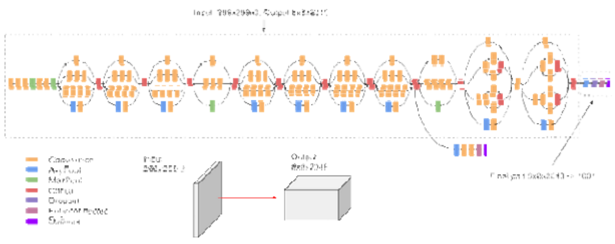


Fig. 8. Architecture of InceptionV3

V. RESULTS AND DISCUSSION

This research uses the VGG16, VGG19, InceptionV3 and DenseNet201 models, and the findings are presented and discussed in this part. To accommodate the input of the VGG, DenseNet, and Inception models, all pictures are shrunk to 224 * 224* 3 pixels for the experimental situations. The models were run for various numbers of epochs, and the training and validation accuracy and loss for the models are shown below.

Table I shows the results of the VGG16 model for different number of epochs. Initially, the VGG16 model has been trained and tested with 1 epoch. Then, gradually increased the number of epochs as there has been an increase in accuracy while increasing the number of epochs. For 20 epochs, VGG16 model achieves 92.4% training accuracy. But, after 20 epochs, there was no significant improvement in accuracy. This is so because, there aren't sufficient number of images.

TABLE I. PERFORMANCE OF VGG16 MODEL

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.8674	0.0432	0.9643	0.0139
5	0.9001	0.0333	0.9557	0.0157
10	0.9105	0.0284	0.9607	0.0152
15	0.9197	0.0256	0.9732	0.0105
20	0.9247	0.0246	0.9721	0.0106

The performance of the DenseNet201 model is shown in Table II. The number of epochs is gradually increased, as it was in VGG16, and the accuracy is documented. The rate of learning is set at 0.0001. For 15 epochs, the training accuracy was 93.8%.

TABLE II. PERFORMANCE OF DENSENET201

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.8135	0.5173	0.8768	0.3260
5	0.9204	0.2367	0.9039	0.3003
10	0.9324	0.1993	0.9104	0.2786
15	0.9387	0.1800	0.9214	0.2603

The VGG19 model's performance is shown in Table III. The rate of learning is set at 0.0001. For 20 epochs, the training accuracy was 92%.

TABLE III. PERFORMANCE OF VGG19

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.6759	1.0434	0.7829	0.6551
5	0.8668	0.4046	0.8725	0.3795
10	0.8949	0.3170	0.8929	0.3098
15	0.9080	0.2709	0.9029	0.2783
20	0.9208	0.2432	0.9018	0.2748

The performance of the InceptionV3 model is shown in Table IV. The rate of learning is set at 0.001. The training accuracy acquired for 20 epochs was 90%.

TABLE IV. PERFORMANCE OF INCEPTIONV3

EPOCH	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.7125	2.6478	0.75006	3.3374
5	0.8461	2.2179	0.8564	2.1803
10	0.8741	2.1769	0.8593	2.6370
15	0.8935	1.9161	0.8475	3.1476
20	0.8998	1.8438	0.8675	3.2558

The confusion matrix obtained for VGG16 model is depicted in Fig. 9. The diagonal elements of the confusion matrix represent correct classifications. The remainder, on the other hand, are misclassifications. The X axis depicts predicted classes, whereas the Y axis depicts actual classes. Individual class measures like as accuracy, precision, recall, and F1 score are determined using the confusion matrix. In order to calculate the values of these metrics, indices such as TP, FP, TN and FN are used.

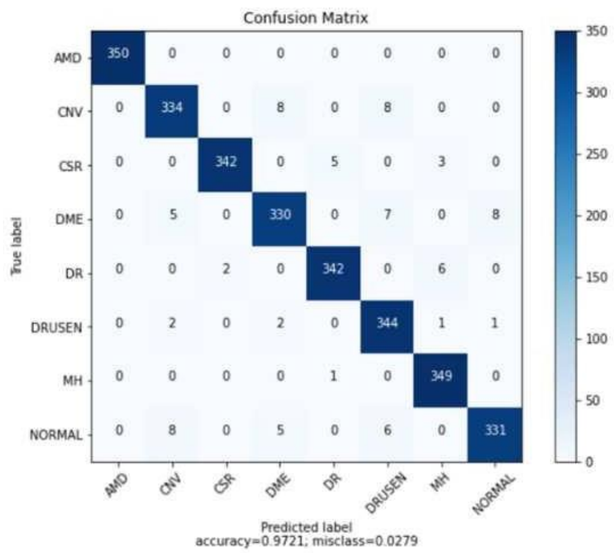


Fig. 9. Confusion Matrix for VGG16 Model

Equations (1) to (4) calculate TP, FP, TP, and TN where i=1,2,3 and 4 representing four classes.

$$tp_i = c_{ii} \tag{1}$$

$$fp_i = \sum_{l=1}^{n} c_{li} - tp_i \tag{2}$$

$$fp_i = \sum_{l=1}^{n} c_{il} - tp_i \tag{3}$$

$$tn_i = \sum_{l=1}^{n} \sum_{k=1}^{n} c_{lk} - tp_i - fp_i - fn_i \tag{4}$$

The metrics Accuracy, Precision, Recall, and F1-Score are defined as follows using these indices: Accuracy is defined as the number of samples correctly identified as a specific class out of total number of samples in that class and given by the Equation (5).

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \tag{5}$$

Recall (also called Sensitivity or True Positive Rate) is defined as the number of samples correctly classified as a specific class out of the total number of actual samples in that class and can be calculated using Equation (6).

$$\text{Recall} = TP/(TP+FN) \tag{6}$$

The number of samples accurately categorised as a certain class out of the total number of samples classified as that class is known as precision (Positive Predictive Value) and given by Equation (7).

$$\text{Precision} = TP/(TP+FP) \tag{7}$$

The F1-Score is a weighted average of precision and recall, or the harmonic average of precision and recall. Equation (8) is used to compute it.

$$\text{F1 Score} = (2 * \text{precision} * \text{recall}) / (\text{precision} + \text{recall}) \tag{8}$$

The performance of VGG16 model for individual classes is shown in Table V. From the experiments, the highest accuracy that is 97.2% was seen in the VGG16 model.

TABLE V. PERFORMANCE EVALUATION FOR DIFFERENT CLASSES

CLASS	ACCURACY	PRECISION	RECALL	F1 SCORE
AMD	100%	1	1	1
CNV	98.89%	0.95	0.96	0.96
DR	99.64%	0.98	0.99	0.99
MH	98.75%	0.94	0.96	0.95
CSR	99.50%	0.98	0.98	0.98
DRUSEN	99.04%	0.98	0.94	0.96
DME	99.61%	1	0.97	0.98
NORMAL	99%	0.95	0.97	0.96

VI. CONCLUSION

In recent years, retinal disorders have become a serious public health concern. Retinal disorders develop slowly and without obvious signs. Each year millions of people are identified with retinal disease all over the world. Early and accurate detection of retinal diseases is difficult because the symptoms are not visible in the early stage. The latest advancements in the technologies in the field of Deep learning has made it easier and more effective in achieving desired results. Small, low-latency, low-power VGG16 models have been the most commonly used, as they may be customized to meet the resource constraints of a wide variety use cases that can be deployed in various contexts. While comparing favorably with common models from the literature, they tradeoff between latency, size, and accuracy. Various models and implementation approaches have been tried and tested. It is shown that the proposed method using VGG16 has a high recognition rate and yields an average accuracy of around 98%. This can be extensively used to detect retinal diseases and prevent devastating consequences. With the use of OCT images, the model can diagnose retinal disorders in real time..

REFERENCES

[1] L. Huang, X. He, L. Fang, H. Rabbani and X. Chen, "Automatic Classification of Retinal Optical Coherence Tomography Images With Layer Guided Convolutional Neural Network", in IEEE Signal Processing Letters, vol. 26, no. 7, pp. 1026-1030, July 2019. doi : 10.1109/LSP.2019.2917779

[2] Sunija A P, Saikat Kar, Gayathri S, Varun P. Gopi, P. Palanisamy, "A Lightweight CNN for Retinal Disease Classification from Optical Coherence Tomography Images",Volume 200,2021,105877,ISSN 0169-2607. doi : https://doi.org/10.1016/j.cmpb.2020.105877

[3] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep learning is effective for classifying normal versus age-related macular degeneration OCT images," Ophthalmology Retina 1(4), 322–327 (2017). Rajagopalan, N., Narasimhan, V., Kunnavakkam Vinjimoor, S. et al. Deep CNN framework for retinal disease diagnosis using optical coherence

tomography images. J Ambient Intell Human Comput 12, 7569–7580 (2021). doi : https://doi.org/10.1007/s12652-020-02460-7

[4] Feng Li, Hua Chen, Zheng Liu, Xue-dian Zhang, Min-shan Jiang, Zhi-zheng Wu, and Kai-qian Zhou, "Deep learning-based automated detection of retinal diseases using optical coherence tomography images," Biomed. Opt. Express 10, 6204-6226 (2019). doi : https://doi.org/10.1364/BOE.10.006204

[5] P. P. Srinivasan et al., “Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images,” Biomed. Opt. Express, vol. 5, no. 10, pp. 3568–3577, 2014.

[6] Eladawi N, Elmogy MM, Ghazal M, Helmy O, Aboelfetouh A, Riad A, Schaal S, El-Baz A. “Classification of retinal diseases based on OCT Images”. Front Biosci (Landmark Ed). 2018 Jan 1;23:247-264. doi: 10.2741/4589. PMID: 28930545.

[7] Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), “Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images”, Mendeley Data, V3, doi: 10.17632/rscbjbr9sj.3 Y. Sun, S. Li, and Z. Sun, “Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning,” J. Biomed. Opt., vol. 22, no. 1, pp. 12–21, 2017.

[8] Gholami, Peyman, and Lakshminarayanan, Vasudevan. Optical Coherence Tomography Image Retinal Database. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-02-20. https://doi.org/10.3886/E108503V1

[9] R. Koprowski, S. Teper, Z. Wrbel, and E. Wylegala, “Automatic analysis of selected choroidal diseases in OCT images of the eye fundus,” BioMed. Eng. OnLine, vol. 12, no. 1, 2013, Art. no. 117.

[10] KAGGLE – Retinal OCT Images

[11] https://www.kaggle.com/hadeersaif/retinal-oct5classes