I.J. Image, Graphics and Signal Processing, **2025, 4, 105-127**
Published Online on August 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2025.04.07

Evaluation of Deep Learning Approaches to Detect Choroidal Neovascularization

Sahil Chukka
Pillai College of Engineering, New Panvel, India
Email: schukka20it@student.mes.ac.in
ORCID iD: https://orcid.org/0009-0003-1784-9617

Vardhanika Jagtap
Pillai College of Engineering, New Panvel, India
Email: vjagtap20ite@student.mes.ac.in
ORCID iD: https://orcid.org/0009-0000-6009-1133

Naveen Patel
Pillai College of Engineering, New Panvel, India
Email: npatel20ite@student.mes.ac.in
ORCID iD: https://orcid.org/0009-0009-8206-0560

Sudiksha Jadhav
Pillai College of Engineering, New Panvel, India
Email: jsud2020ite@student.mes.ac.in
ORCID iD: https://orcid.org/0009-0004-6602-6761

Mimi Cherian*
Pillai College of Engineering, New Panvel, India
Email: mcherian@mes.ac.in
ORCID iD: https://orcid.org/0000-0002-1608-9682
*Corresponding author

Jinesh Melvin Y. I.
Pillai College of Engineering, New Panvel, India
Email: yijmelvin@mes.ac.in
ORCID iD: https://orcid.org/0000-0003-1515-8487

Received: 15 March, 2024; Revised: 16 July, 2024; Accepted: 20 May, 2025; Published: 08 August, 2025

Abstract: In ophthalmology, Choroidal Neovascularization (CNV) is a serious medical disease that, if left untreated, frequently results in significant vision loss. In this investigation, we investigate the evaluation and working of deep learning models, notably basic Convolutional Neural Networks (CNN), ResNet18, ResNet50, VGG16, VGG19, Vision Transformers, EfficientNetV2L, MobileNetV2 and InceptionV3 for identification and classification of CNV in Optical Coherence Tomography (OCT) images. The Kermany dataset, which includes OCT images of both CNV-patients and non-CNV patients (Normal OCT images) are utilized for this paper. The dataset was further used in three different versions based on validation and training split. The images from the dataset are already pre-processed and labelled so no pre-processing operations were performed, how- ever resizing of images have been performed according to the models. The deep learning models are trained and evaluated on standard performance metrics such as precision, recall, accuracy, F1-score, etc. All things considered, our work shows the evaluation of deep learning models to classify OCT images that show the presence of CNV. Based on all three dataset versions, the findings of our study confirm that ResNet18, VGGNet19, and MobileNetV2 beat all other approaches and achieved an average accuracy of 1. Additionally, Vision Transformer and Effi- cientNetV2L demonstrated strong performance, averaging 0.99 and 0.96 accuracy on each of the three dataset versions, respectively. These models have the potential to help ophthalmologists detect CNV early and monitor it, which may lead to prompt treatment and better vision preservation for patients.

Index Terms: Choroidal Neovascularization, CNN, ResNet, Vision Transformer, OCT images, Deep learning

1. Introduction

The human eye, a complex work of evolution, provides us with the priceless gift of sight and acts as a window to the outside world. This intricate organ plays a crucial part in our everyday interactions, learning, and comprehension of the world in addition to enabling us to experience the complexity of our surroundings. This wonderful technology is not, however, immune to difficulties. Vision and quality of life can be affected by ocular disorders, which can range from mild problems like frequent refractive errors to serious ones like Choroidal Neovascularization (CNV).

1.1 Choroidal Neovascularization (CNV)

Neovascularization is the technical word for a new blood vessel. The choroid, a layer of blood vessels behind the retina, is where these brand-new, abnormal blood vessels begin. Vascular endothelial growth factor (VEGF) is overproduced in the retinas of AMD patients, leading to the development of new blood vessels from the choroid into the retina. Exudative macular degeneration, associated with CNV, manifests as choriocapillaris thinning, alterations in Bruch’s membrane thickness, and deposition of insoluble material in the retinal layers. Lipoprotein accumulation results in hypoxia, para-inflammation, and atrophy of the photoreceptors and retinal pigment epithelium (RPE) [1].

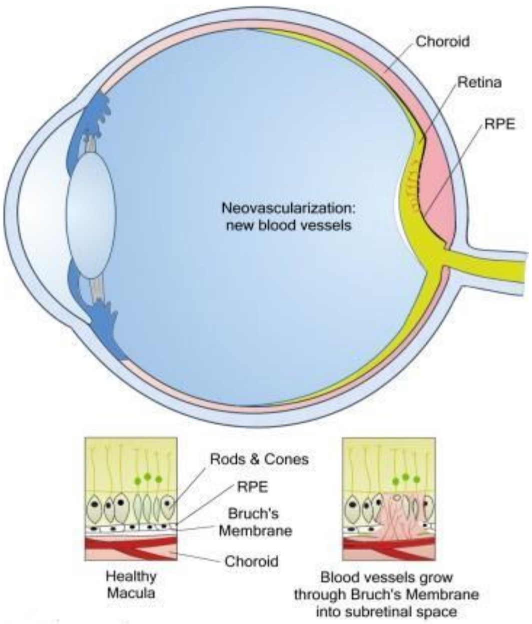


Fig. 1. Diagram of the CNV affected eye depicting the growth of new blood vessels through Bruch's membrane into sub retinal space. [2]

A distortion or waviness in the center of the field of vision, along with a gray, black, or empty area, are further signs of CNV. This should trigger an immediate phone call to an ophthalmologist to schedule an urgent emergency visit. By injecting a medicine that blocks a protein called VEGF into the eye, the ophthalmologist can stop the development and leaking of blood vessels, but only if they can do it as soon as possible—within hours or days—after the moment you detect the change in your vision. A vision lost is time lost!

Refer to Fig.1 for better understanding. There are multiple imaging techniques which help in the detection of CNV. Few are listed below,

- Optical Coherence Tomography (OCT): Using light waves, OCT [3] is a non-invasive imaging technique that creates detailed cross-sectional images of the retina. It is commonly used to detect abnormal blood vessel existence and localization on the retina. [4] employed a new architecture CorNet to segment the corneal tissue in OCT images. They computed Mean Absolute Difference in Layer Boundary Position (MADLP) for Epithelium (EP) 0.33 ±0.21, $Bowman's Layer$ (BL) 0.42 $\pm$0.13$andEndothelium$ (EN) 0.79 $\pm$0.19.
- Fluorescein Angiography (FA): In FA, a fluorescent dye is injected into the circulation, and as the dye circulates, retinal pictures are taken. This method aids in identifying aberrant blood arteries since they frequently leak dye and show up as patches of hyperfluorescence.[5] suggest a system for automatically recognising microaneurysms in FA using optimal wavelet transform, Microaneurysms were detected with sensitivity of 89.62%, 90.24%, and 93.74%, and positive predictive values of 89.50%, 89.75%, and 91.67%—better than previously reported methods.
- Indocyanine Green Angiography (ICG): ICG is comparable to FA, but it makes use of a different dye that may be able to reveal additional details about the choroidal blood vessels that may be crucial for CNV diagnosis [6], employed a generative adversarial network to generate Indocyanine Green Angiography (ICG)

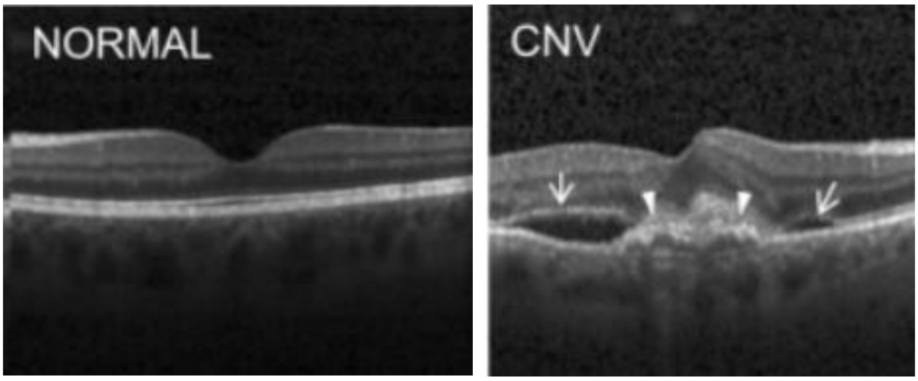
images using Fundus Images. When compared to the original colored fundus, the average PSNR of the max and mean function of the generated ICGA is 16.35, whereas it is 16.25 for the original ICGA.

- Fundus Photography: High-resolution color photographs of the retina and other internal structures of the eye are taken using fundus photography. Hemorrhages or exudates in the retina could be signs that CNV is present.[7] suggested their own design, the clinically oriented fundus enhancement network, or ”cofeNet,” to improve fundus pictures and reduce global degradation factors. With their architecture, they were able to precisely adjust the quality of the image.

Our research study primarily focuses on the detection of Choroidal Neovascularization (CNV) using Optical Coherence Tomography (OCT) images. Sub section 1.2 explains OCT images and its relevance in this paper.

1.2 OCT images

OCT imaging technology is based on the utilization of low-coherence light for micro- and sub-micron-scale images, and can be used to acquire three-dimensional images. OCT imaging technique produces cross-sectional images of a medium that exhibits optical scattering utilizing light waves, typically with long wavelengths, and is connected to certain processes [3]. In ophthalmology, optical coherence tomography (OCT) has become a ground-breaking imaging method. OCT offers highly detailed cross-sectional images of the retina, enabling medical professionals to see the retinal layers and spot anomalies with astounding accuracy. The detection and treatment of retinal illnesses, including CNV, have been revolutionized by this non-invasive imaging technique. The following information explains how to tell a normal OCT scan from one that has CNV.



(a) (b)

Fig. 2. Depiction of OCT images. Fig. 2(a) shows a normal OCT image which depicts the absence of intraretinal fluid. Fig. 2(b) shows a CNV affected OCT image which depicts the presence of intraretinal fluid (marked by arrows in the figure)

As shown in Fig.2(a), The OCT image depicts an absence of any intraretinal fluid with a combination of multiple pigment epithelial detachments and sub retinal hyper reflective material stating that it is an Normal OCT image. While Fig.2 (b) indicates the presence of intraretinal fluid, pigment epithelial detachments and sub retinal hyper reflective material (marked by arrows in the figure) stating that the OCT image is of a CNV affected patient.

1.3 Contribution of the Study

This section highlights the study’s key contributions:

- This study shows the individual capabilities of the selected nine models. This study demonstrates their raw effectiveness for the classification of OCT images diagnosed with CNV.
- The same dataset was used in three different iterations. Every version has a distinct validation and training split. Hence, this study has attempted to put a new idea to use by doing this.
- This research has provided a thorough analysis of the prior research in the literature review. This study has discussed their excellent findings as well as the gaps in their research.
- Compared to other research studies, this research study is one of the few using the Vision Transformer model for CNV detection. The application of Vision Transform- ers in clinical and other fields of research may be encouraged by the results of this study.

1.4 Organization of the paper

The study began with an overview of CNV illness, OCT pictures, and several CNV imaging techniques. In Section 2, we have mentioned the motivation and objectives behind this research. A brief overview of earlier efforts will be provided in Section 3. The methodology will be covered in Section 4, the results will be covered in Section 5, Section 6 provides a discussion on the topic while comparing this research’s results to other research works and a conclusion will be provided in Section 7.

2. Motivation and Objectives

Several reputable organizations, including the World Health Organization, have warned that the rapid increase of glaucoma, which causes irreversible blindness in human society, would create a worrying scenario in the years to come. The World Health Organization (WHO) estimates that 2.2 billion people worldwide have vision impairments. At least 1 billion of these incidents, or close to half of them, resulted in aesthetic damage that was either avoidable or ignored.[8].Refer to Fig.4 for additional information.

The motivation behind writing a research paper on the detection of Choroidal Neo- vascularization (CNV) using DL models, such as CNNs, ResNets, Vision Transformers, VGGNets etc stems from several important factors which are mentioned below. The models were often chosen and optimized for tasks like detecting Choroidal Neovascularization (CNV) due to their specialized architecture and capabilities. CNNs are inherently designed for image analysis. Their convolutional layers can extract spatial features such as edges, textures, and patterns, which are critical for identifying abnormal structures like CNV in medical images. The Optimized part of CNV detection is Pretrained Models, Fine-Tuning, Data Augmentation. ResNets address the vanishing gradient problem, making it easier to train very deep networks. This is crucial for detecting subtle anomalies like CNV in high-resolution images. ResNet allows the model to focus on both high-level and low-level features simultaneously. Freezing initial layers during fine-tuning can help focus training on disease-specific patterns. Vision Transformers by analyzing them as a sequence of patches. Patch Size Selection, Pretraining on Medical Datasets, Hybrid Architectures are optimized in Vision Transformers. VGGNets have a simple yet effective architecture that provides robust feature extraction with consistent receptive field sizes and reduces the complexity.

- Clinical Importance: AMD and myopia degeneration are two retinal diseases that are associated with CNV, a serious and sight-threatening consequence. In order to stop permanent vision loss, early identification and management are essential. Therefore, the need for precise and effective diagnostic techniques is urgent.
- Advancements in Deep Learning: In a variety of domains, including image analysis and medical picture interpretation, deep learning approaches, notably CNNs and

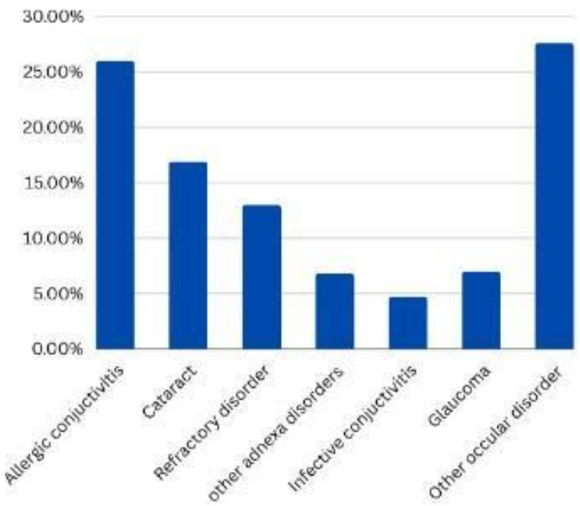


Fig. 3. A graphical representation of the recent cases related of ocular disorders, illustrating Allergic Conjunctivitis(20%-30%), Cataract(15%-20%), Refractory Disorder(10%-15%), other Adnexa Disorders(0%-10%), Infective Conjunctivitis(0%- 10%), Glaucoma(0%-10%) and other ocular disorders (25%-30%).

Sophisticated architectures like ResNet, have displayed astounding capabilities. The accuracy and speed of diagnosis might be increased by using these technologies for CNV identification.

- Automation and Efficiency: It might take time and be prone to mistake to manually diagnose CNV in optical coherence tomography (OCT) pictures. Deep learning algorithms might greatly increase the efficiency of CNV diagnosis by automating this procedure, freeing up doctors to devote more time to patient care and treatment options.
- Potential for Early Intervention: Timely detection of CNV is critical for initiating prompt treatment and preserving patients' vision. Deep learning algorithms can aid in the early identification of CNV, enabling interventions at earlier stages of the disease when treatments are often more effective.
- Research Gap: While there has been substantial research in the field of medical image analysis using deep learning, there may be a gap in specific applications such as CNV detection.

- Some identified research gaps are: Small dataset used comparisons with other models have not been presented Further, research gaps are mentioned in Table 1 in Section 3.Also, it's important to clearly outline the objectives of our study. Here are some specific objectives we would like to mention:
- Create, develop, and put into use deep learning models, such as CNNs and ResNet, for automatically detecting CNV in optical coherence tomography (OCT) pictures.
- To compile an extensive and well-annotated collection of OCT pictures, including those with and without CNV, and to carry out the required preprocessing operations to improve image quality and eliminate artefacts.
- Utilizing the provided dataset to train and optimize the deep learning models, including choosing the best hyper parameters, regularization methods, and data augmentation approaches to boost model performance.
- To rigorously evaluate the performance of the developed models in CNV detection using established metrics such as sensitivity, specificity, accuracy.

3. Literature Review

This section will provide a thorough analysis of earlier efforts done to detect multiple ocular diseases. This provides a comprehensive overview and background of existing algorithms and knowledge in this field. The two primary methods used in the research to detect ocular diseases are image classification and image segmentation.

The study by Venkatesan et al. IEEE (2021) [9] uses the Machine Learning Scheme (MLS) to classify OCT images into normal and CNV affected classes. Their approach involves gathering photos, applying image scaling, pre-processing, and removing bespoke features. The Mayfly Optimization Algorithm (MOA) is used to reduce features and operationalize a two-class classifier. The study used 484 retinal OCT images, with 242 healthy and 242 featuring CNV issues. The top three classification methods were SVM-FG (92.86%), SVM-SG and CT (91.67%), and FT (90.48%). The study uses a small number of photos, but future improvements could increase the number of images.

Jinyoung et al. MDPI (2023) [10] proposes a deep convolutional network to analyze OCT images for diagnosing macular diseases. The study uses a dataset from Hangil Eye Hospital, including SD-OCT images of non-affected volunteers, nAMD and CSC. Data augmentation and transfer learning were used to address training data shortages. The VGG19 model with four fully connected layers achieved the maximum accuracy in classifying normal, nAMD, and CSC classes. The results achieved showed great accuracy in detecting ocular disorders. The maximum classification accuracies (99.7%, 91.1%) were attained by the VGG19 model with four fully connected layers while classifying the normal, nAMD, and CSC classes (3-class classification).

ROCT-Net, a newly proposed ensemble model developed by Mohammad et al. IEEE (2021) [11], uses a mixture of neural models and transfer learning to identify common diseases in retinal OCT images. The model achieved high accuracy on both the OCTID and Kermany datasets, also Mean Sensitivity: 0.9870 and Specificity: 0.9957. The study by Baharlouei et al. IEEE (2022) [12] discusses the use of a Wavelet Scattering Network (WNS) approach for detection of retinal abnormalities in OCT images. The researchers highlight the limitations of conventional techniques, such as noise, poor image quality, and complex abnormal shapes. They use the OCTID open-access dataset, which includes 572 OCT pictures divided into five classes: Normal, CSR, Macular Hole, AMD, and Diabetic Retinopathy. The study achieved 97.4% accuracy in normal/abnormal diagnosis and 100% accuracy in DR pathology.

Anna Lathe M et al. IEEE (2018) [13] study uses color picture segmentation of fundus blood vessels to identify hypertensive retinopathy, a condition linked to high blood pressure. Fundus scans provide detailed images of the retina, and a dataset of 7 patients is used. The process entails morphological structure, contrast limited Adaptive Histogram Equalization, and removing the optic disc from the input image by transforming it into a complementary image. The output array is then converted into a binary image, and the color-segmented retinal blood vessels are revealed by concatenating three photos.

The study by Bino N et al. IEEE (2022) [14] explores feature extraction and quantification in fundus auto fluorescence (FAF) images of patients with central serous retinopathy (CSR). The FAF imaging technique provides information about retinal cell health and metabolic activity. The study uses 30 patients with CSR, resizing the images to 584 x 565 pixels and transforming them to grayscale. Gaussian noise is eliminated using a 3 x 3 wiener filter. The study accurately segmented blood vessels, macular region, and optic disc to determine the extent of CSR.

Pawan et al. Springer (2021) [15] created structures based on capsule networks to separate sub-retinal serous fluid in CSR affected in OCT images. The architecture uses an encoder-decoder architecture with strided convolutions and three stages of convolutional capsules. The model uses masked reconstruction for regularization. The proposed method outperformed existing methods using CSCR data from the Pink City Eye and Retina Centre in Jaipur, India. They were able to outperform other architectures using their architecture. Capsule-UNet & Capsule-Att-UNet outperform existing methods (DSC: 0.869 & 0.874).

The study by Abirami M.S et al.ResearchSquare (2021) [16] used two deep convolution neural network models, VGGNet-16 and DenseNet to detect CNV. The study used a private dataset for model training and testing. A total of 10000 OCT images were used in this dataset. CNV and Normal are the two classifications in the dataset. Each class has 1500 testing images and 3500 training images. On the train- ing dataset, VGGNet-16 obtained an accuracy of 99.03%, and on the testing dataset, 97.53%. On the training dataset, DenseNet obtained an accuracy of 99.14%, and on the testing dataset, 96.87%. DenseNet underperformed on the testing dataset despite outperforming VGGNet-16 on the training dataset.

The study by Jie Wang et al. tvst (2023) [17] used deep learning models for diagnosing and segmenting CNV using OCTA images. The entire architecture can be divided into two main halves. The first half uses multiple subnets for 2D and 3D feature extraction. The other half has two models for diagnosis and segmentation i.e. ResNet for diagnosis and U-Net for segmentation of the affected region in the OCTA images. The authors used a private dataset collected from 5 clinics. They gathered 4701 CNV affected OCTA images and 5865 normal OCTA images. They achieved a sensitivity of 95% at 95% specificity and AUROC = 0.97.

The study by Wei Feng et al. Springer (2023) [18] employed deep learning models to separate the CNV-affected area from the OCTA pictures of patients with nAMD. The authors developed an architecture which holds traditional U-Net as the back- bone for segmentation. They replace the encoder and decoder block of U-Net with Resets blocks, which helps it learn feature representation. They add an adapted spatial pyramid pooling module for the encoders. They applied this architecture on a private dataset gathered by them and the results they achieved are of great accuracy, AUC=0.9476, Sensitivity=0.9950 and Specificity=0.7271.

The study by G.V. Vinod et al. JETIR 2023 uses the publicly available Kermany dataset to detect various ocular diseases using deep learning approaches. The architecture developed by the authors employs segmentation of OCT and then classifies into the 4 classes provided in the dataset. They were able to classify images with great results.

The study by Takahiro Sogawa et al. PONE (2020) [19] explores the accuracy of deep convolutional neural networks to detect myopic macular diseases using OCT images. They tested nine different deep learning models on a private dataset. Nine Deep Neural Network models, namely VGG16, VGG19, ResNet50, InceptionV3, InceptionResNetV2, Exception, DenseNet121, DenseNet169, and DenseNet201, were built and trained. The model that performed the best as an ensemble of VGGNet-16, VGGNet-19, DenseNet121, InceptionV3, and ResNet50 was as follows: 0.970 for Area under the ROC Curve, 90.6% for sensitivity, and 94.2% for specificity.

Mahsa Vali et al. MDPI 2023[20] explored deep learning models to segment and classify CNV affected regions from OCTA images. The suggested model by the authors determines if the CNV activity criteria—branching, peripheral arcade, dark halo, shape, loop, and anastomoses—are present in OCTA pictures or not. A dataset of 130 images were used on a U-Net model and later five binary classification networks were used. The segmentation phase yielded a Dice coefficient of 0.86, whereas the five activity criteria were represented by individual classifiers with matching accuracies of 0.84, 0.81, 0.86, 0.85, and 0.82, respectively.

Parsa Riazi Esfahani et al. cureus (2023) [21] developed a robust deep learning model to detect CNV, DME and Diabetic Retinopathy. They incorporated a dataset from Kaggle.com, a total of 3133 OCT images were incorporated. They achieved an accuracy of 0.98, recall of 0.98, and precision of 0.95 and F1-score of 0.97.

OCT imaging and fundus imaging were the two primary imaging modalities employed in the literature review publications. Only [13] and [14] utilized Fundus photos to train their models; the remaining [9, 16, 17, 18, 19, 20, 21, 22,] used OCT images.

Three datasets were primarily used in all of these papers: the Kermany, OCTID, and private datasets. While [11] utilized both the OCTID and Kermany datasets for their model, [9, 16, 21, 22] used the Kermany dataset. Similarly, [12] also made use of the OCTID dataset. The following [10, 13, 14, 15, 17, 19, 20] utilized private datasets.

The provided Table 1 below offers a comprehensive overview of prior research efforts in the field of ocular diseases, encompassing various disease types and modeling approaches. It comprises essential columns, including literature references, models employed, resultant accuracy metrics, and the identification of research gaps. This summary effectively encapsulates the diverse body of work done in the domain of ocular diseases, providing valuable insights for researchers and practitioners seeking to further our understanding and treatment of these conditions.

Table 1. Summary of previous work done on different ocular diseases and different models (Continued)

Literature	Methods	Results Accuracy (%)	Research Gap
Venkatesan et al. IEEE 2021[9]	MLS	SVM-L: 84.52 SVM-Q: 88.09 SVM-FG: 92.86 SVM-CG: 91.67 CT: 91.67 FT: 90.48	Small dataset used.
Jinyoung et al. MDPI 2023[10]	DCNN	Class Classifica tion: VGG–16: 99.1 VGG–19: 99.7 Resnet: 98.5 Class Classifica- tion: VGG–16: 90.3 VGG–19: 91.1 Resnet: 87.4	Small dataset used. The model's perfor mance assessment relied on a single OCT image.
Mohammad Rahimzadeh et al. IEEE 2021[11]	ROCT-Net	0.9870 0.9646.	For OCTID dataset, Even with very less number of dataset it exhibited high accuracy, which can lead to overfitting of the model.
Zahra Baharlouei et al. IEEE 2022[12]	WSN	82.5	Retinal disease diagnosis can be improved by more Wavelets function.
Anna Latha M et al. IEEE 2018[13]	Colour Image Segmentation	92	Resilience of approach Not tested.
Mr. Bino. N et al. IEEE 2022[14]	Feature Extraction	Algorithm extract CSR in retinal images.	Models can be created to forecast whether images exhibit signs of CSR
S. J. Pawan et al. Springer 2021[15]	Capsule Networks	Capsule-UNet: DSC: 0.869 Capsule-Att-Unet: DSC: 0.874	Small dataset used. Capsule network-based Architectures can be More improved by optimizing the routing algorithm.
Abirami M.S et al. ResearchSquare 2021[16]	DCNN	VGG16: 99.03 DenseNet: 96.87	Image segmentation can Be improved Hough transforms and heatmap variance comparing with Canny edge and contour Marking outputs. More Deep learning
Mahsa Vali et al. MDP 2023[20]	DCNN	Classifiers: Branch: 0.60 Shape: 0.74 Anastomosis and loops: 0.70 Peripheral Arcade: 0.79 Dark Hola: 0.65	Small sample and cross- sectional design. Large datasets can be used, data augmentation can be employed.

4. Proposed Methodology for Detecting CNV

Medical imaging, especially the diagnosis of eye disorders, has increasingly used deep learning approaches. Employing artificial neural networks, these approaches analyze vast volumes of medical picture data to find patterns linked to different illnesses. This research presents a thorough investigation of deep learning approaches' use in the field of detecting CNV. The architectures used are CNN, ResNet, VGGNet and Vision Transformers. ResNet and VGGNet are further divided into two architectures, ResNet18, ResNet50, VGGNet-16(VGG16 in short), VGGNet-19(VGG19 in short) respectively. The process takes place as shown in Fig.4

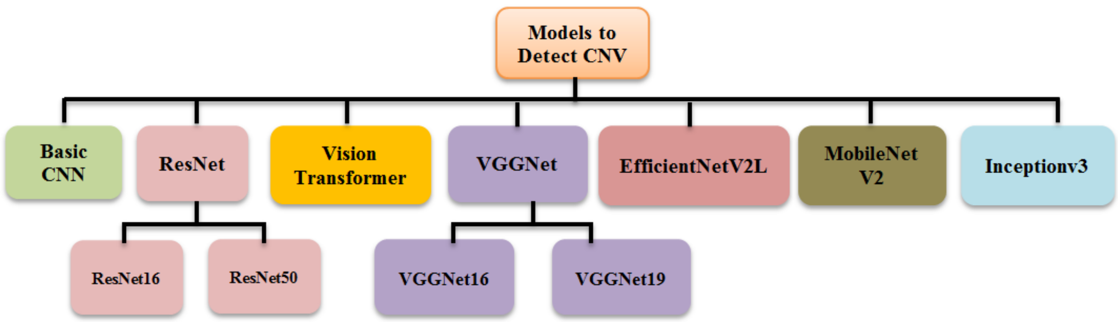


Fig. 4. All the deep learning models used to detect CNV. CNN, ResNets, VGGNets, Vision Transformers, MobileNetV2 and Inceptionv3.

4.1 Dataset

A series of medical photos is utilized as the dataset for the identification of Choroidal Neovascularization (CNV), which is used to train, test and validate deep learning models for detection. The images were taken from the Kermany dataset [23].

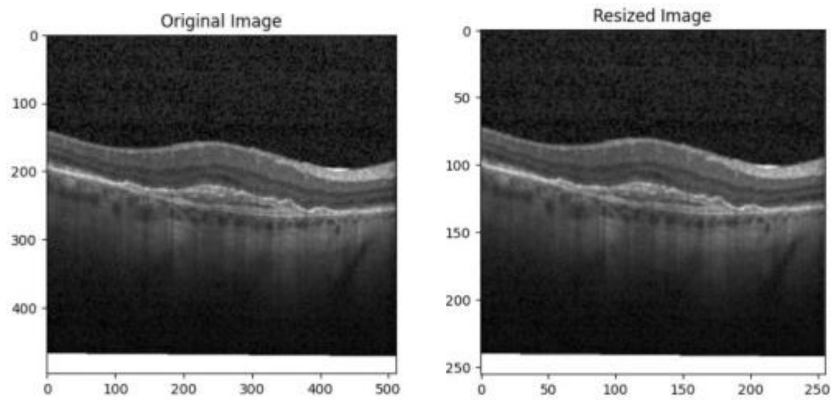
Table 2. Kermany [23] Dataset used in this paper which is further split into three versions

Dataset	**Dataset 1(60% & 40% split)**		**Dataset 2(70% & 30% split)**		**Dataset 3(80% & 20% split)**	
	CNV	Normal	CNV	Normal	CNV	Normal
Training	4810	4810	5612	5612	6413	6413
Validation	3206	3206	3206	3206	1603	1603
Testing	242	242	242	242	242	242

For the 60-40 dataset split, the training dataset consists of a total 9620 OCT images which were further divided into two classes: CNV affected (4810 images) and Normal (4810 images). For validation dataset, 3206 images were used from both classes each, for the 70-30 dataset split, the training dataset consists of total 11224 OCT images which were further divided into two classes: CNV affected(5612 images) and Nor- mal(5612 images). For the validation dataset, 2404 images were used from both classes each and for the 80-20 dataset split, the training dataset consists of total 12556 OCT images which were further divided into two classes: CNV affected(6413 images) and Normal(6413 images). For the validation dataset, 1603 images were used from both classes each. For the testing dataset, a total of 242 CNV oct images and 242 Normal oct images were used.

4.2 Pre-processing

All of the photos from the three datasets were combined into a single pattern during the preprocessing stage. The photographs' colours haven't been altered in any way. The 16032 photos were all reduced in size to 224x224.The test dataset's photos were likewise downsized to 224x224 and cropped. Fig. 5(a) shows the OCT image before resizing and fig. 5(b) depicts the resized OCT image.



(a) (b)

Fig. 5. Depiction of the pre-processing stage. Fig.5(a) depicts the original OCT image from the dataset. Fig.5(b) depicts the resized OCT images.

4.3 Architectures tested for the detection of CNV

The architectures used are CNN, ResNet, VGGNet, Vision Transformers, EffiencetNetV2L, MobileNetV2 and InceptionV3. ResNet and VGGNet are further divided into two architectures, ResNet18, ResNet50, VGG16, and VGG19 respectively. Images are pre-processed for each architecture.

In brief, Fig.6 shows the working of our proposed model. As we can see, the dataset contains OCT images of two classes (CNV & Normal). The OCT images are then pre-processed, they are then fed to the models for training. Further, the models use the testing dataset for evaluation. Finally, the models predict if the OCT images that have been given as an input belong to either of the two classes. Based on that accuracy, recall, precision and F1-score are achieved.

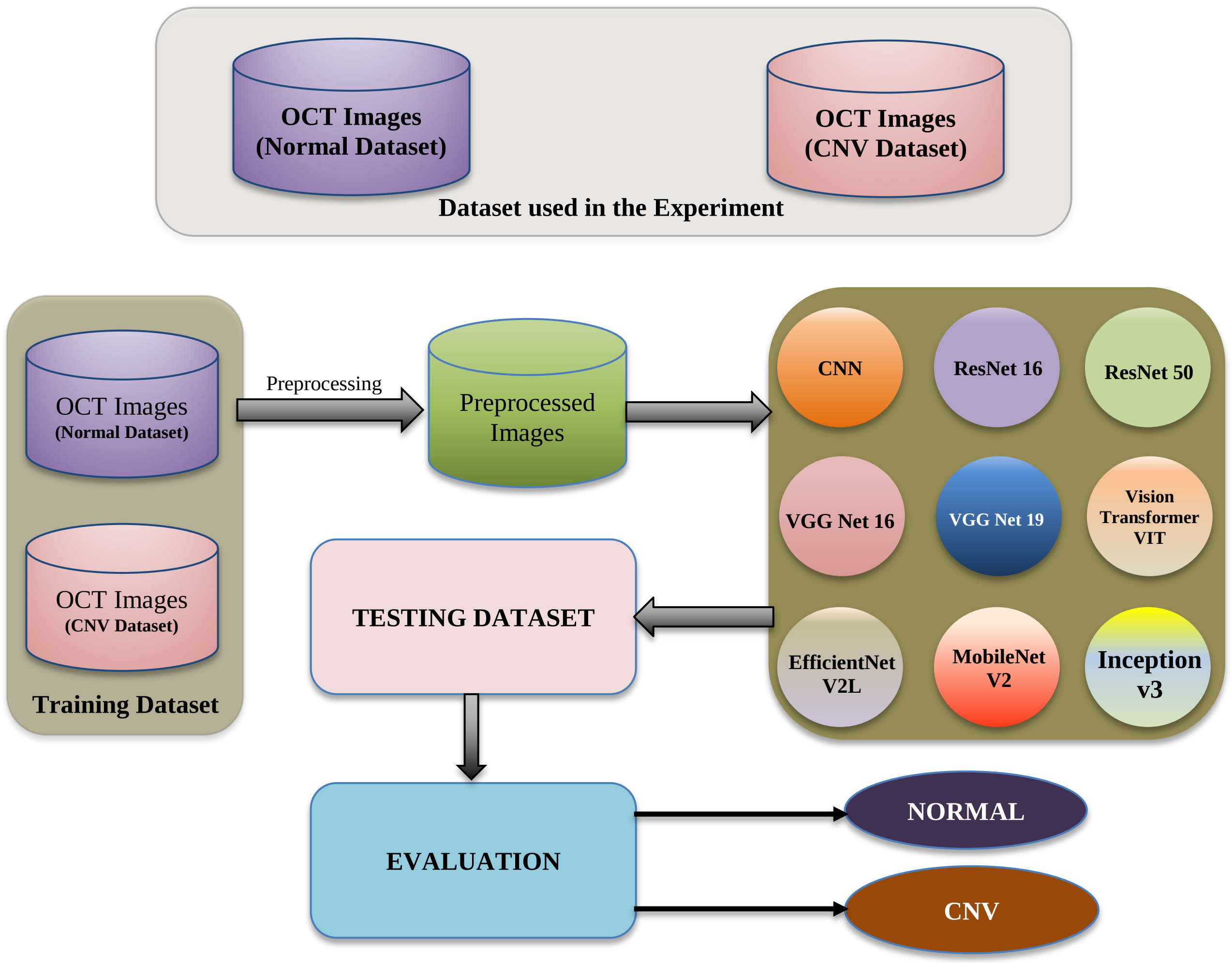


Fig. 6. Illustration of the proposed system. Displays all the different stages of the proposed system.

4.3.1 *Basic Convolution Neural Network (CNN)*

A deep learning architecture specifically designed for the purpose is the Convolutional Neural Network (CNN)[24] used to identify Choroidal Neovascularization (CNV) in optical coherence tomography (OCT) pictures. As shown in Fig.7 with initial convolutional layers starting with basic characteristics like edges and textures and moving up to more abstract representations, these CNNs are skilled at automatically extracting hierarchical properties from OCT pictures. Max-pooling layers down sample the features while Rectified Linear Unit (ReLU) activation functions create non-linearity, resulting in further fully linked layers for classification. The output neuron in the final layer calculates the probability of CNV occurrence in the context of CNV detection using the learnt characteristics.

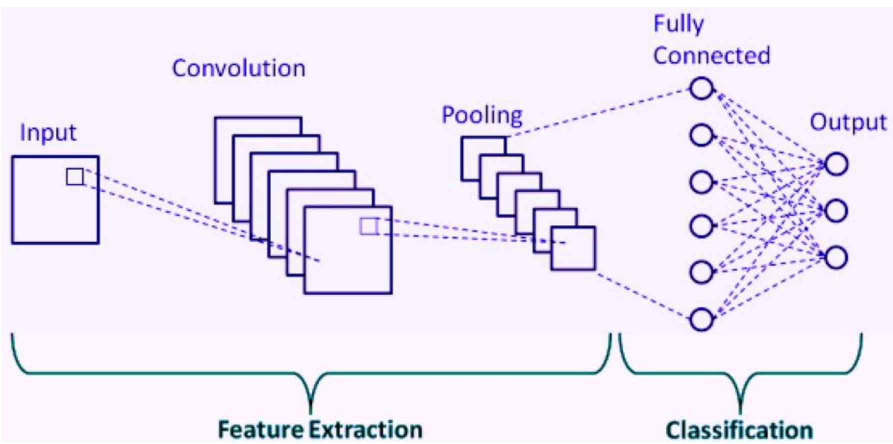


Fig. 7. Basic Convolution Neural Network (CNN) architecture. Illustrates all the working layers and their tasks in CNN.

4.3.2 ResNet

A deep convolutional neural network architecture called ResNet[25] is used to identify Choroidal Neovascularization (CNV) in optical coherence tomography (OCT) pictures. ResNet is renowned for its ability to handle challenging image recognition tasks. ResNet is distinct because it makes use of residual connections, which enable the net- work to learn and optimize the difference (or residual) between the input and output of each layer. This discovery mitigates the problem of vanishing gradients and facilitates the training of extremely deep networks. Complex and valuable properties may be extracted from OCT images for CNV detection thanks to ResNet's deep architecture. It is composed of many residual blocks with multiple convolutional layers, batch normalization, and ReLU activation algorithms in each block. The characteristics that are essential for CNV classification are gradually extracted and refined by these blocks. The network learns and generalizes these properties by training on annotated OCT datasets, finally obtaining a high degree of accuracy in differentiating between CNV and non-CNV OCT images. A potent method for automating the identification of this serious retinal disorder, the use of ResNet in CNV detection might improve patient outcomes by enabling early intervention and therapy.

A more complex version of the ResNet architecture called ResNet50 [25] shown in Fig. 8 has 50 weight layers. It is well renowned for its capacity to capture nuanced picture characteristics and patterns, which makes it particularly well suited for challenging computer vision tasks like object and image categorization. ResNet50's depth and feature-extraction skills may be used to successfully analyse OCT images for CNV detection.

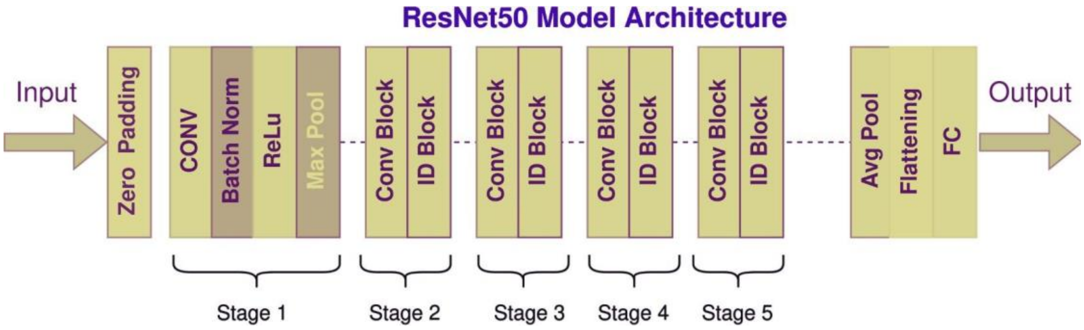


Fig. 8. Working architecture of ResNet50. Displays all the stages performed and the activities performed in those stages.

Similar to previous ResNet variations, ResNet18 [25] is a deep neural network design with residual connections. Convolutional layers, batch normalization layers, and fully linked layers are among the 18 weight layers. ResNet18 is intended to strike a compromise between computational complexity and model depth. When a reasonably deep model is sought for CNV detection, ResNet18 might be a useful option since it offers a reasonable mix between accuracy and computational efficiency.

4.3.3 VGGNet

The University of Oxford's Visual Geometry Group established the deep neural net- work VGGNet [26]. VGGNet is renowned for its consistent design and ease of use. Its fundamental building blocks are stacked convolutional layers with max-pooling layers and tiny 3x3 convolutional filters. VGG16 [26] and VGG19 [26], the two most well- known iterations of VGGNet, have different depths; VGG19 is deeper than VGG16. A summary of VGG16 and VGG19 is given below:

A variation of the VGGNet architecture known as VGG16 (Refer to Fig. 9) has 16 weight layers, including 13 convolutional layers and 3 fully linked layers. It utilizes max-pooling layers with a 2x2 window and a stride of 1 and 3x3 convolutional filters with a stride of There are over 138 million parameters in VGG16. VGG16 has demonstrated good performance in a number of computer vision applications, including picture classification and object identification, while being less deep than VGG19.

A more intricate variant of the VGGNet architecture is the VGG19 architecture (see Fig. 10), which consists of 19 weight layers, including 16 convolutional layers and 3 fully linked layers. Similar to VGG16, it makes use of 3x3 convolutional filters with a stride of one and max-pooling layers with a 2x2 window and a stride of two. There are almost 143 million more parameters in VGG19 than in VGG16. Because VGG19 has more layers, it can capture aspects of more complicated images, making it appropriate for situations where a more sophisticated architecture could be useful.

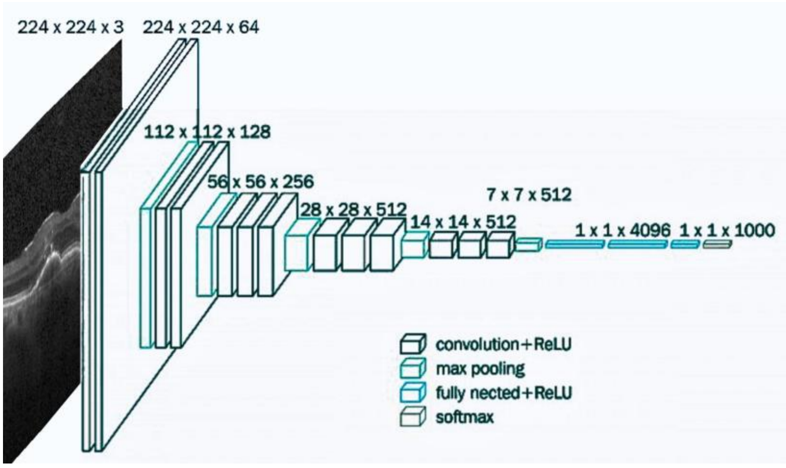


Fig. 9. Working architecture of VGG16. Displays all the layers present and their sizes as well as their tasks.

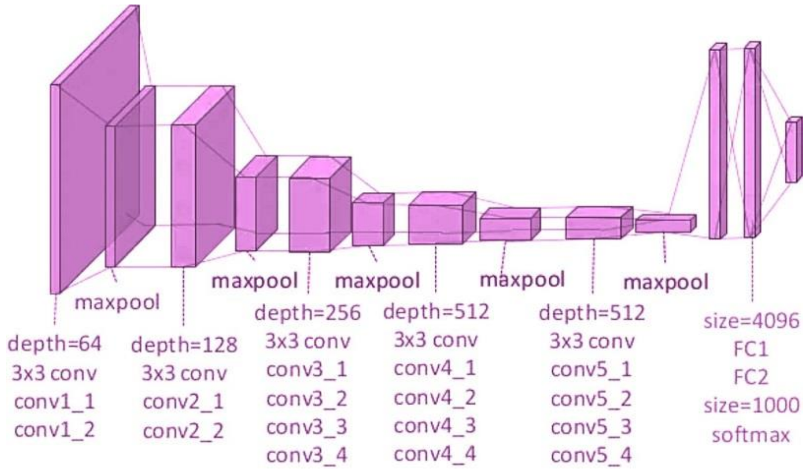


Fig. 10. Working architecture of VGG19. Displays all the layers present and their sizes as well as their tasks.

4.3.4 Vision Transformer

A Vision Transformer (ViT)[27] is a deep learning architecture that applies the fundamental ideas of the Transformer[28] model, which was created for natural language processing (NLP), to computer vision tasks. Due to the Transformer architecture's exceptional capacity to recognize long-range dependencies in text sequences, operations like language translation and text production benefit greatly from its superior performance. Similar concepts are used by vision transformers to interpret and comprehend vision data, such as photographs and movies. Fig. 11 shows the depiction of a Vision Transformer.

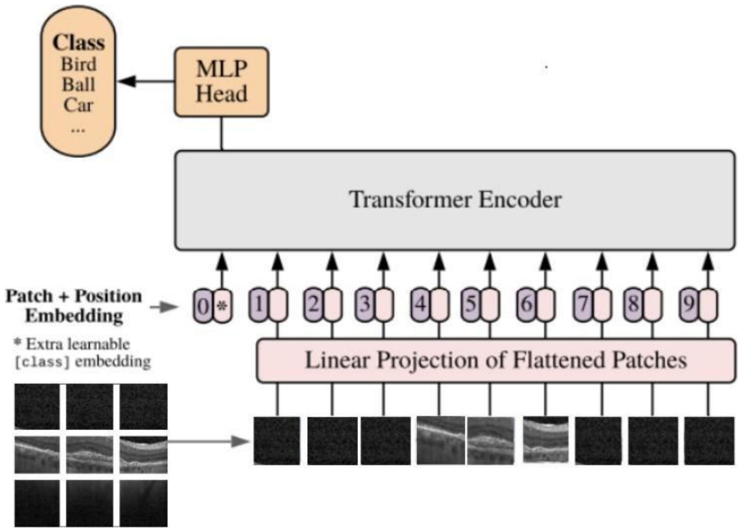


Fig. 11. Working architecture of Vision Transformer. Displays all the components of the model and their working.

In a number of computer vision tasks, Vision Transformers has produced cutting- edge results, challenging the dominance of conventional CNN-based methods. They have emerged as a promising option in deep learning for computer vision due to their capacity to manage a variety of spatial relationships and adapt to different vision tasks.

4.3.5 EfficientnetV2L

A very effective and potent CNN architecture for image classification applications is EfficientNetV2-L [29]. It is a member of the EfficientNet family, which is renowned for producing cutting-edge results with fewer parameters. The "L" variety denotes its deeper and bigger design, which provides more accuracy than its predecessors. EfficientNetV2-L uses cutting-edge architecture improvements and unique scaling strategies to enhance model performance while preserving computational effectiveness. By establishing a balance between model complexity and accuracy, it is perfect for a variety of computer vision applications, from object detection to picture analysis, making it an invaluable tool for deep learning practitioners. Fig. 12 depicts a basic understandable block diagram of EfficientNetV2L.

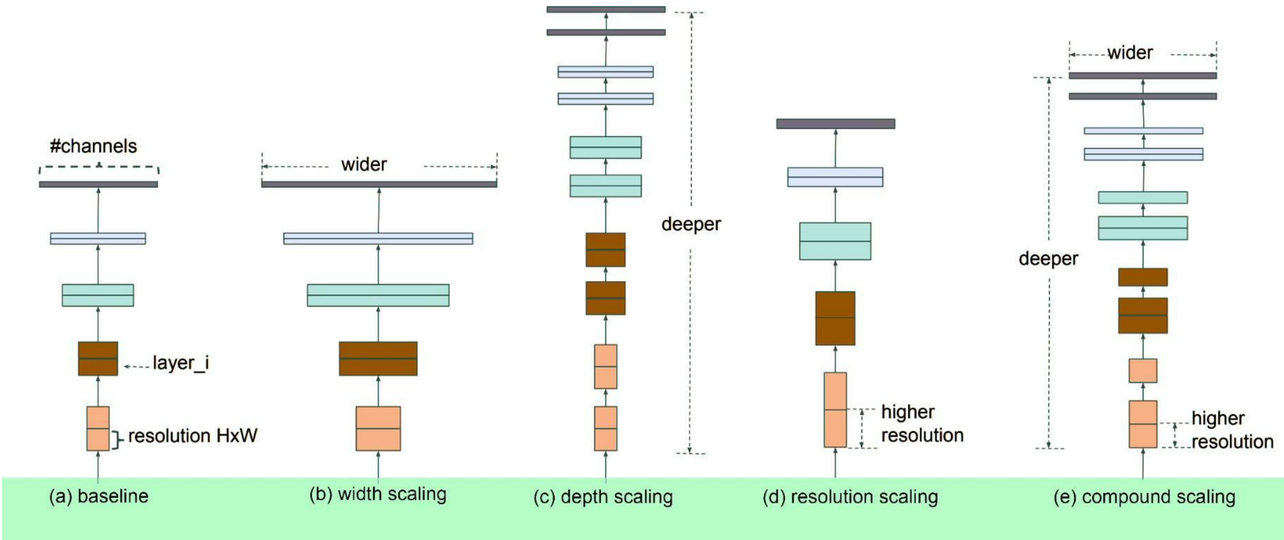


Fig. 12. Working architecture of EfficientnetV2L. Depicts all 5 scaling methods of EfficientnetV2L and the layers they consist.

4.3.6 MobileNetV2

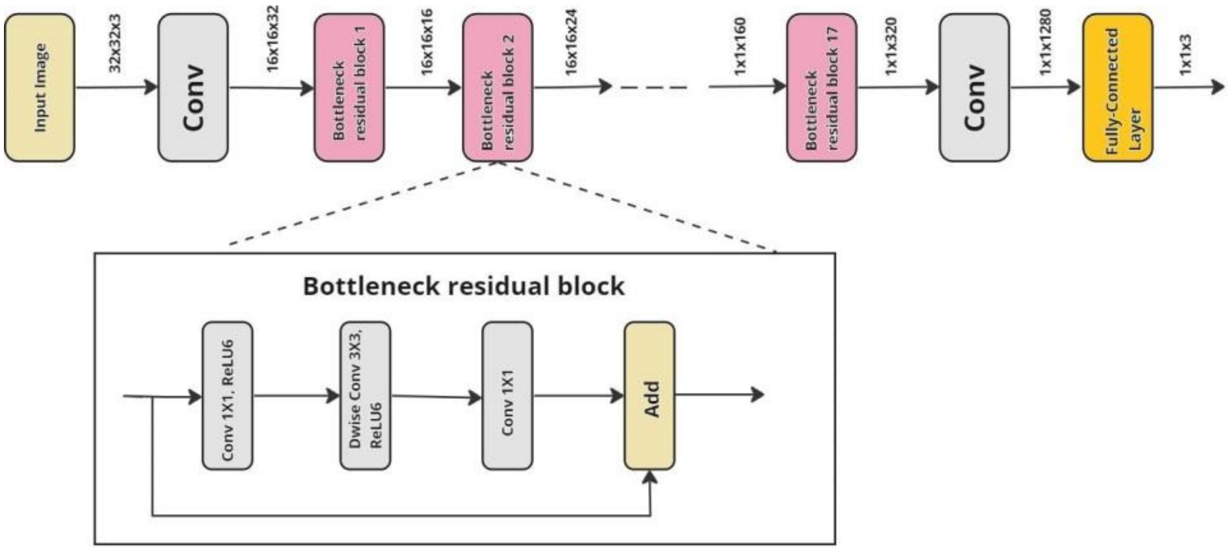


Fig. 13. Working architecture of MobileNetV2. Illustrates all the layers present and the tasks they perform.

To balance model accuracy and computational efficiency, the highly effective MobileNetV2 [30] convolutional neural network architecture was created for mobile and embedded devices. It adds significant advancements over MobileNetV1, which it replaces. Depth wise separable convolutions are used by MobileNetV2 to provide performance preservation while reducing the number of parameters and calculations. To improve the flow of gradients during training, the architecture features inverted residual blocks and makes use of linear bottlenecks and shortcut links. This allows MobileNetV2 to efficiently learn complicated features while minimizing model size and processing demands. Fig. 13 depicts a basic understandable block diagram of MobileNetV2.

4.3.7 InceptionV3

For picture classification and object recognition applications, the deep convolutional neural network (CNN) architecture InceptionV3 [31] was developed. It marks a huge improvement in computer vision and was created by Google in 2015. With the use of a special "inception module" found in InceptionV3, which employs numerous convolutional filters of varied sizes within the same layer, it is possible to collect features of diverse scales and complexity. InceptionV3 architecture outperforms ImageNet [32] in picture classification datasets, using factorized convolutions, batch normalization, and global average pooling to reduce computational complexity and accuracy.

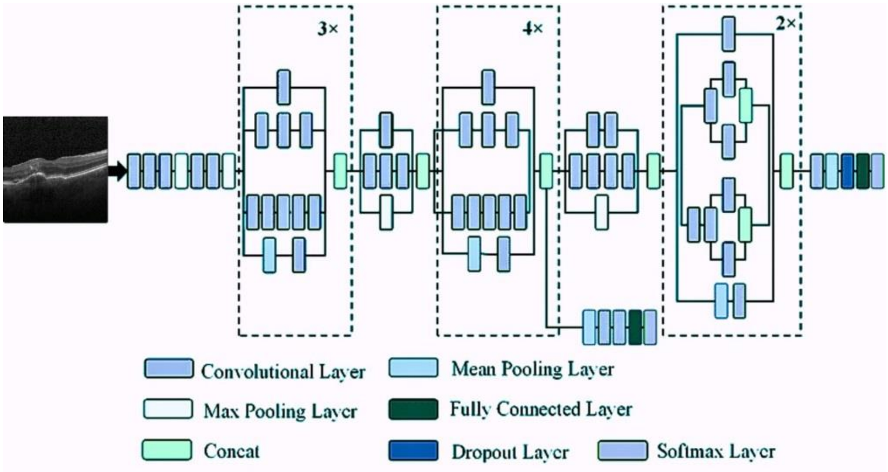


Fig. 14. Working architecture of InceptionV3. Illustrates all the layers present and the tasks they perform.

Beyond picture classification, InceptionV3's adaptability includes object recognition and other computer vision tasks. Fig. 14 depicts the architecture of InceptionV3

4.4 Hyperparameters

Hyperparameters are settings for a machine learning model that are defined in advance and are not learnt via training data. They control how the training algorithm behaves and have a big effect on how well the model works. Hyperparameters are determined by the machine learning engineer or researcher based on domain expertise, experience, or by carrying out experiments, in contrast to the model's weights and biases, which are learned during training. There are various parameters used in our deep learning models like Batch size, Epoch, Loss functions. In the case of this research paper, most of the models were given a batch size of 32, while it varies from other models. Epochs have been kept the same i.e. 20 epochs for all the models. The most used optimizer is Adam. The detailed information of the hyperparameters used by the models in this paper are presented in Table 3 and Table 4.

Table 3. Various parameters used during training

	Models				
Various Parameters	**Basic Convolutional Neural Network**	**EfficientnetV2L**	**MobilenetV2**	**InceptionV3**	**ResNet18**
Batch size	32	32	16	20	30
Epoch	20	20	20	20	20
Learning rate	0.001	0.001	0.0001	0.0001	3e-4
Dropout rate	0.1	0.2	0.2	0.2	0.5
Optimizers	Adam	Adam	RMSprop	RMSprop	Adam
Loss functions	Binary CrossEntropy	Binary CrossEntropy	Categorical CrossEntropy	Binary CrossEntropy	Binary CrossEntropy

Table 4. Various parameters used during training

height	**Models**			
Various Parameters	**ResNet50**	**VGG16**	**VGG19**	**Vision Transformer**
Batch size	32	64	30	32
Epoch	20	20	20	20
Learning rate	0.001	0.001	0.001	1e-3
Dropout rate	0.2	0.2	0.2	0.2
Optimizers	Adam	Adam	Adam	Adam
Loss functions	Binary CrossEntropy	Binary CrossEntropy	Sparse Categorical CrossEntropy	Binary CrossEntropy

4.5 Evaluation Metrics

Different evaluation metrics are used in AI and ML to evaluate the performance of models on distinct task types. The exact situation at hand and the characteristics of the data will determine the evaluation metrics to be used. The following list of evaluation metrics performed in this paper.

4.5.1 Confusion Matrix

A table that provides a thorough analysis of the model's predictions and the actual results for a classification task is called a confusion matrix. Although it can be expanded to multiclass issues, it is primarily utilized in binary classification jobs (two classes). The confusion matrix includes four crucial parts:

- True Positives (TP): The predicted value and the actual value are both positive.
- False Positives (FP): The prediction is positive but the actual result is negative.
- True Negatives (TN): The prediction and the actual value are both negative.
- False Negatives (FN): The actual result is positive but the prediction is negative.

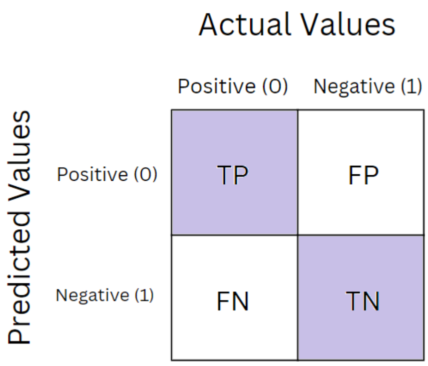


Fig. 15. Illustration of a Confusion Matrix. Shows the Positive and Negative class, and shows the appropriate location of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN)

Our paper illustrates the confusion matrices of models in Section 5.1. The confusion matrices are used to compare the results of the models achieved in classifying CNV affected OCT images and Normal OCT images

4.5.2 Accuracy

One of the most simple classification measures is accuracy. It measures the overall accuracy of a model's predictions. A higher accuracy denotes that more predictions came true on average. In this case, the models must classify the correct classes provided i.e. CNV and Normal. The higher the accuracy of classifying OCT images the robust the model is. The formula to calculate accuracy is given below,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

4.5.3 Precision

A model's precision is measured by how well it can produce reliable positive predictions. It provides the answer to the question: What proportion of the cases the model predicted as positive were actually positive? Precision is essential when false positives are expensive or unwanted. For instance, in medical diagnostics, false positive predictions can result in treatments or interventions that are not essential. In this case, the models must classify CNV affected OCT images in its respective class i.e. CNV, how- ever if it fails to do so it's a false positive. The formula to calculate precision is given below,

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

4.5.4 Recall

The capacity of a model to properly identify all positive cases is measured by recall, also known as sensitivity or the true positive rate. How many of the actual positive instances did the model anticipate would be positive? When missing positive cases can have serious repercussions, as when discovering flaws in manufacturing processes, recall is especially crucial. In this case, the models must properly identify all CNV affected images. The formula to calculate recall is given below,

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4.5.5 F-1 Score

A balanced metric, the F1-Score combines recall and precision into a single number. It provides a measure that takes into account both false positives and false negatives, offering a trade-off between the two metrics. When attempting to balance reducing false positives and false negatives, the F1-Score is useful. It is particularly helpful when there is an unequal distribution of classes. In this case, the models would com- bine Normal images classified as CNV affected and CNV affected images classified as Normal. The formula to calculate F1-score is given below,

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

4.5.6 ROC Curve

A graphical tool used in machine learning and statistics to evaluate a classification model's performance at different threshold settings is the Receiver Operating Characteristic (ROC) curve. It shows how different threshold values affect the trade-off between the true positive rate (sensitivity) and the false positive rate (specificity). An ROC curve's main components are:

- True Positives Rate (Sensitivity): The predicted value and the actual value are both positive.

$$True\ Positive\ Rate = \frac{TP}{TP + FN} \tag{5}$$

Where TP is True Positive and FN is False Negative

- False Positives Rate (1-Specificity): The prediction is positive but the actual result is negative.

$$False\ Positive\ Rate = \frac{FP}{FP + TN} \tag{6}$$

Where FP is False Positive and TN are true negative.

The ROC curve is created by plotting true positive rate and false positive rate at different thresholds. A ROC curve closer to the top-left corner indicates better classifier performance. AUC-ROC is another metric used to measure overall performance. In our case, the ROC curves of the models are presented in Section 5.1 in Fig. 25

The performance of models like **CNN (Convolutional Neural Networks)**, **ResNet (Residual Networks)**, and **Vision Transformers (ViTs)** can significantly impact real-world clinical settings for tasks like detecting Choroidal Neovascularization (CNV). Efficient, well-understood, good for small datasets as strength and the best use case as routine screening or basic clinical workflows. In ResNet the strength is Robust, handles complex patterns well and the best case is versatile use in varied clinical settings.

5. Results

In the Results section, we give a thorough analysis of different deep learning algorithms in the context of our research goals, including Convolutional Neural Networks (CNN), ResNet18, ResNet50, VGG16, VGG19, Vision Transform- ers,MobileNetV2,EfficientNetV2L and InceptionV3. To evaluate the effectiveness of the algorithms, we used a large Open access “Kermany dataset [23]” using accepted evaluation metrics. The performance comparison chart that is being provided shows that Vision Transformers has shown significant promise by outperforming well-known CNN designs. Notably, Vision Transformers demonstrated an ability to successfully capture global context and spatial linkages, highlighting their potential benefits in our particular research field.

Using three different datasets with various train-test splits, we thoroughly evaluated the effectiveness of each model for detecting choroidal neovascularization (CNV). 60% of the data were used for training and 40% for validation in the first dataset. The second dataset used to allocate 30% of the data for validation and 70% for training, while the third dataset used to allocate 80% of the data for training and 20% for validation. For the testing dataset, a total of 242 CNV oct images and 242 Normal oct images were used. All the models tested in this paper are individually tested and are not fine-tuned as we wanted to test the individual efficiency of the model, hence the results may be different than expected. The study was performed on Google Colab using T4 GPU, an online platform provided by Google.

Table 5. Results achieved after evaluation of first dataset (Kermany dataset [23])

Model	Accuracy	Precision	Recall	F1-score	Computation Time (mins)
CNN	0.52	0.52	0.53	0.53	83
ResNet18	1.00	1.00	1.00	1.00	80
ResNet50	0.50	0.50	1.00	0.67	113
VGG16	0.52	0.52	0.52	0.52	67
VGG19	1.00	1.00	1.00	1.00	110
Vision Transformer	0.99	0.99	0.98	0.99	105
MobileNetV2	1.00	1.00	1.00	1.00	53
EfficientNetV2L	0.97	0.95	0.98	0.97	62
InceptionV3	0.49	0.48	0.29	0.36	59

Table 6. Results achieved after evaluation of second dataset (Kermany dataset [23])

Model	Accuracy	Precision	Recall	F1-score	Computation Time (mins)
CNN	0.49	0.49	0.48	0.48	150
ResNet18	1.00	1.00	1.00	1.00	153
ResNet50	0.48	0.47	0.34	0.39	130
VGG16	0.51	0.51	0.51	0.51	140
VGG19	1.00	1.00	1.00	1.00	170
Vision Transformer	0.99	0.99	0.98	0.99	110
MobileNetV2	1.00	1.00	1.00	1.00	143
EfficientNetV2L	0.96	0.99	0.93	0.96	123
InceptionV3	0.49	0.44	0.05	0.09	121

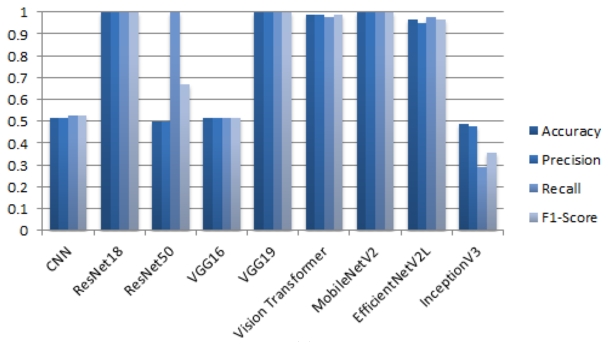
Table 7. Results achieved after evaluation of third dataset (Kermany dataset [23])

Model	Accuracy	Precision	Recall	F1-score	Computation Time (mins)
CNN	0.50	0.50	0.50	0.50	190
ResNet18	1.00	1.00	1.00	1.00	180
ResNet50	0.53	0.54	0.36	0.43	230
VGG16	0.50	0.50	0.50	0.50	178
VGG19	1.00	1.00	1.00	1.00	195
Vision Transformer	0.99	0.99	0.98	0.99	216
MobileNetV2	1.00	1.00	1.00	1.00	181
EfficientNetV2L	0.97	0.97	0.97	0.97	197
InceptionV3	0.45	0.46	0.50	0.48	130

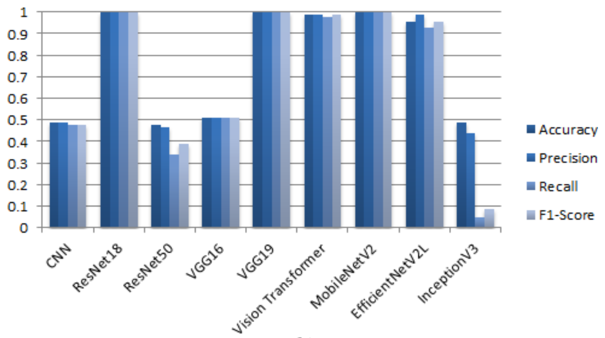
Table 5, Table 6 and Table 7 depicts the various evaluation metrics achieved by each model tested on the three datasets. Their accuracy, precision, recall, and F1-score help to tune the model further such that the detectional CNV can improvise. The computational time has to be reduced further as making clinical decisions, efficiency in high-volume clinics, early detection and prompt treatment, handling the large data volumes to be processed efficiently, and avoiding bottlenecks. Faster detection systems reduce the time a patient spends undergoing testing and waiting for results, improving their overall experience and adherence to follow-up appointments and it works in emergency situations. Through the help of these tables it is easy to verify which models outperforms other models.

5.1 Comparison of models tested

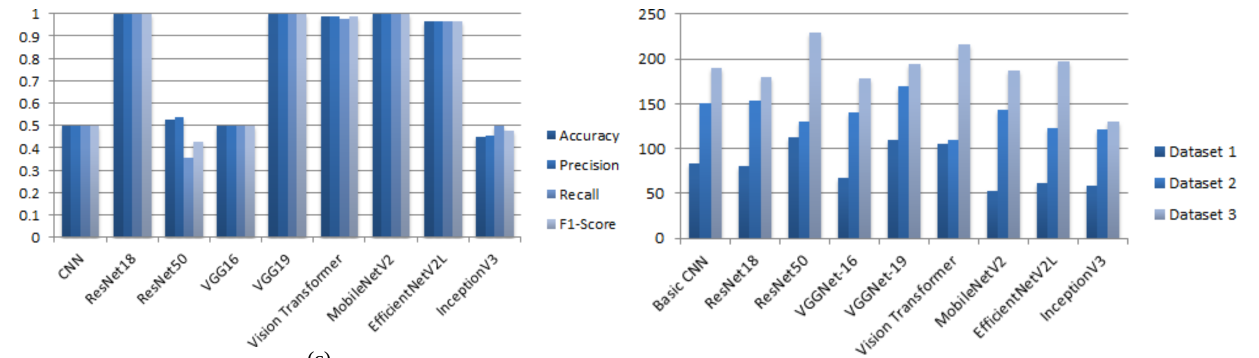
We give a thorough evaluation and comparison of the results attained using various models used in our research. Our main goal is to assess how well each model performs in relation to the unique difficulties presented by our research subject. The essential metrics and findings from each method, including accuracy, precision, recall, F1-score. Graphs are presented to compare the accuracy of all algorithms on all three datasets. From Fig.16(a),Fig.16(b) and Fig.16(c) it is clear that the best models individually are ResNet18, VGG19, Vision transformer, MobileNetV2 and EfficientNetV2L. These 4 models show high accuracy in terms of individual efficiency on all the three datasets created. Our experiment further found out that CNN, ResNet50, VGG16 and InceptionV3 did not achieve satisfactory results. Major reason for this to occur is may be because of no fine tuning of models or due to no data augmentation. Further researchers can combine all these models and find which works best for classification of OCT images.



(a)



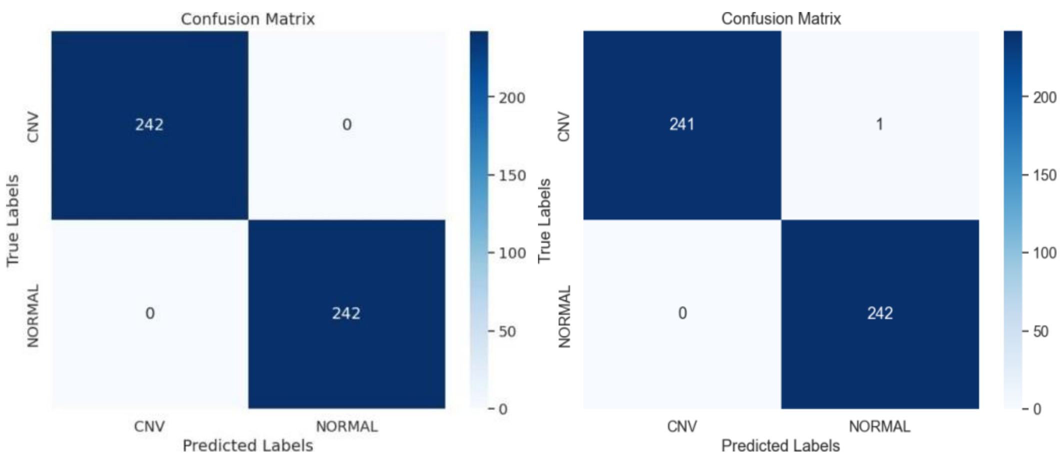
(b)



(c) (d)

Fig. 16. Fig. 16(a) compares the evaluation metrics achieved on the first dataset(60% & 40% split).Fig. 16(b) compares the evaluation metrics achieved on the second dataset(70% & 30% split).Fig. 16(c) compares the evaluation metrics achieved on the third dataset(80% & 20% split).Fig. 16(d) compares the computational time of Dataset 1(60% & 40% split),Dataset 2(70% & 30% split) and Dataset 3(80% & 20% split).Fig. 16(d) depicts the graphical representation of the computational time over the three datasets. It clearly shows that with increasing training dataset and decreasing testing dataset the computational time increases. The more samples and more complicated data used in the model training process are what cause the rise in processing time. More calculations and iterations are needed to optimize performance as the model gains knowledge from a bigger training dataset.

The models shortlisted are ResNet18, VGG19, Vision Transformer, MobileNetV2 and EfficentNetV2L as these show high performance.

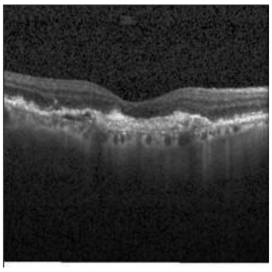


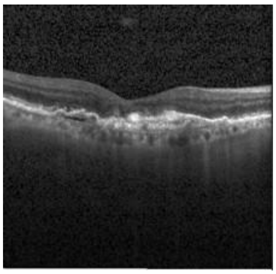
(a) (b)

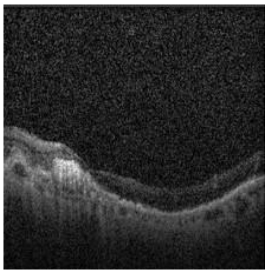
Fig. 17. Fig 17(a) displays the confusion matrix of ResNet18 on the testing dataset. Fig.17 (b) displays the confusion matrix of VGGNet-19 on the testing dataset.

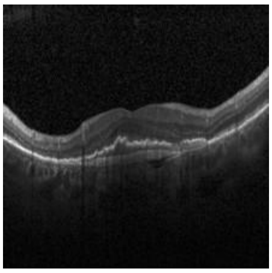
In Fig. 17, the confusion matrices of the top 2 best working models in classifying CNV affected OCT images are presented i.e., ResNet18 and VGGNet19. These confusion matrices are obtained from the testing dataset. The figure illustrates the capacity of both models to correctly predict the CNV-affected OCT image and the Normal OCT image. The capacity of ResNet18 and VGGNet19 to distinguish between OCT images with CNV and normal OCT images is seen in the figure. Through the visualization of true positives, true negatives, false positives, and false negatives, the confusion matrices provide important information on how well the models perform in various classes of data.

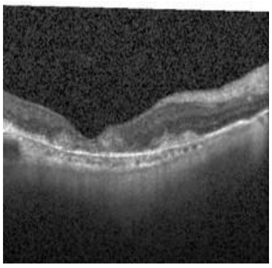
The True Positives (TP) are the cases where ResNet-18 accurately recognizes CNV- affected OCT images; Figs. 18(a), 18(b), and 18(c) show these events. The accuracy with which the model can identify positive instances may be understood from these figures. On the other hand, Figs. 18(d), 18(e), and 18(f) show the False Positives (FP), which are instances in which ResNet-18 misclassified normal OCT images as CNV-affected. To evaluate the specificity of the model and reduce misclassifications, it is crucial to comprehend the frequency and distribution of false positives. The True Negatives (TN) in Figures 18(g), 18(h), and 18(i) demonstrate examples of when ResNet-18 properly classifies normal OCT images as not being impacted by CNV. These numbers demonstrate how well the model can identify negative instances. Lastly, the False Negatives (FN) are shown in Figs. 18(j), 18(k), and 18(l), where ResNet- 18 is unable to accurately detect OCT images impacted by CNV. Understanding the sensitivity of the model and pinpointing opportunities for enhanced CNV identification need a thorough analysis of false negatives.

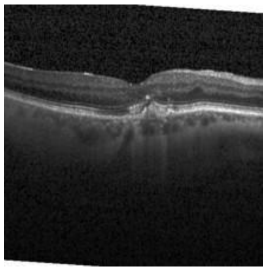

(a)

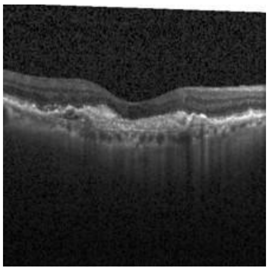

(b)

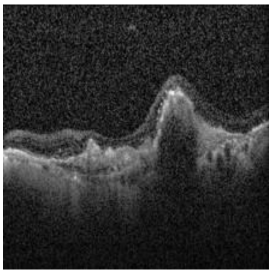

(c)

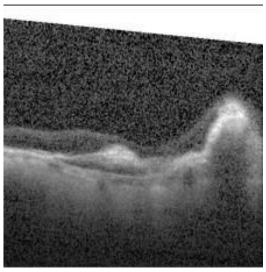

(d)

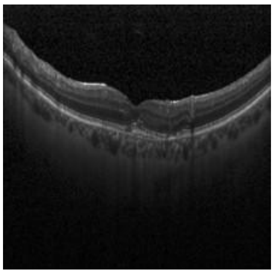

(e)

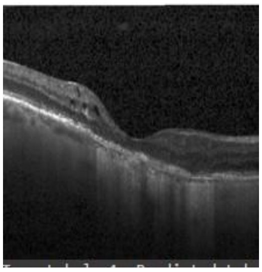

(f)


(g)


(h)


(i)


(j)


(k)

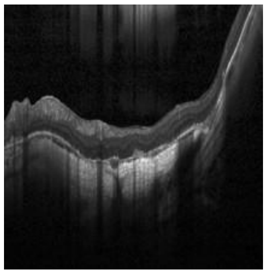
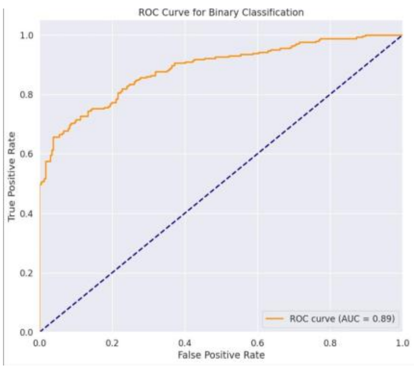
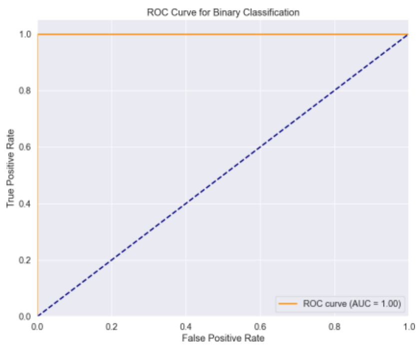

(l)

Fig. 18. Fig. 18(a)-Fig. 18(c) shows True Positives (TP), Fig. 18(d)-Fig. 18(f) shows False Positives (FP), Fig. 18(g)-Fig. 18(i) shows True Negatives (TN), Fig. 18(j)-Fig. 18(l) shows True Negatives (TP) of ResNet-18



(a)



(b)

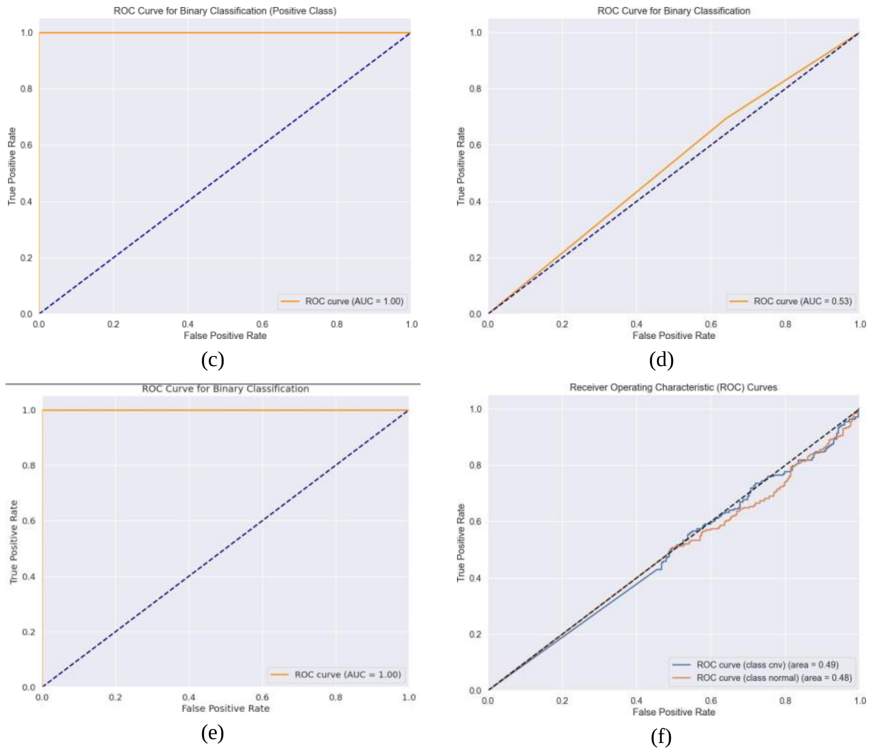


Fig. 19. Fig.19 (a) illustrates the ROC curve achieved by InceptionV3 (AUC=0.89). Fig.19 (b) illustrates the ROC curve achieved by CNN (AUC=1.00).Fig.19(c) illustrates the ROC curve achieved by MobileNetV2 (AUC=1.00). Fig.19 (d) illustrates the ROC curve achieved by ResNet50 (AUC=0.53). Fig.19 (e) illustrates the ROC curve achieved by ResNet18 (AUC=1.00).Fig.19 (f) illustrates the ROC curve achieved by VGGNet16 (Class CNV=0.49, Class Normal=0.48)

For further understanding of models tested we have presented the ROC curves of CNN, MobileNetV2, InceptionV3, InceptionV3, ResNet50, ResNet18, and VGG16 in Fig.19, respectively.

6. Discussion

This work classified CNV-affected OCT images and tested and evaluated nine deep learning models for CNV detection. The purpose of this study is to present an extensive study of these models' operations on three different versions of the same dataset. In this study, 13,412 retinal OCT pictures were utilized. The open-source Kermany [23] dataset is where all of these photos were obtained. Diabetic macular edema (DME), Drunsen, and normal retinal OCT pictures are among the various retinal OCT images included in the Kermany dataset, in addition to CNV images. During this process, optimizers like Adam were used. Adaptive Moment Estimation or Adam for short, is an algorithm best known for minimizing loss function during the training of neural networks through iterative optimization. Adam optimizer was paired with the cross entropy loss function to provide the best results. All of the processes of training, testing, validation, and evaluation were done on Google Colab with the GPU provided by Google Colab.

When we compared the findings of this study to those of other studies that also used the same dataset, we discovered that, on the Kermany dataset, Mohammad et al. IEEE (2021)[11] obtained a sensitivity of 0.98, whereas our evaluations on the same dataset revealed an average recall of 0.992 on the best working models. Using the same dataset, Parsa Riazi Esfahani et al. Cureus (2023) [21] obtained an accuracy of 0.98 for their model, whereas the top working models evaluated in this study had an average accuracy of 0.99 on the same dataset. To categories OCT-A pictures impacted by CNV, Choudhary A et al. Healthcare (Basel) (2023) [31] integrated the VGG-19 model. They made use of the same dataset that was utilized in this study. While our test of the VGG-19 model on the three iterations of the same dataset yielded an accuracy of 100%, they were able to reach an accuracy of 99.17% in classification. With their "ConvViT" model, Dutta P et al. J Imaging (2023) [33] achieved 94% accuracy. Vision transformers were one of the models in their hybrid model that was utilized to diagnose CNV. Based on our testing of the Vision Transformer Model, we have achieved an average accuracy across the three dataset versions of 99%.

7. Conclusion and Future Scope

In conclusion, our study has significantly advanced the field of medical image analysis, especially in the recognition of the serious ocular condition Choroidal Neovascularization (CNV). For effective treatment, this illness must be accurately and quickly diagnosed. It poses a serious threat to vision. We conducted a thorough analysis to determine the most effective model for the task using cutting-edge deep learning models, finally revealing interesting insights into their individual performances.

ResNet18, Vision Transformer, EfficientNetV2L, MobileNetV2, and VGG19 stood out among the group of models studied, regularly surpassing their rivals. Their exceptional memory rates, accuracy, and precision demonstrate their ability to help medical practitioners quickly identify CNV. These models demonstrated their aptitude for identifying detailed patterns in medical images as well as their adaptability in dealing with fluctuations in image data, highlighting their relevance for practical clinical applications.

Furthermore, future work may focus on fine-tuning models and exploring hybrid approaches that combine their strengths with traditional architectures, ultimately leading to more effective solutions in the research domain requirements. In order to enhance the accuracy and efficiency of CNV detection, future work may focus on designing and training custom deep learning models tailored specifically for this purpose. Developing novel architectures that leverage domain-specific knowledge can potentially lead to improved diagnostic capabilities.

Expanding the scope of our project to include a broader spectrum of ocular dis- eases, such as diabetic retinopathy or glaucoma, can further demonstrate the versatility of deep learning models in ophthalmology. Investigating the models' ability to detect multiple conditions concurrently can have significant clinical implications. Integrating advanced segmentation techniques into the CNV detection process can refine the model's precision and localization capabilities. Investigating methods like instance segmentation or graph-based segmentation can offer more detailed insights into the CNV lesions.

Acknowledgement

We would like to sincerely thank each and every contributor for their insightful contributions to this study report. Their combined knowledge and efforts greatly improved the caliber and substance of this work.

References

[1] Ghanchi, Faruque D, et al. "Optical Coherence Tomography Angiography for Identifying Choroidal Neovascular Membranes: A Masked Study in Clinical Prac- tice." Eye (London, England), U.S. National Library of Medicine, Jan. 2021, www.ncbi.nlm.nih.gov/.

[2] "Wet AMD - Choroidal Neovascularization. "StackPath, Natural Eye Care, www.naturaleyecare.com/eye-conditions/choroidal-neovascularization/. Accessed 28 Sept. 2023.

[3] L. Milanovic, S. Milenkovic, N. Petrovic, N. Grujovic, V. Slavkovic and F. Zivic, "Optical Coherence Tomography (OCT) Imaging Technology," 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE), Kragujevac, Serbia, 2021, pp. 1-5, doi: 10.1109/BIBE52308.2021.9635099.

[4] T. S. Mathai, K. L. Lathrop and J. Galeotti, "Learning To Segment Corneal Tissue Interfaces In Oct Images," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 2019, pp. 1432-1436, doi: 10.1109/ISBI.2019.8759252.

[5] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener and C. Roux, "Optimal Wavelet Transform for the Detection of Microaneurysms in Retina Photographs," in IEEE Transactions on Medical Imaging, vol. 27, no. 9, pp. 1230-1241, Sept. 2008, doi: 10.1109/TMI.2008.920619.

[6] P. Tanachotnarangkun et al., "A Framework for Generating an ICGA from a Fundus Image using GAN," 2022 19th International Conference on Electri- cal Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Prachuap Khiri Khan, Thailand, 2022, pp. 1-4, doi: 10.1109/ECTI-CON54298.2022.9795543.

[7] Z. Shen, H. Fu, J. Shen and L. Shao, "Modeling and Enhancing Low-Quality Retinal Fundus Images," in IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 996-1006, March 2021, doi: 10.1109/TMI.2020.3043495.

[8] "Vision Impairment and Blindness." World Health Organization, World Health Organization, www.who.int/news-room/fact-sheets/detail/blindness-and-visual- impairment. Accessed 28 Sept. 2023.

[9] V. Rajinikanth, S. Kadry, R. Dama˘sevi˘cius, D. Taniar and H. T. Rauf, "Machine-Learning-Scheme to Detect Choroidal-Neovascularization in Retinal OCT Image," 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 2021, pp. 1-5, doi: 10.1109/ICB- SII51839.2021.9445134.

[10] Han, J.; Choi, S.; Park, J.I.; Hwang, J.S.; Han, J.M.; Ko, J.; Yoon, J.; Hwang, D.D.-J. Detecting Macular Disease Based on Optical Coherence Tomog- raphy Using a Deep Convolutional Network. J. Clin. Med. 2023, 12, 1005. https://doi.org/10.3390/jcm12031005.

[11] M. Rahimzadeh and M. R. Mohammadi, "ROCT-Net: A new ensemble deep convolutional model with improved spatial resolution learning for detecting common diseases from retinal OCT images," 2021 11th International Conference on Com- puter Engineering and Knowledge (ICCKE), Mashhad, Iran, Islamic Republic of, 2021, pp. 85-91, doi: 10.1109/ICCKE54056.2021.9721471.

[12] Z. Baharlouei, H. Rabbani and G. Plonka, "Detection of Retinal Abnormal- ities in OCT Images Using Wavelet Scattering Network," 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Soci- ety (EMBC), Glasgow, Scotland, United Kingdom, 2022, pp. 3862-3865, doi: 10.1109/EMBC48229.2022.9871989

[13] M. Anna Latha, N. Christy Evangeline and S. SankaraNarayanan, "Colour Image Segmentation of Fundus Blood Vessels for the Detection of Hypertensive Retinopathy," 2018 Fourth International Conference on Biosignals, Images and Instrumentation (ICBSII), Chennai, India, 2018, pp. 206-212, doi: 10.1109/ICB- SII.2018.8524781.

[14] B. N, H. P. A and S. O, “Feature Extraction and Quantification in Retinal Fundus Autofluorescence Images of CSR Patients,” 2022 International Conference on Innovations in Science and Technology for Sustainable Development (ICISTSD), Kollam, India, 2022, pp. 74-79, doi: 10.1109/ICISTSD55159.2022.10010533.
[15] Pawan, S.J., Sankar, R., Jain, A. et al. Capsule Network–based architectures for the segmentation of sub-retinal serous fluid in optical coherence tomography images of central serous chorioretinopathy. Med Biol Eng Comput 59, 1245–1259 (2021). https://doi.org/10.1007/s11517-021-02364-4
[16] Abirami, M.S., Vennila, B., Suganthi, K. et al. Detection of Choroidal Neovas- cularization (CNV) in Retina OCT Images Using VGG16 and DenseNet CNN. Wireless Pers Commun 127, 2569–2583 (2022). https://doi.org/10.1007/s11277- 021-09086-8.
[17] Jie Wang, Tristan T. Hormel, Kotaro Tsuboi, Xiaogang Wang, Xiaoyan Ding, Xiaoyan Peng, David Huang, Steven T. Bailey, Yali Jia; Deep Learning for Diagnosing and Segmenting Choroidal Neovascularization in OCT Angiogra- phy in a Large Real-World Data Set. Trans. Vis. Sci. Tech. 2023; 12(4):15. https://doi.org/10.1167/tvst.12.4.15.
[18] Feng, W., Duan, M., Wang, B. et al. Automated segmentation of choroidal neovascularization on optical coherence tomography angiography images of neovascular age-related macular degeneration patients based on deep learning. J Big Data 10, 111 (2023). https://doi.org/10.1186/s40537-023-00757-w
[19] Sogawa T, Tabuchi H, Nagasato D, Masumoto H, Ikuno Y, Ohsugi H, Ishitobi N, Mitamura Y. Accuracy of a deep convolutional neural network in the detection of myopic macular diseases using swept-source optical coherence tomography. PLoS One. 2020 Apr 16;15(4):e0227240. doi: 10.1371/journal.pone.0227240. PMID: 32298265; PMCID: PMC7161961.
[20] Vali, M.; Nazari, B.; Sadri, S.; Pour, E.K.; Riazi-Esfahani, H.; Faghihi, H.; Ebrahimiadib, N.; Azizkhani, M.; Innes, W.; Steel, D.H.; et al. CNV-Net: Segmentation, Classification and Activity Score Measurement of Choroidal Neovascularization (CNV) Using Optical Coherence Tomography Angiography (OCTA). Diagnostics 2023, 13, 1309. https://doi.org/10.3390/diagnostics13071309.
[21] Riazi Esfahani P, Reddy A J, Nawathey N, et al. (July 09, 2023) Deep Learning Classification of Drusen, Choroidal Neovascularization, and Diabetic Macular Edema in Optical Coherence Tomography (OCT) Images. Cureus 15(7): e41615. doi:10.7759/cureus.41615.
[22] “DEEP LEARNING MODEL FOR OPTICAL DIAGNOSIS OF AN RETINA”, International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.10, Issue 2, page no.c521-c527, February- 2023, Available :http://www.jetir.org/papers/JETIR2302271.pdf.
[23] 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
[24] O’Shea, Keiron, and Ryan Nash. “An introduction to convolutional neural networks.” arXiv preprint arXiv:1511.08458 (2015).
[25] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recog- nition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
[26] Simonyan, Karen, and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition.” arXiv preprint arXiv:1409.1556 (2014).
[27] Dosovitskiy, Alexey, et al. “An image is worth 16x16 words: Transformers for image recognition at scale.” arXiv preprint arXiv:2010.11929 (2020).
[28] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (pp. 5998–6008).
[29] Tan, Mingxing, and Quoc Le. “Efficientnetv2: Smaller models and faster train- ing.” International conference on machine learning. PMLR, 2021.
[30] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).
[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” 2016 IEEE Conference on Com- puter Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.
[32] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” 2009 IEEE Conference on Com- puter Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.
[33] Choudhary A, Ahlawat S, Urooj S, Pathak N, Lay-Ekuakille A, Sharma N. A Deep Learning-Based Framework for Retinal Disease Classification. Health- care (Basel). 2023 Jan 10;11(2):212. doi: 10.3390/healthcare11020212. PMID: 36673578; PMCID: PMC9859538.

Authors’ Profiles

Sahil Chukka has recently graduated with a degree in Bachelors of Technology(B.Tech) in Information Technology, specializing in Artificial Intelligence and Machine Learning from Pillai College of Engineering, situated in New Panvel, 410206. His academic journey represents a deep passion and enthusiasm for new age technology with particular interests in the field of Artificial Intelligence(AI), Machine Learning(ML) and Deep Learning(DL).

Vardhanika Jagtap has recently graduated with a degree in Bachelors of Technology(B.Tech) in Information Technology, specializing in Artificial Intelligence and Machine Learning from Pillai College of Engineering, situated in New Panvel, 410206. Her academic and professional journey showcases her experience in Artificial Intelligence(AI), Machine Learning(ML), Frontend Development and UI/UX development.

Naveen Patel has recently graduated with a degree in Bachelors of Technology(B.Tech) in Information Technology, specializing in Artificial Intelligence and Machine Learning from Pillai College of Engineering, situated in New Panvel, 410206. His work centers around Artificial Intelligence (AI), Machine Learning (ML), as well as data innovation and development.

Sudiksha Jadhav has recently graduated with a degree in Bachelors of Technology(B.Tech) in Information Technology, specializing in Cloud Computing and Cybersecurity from Pillai College of Engineering, situated in New Panvel, 410206. She has actively participated in the field of Cloud Computing and has hands -on experience in Full Stack Development.

Mimi Cherian received the degree of Bachelor in Engineering and Master's in Engineering from Mumbai University. Currently pursuing PhD in Computer Engineering from Mumbai University. Have published multiple research papers, copyrights, and patents related to the field of the Internet of Things. Areas of interest are the Internet of Things, Network Security, and Artificial Intelligence. Has been awarded for Mentoring students for NPTEL certifications and Avishkar University competitions.

Jinesh Melvin Y. I. has been awarded a Ph.D degree in Computer Engineering from Pacific Academic of Higher Education and Research University. I received a BE and ME degree in Computer Engineering from Anna University Chennai, Tamil Nadu, India in June 2012 and May 2014 respectively. I have published various book chapters, Scopus, SCI, WOS, UGC care list Journalpapers and present the papers in referred national as well as international conferences including IEEE, ACM, Elsevier, Springer, CRC Press-Taylor & Francis, Atlantis press,with copyrights. I have been awarded 2nd Place in National level XZ1BIT2.0 digital national level poster making competition, also recognize various awards, appreciation and excellence in different events conducted by various organizations. My research activities involve Machine Learning, Data Mining, Advanced DBMS and Internet of Things.

How to cite this paper: Sahil Chukka, Vardhanika Jagtap, Naveen Patel, Sudiksha Jadhav, Mimi Cherian, Jinesh Melvin Y. I., "Evaluation of Deep Learning Approaches to Detect Choroidal Neovascularization", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.17, No.4, pp. 105-127, 2025. DOI:10.5815/ijigsp.2025.04.07