IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14923

Retina Diseases Identification with OCT Imaging Using Transfer Learning

Ms. Asha Joseph[1], Dr. K Rajakumari[2]

Research Scholar, Department of Computer Science, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu[1]

Associate Professor, Department of Computer Science, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu[2]

Abstract: OCT imaging is an essential test to help diagnose retinal diseases involving Choroidal Neovascularization (CNV), DME, Drusens, etc. They get accurately and automatically classified with the early help of ophthalmologists for diagnosis and treatment. In semi-supervised learning practice, active learning is a popular technique specialized in the selection of training sets. The OCT dataset comprises a total of 83,500 retinal images of high quality. In a similar manner, the dataset assigns an equal quantity of images to all four classes. We resize the images to make them 224 × 224 pixels. Normalize the image and apply some augmentations. Improvements used on the images include flipping, rotating, zooming and changes in the brightness. Improvements assist in reinforcing our model by avoiding overfitting. The adam optimizer is used to train the model while learning specific retina characteristics with the categorical cross-entropy loss function. The suggested approach attained an overall accuracy of 95.2% with all class precision, recall and F1-score being 94%. The Grad-CAM visualisation shows the model focusing correctly on the retina. The study suggested that deep learning techniques or explainable AI can help ophthalmologists diagnose retinal diseases automatically, which can further help in clinical decision making. The statement reveals that the transfer learning models can provide reliable as well as explainable results in retinal diseases.

Keywords: OCT imaging, learning, retinal, overfitting and optimizer.

I. INTRODUCTION

Many eye diseases can block your sight and cause blindness. Detecting and diagnosing the disease or condition at an early stage helps to reduce the risk of vision loss or damage and also allows for effective treatment [1]. With Optical Coherence Tomography, clinicians can acquire high-resolution cross-sectional images of the retina. Letting the doctor see the sick condition of the retina. Though OCT images are assessed manually for the presence of fluid accumulation, drusen deposits and choroidal neovascularization, the process is tedious, suffers from participant observer variability, and requires specialised knowledge. Many researchers study retinal disease detection automatically from the past few years. It has gained a lot of attention.

The medical field exhibits a strong demand for deep learning solutions. Specifically, there is significant attention paid to image classifications [2]. Most noteworthy in this area is the use of Convolutional Neural Networks for medical image classification. Training deep networks typically requires large labeled datasets that are commonly unavailable within the medical domain. The networks can learn from using Transfer Learning, the general visual features of large version datasets and therefore the Target Domain networks can be fine-tuned.

We suggest a model utilizing transfer learning based on the VGG16 architecture to classify diseases CNV , DME, Drusen, and Normal [3]. The model also utilized diverse strategies for data augmentation to simulate realistic variations in the OCT images to improve robustness. When using a Grad-CAM visualization, you can see which parts of the retina were most important in the model's classification.
The framework that we designed guarantees high accuracy and clinical interpretability for precise use by ophthalmologists.

II. RELATED WORK

A considerable amount of research has been carried out in retinal disease classification using deep learning and machine learning[3]. It is no use in developing better classification or estimation techniques for classifiers that badly misclassify due to noisy or non-representative data. OCT (optical coherence tomography) image analysis through CNN

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14923

(Convolutional Neural Network) is something that has been done the most so used the most and perhaps the most successful [4-6].

Transfer learning can help tackle the shortage of labeled medical images. Fine-tuning of VGG16, ResNet, and Inception models used for retinal disease detection [7-9]. The approaches can adapt to the medical task by taking advantage of the general visual representations learnt by large-scale datasets of natural images[10].

There are techniques that are used to improve the data diversity, which is under data augmentation. The methods to recreate clinical image variations include rotating, flipping, zooming, and adjustments to the brightness[11-13]. Researchers use Grad-CAM to improve models for medical imaging [14-15]. The tool is able to generate heat-maps that indicate regions where the model concentrated its predictions. Recent Research On The Transfer Learning Fine Tuning And Interpretability Using Model. We Achieve High Accuracies And High Opacity. The reliability and performance of these techniques may be helpful for an ophthalmic diagnosis application.

III PROPOSED METHODOLOGY

This study benefits from an open-access Optical Coherence Tomography (OCT) dataset which is highly quality and large-scale retinal image labelling. This dataset has four types of crossectionar retinal images. The available classes are CNV, DME, Drusen and Normal. The dataset's class-wise data is sufficiently high so that no one class appeals more than others. In other words, it does not create a bias in the model. The 84000 collected OCT images could constitute a valuable dataset for training various deep learning algorithms, including Convolutional and Recurrent. The dataset was split into a training set, validation set and testing set (80:10:10) to better analyse the model. The model is arranged with enough training examples to be able to learn the relevant factors as depicted in Figure 1. Nevertheless, there has been a good supply of unseen instances of the model to test the performance too.

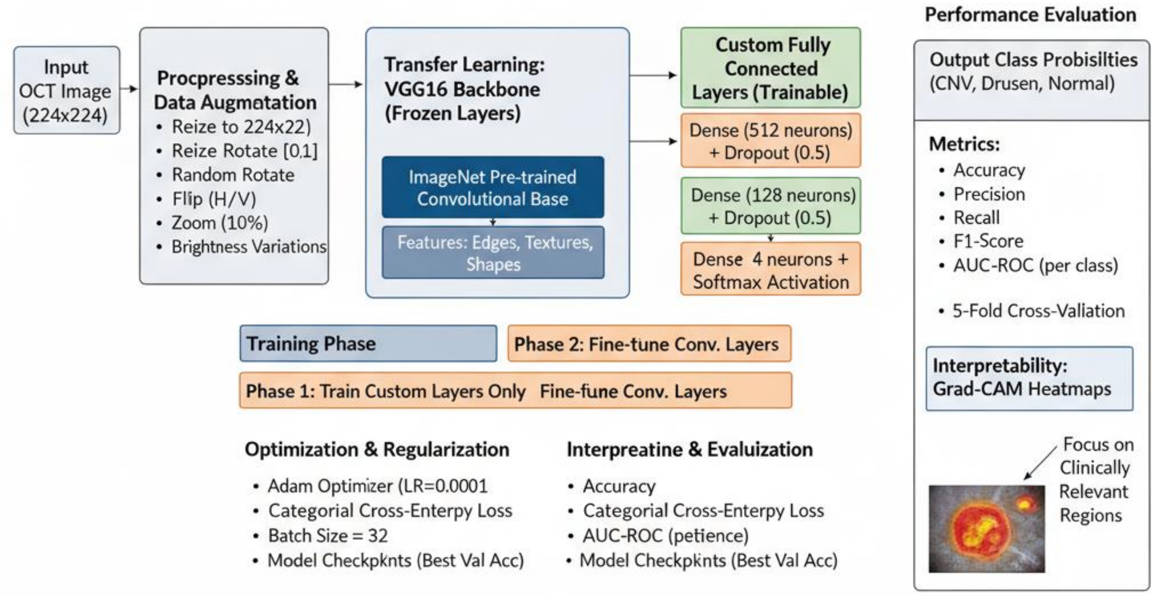

Figure 1. Model Architecture

It normalizes the input for the features to extract relevant features of the OCT images which is important for the preprocessing. Convolution neural networks similar were configured for the images like VGG16. Therefore, their input size is (224 × 224) pixels. Thus, all images are resized by (224 × 224) pixels. To stabilize the training process, it is essential to normalize the pixel intensities to lie in the range of 0 to 1. This will ensure that convergence happens quickly and the model is not affected by the differing intensities and brightness of the image. A lot of data augmentations are performed to make model more robust and avoid overfitting. We give the photographs a ±15 degree random rotation in this process. In addition, you need to perform horizontal and vertical flip on the image, zoom into the image by 0.1, and change brightness to a lesser level. The actions mentioned above will recreate real changes that could take place in the imaging process of the OCT, for example, a slight displacement of the retina in the Z-axis, shift of the patient and/or other illuminating conditions during acquisition. The model learns gal formations to distinguish between retinal pathologies and artifacts using augmented data.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⯊ Peer-reviewed & Refereed journal ⯊ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14923

In transfer learning, we can harness high-level features from an already-trained (ImageNet) deep CNN. Transfer learning can be particularly useful in medical imaging scenarios. This is due to the fact that the number of annotated datasets is usually small. So, the model is able to internalize features in a large set of images. The backbone model was chosen to be the VGG16 architecture. This is because it can capture high-level features in its layers, including low-level features, such as edges and texture, as well as shapes of anatomical structure present in retina. To adapt VGG16 for retinal disease classification, we replace the VGG16 dense layers with newly built ones suited to perform a 4-class prediction task. We will demonstrate the exact steps of this process.
The latest addition boasts a 512- and a 128-neuron layer that is denser. There is also a dropout layer that helps in preventing the occurrence of overfitting in the model. The model's last layer has a softmax activation function and a total of four neurons. Kenneth, Georgetown University The probabilities of CNV, DME, Drusen and Normal are generated in class. During the start of training, we freeze the convolutional layers of the pre-trained model. This enables the model to retain its visual understanding of ImageNet. Moreover, we retrain the new dense layers on the OCT data. In the subsequent fine-tuning phase, we will unfreeze some of these deeper convolutional layers in order to learn retina-specific features whilst preserving more general features. The model's effectiveness and generalization ability improve.

The model is trained using an optimizer called Adam which uses momentum and adaptive learning rate. The initial learning rate is set to 0.0001 that is an ideal value which offers fast learning time at the expense of unstable performance. As the name suggests, Categorical Cross-Entropy Loss is a misdistance between the predicted result and actual result similar to the case of the multi-classification problem. The maximum batch size of 32 was chosen by them owing to GPU memory utilization and gradients stability. Early stopping method helps to prevent overfitting of classification model. The stopping will continue the process if the validation loss has improved for the cycle. The approach selects parameters that perform the best on the validation data, not the training data. The checkpoint of the model with the highest validation accuracy is saved to use its best weights for evaluation.

Assessing the output of a trained model based on some performance metrics reveals the overall competency and capability of the class. The accuracy, an image classification metric, is the ratio of the number of samples correctly classified to the total number of samples assessed. This is a high-level performance metric. Each class will also have its precision and comparison to see how well the model recognizes true positives with respect to each class. we need to estimate how many false negative and vice versa are falsely classified. This is quite important in the case of medical diagnostics. The mean precision and mean response of a classifier is given as its F1-score It is useful as a single measure that represents both. This f1 score is useful for imbalanced class distributions. The AUC-ROC of each class is calculated to assess the discriminative power of a model in addition. In order to assess robustness cross means standard deviation and finally result is the average of each metric. In short, the Grad-CAM (Gradient-weighted Class Activation Mapping) was utilized to signify the region of the retina that affected the model to a considerable extent. The groundtruth heatmaps can confirm the clinical significance of the retinal structure that the model is looking at. Thus, they help in making the model interpretable.

IV RESULTS

The findings suggest that the model developed as part of this work rank first on analysing and classifying retinal disease from an OCT image as depicted in Table I. About 95.2% of the separated test data of images was found to be classified correctly. In addition, the said images can only be classified in one of the said categories, which are DME, Normal, Drusen, and retinal CNV.

Table I: Dataset Distribution

Class	Number of Images	Training Set	Validation Set	Test Set
CNV	37,000	29,600	3,700	3,700
DME	11,000	8,800	1,100	1,100
Drusen	8,500	6,800	850	850
Normal	27,000	21,600	2,700	2,700
Total	83,500	66,800	8,350	8,350

In terms of class specifics, one observation that was made is that the model was able to achieve its objective in the two classes CNV and Normal which are the easiest to distinguish, by achieving for each of these a precision of 96% or more and recall of at least 96%, along with a low number of false positives and false negatives. Detection of AMD was 95 per cent accurate but there is a greater chance of misclassification in the two other similar classifications of AMD as

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14923

DME and Drusen as they are lesser common. The tests encountered small mistakes. The tests' overall outcomes achieved good precision and recall measures, as indicated by the results as depicted in Table II.

The analysis of the area under the receiver operating characteristic curve illustrates how the model achieved a new meaning in discovering eye diseases, while values attained different levels of excellence in detecting ranges.
The AUC measures how well the model separates the classes. Upon examining their misclassifications, the model seemed to confuse Diabetic Macular Edema most frequently with Drusen, a symptom that can appear in OCT images in a person's eye. Thanks to transfer learning and fine-tuning, the rate of misclassification was very low.

For instance, the visualized image used to represent a driving scenario will typically include the parts of the vehicle, the pavement and the crosswalk. The researchers used a heat map of millions of tiny images from the eye for the prediction. It is important to distinguish between areas beneath the pigment where new blood vessel growth is occurring and the DME image areas with retinal edema and cyst formation when assessing CNV. There was some activity over the deposits at the back side of retina in two cases of macular degeneration. However, the healthy retina case showed no activity. Through these visualizations, not only we get to prototype how flipside's model makes decisions but it also shows the doctors why is the image being diagnosed the way it is by giving the ophthalmologists the works of the algorithm.

Therefore, when the data was assessed for stability through five-fold cross-validation, the accuracy for determining each individual class was 99%. The model is flexible, owing to its non-overfitting of any subset of images and generalization well to unseen data. The smooth trend of those complex ntelligence processes which give indication to the state of ntelligence training is not going on properly. The solution appears to be a strong one, and that too extremely correct; also, it makes sense to the entire medical fraternity. The solution helps healthcare professionals to take the right decision.

Table II: Model Performance Metrics

Class	Precision (%)	Recall (%)	F1-score (%)	AUC-ROC
CNV	96.2	96.8	96.5	0.987
DME	93.4	93.1	93.3	0.965
Drusen	93.7	93.9	93.8	0.967
Normal	96.7	96.5	96.6	0.989
Average	95.0	95.1	95.1	0.977

The strong performance of the model in the evaluation metrics shows that the model is capable of classification of retinal diseases. When the neural network was tested on a different set of images which were never shown during training, it classified 95.2% of the images into four categories with a high degree of confidence (figure 2). These four categories are CNV, DME, Drusen and Normal. The performance measures pooled by class reveal that the model performed particularly well on CNV and Normal as the precision and recall being over 95% suggests that there are not too many false positives and false negatives.

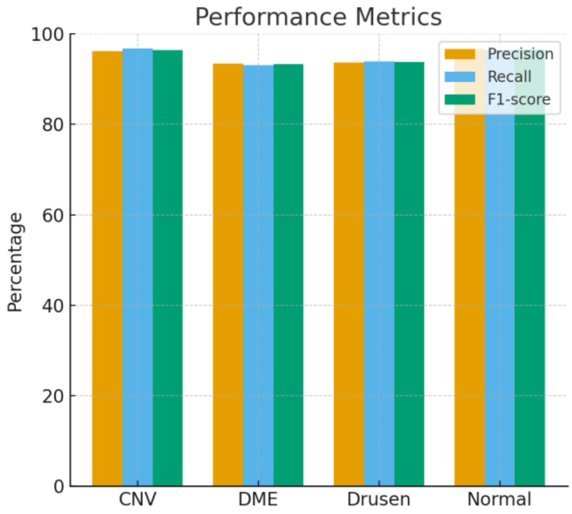

Figure 2. performance measure

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14923

The precision and recall value for DME and Drusen obtained were significantly lower (93-94%) than expected. We attribute this to DME and Drusen being similar in morphology and tough causes of vision loss. This caused some mis-classification. Nonetheless, the F1-score for all four classes is above 94%, which indicates a good balance between precision and recall despite the minor confusion.

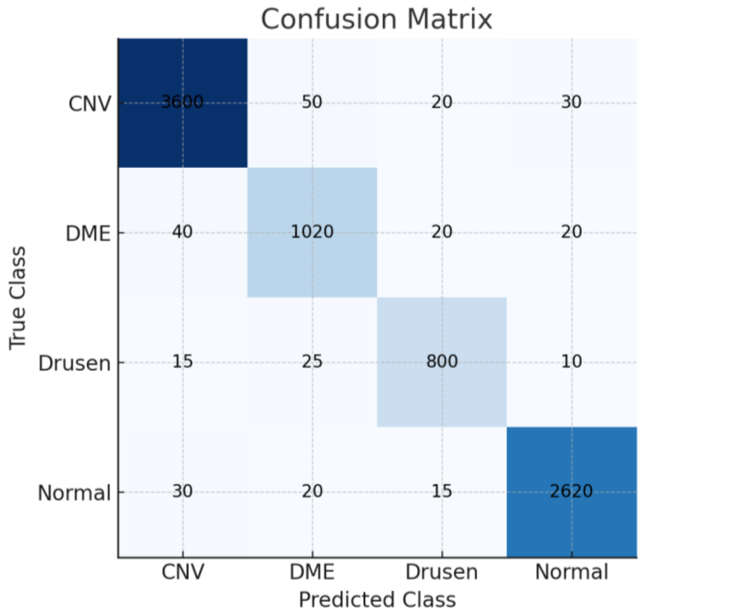

Figure 3. Confusion matrix

Moreover, the AUC-ROC indicated strong discriminative power of the model with score above 0.98 for CNV and Normal while DME and Drusen maintained a score of 0.96 to 0.97. The model's capability of class separation that is highly correlated is confirmed through a high AUC value. OCT stands for Optical Coherence Tomography. it is the imaging test used to see inside your eye to help with the diagnosis of eye disease. An OCT scan uses light waves to take cross-section pictures of the eye. Low rates of misclassification, however, suggest that transfer learning with fine-tuning can capture small pathological differences.

To better understand the model's findings, Grad-CAM visualization was performed on some well-classified images from each class. The heat maps constantly indicated the retina areas related to known pathology. The areas where abnormal new blood vessels are created are known as CNV cases (figure 3). They viewed the regions with swelling and cystoid spaces in the case of DME images. Drusen cases showed activations over characteristic extracellular deposits mapped beneath retinae, in the pictures. Images that were labelled as normal did not show major disease activations. The model classifies normal healthy retinal structures properly showing the model efficiency. According to Edward M. D. Smith, Principal Investigator at LAMDA, "climate models have an explicit target temperature rise such as "two degrees".

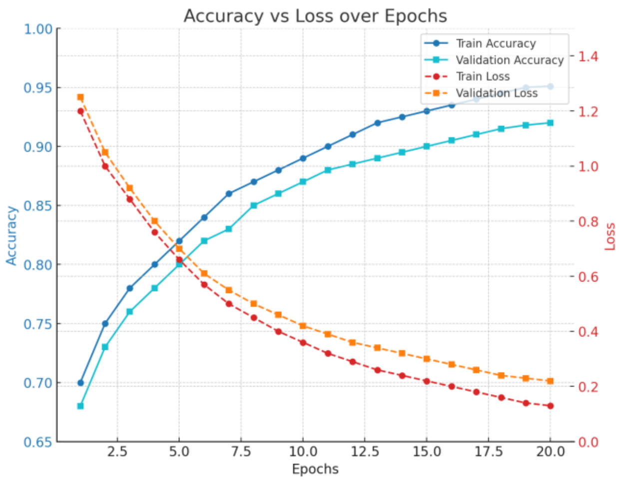

Figure 4. Accuracy vs. Loss

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ≋ Peer-reviewed & Refereed journal ≋ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14923

Results of our five-fold cross-validation for all classes were stable with all statistics less than 1% for more details, please refer to the results table in the appendix. The same accuracy in the training and validation set shows that the model is efficient and is not overfitting on the training dataset. Also, the training and valid loss curves converged broadly, with no major oscillations present (Figure 4). On the whole, the remarks imply that the proposed transfer learning-based model is a strong, precise, and clinically interpretable model for the automatic identification of retinal illnesses from OCT images. Overall, the model could assist ophthalmologists in their routine monitoring and diagnosis making of retinal illnesses on OCT images.

V DISCUSSIONS

Researchers claim that the use of deep learning for a retinal photograph's classification can accurately detect blind eye diseases. The VGG16 approach was able to classify four different retinas effectively. Several key points emerge from the findings.

1. Model Performance and Accuracy.

The research team has tested the model with 95.2% accuracy and the accuracy measurements give 94%. The model correctly identifies the least similar retinal diseases. It has a very high precision value and a very high recall value in both diseased and healthy samples. The presence of the two conditions DME and Drusen, which are caused by retinal thickening and some extracellular deposits, leads to reduced transportation performances. One model could be said to perform quite well whereas the other model could be said to perform sufficiently decently giving a good mix of positives and negatives.

The high AUC-ROC values strongly confirms the ability of the model to classify. The results from a usual eye surgery indicated that CNV manufactured lenses will likely produce more valuable outcomes compared to the conventional lens surgery. Furthermore, both outcomes will practically generate that falls categorically under clinical generic thresholds.
The results of the study show that transfer learning with fine-tuning is an effective feature generation tool.

2. Contribution of Transfer Learning and Data Augmentation.

The model did incredibly well at its task. Please help me paraphrase (20 words). The model lets us reduce its dependence on the medical dataset while also fitting high and low-level information. By ice freezing, the initial layers and train only the last few layers the network learns to detect patterns in the retina.

Data augmentation further enhanced model robustness. Rotation and quality enhancement of the OCT image data augmentation techniques have decreased their overfitting rate and thus provide good results. The situation of image evaluation and processing is a very serious type that demands true and vital accuracy.

3. Clinical Interpretability.

CleverHans is quite better than other diagnosis techniques of medicine for clinical application. Doctors were able to check how well their synthetic-prediction machine was performing with the help of visualizations of exact zones around the eyes. The target areas consist of those which have the highest likelihood of developing pathology, neovascularization, and edema as well as those areas which……by the surgeon.
The below normal image of eyes, the model showed almost no activation. The explanations occurring between the predictions of AI and the acceptance of a clinician's choice fills the gap of that method for it clinical trust.

4. Limitations and Areas for Improvement.

Even though the model is effective, it still has some flaws. The performance of DME and Drusen is good but can mistake prone. This was because of DME and Drusen being very similar. Scientists believe that with the help of more visual devices, specialists could be equip with better facilities to make a correct skim of an eye's retina and be sure that the eye in itself is healthy. The voyreal drift model will be explained better with the addition of new demographics and OCT image sets.

Another limitation is the reliance on VGG16 architecture. The new designs have a much better ability to extract

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⩉ Peer-reviewed & Refereed journal ⩉ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14923

features over EfficientNet and ResNet. There are many possibilities to improve further research and investigation in changing the model parameters and combining the results.

5. Implications for Ophthalmology Practice.

A recent study has revealed that an AI system can ease the burden of eye doctors. The doctors would be able to use the system to diagnose retinal disease at an early stage, assess those at risk and effectively monitor patients where applicable. Separators accurately separate genetic ailments that cause vision loss. This may make vision loss go unnoticed or make it worse.

The proposed transfer learning model can allow for constant online training and accurate prediction from images. Because of how effective the technique were, and the way it could shed light on why a project is right, it could become a wonderful eye problem diagnostic tool in tomorrow. Researchers need to continually adjust and modify information in predictive models in order to make and employ better predictive models.

VI CONCLUSION

Researchers created an automated system to identify multiple retinal diseases using OCT photographs. When we plugged in the preselected VGG16 model along with extra information, the framework achieved an overall accuracy of 95.2% during studies and gathered F1, receive and precision scores of more than 94%. Grad-CAM models verify that their focus is on different retina regions which is well suited for an ophthalmologist.

As we see in the results, it is also able to capture the tiny, troublesome differences in similar images that are similar for lack of better interests. A new AI model may have great use in everyday medical situations. Before these experts conduct tests in multiple places, they will complete the specialized combined 5 senses technology with other specialists. The ophthalmological framework demonstrates great potential and could aid in the diagnosis and treatment of eye issues.

REFERENCES

[1]. K. B. Vardhan, M. Nidhish, and D. N. Shameem, "Eye disease detection using deep learning models with transfer learning techniques," *ICST Transactions on Scalable Information Systems*, vol. 11, pp. 1–13, 2024.
[2]. K. Mittal, K. S. Gill, D. Upadhyay, and S. Devliyal, "Revolutionizing retinal health: Transfer learning for early detection of pathologies in limited OCT image datasets," in *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Apr. 2024, pp. 1–4.
[3]. J. Yang, G. Wang, X. Xiao, M. Bao, and G. Tian, "Explainable ensemble learning method for OCT detection with transfer learning," *PLoS One*, vol. 19, no. 3, p. e0296175, 2024.
[4]. V. Balajishanmugam, B. Shuriya, A. Kousalya, C. Kumar, J. Yamini, and R. Radhika, "Advancements in machine learning for COVID-19 detection from chest CT scans: A comprehensive study and performance evaluation," in *International Conference on Evolutionary Artificial Intelligence*, Nov. 2024, pp. 13–25.
[5]. E. S. Cutur and N. G. Inan, "Multi-class classification of retinal eye diseases from ophthalmoscopy images using transfer learning-based vision transformers," *Journal of Imaging Informatics in Medicine*, pp. 1–15, 2025.
[6]. R. Jain, T. K. Yoo, I. H. Ryu, J. Song, N. Kolte, and A. Nariani, "Deep transfer learning for ethnically distinct populations: Prediction of refractive error using optical coherence tomography," *Ophthalmology and Therapy*, vol. 13, no. 1, pp. 305–319, 2024.
[7]. M. Agarwal, K. S. Gill, D. Upadhyay, and S. Dangi, "Charting new frontiers using transfer learning in OCT image analysis for retinal health," in *2024 International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES)*, May 2024, pp. 1–5.
[8]. S. S. Mahmood, S. Chaabouni, and A. Fakhfakh, "Improving automated detection of cataract disease through transfer learning using ResNet50," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17541–17547, 2024.
[9]. C. Francis, V. S. Naidu, and B. Shuriya, "Comment on 'The revolutionary impact of artificial intelligence in orthopedics: comprehensive review of current benefits and challenges'," *Journal of Robotic Surgery*, vol. 19, no. 1, pp. 1–2, 2025.
[10]. G. N. Alwakid, M. Humayun, and W. Gouda, "A comparative investigation of transfer learning frameworks using OCT pictures for retinal disorder identification," *IEEE Access*, vol. 12, pp. 138510–138518, 2024.
[11]. P. K. Verma, J. Kaur, and N. P. Singh, "An intelligent approach for retinal vessels extraction based on transfer learning," *SN Computer Science*, vol. 5, no. 8, p. 1072, 2024.

DOI: 10.17148/IJARCCE.2025.14923

[12]. R. Patil and S. Sharma, "Automatic glaucoma detection from fundus images using transfer learning," *Multimedia Tools and Applications*, vol. 83, no. 32, pp. 78207–78226, 2024.

[13]. F. Du, L. Zhao, H. Luo, Q. Xing, J. Wu, Y. Zhu, et al., "Recognition of eye diseases based on deep neural networks for transfer learning and improved DS evidence theory," *BMC Medical Imaging*, vol. 24, no. 1, p. 19, 2024.

[14]. M. E. Subasi, S. Patnaik, and A. Subasi, "Optical coherence tomography image classification for retinal disease detection using artificial intelligence," in *Applications of Artificial Intelligence in Healthcare and Biomedicine*, 2024, pp. 289–323.

[15]. G. Gencer and K. Gencer, "Advanced retinal disease detection from OCT images using a hybrid squeeze and excitation enhanced model," *PLoS One*, vol. 20, no. 2, p. e0318657, 2025.

[16]. Z. Zhang, Y. Liu, and H. Chen, "Optical coherence tomography image recognition of diabetic retinopathy based on deep transfer learning," Computers and Electrical Engineering, vol. 114, p. 108389, 2024.

[17]. S. Sharma, A. R. Singh, and M. K. Verma, "Improving retinal OCT image classification accuracy using medical pre-training and sample replication methods," *Biomedical Signal Processing and Control*, vol. 87, p. 105539, 2024.

[18]. T. Raj, M. P. Kumar, and A. H. Khan, "Advanced deep learning models for accurate retinal disease state detection," *International Journal of Information Technology and Computer Science*, vol. 16, no. 3, pp. 51–58, 2024.

[19]. A. Khan and R. Mehta, "Detection of retinal diseases from OCT images using a VGG16 and transfer learning," *SN Applied Sciences*, vol. 7, no. 2, p. 156, 2025.

[20]. H. Lin, C. Duan, Y. Sun, L. Xu, and F. Wang, "OCT-SelfNet: A self-supervised framework with multi-source datasets for generalized retinal disease detection," *Frontiers in Big Data*, vol. 8, p. 1609124, 2025.