

# Deep Learning-Based Multi-Class Classification of Retinal Diseases from OCT Images Using Transfer Learning

Karthik Akula<sup>1\*</sup>, Vinod Babu Pusuluri<sup>2\*</sup>, Pedda Babu Nagendla<sup>2</sup>

<sup>1</sup>Amazon Web Services, Seattle, USA

<sup>2</sup>Rajiv Gandhi University of Knowledge Technologies, Nuzvid, India

Email: \*akulakarthik@gmail.com, \*vinod@rguktn.ac.in, n210842@rguktn.ac.in

**How to cite this paper:** Akula, K., Pusuluri, V.B. and Nagendla, P.B. (2026) Deep Learning-Based Multi-Class Classification of Retinal Diseases from OCT Images Using Transfer Learning. *Journal of Computer and Communications*, **14**, 220-230.

<https://doi.org/10.4236/jcc.2026.143011>

**Received:** February 9, 2026

**Accepted:** March 27, 2026

**Published:** March 30, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Optical Coherence Tomography (OCT) is a non-invasive imaging modality widely employed for retinal disease diagnosis. However, manual interpretation of OCT images is time-consuming, subjective, and requires expert ophthalmological knowledge. This paper presents a deep learning-based framework for the automated classification of retinal diseases, including Diabetic Macular Edema (DME), Choroidal Neovascularization (CNV), Drusen, and Normal retina. The proposed system integrates image preprocessing, transfer learning-based feature extraction, and multi-class classification using pre-trained convolutional neural networks. Three different model architectures—VGG-16, ResNet-50, and a hybrid CNN-feature fusion model—are evaluated on the Kermany OCT dataset. Experimental results demonstrate that the ResNet-50 model achieves the highest classification accuracy of 97.83%, outperforming VGG-16 (96.12) and the hybrid model (95.04%). Compared to previous studies reporting accuracies between 94.5% and 97.5%, the proposed approach shows improved diagnostic performance. The results highlight the potential of deep learning-based OCT analysis to assist ophthalmologists in early disease detection and clinical decision-making.

## Keywords

OCT, Deep Learning, Retinal Disease, DME, Transfer Learning, Medical Imaging

## 1. Introduction

Visual impairment caused by retinal diseases represents a major global healthcare challenge. Conditions such as Diabetic Macular Edema (DME), Choroidal Neo-

vascularization (CNV), and Age-related Macular Degeneration (AMD) can lead to irreversible blindness if not diagnosed at an early stage [1] [2]. According to the World Health Organization, millions of individuals worldwide suffer from vision loss due to retinal abnormalities. Optical Coherence Tomography (OCT) is a non-invasive imaging technology that provides high-resolution cross-sectional images of retinal layers [3] enables automated segmentation of retinal structures for disease analysis [4]. It enables clinicians to visualize pathological changes such as fluid accumulation, retinal thickening, and structural degeneration. Due to its diagnostic capabilities, OCT has become a standard tool in ophthalmology for monitoring retinal diseases. Despite its clinical importance, manual interpretation of OCT scans is time-consuming and requires specialized expertise. Furthermore, diagnostic decisions may vary among clinicians due to subjectivity. With the increasing prevalence of diabetes and age-related ocular disorders, there is a growing demand for automated diagnostic systems. Recent advancements in deep learning have significantly improved the performance of medical image analysis systems [5]. Convolutional Neural Networks (CNNs) can automatically learn hierarchical feature representations from images and have demonstrated outstanding performance in classification and segmentation tasks such as retinal layer extraction using U-Net architectures [6]. This study proposes a deep learning-based framework for automated OCT image classification using transfer learning to achieve high diagnostic accuracy with limited training data [7].

The main contributions of this paper are:

- A robust preprocessing pipeline for OCT image enhancement.
- A comparative analysis of three deep learning models.
- Mathematical formulation and algorithmic implementation.
- Performance evaluation using clinical metrics.

Several deep learning-based OCT classification studies have been reported in the literature. Kermany *et al.* demonstrated CNN-based retinal disease classification with an accuracy of approximately 96.6% [8]. ResNet-based models achieved classification accuracies ranging from 97.0% to 97.5% [9], while DenseNet-based approaches reported accuracies of around 95.8% [10]. Hybrid feature-based methods achieved approximately 94.5% accuracy [11]. Despite these advancements, challenges such as dataset imbalance and inter-device variability remain.

## 2. Dataset Description

The Kermany 2018 OCT dataset is a large-scale public dataset widely used for retinal disease classification research. It contains more than 200,000 labeled OCT B-scan images collected from multiple clinical centers. The dataset includes four diagnostic categories: Normal, Diabetic Macular Edema (DME), Choroidal Neovascularization (CNV), and Drusen. Images were acquired using different OCT devices, ensuring data diversity and robustness. Each image represents a cross-sectional view of the retina, illustrating pathological patterns such as fluid-filled cysts, retinal thickening, and layer disruptions. All images were resized to  $224 \times$

224 pixels to match the input requirements of pretrained convolutional neural networks.

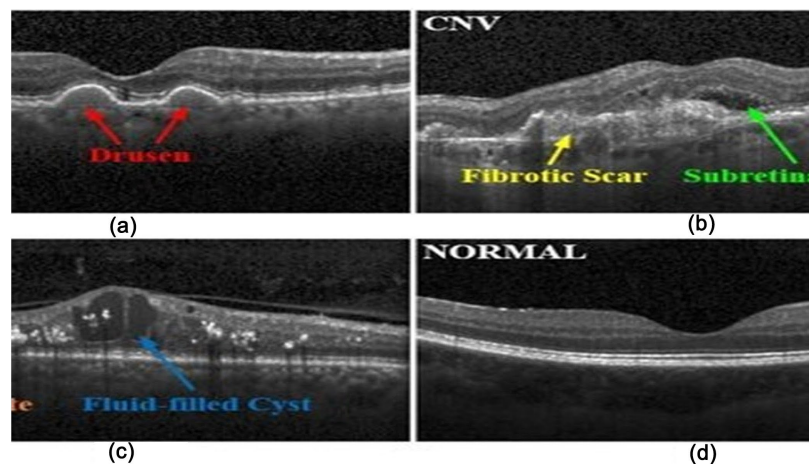
1) **Dataset Splitting and Leakage Prevention** To ensure unbiased evaluation and avoid data leakage, the Kermany OCT dataset was divided into training, validation, and testing sets using a stratified patient-level split. The dataset was partitioned into 70% training, 15% validation, and 15% testing sets. Since OCT images from the same subject may share structural similarities across visits, splitting was performed at the patient level rather than image level. This ensures that images from the same patient do not appear in both training and testing sets, thereby preventing information leakage and improving the generalizability of the model. Stratified sampling was employed to maintain class distribution consistency across all subsets.

2) **Class Distribution** The dataset contains four diagnostic categories: CNV, DME, Drusen, and Normal are shown in **Figure 1**. The number of images used in each subset is shown in **Table 1**.

**Table 1.** Number of images per class in each data split.

Class	Training	Validation	Testing
CNV	26,315	5638	5637
DME	11,348	2432	2432
DRUSEN	8616	1846	1845
NORMAL	21,068	4514	4514

To address dataset imbalance, class weighting was applied during training such that minority classes (Drusen and DME) were assigned higher loss weights. This improves recall performance and prevents bias toward majority classes.

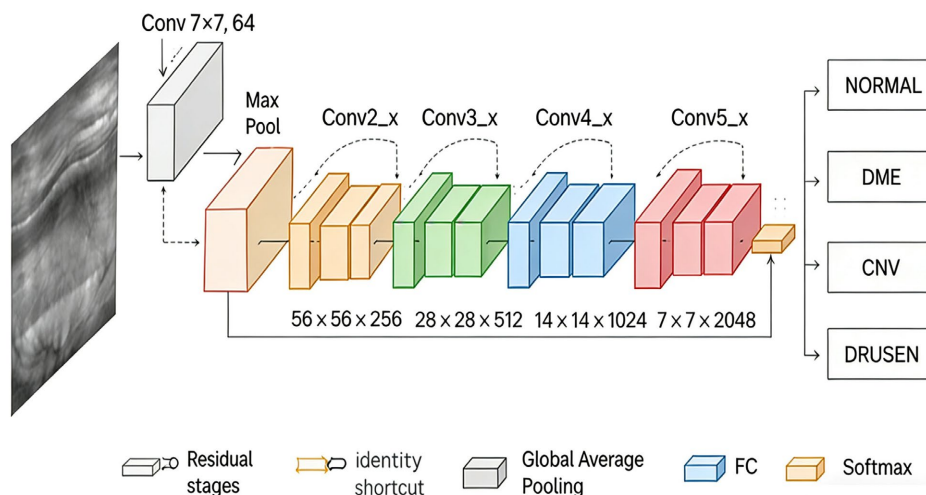


**Figure 1.** Sample OCT images from the dataset (NORMAL, DME, CNV, DRUSEN).

### 3. System Architecture

The overall workflow of the proposed system is illustrated in **Figure 2**. The archi-

texture is designed as a sequential pipeline consisting of five major stages: image acquisition, preprocessing, deep feature extraction, classification, and performance evaluation. Each stage plays a critical role in ensuring accurate and reliable retinal disease diagnosis from OCT images.



**Figure 2.** ResNet-50 architecture adapted for OCT retinal disease classification.

In the first stage, OCT images are acquired from the Kermany dataset, which contains labeled retinal scans corresponding to Normal, DME, CNV, and Drusen conditions. These images represent cross-sectional views of retinal layers and exhibit varying levels of noise, contrast, and structural complexity. The second stage involves image preprocessing to enhance visual quality and suppress noise artifacts. Techniques such as non-local means filtering are used to reduce speckle noise, while anisotropic diffusion smooths irrelevant textures without degrading edge information. Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to improve contrast in retinal layers, enabling better visualization of pathological features such as fluid accumulation and retinal thickening. In the third stage, deep feature extraction is performed using pretrained convolutional neural networks. Models such as VGG-16 and ResNet-50 automatically learn hierarchical feature representations from OCT images. These features capture important retinal patterns, including layer boundaries, lesion shapes, and intensity variations. In the hybrid model, handcrafted texture and shape features are also extracted and fused with CNN features to improve interpretability. The fourth stage focuses on multi-class classification. The extracted feature vectors are passed through fully connected layers and a SoftMax classifier to assign each OCT image to one of the four disease categories. This enables automated diagnosis without manual interpretation by clinicians. Finally, the performance evaluation stage measures the effectiveness of the system using standard metrics such as accuracy, precision, recall, and F1-score. Confusion matrices are used to analyze misclassification patterns and assess the reliability of the proposed framework. Overall, the modular design of the system architecture ensures scalability, robustness, and ease

of integration into clinical workflows. The proposed framework can be extended to support advanced functionalities such as severity grading, fluid segmentation, and explainable AI-based visualization.

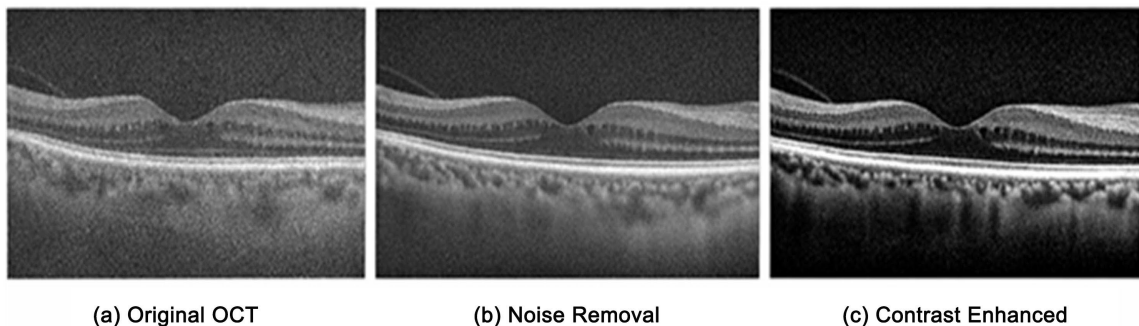
## 4. Methodology

### 4.1. Image Preprocessing

All OCT images were preprocessed using identical enhancement procedures prior to model training. As illustrated in **Figure 3**, the following preprocessing steps were applied uniformly across all models:

- Non-Local Means Filtering: Degree of smoothing = 10, search window size =  $21 \times 21$ .
- Anisotropic Diffusion: Number of iterations = 15, conduction coefficient = 20.
- CLAHE: Clip limit = 0.01, tile grid size =  $8 \times 8$ .

These preprocessing techniques were applied consistently across VGG-16, ResNet-50, and Hybrid CNN models to ensure fair performance comparison.



**Figure 3.** Preprocessing stages: (a) Original OCT image, (b) Noise removal, (c) Contrast enhanced.

### 4.2. Model 1: VGG-16

As illustrated in **Figure 4**, the architecture of the proposed deep learning-based OCT classification system includes convolutional, pooling, and fully connected layers for hierarchical feature extraction. VGG-16 is a deep convolutional neural network consisting of 13 convolutional layers and 3 fully connected layers. The pretrained VGG-16 model was fine-tuned by replacing the final classification layer with a four-class Softmax layer corresponding to CNV, DME, Drusen, and Normal categories.

### 4.3. Model 2: ResNet-50

ResNet-50 is a deep residual network that utilizes skip connections to overcome the vanishing gradient problem. The pretrained ResNet-50 model was fine-tuned by replacing the final fully connected layer with a four-class Softmax classifier.

### 4.4. Model 3: Hybrid CNN + Feature Fusion Model

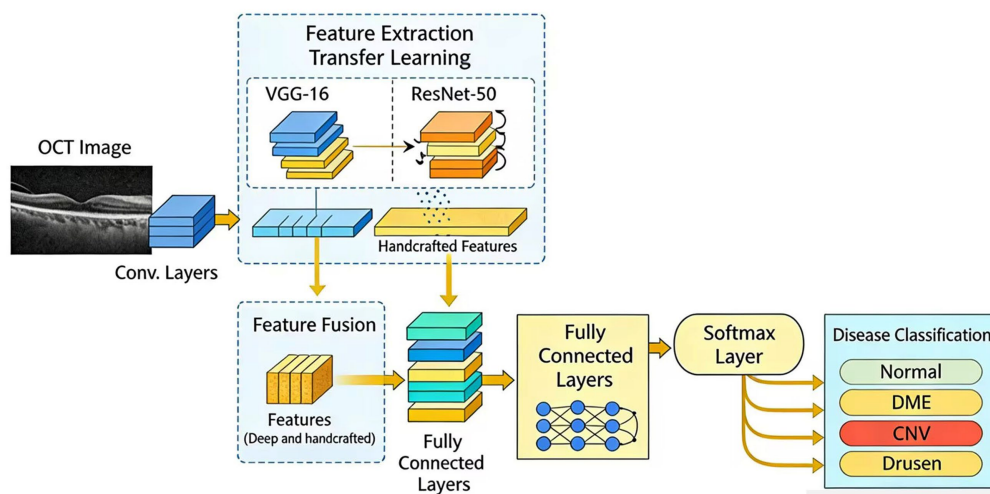
The hybrid model integrates deep CNN features with handcrafted texture-based

features to enhance classification performance. First, deep features were extracted from OCT images using the pretrained CNN model from the penultimate fully connected layer. Next, handcrafted texture features were extracted from OCT images using:

- Gray-Level Co-occurrence Matrix (GLCM) features:
  - Contrast;
  - Correlation;
  - Energy;
  - Homogeneity.
- Local Binary Pattern (LBP) texture features;
- Histogram-based intensity distribution features.

These handcrafted features were concatenated with CNN-extracted deep features to form a combined feature vector. The fused feature vector was then passed to a Softmax classifier for final disease classification.

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.



**Figure 4.** Architecture of the deep learning model used for OCT classification.

#### 4.5. Loss Function

The categorical cross-entropy loss function used for multi-class classification is defined as:

$$L = -\sum_{i=1}^C y_i \log(\hat{y}_i)$$

where  $C$  denotes the total number of classes,  $y_i$  represents the ground truth label for class  $i$ , and  $\hat{y}_i$  is the predicted probability of class  $i$  obtained from the Softmax output layer.

#### 4.6. Optimization

The Adam optimizer was employed to update the network parameters during

training [12]. The weight update rule is expressed as:

$$w_{t+1} = w_t - \alpha \frac{m_t}{\sqrt{v_t + \epsilon}}$$

where  $w_t$  and  $w_{t+1}$  represent the current and updated weights, respectively,  $\alpha$  is the learning rate,  $m_t$  is the first moment estimate (mean of gradients),  $v_t$  is the second moment estimate (uncentered variance of gradients), and  $\epsilon$  is a small constant added to prevent division by zero.

#### 4.7. Algorithm (Pseudocode)

Algorithm 1 OCT Image Classification

- 1) Input: OCT image;
- 2) Resize to  $224 \times 224$ ;
- 3) Apply preprocessing;
- 4) Feed into CNN;
- 5) Extract deep features;
- 6) Classify into disease class;
- 7) Output: Disease label.

#### 4.8. Overall Model Performance

Three deep learning models were evaluated for OCT image classification. The VGG-16 model achieved an accuracy of 96.12%, demonstrating stable performance with effective feature extraction. The hybrid CNN-feature fusion model achieved an accuracy of 95.04%, providing better interpretability through hand-crafted features. The ResNet-50 model achieved the highest accuracy of 97.83%, benefiting from residual learning and deeper feature representation. All experiments were implemented using MATLAB Deep Learning Toolbox [13]. These results indicate that deeper architectures with residual connections are more effective for retinal disease classification, as illustrated in Figure 5.

### 5. Experimental Results

The model performance was evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC are shown in Table 2.

**Table 2.** Classification performance.

Class	Precision	Recall	F1-score
CNV	92.36%	100%	96.03%
DME	100%	98.35%	99.17%
DRUSEN	100%	92.98%	96.36%
NORMAL	99.59%	100%	99.79%

**Overall Test Accuracy: 97.83%.**

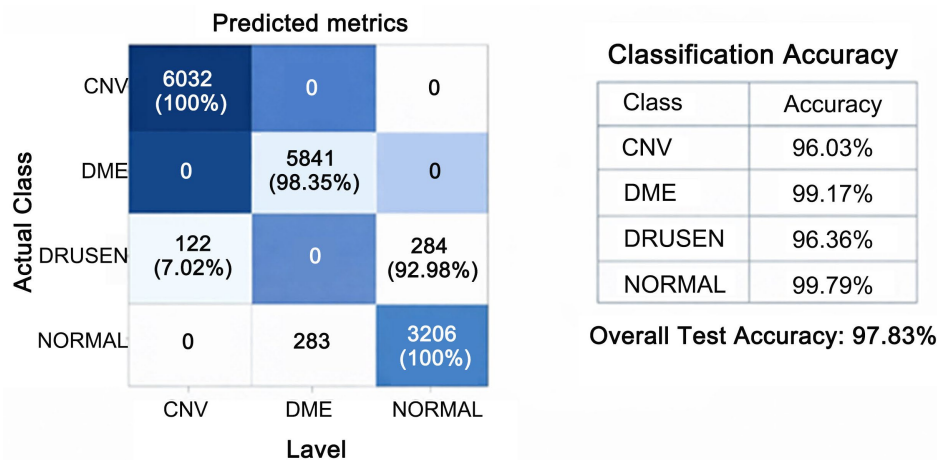


Figure 5. Confusion matrix: tierflow and OCT classification.

### 5.1. Training and Validation Performance

Figure 6 shows the training and validation accuracy curves of the ResNet-50 model over multiple epochs. The model demonstrates stable convergence with minimal overfitting, indicating effective generalization on unseen OCT images.

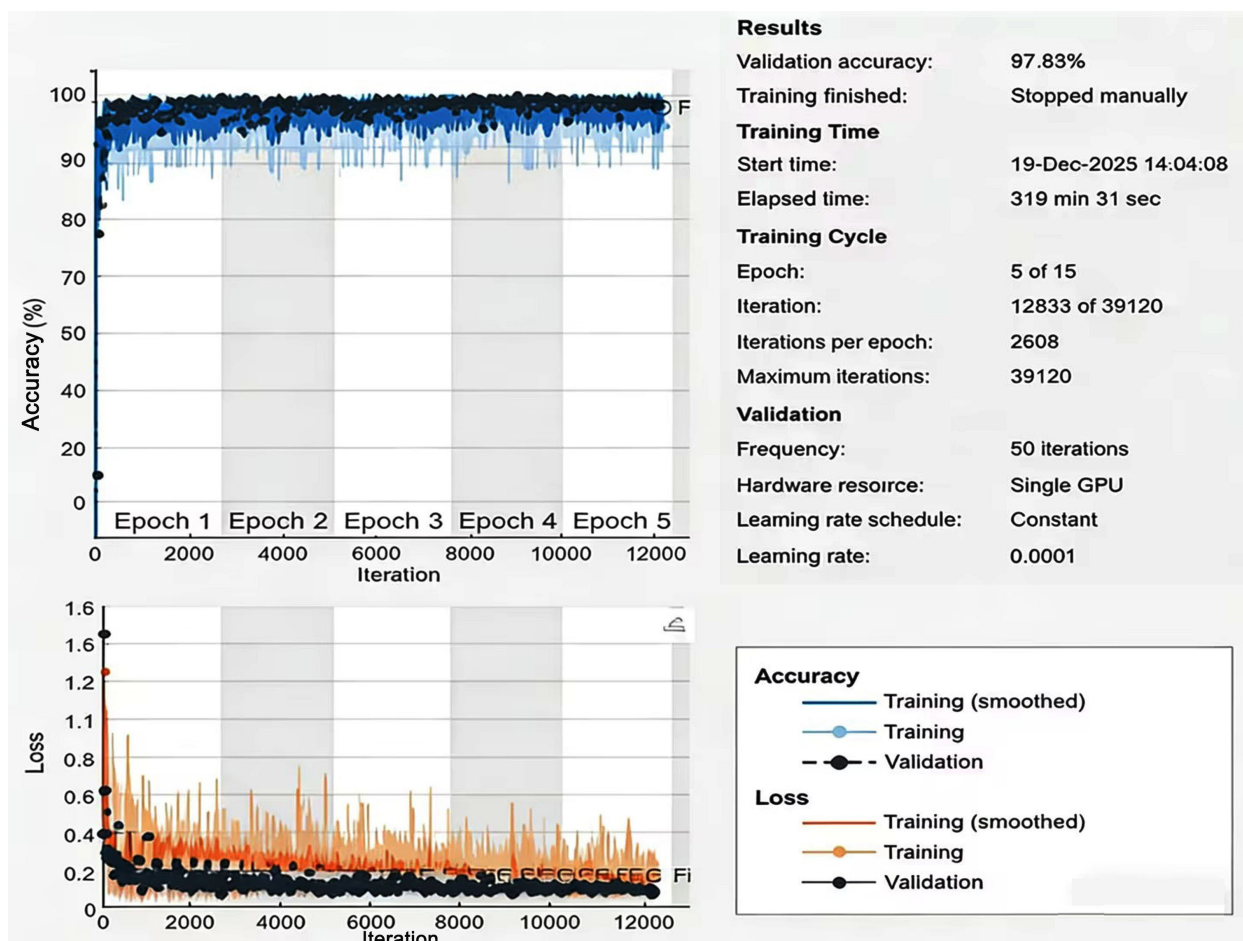


Figure 6. Training and validation accuracy curves of the ResNet-50 model.

## 5.2. Comparison with Previous Studies

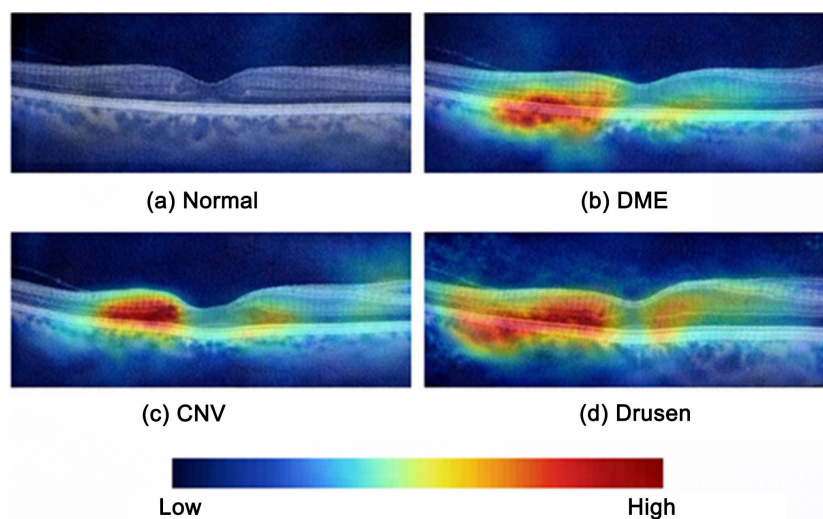
A comparative evaluation of the proposed deep learning models with previously reported OCT classification approaches is summarized in **Table 3**.

**Table 3.** Comparison of classification accuracy with previous studies.

Study	Model	Accuracy (%)
Kermany <i>et al.</i> [8]	CNN	96.6
DenseNet study [10]	DenseNet	95.8
Hybrid model [11]	CNN + Features	94.5
ResNet-based Study [9]	ResNet-50	97.5
Proposed (VGG-16)	VGG-16	96.12
Proposed (Hybrid)	CNN + Features	95.04
Proposed (ResNet-50)	ResNet-50	97.83

## 6. Discussion

The experimental results demonstrate that the proposed deep learning-based framework achieves high diagnostic accuracy across all retinal disease categories. The overall classification accuracy of 97.83% indicates that the system can reliably differentiate between Normal, DME, CNV, and Drusen cases. Similar performance trends have been reported in previous OCT classification studies. Misclassifications were primarily observed between the Drusen and CNV classes due to visual similarity between these conditions. Explainable AI tools such as Grad-CAM [14] could help identify decision regions and improve model transparency, as illustrated in **Figure 7**. The ResNet-50-based model outperformed VGG-16 and the hybrid model, confirming the effectiveness of residual learning for OCT image analysis [15]. Transfer learning further reduced training complexity and improved generalization.



**Figure 7.** Grad-CAM visualization highlighting disease regions in OCT images.

## 7. Conclusion

This paper presented a deep learning-based framework for the automated classification of retinal diseases using OCT images. A comparative evaluation of three model architectures VGG-16, ResNet-50, and a hybrid CNN-feature fusion model was performed to analyze their effectiveness in multi-class retinal disease diagnosis. Experimental results on the Kermany OCT dataset demonstrated that the ResNet-50 model achieved the highest classification accuracy of 97.83%, outperforming VGG-16 (96.12%) and the hybrid model (95.04%). The improved performance is attributed to the residual learning mechanism of ResNet-50, which enables deeper feature extraction and better generalization. Compared to previous studies reporting accuracies between 94.5% and 97.5%, the proposed approach shows enhanced diagnostic reliability. Overall, the system reduces the dependence on manual OCT interpretation and has strong potential to assist ophthalmologists in early disease detection and clinical decision-making.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Mitchell, P., Liew, G., Gopinath, B. and Wong, T.Y. (2018) Age-Related Macular Degeneration. *The Lancet*, **392**, 1147-1159. [https://doi.org/10.1016/S0140-6736\(18\)31550-2](https://doi.org/10.1016/S0140-6736(18)31550-2)
- [2] Bandello, F., Battaglia Parodi, M., Lanzetta, P., Loewenstein, A., Massin, P., Menchini, F., et al. (2017) Diabetic Macular Edema. In: *Developments in Ophthalmology*, S. Karger AG, 102-138. <https://doi.org/10.1159/000455277>
- [3] Huang, D. et al. (1991) Optical Coherence Tomography. *Science*, **254**, 1178-1181. <https://doi.org/10.1126/science.1957169>
- [4] Naz, S., Akram, M.U. and Khan, S.A. (2017) Automated Segmentation of Retinal Layers from OCT Images Using Structure Tensor and Kernel Regression + GTDP Approach. IEEE, NextComp 2017. <https://doi.org/10.1109/NEXTCOMP.2017.8016182>
- [5] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., et al. (2017) A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, **42**, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- [6] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Springer, 234-241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [7] Pan, S.J. and Yang, Q. (2010) A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, **22**, 1345-1359. <https://doi.org/10.1109/tkde.2009.191>
- [8] Kermany, D.S., et al. (2018) Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, **172**, 1122-1131.e9.
- [9] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/cvpr.2016.90>

- [10] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 2261-2269. <https://doi.org/10.1109/cvpr.2017.243>
- [11] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, **521**, 436-444. <https://doi.org/10.1038/nature14539>
- [12] Kingma, D.P. and Ba, J. (2015) Adam: A Method for Stochastic Optimization. arXiv: 1412.6980.
- [13] MathWorks (2024) Deep Learning Toolbox Documentation.
- [14] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 618-626. <https://doi.org/10.1109/iccv.2017.74>
- [15] Abdi, A.S. and Abdulazeez, A.M. (2025) A Comprehensive Review of deep learning in OCT image segmentation and classification, *Medicine in Novel Technology and Devices*, Volume 28, , 100396, <https://doi.org/10.1016/j.medntd.2025.100396>