Electronic Letters on Computer Vision and Image Analysis 25(1):43-59, 2025

A Hybrid Feature Fusion for Retinal OCT Image Classification using Traditional and Deep Learning Methods

Mithilesh Kumar Singh Yadav* and Nagendra Pratap Singh+

* + *Department of Computer Science and Engineering, Dr B R Ambedkar National Institute of Technology, Jalandhar, Punjab, India*
Received 16th of July 2025; accepted 6th of March 2026

Abstract

Early and accurate detection of diabetic macular edema (DME) is essential to avoid permanent loss of vision. This paper introduces Fusion-WideNet, a new hybrid classification model that combines handcrafted and deep features for the analysis of retinal OCT images. Handcrafted features—Gray Level Co-occurrence Matrix (GLCM), Histogram of Oriented Gradients (HOG), and Local Binary Pattern (LBP) are learned to extract local textural information, and deep semantic features are extracted from a pretrained ResNet50 convolutional neural network. These two feature sets are combined in a high-dimensional space and fed into a Wide Neural Network (WideNet), a shallow yet high-capacity network designed to handle large feature vectors. The model achieves a classification accuracy of 99.5%, outperforming traditional models and deep-only baselines. The proposed Fusion-WideNet framework not only demonstrates high diagnostic performance but also provides interpretability, which is essential for real-world ophthalmic screening and clinical decision support.

Key Words: Retinal OCT images, Feature extraction, Classification, Machine learning, Deep learning.

1 Introduction

Over the last ten years, optical coherence tomography (OCT) has been a remarkable development as a medical imaging tool. The imaging method may record various properties of biological tissue, such as molecular content, structural information, elastic parameters, polarization status, and blood flow [1]. Today, retinal diseases are prevalent, widespread, and profound. The eyes are the most sensitive and crucial part of the human body that should be cared for regularly to avoid retinal problems. The internal structure of the eye is illustrated in Figure 1. The most rapidly spreading retinal diseases are diabetic macular edema, age-related macular degeneration, glaucoma, hypertension, obesity, and choroidal neovascularization (CNV). In 2018, the World Health Organization (WHO) estimated that 2.2 billion people globally experienced some form of visual impairment. The predominant cause of blindness and visual impairment among the older population in the United States is age-related macular degeneration (AMD). The World Health Organization estimates that 196 million individuals will have AMD globally in 2020, with projections indicating an increase to 288 million by 2040.
Finding diseases of the eye is a common classification problem for machine learning. This study solves the problem by automatically finding diseases from scanned optical coherence tomography (OCT) retinal images.

Correspondence to: mithileshksy.cs.23@nitj.ac.in,singhnp@nitj.ac.in

Recommended for acceptance by Angel D. Sappa
https://doi.org/10.5565/rev/elcvia.2280
ELCVIA ISSN:15108-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

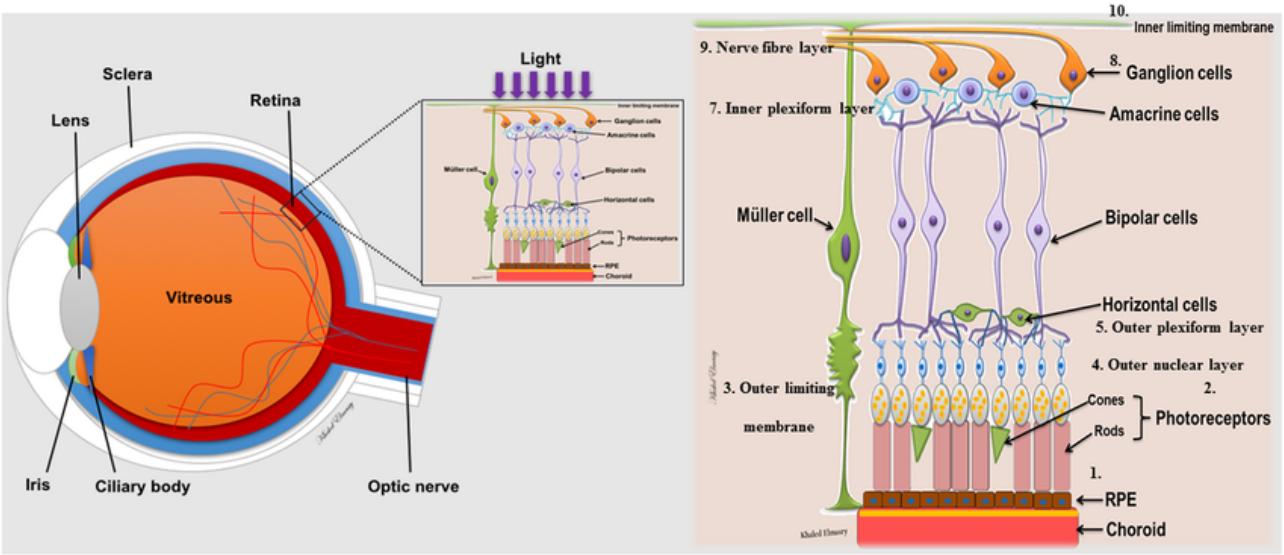

Figure 1: Anatomy of the eye and organization of retinal layers [2]

These images show the retina in two different states: Healthy and Diabetic macular edema(DME), as shown in Figure 2. Because optical coherence tomography (OCT) imaging methods are increasingly used in medicine, an automated computer-aided diagnosis system may be very useful for automatically finding retinal diseases and helping with their treatment and follow-up. Image pre-processing is a big part of standard methods for automatically finding retinal diseases. The processed images are then fed into a shallow neural network. Because of these limitations, we will use machine learning methods with shallow networks. Many researchers focus on using OCT to analyze medical data to make better medical choices. This can be done by using and improving AI techniques like machine learning (ML) and deep learning (DL), alone or together. OCT has been used as a clinical instrument in various clinical disciplines, including ophthalmology [3, 4], cardiology [5, 6], endoscopy [7, 8], oncology [9, 10], and dermatology [11]. In the field of developmental biology, OCT has been employed to examine and analyze the morphological and functional progression of various organs, including but not limited to the eyes [12], the brain [13], the limbs [14], the reproductive organs [15] and the heart [16, 17].

Deep learning, particularly Convolutional Neural Networks (CNNs), has demonstrated remarkable performance in medical image analysis by learning hierarchical representations directly from raw image pixels. Pretrained networks such as ResNet, VGG, and Inception have been applied to OCT image classification using transfer learning. However, these models can be opaque, requiring large labeled datasets, and may overlook subtle but clinically significant texture variations.

To overcome the limitations of both approaches, we propose Fusion-WideNet, a novel hybrid framework that combines the strengths of handcrafted and deep features through a unified architecture. Fusion-WideNet extracts handcrafted descriptors (GLCM, HOG, LBP) to capture low-level texture patterns and fuses them with high-level features obtained from a pretrained ResNet50 network. The resulting feature vector is passed into a Wide Neural Network (WNN) classifier designed with broad but shallow layers to efficiently learn from the fused representation.

The key contributions of this study are as follows:

- We present Fusion-WideNet, a hybrid classification framework combining traditional handcrafted descriptors with deep ResNet-50 features to improve retinal OCT image classification.
- We design a wide-shallow neural network model that efficiently manages the high-dimensional fused features and minimizes the overfitting risk.
- We validate our framework using stratified k-fold cross-validation, achieving superior performance over

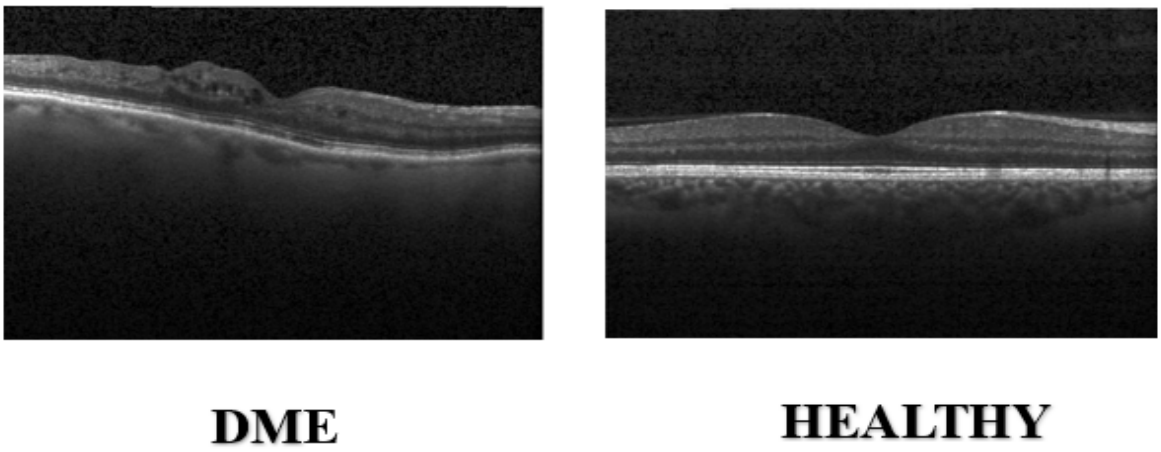

Figure 2: Sample image of 2 class

state-of-the-art methods.

2 Related Work

Earlier OCT classification models predominantly relied on handcrafted texture descriptors. GLCM features were used by Haralick et al. [18] to capture spatial co-occurrence relationships between pixels, which proved effective in texture classification tasks. Similarly, HOG [19] and LBP [20] descriptors have been widely used for edge-based and local structure analysis, respectively. With traditional classifiers like SVM or RF, the corresponding features exhibited good performance, supporting the conclusion above. However, handcrafted methods inherently lack adaptability and fail to generalize across diverse datasets or imaging conditions.

Following the arrival of deep learning, CNNs now form the norm for image-based diagnostics. Kermany et al. [21] presented an OCT dataset with massive numbers and utilized transfer learning in the form of pre-trained models such as Inception and ResNet to show dramatic performance gains in the classification of retinal disease. Recent studies have also made use of fine-tuned deep networks for OCT segmentation and classification to show high sensitivity and specificity. However, CNN-based approaches tend to need big annotated datasets and are black-box models, which restricts their clinical application due to a lack of interpretability.

Hirota et al.[22] presented an ML model for screening retinal abnormalities using lower-quality OCT images taken during routine eye exams as a preventive measure against eye problems. The ensemble mode integrating CNNs and random forest architectures performed well in screening single-shot OCT images taken during actual eye exams. The study found that the ensemble approach combining CNN and random forest algorithms showed great potential for accurately diagnosing abnormalities during eye exams.

A. Garcia et al.[23] introduced an effective OCT image classification technique for the early identification of AMD. The approach employed SVM for classification, with an accuracy of 92.12%, illustrating its efficacy in detecting early AMD characteristics in OCT pictures.

The author Wali et al.[24] introduced an improved diagnostic method for retinal diseases, emphasizing the crucial significance of machine learning. Employing a Random Forest classifier, researchers attained 80% accuracy in differentiating diseases such as DME, CNV, and drusen. Wali et al.[25] presented an ensemble model, which uses many deep learning models to maximize their strengths. This method tries to classify OCT images more reliably and effectively. CNN architecture, DenseNet121, and InceptionV3 models improve classification accuracy. The study classifies OCT pictures as CNV, DME, drusen, and normal. The proposed ensemble model was 97.5% accurate.

The incidence of diabetic retinopathy was determined using various classifiers, including alternating decision tree, AdaBoost, Naive Bayes, Random Forest, and SVM [26]. The proposed system extracted textural features from the GLCM. These features were utilized for classifying normal and papilledema images using a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel. The proposed system demonstrated an average accuracy of 92.86% [27]. In this experiment, analysis was conducted on the arteriovenous ratio and the area of the optic nerve head (ONH). A hybrid set of structures was employed in the classification of Artery or Vein (A/V) using an SVM with its RBF kernel for arteriovenous ratio.

The study [28] uses OCT images to identify CNV. The study encompassed 484 OCT images, 242 healthy images, and 242 CNV photos. The findings indicated that the Fine-Gaussian Support Vector Machine (SVM-FG) classifier achieved a classification accuracy of 92%.

The author Rasti et al.[29] introduced an innovative method for identifying macular abnormalities with modern imaging techniques. It utilizes a random forest classifier in the second phase of its suggested methodology. The algorithm was evaluated using two distinct retinal SD-OCT datasets, achieving precision rates of 99.33% for two-class classification and 98.67% for three-class classification.

Researchers in this study[30] used OCT to automatically detect and measure areas with pixel-accuracy in geographic atrophy (GA) in late-stage AMD. They used novel, ML-based methods for this purpose.Both the detection and GA area measurement accuracy were 91% and 94%, respectively, for the B-scan model.

Mathew et al. [31] introduce a lightweight deep learning model designed for the classification of retinal OCT images, primarily focusing on DME and drusen macular degeneration (DMD). The model achieved a classification accuracy of 99.5% on the Mendeley OCT dataset and 94.9% on the Duke dataset by utilizing residual blocks and channel attention modules.

Author Aslam et al. [32] employed machine learning methods to categorize diabetic states based on optical coherence tomography angiography (OCT-A) pictures. Automated algorithms conducted expert analysis of OCT-A scans from 182 patients, producing clinically significant outcome indicators. The measures were utilized as inputs for classifiers, with the Random Forest classifier attaining the highest AUC of 0.8 for differentiating diabetic from non-diabetic patients, whereas logistic regression achieved an AUC of 0.91 for distinguishing diabetic patients.

Phu et al. [33] presents a straightforward and efficient model for the automatic classification of macular hole and diabetic macular edema illnesses utilizing optical coherence tomography pictures through Support Vector Machine methodology. The classification accuracy for macular holes reaches 95.8%, whereas that for diabetic macular edema is 86.7%.

Recent progress in OCT image processing has increasingly adopted hybrid deep learning frameworks that integrate complementary architectures to enhance diagnostic precision, resilience, and clinical relevance. Researchers have concentrated on lightweight and broadly applicable designs. Khan et al. [34] introduced a hybrid model combining ResNet50 and EfficientNetB0, demonstrating robust generalization and interpretability, achieving 97.5% accuracy on 84,452 OCT images while reducing overfitting via feature concatenation. Similarly, Prabha et al. [35] showed that RD-OCT Net is a lightweight hybrid framework for multiclass retinal illness classification (nAMD, DME, RVO, and normal). It achieved class-wise accuracies over 97% and had fewer trainable parameters, making it acceptable for clinical use. Agrawal et al. [36] introduced HDL-ACO, which integrates Convolutional Neural Networks (CNNs) with Ant Colony Optimization (ACO) and Transformer-based embeddings, with 93% validation accuracy while improving computing efficiency and hyperparameter optimization. Dutta et al. [37] introduced a novel approach for classifying retinal illnesses by extracting texture information from OCT images, employing InceptionV3 and ResNet-50 deep neural networks, with shape characteristics obtained from vision transformers (ViT), achieving an accuracy of 94%.

In this study, our suggested Fusion-WideNet enhances the field by methodically integrating handmade descriptors (GLCM, HOG, LBP) with deep features extracted from ResNet50 into a cohesive high-dimensional representation. In contrast to previous models that depend exclusively on deep or transformer architectures, Fusion-WideNet utilizes a wide-shallow architecture for fast learning, integrates attention-based weighting to emphasize discriminative features, and achieves a balance among accuracy, interpretability, and scalability.

This facilitates enhanced diagnostic efficacy for OCT-based retinal disease classification while preserving computational efficiency and clinical feasibility.

3 Dataset

In this study, we used retinal OCT images from the Kermany OCT dataset, which was made public by Kermany et al.(2018)[21] for retinal disease classification. The dataset is downloaded on Kaggle at the following link: `https://www.kaggle.com/datasets/paultimothymooney/kermany2018/data`. The dataset is freely available for academic and research purposes, ensuring transparency and reproducibility of the proposed methodology. The original dataset comprises OCT B-scan images of various retinal conditions, including CNV, DME, Drusen, and healthy cases. For this study, we selected a balanced subset of images that included only DME and healthy retinal images. This subset was focused on binary classification. Specifically, 6,000 DME images and 6,000 normal images were utilized, totaling 12,000 OCT images. We obtained all the images using spectral-domain OCT imaging systems and provided them to us as grayscale B-scan images. To ensure consistency and compatibility with the deep learning backbone, each image was resized to 224 × 224 pixels, and z-score normalization was applied to standardize pixel intensity values. To improve model generality and mitigate overfitting, we employed data augmentation approaches including random rotation (±15°), flipping, and scaling. The dataset was divided utilizing a 5-fold stratified cross-validation method, guaranteeing class balance in each fold, with 80% allocated for training and 20% for testing.

3.1 Retinal Abnormalities

Retinal disease is a common and serious condition that affects the human retina. Retinal illnesses are a broad category of conditions that affect the retina. The light-sensitive tissues at the back of the retina capture images and transmit visual information to the brain. These disorders can range in severity and, if not addressed, can result in vision impairment or blindness [38]. A variety of factors, including hypertension and vitamin insufficiency, cause retinal disease. Many retinal illnesses are discussed as follows:
Choroidal neovascularization (CNV): Neovascularization refers to developing novel blood veins. This depicts the unusual formation of recently developed veins on the tissue layer, those residing on the back side of the retina. The symptoms of CNV include vision malformation or the production of waves in central sight, which leads to the emergence of either a grey or black hole in the center. One of the most frequent disorders is CNV, which results from age-related macular degeneration, as shown in fig.3.

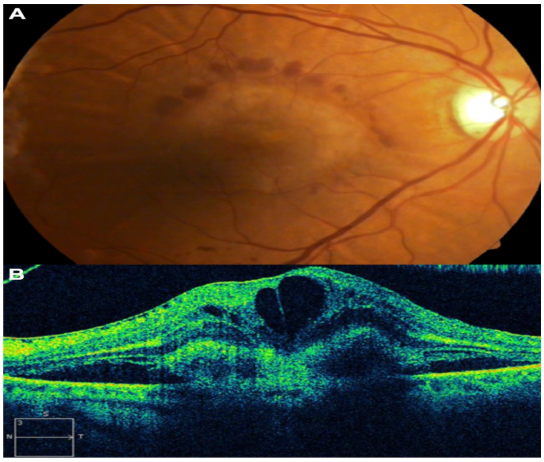

Figure 3: CNV in age related macular degeneration [39]

Diabetic macular edema (DME): Diabetic macular edema arises from a diabetic disease, comparable to dia-

betic retinopathy, as shown in Fig. 4. DME is caused by the collection of liquid in the macula, which is located behind the retina. Edema occurs when the macula becomes enlarged due to the accumulation of fluid. If not addressed, the vision will be permanently lost. DME might be caused by a different type of factors, which include improper glucose management, smoking, increased blood pressure, uncontrolled cholesterol, and pregnancy [40].

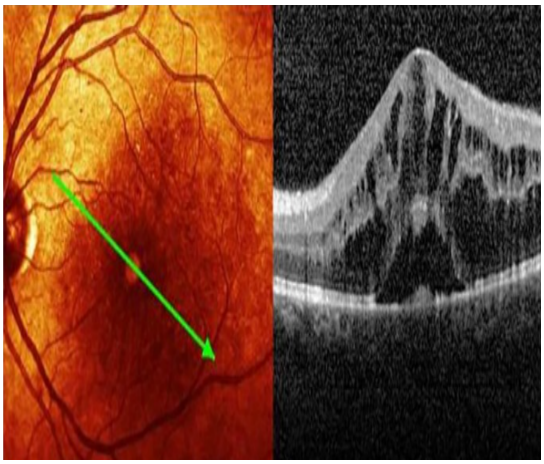

Figure 4: DME [41]

Drusen: Drusen are yellow-colored, microscopic fatty protein accumulations that form beneath the retina, as shown in fig. 5. The retina is an extremely thin film made up of cells that run along the behind the eye, around the optic nerve. Drusen seems like pebbles of waste matter that gather after a while. Drusen is mainly divided into two types: soft drusen and hard drusen. The soft drusen are mostly velvety, abundant, and huge. Hard drusen are smaller in size. It is usual for people to develop a few hard drusen as they age, and the majority of adults have one or more. The major cause of soft drusen is eye disorders such as AMD, since it primarily has an impact on people over the age of 60.

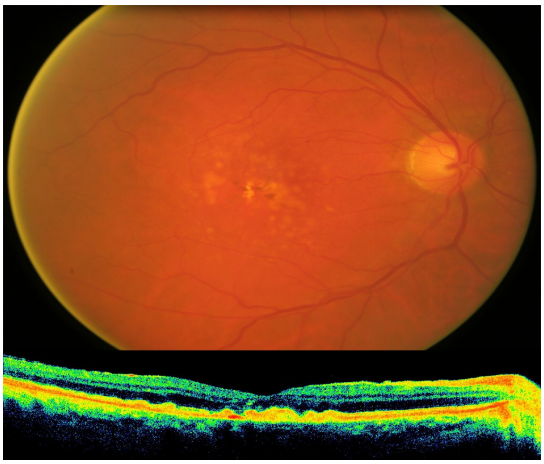

Figure 5: Drusen [42]

Age-related macular degeneration(AMD): AMD is a productive eye disorder affecting mostly the back of the eye, which is the macula. AMD is a primary cause of vision loss and blindness among older adults, particularly people aged 50 years and over. AMD consists of two main types: dry AMD and wet AMD, which are less frequent but account for the bulk of severe visual loss caused by AMD [43]. AMD includes risk aspects, for instance, advancing age, history of the disease in the family, smoking, obesity, and certain genetic factors.

We prioritize DME over Drusen, CNV, and AMD since it is a diabetic consequence affecting many patients worldwide. DME must be detected early to prevent irreparable visual loss. DME is fast-progressing and requires quick treatment, unlike Drusen or CNV. Due to its evident fluid-filled cystic alterations, OCT images are suited for machine learning-based DME categorization.

4 Proposed Methodology

A unified, efficient-performance classification system for retinal OCT image processing is provided by the suggested Fusion-WideNet framework, which combines deep convolutional features with handcrafted texture descriptors. 5 main steps comprise the methodology: (i) Preprocessing the data, (ii) Handcrafted feature extraction, (iii) Deep feature extraction using ResNet50, (iv) Hybrid feature fusion, (v) Classification using a Fusion-WideNet. The proposed architecture is illustrated in its entirety in Figure 6.

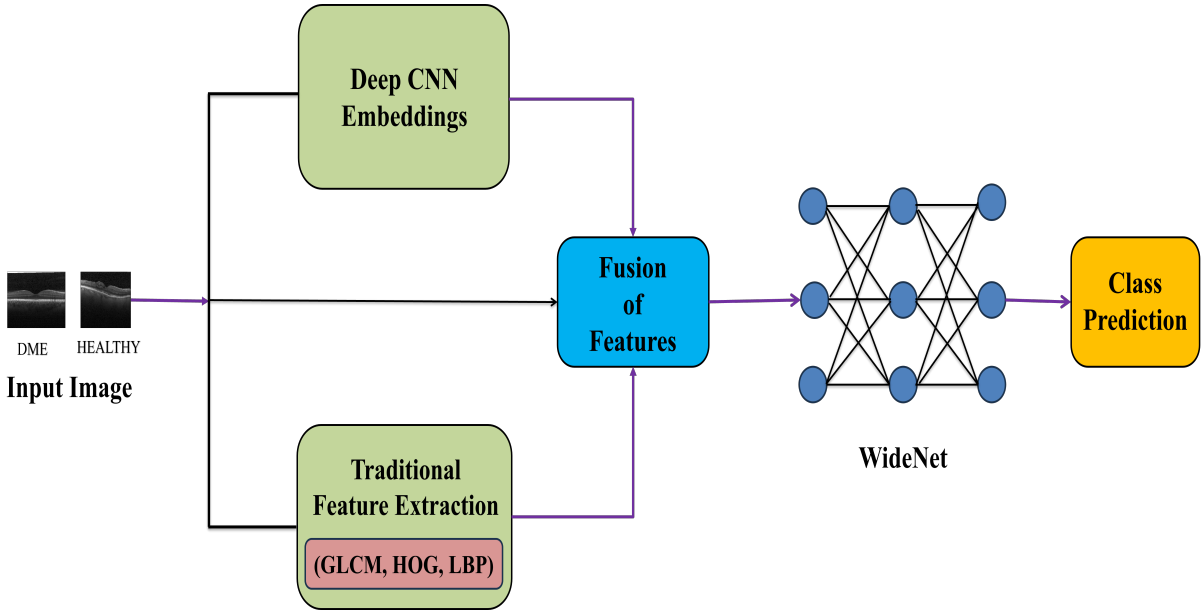

Figure 6: Architecture of the Proposed Fusion-WideNet framework

4.1 Preprocessing

A consistent resolution of 224 × 224 pixels is applied to all OCT images to ensure they are compatible with the input dimensions required by deep learning models. The images are prepared for handmade texture analysis by performing grayscale conversion when needed. For binary categorization (DME vs. Healthy), labels are encoded. The dataset is balanced, and to provide a strong model assessment, stratified k-fold cross-validation is used.

4.2 Handcrafted Feature Extraction

The following manually created features are taken from every preprocessed OCT image in order to record low-level textural and spatial information:

- **GLCM**: GLCM is an effective texture analysis method commonly employed in the classification of retinal diseases to derive spatial correlations among pixel intensities. Statistical texture characteristics are calculated from normalized co-occurrence matrices and include contrast, energy, homogeneity, correlation, and ASM.

- **HOG**: HOG is a feature extraction method extensively employed in object recognition and texture analysis, including the classification of retinal diseases. HOG operates by collecting the distribution of gradient orientations within an image, rendering it particularly effective for finding structural patterns in retinal images. **Gradient computation**: Compute the image gradients (typically using techniques like Sobel or Scharr filters) to capture both edge and texture information.

$$g_m(x,y) = \sqrt{(g_x(x,y))^2 + (g_y(x,y))^2} \tag{1}$$

$$g_0(x,y) = arctan\left(\frac{g_y(x,y)}{g_x(x,y)}\right) \tag{2}$$

arctan is a tangent function that computes the angle (in radians) from the positive x-axis to the point (x, y) in the Cartesian plane.
where:
$g_x(x,y)$ = horizontal gradient at pixel (x, y) , $g_y(x,y)$ = vertical gradient at pixel (x, y), $g_m(x,y)$ = gradient magnitude at pixel (x,y), $g_0(x,y)$= gradient orientation (in radians)

- **LBP**: In order to represent local structures and micro-textures, rotation-invariant histograms are generated.

$$LBP_{P,R}(x,y) = \sum_{p=0}^{P-1} S(I_p - I_c) * 2^p \tag{3}$$

where:
$I_c = I(x,y)$ = intensity of the central pixel, I_p = intensity of the pth neighboring pixel located on a circle of radius R, p = total number of sampling points , R is the radius of the circle and $S(z)$ = thresholding fuction .

$$S(z) = \begin{cases} 1 & \text{if } z \geq 0, \\ 0 & \text{if } z < 0 \end{cases}$$

4.3 Deep Feature Extraction Using ResNet50

Important deep features are extracted from every OCT image using a pretrained ResNet50 model. ImageNet is a sizable image dataset that was previously used to train this model. After removing the last classification layers, we obtain a 2048-length feature vector for every picture using global average pooling. We initially transform the grayscale OCT pictures into a 3-channel RGB format before feeding them into the ResNet50 model. Traditional texture approaches are unable to adequately represent the general structure and patterns in the images, but these deep characteristics are supported.

4.4 Hybrid Feature Fusion

We employed a hybrid feature fusion technique to utilize both handcrafted and deep representations. For each OCT image, custom descriptors, including GLCM, HOG, and LBP, were extracted, resulting in a total of 117 features. Simultaneously, deep features were extracted from the ResNet50 network, yielding a 2048-dimensional feature vector from the global average pooling layer. The fusion technique involved concatenating handcrafted and deep features, producing a final fused vector of 2165 dimensions.
Before classification, z-score standardization was used to standardize all characteristics. This made sure that each feature contributed equally, without being affected by changes in scale. There were 12,000 OCT images in the dataset (6,000 DME and 6,000 Normal), and they were evenly spread out among classes. We used a 5-fold stratified cross-validation method, where 80% of the data was used for training and 20% for testing in each fold. This preserved the class balance.

Dimensionality reduction methods such as PCA were deliberately avoided, as the WideNet classifier is designed to work well with high-dimensional fused vectors. After that, the normalized fused features were sent to WideNet. We have the ability to utilize an attention-based weighting system to further enhance performance. This method assigns greater weights to the most beneficial traits and lower weights to those that are not as important. It helps the model focus on the features that are most helpful for distinguishing between healthy and DME images.

4.5 Classification Using WideNet

A specifically designed classifier, known as a WideNet, receives the fused feature vector, which comprises both deep and constructed features. WideNet employs a shallow yet wide architecture, in contrast to conventional deep learning models that utilize several stacked layers, which can be slow to train and prone to overfitting. It can successfully handle high-dimensional input since it has fewer layers but more neurons in each layer.
The structure of the WideNet includes:

- An input layer with the same size as the fused vector's feature count.
- To provide non-linearity and assist in the model's learning of complex structures, two hidden layers, one with 512 neurons and the other with 256 neurons, use the ReLU activation function.
- The estimated probabilities for each class are provided via a softmax output layer with two units (for binary classification: Healthy or DME).

We train the model with the categorical cross-entropy loss function, which is appropriate for multi-class (or binary) classification tasks. The Adam optimizer, a commonly used and effective algorithm for adjusting model weights during training, is employed for optimization.
Training is performed with a batch size of 32, and the training runs for up to 100 epochs. This shallow but wide design is chosen intentionally—it helps the model avoid overfitting on limited data, while still being powerful enough to learn from the rich, high-dimensional feature space generated by the fusion of handcrafted and deep features.

5 Result and Analysis

This section presents the experimental assessment of the Proposed Fusion-WideNet model with a retinal OCT image database. The task involved a binary classification problem, distinguishing between DME and Healthy cases based on a fused handcrafted texture feature(117 features) and deep semantic features.
To assess the performance of the classification system, we employed several standard evaluation metrics. Accuracy was used to measure the proportion of correctly classified samples out of the total. Precision was calculated to indicate the percentage of true positive predictions among all positive predictions made by the model, while Recall (or Sensitivity) measured the model's ability to identify all actual positive (DME) cases. To balance precision and recall, we computed the F1- score, which is their harmonic mean and provides a single metric that accounts for both false positives and false negatives. Lastly, we used the AUC as a measure of the model's ability to distinguish between the two classes. Applying the proposed Fusion-WideNet classifier to the fused features showed remarkable performance. The model demonstrated nearly perfect consistency in correctly identifying both healthy and DME cases, with a mean classification accuracy of 99.5% across all five folds of stratified cross-validation. Furthermore, the model demonstrated exceptional ability to differentiate between the two classes with an AUC (Area Under the Curve) of 0.9997. Additionally, both classes' precision, recall, and F1-score were consistently high, demonstrating the model's efficacy in detecting DME cases.
Its computing efficiency is one of the Wide Neural Network's main advantages. The training time was low and consistent across folds, with no variation in convergence time or performance, because of its shallow yet

Table 1: Fused feature performance of a classifier (5-fold CV average)

Classifier	**Accuracy(%)**	**Precision**	**Recall**	**F1-Score**	**AUC**	**Training Time(s)**
Fusion-WideNet	**99.5**	**0.994**	**0.994**	**0.994**	**0.9997**	**385**
Cubic SVM	93.8	0.938	0.940	0.938	0.9806	204
Fine Gaussian SVM	92.3	0.923	0.923	0.923	0.9767	28
Medium Neural Network	90.9	0.909	0.909	0.909	0.9714	304
Bilayered Neural Network	87.9	0.879	0.879	0.879	0.9519	347
Trilayered Neural Network	88.6	0.887	0.887	0.887	0.9515	368
Narrow Neural Network	86.9	0.869	0.869	0.869	0.9440	273
Quadratic SVM	84.9	0.849	0.849	0.849	0.9241	55
SVM (Linear Kernel)	84.1	0.841	0.841	0.841	0.9192	390
Logistic Regression	83.0	0.831	0.831	0.831	0.9073	378

broad design. This enables Fusion-WideNet to be both accurate and useful in large-scale or real-time screening settings. A comprehensive comparison of classifier performance is shown in Tables 1, 2, 3, and 4.

6 Discussion

The experimental results shown in Tables 1 through 4 clearly show how much superior the suggested Fusion-WideNet architecture is over both deep learning baselines and traditional machine learning models. AUC (0.9997), F1-score (0.994), recall (0.994), accuracy (99.5%), and precision (0.994) were the main evaluation criteria that Fusion-WideNet consistently outperformed when evaluated using 5-fold cross-validation. With an average training time of 385 seconds, the model also maintained effective computing efficiency in spite of its high-dimensional input space.

The fused features were utilized to evaluate several classifiers, as shown in Table 1. Models like the Fine Gaus-

Table 2: The proposed model outperformed all baseline approaches in every evaluation metric

Model	**Accuracy(%)**	**Precision**	**Recall**	**F1-Score**	**AUC**
SVM + GLCM	94.2	0.92	0.93	0.925	0.94
ResNet50 (only Deep Learning)	97.8	0.97	0.97	0.97	0.98
GLCM + HOG+ LBP (Traditional)	96.1	0.95	0.95	0.95	0.96
Fusion-WideNet (Proposed)	**99.5**	**0.995**	**0.995**	**0.995**	**0.997**

Table 3: Performance Comparison

Model	**Feature Type**	**Classifier**	**Accuracy(%)**	**F1-Score**	**AUC**
SVM + GLCM	Handcrafted	SVM	94.2	0.925	0.94
LBP + Random Forest	Handcrafted	Random Forest	93.5	0.915	0.93
ResNet50(Deep Only)	Deep	Dense Classifier	97.8	0.975	0.98
GLCM + CNN (Literature)	Handcrafted + Deep	Multilayer Perceptron	97.2	0.955	0.97
Fusion-WideNet (Proposed)	**Handcrafted + Deep + Attention**	**WideNet**	**99.5**	**0.995**	**0.997**

Table 4: Comparative Analysis of the proposed model with various authors' work

Author'Name	**Dataset**	**Methodology**	**Performance**
Hirota et al.[22]	Low-quality OCT	Ensemble CNN and Random Forest	Accuracy= 94%
A. Garcia et al.[23]	AMD OCT Images	SVM with handcrafted features	Accuracy= 92.12%
Wali et al.[24]	DME, CNV, and drusen OCT image	Random forest	Accuracy= 80%
Wali et al.[25]	CNV, DME, DRUSEN, and NORMAL	DenseNet121 + InceptionV3 Ensemble	Accuracy= 97.5%
Akbar et al.[27]	STARE	GLCM , SVM	Accuracy= 87.25% Sensitivity= 88.12%
Rajinikanth et al. [28]	OCT CNV images	SVM-Fine Gaussian	Accuracy= 92%.
Rasti et al.[29]	SD-OCT	Random forest	Precision= 99.33%
Mathew et al. [31]	Duke OCT	Lightweight DL model	Accuracy=94.9%
Aslam et al. [32]	OCT-A scans	Logistic regression	AUC=0.91
Phu et al. [33]	OCT image	SVM	Accuracy=95.8%
Khan et al. [34]	Kermany OCT dataset	Hybrid model ResNet50 and EfficientNetB0	Accuracy= 97.50%
Prabha et al. [35]	nAMD, DME, RVO, and Normal	RD-OCT Net	Accuracy= 98.08%
Agrawal et al. [36]	Kermany OCT dataset	HDL-ACO	Accuracy=93%
Dutta et al. [37]	CNV, DME, DRUSEN, and NORMAL	Conv-ViT model	Accuracy= 94%
Fusion-WideNet (Proposed)	**Normal And DME**	**GLCM+HOG+LBP + ResNet50 + WNN**	**Accuracy=99.5%**

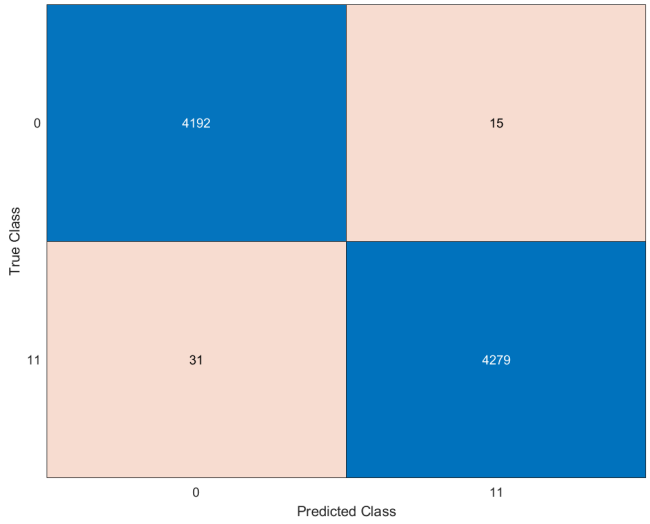

Figure 7: Confusion matrix of the proposed model

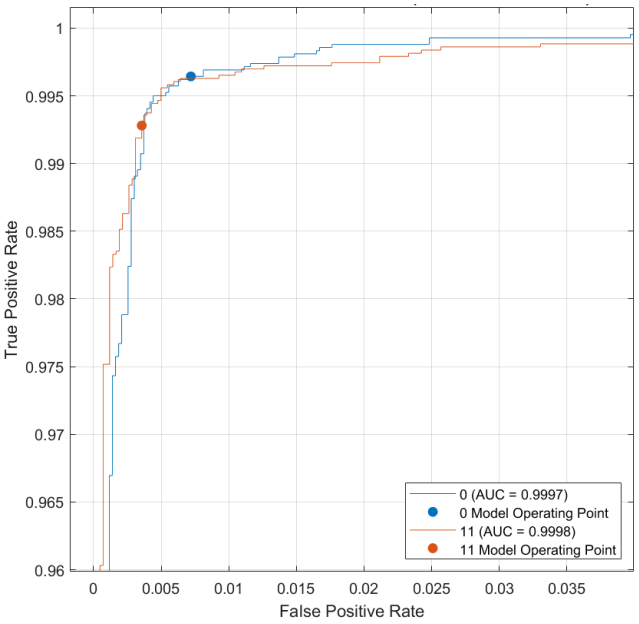

Figure 8: ROC curve of the proposed model

sian SVM and Cubic SVM performed quite well (accuracies of 92.3% and 93.8%, respectively), but Fusion-WideNet outperformed them in terms of F1-scores and AUC values. Likewise, a number of neural network configurations (such as narrow, medium, bilayered, and trilayered architectures) shown deteriorating performance with increasing depth, frequently accompanied by longer training periods and a bias toward overfitting. This demonstrates that the wide but shallow network design was the best option for managing fused feature vectors without compromising generalization.

Fusion-WideNet is compared with alternative feature techniques in Table 2. The suggested hybrid model outperformed conventional texture-based models like GLCM+SVM (accuracy = 94.2%) and GLCM+HOG+LBP (accuracy = 96.1%), which highlighted the importance of handmade features. The advantages of combining low-level textural characteristics with high-level semantic features was shown by the fact that even a deep learning-only baseline utilizing ResNet50 (accuracy = 97.8%) was unable to equal the performance of the fused feature set.

Table 3 illustrates how Fusion-WideNet's Wide Neural Network (WNN) performed better than alternative classifiers, including Multilayer Perceptrons (MLP), Random Forests, and conventional SVMs. The best F1-score (0.995) and AUC (0.997) were obtained by combining attention mechanisms and WNN with a hybrid fusion of handmade and deep features. This demonstrated that expanding the network structure facilitates efficient learning from a vast and varied feature space.

Additionally, Table 4 compares the suggested model with previous studies in the field. Depending on the dataset and methodology, models by Hirota et al., Wali et al., Akbar et al., and others generally obtained accuracies between 80% and 97.5%. Ensemble models and handcrafted + SVM combinations, such as random forests or GLCM with SVM, for instance, performed comparatively poorly. On the other hand, Fusion-WideNet outperformed all previously demonstrated techniques with an accuracy of 99.5% when tested on both normal and DME images. This demonstrates that combining deep ResNet50 features with multi-scale handmade descriptors and then optimizing broad learning produces state-of-the-art results.

7 Conclusion and Future Scope

In this study, we present Fusion-WideNet, a novel hybrid framework for the accurate and interpretable classification of retinal OCT images, with a particular focus on DME and healthy images. The model leverages the complementary advantages of high-level deep features extracted from a pre-trained ResNet50 network and handcrafted texture features (GLCM, HOG, and LBP). After being fused into a high-dimensional representation, these features are run through a WNN, which is made especially to handle fused feature spaces effectively. The proposed Fusion-WideNet framework outperformed traditional handcrafted-only and deep learning-only models, achieving a classification accuracy of 99.5%. The method worked very well by combining deep semantic characteristics and handcrafted textural data into a single high-dimensional representation, then categorizing them using a Wide Neural Network. Our study of comparisons demonstrated Fusion-WideNet's robust performance and generalizability across cross-validation folds, confirming its dominance in terms of accuracy, AUC, and F1-score. The framework may also be utilized in automated retinal disease screening systems, as it is computationally efficient and doesn't require manual annotation or segmentation.
While the proposed Fusion-WideNet structure shows impressive classification performance and flexibility, several limitations must be recognized. The work depends on only one publicly accessible OCT dataset, potentially constraining the model's applicability to images obtained from other devices or clinical environments. Secondly, while feature fusion improves discriminative capability, a combination of handcrafted and deep features yields a high-dimensional representation, potentially elevating processing requirements for extensive deployments. The model now concentrates on four disease categories, whereas clinical practice necessitates the consideration of a broader spectrum of retinal disorders. In future endeavors, we plan to enhance the framework by integrating multi-center datasets to augment generalizability, employing dimensionality reduction or feature selection techniques to optimize computational efficiency, and investigating transformer-based fusion mechanisms to capture intricate feature interactions. The use of longitudinal OCT data may yield significant insights for monitoring disease development and tailoring treatment strategies.

Declaration

Author contribution: All the authors have contributed equally to this article.

Data availability: The Dataset is publicly available from the Kaggle OCT dataset.

Funding statement: This research received no external funding

Conflict of interest: There is no conflict of interest for any of the authors.

Consent to participate: Each participating author has consented to be included in this submitted work.

References

[1] N. Eladawi, M. Elmogy and M. Ghazal, "Optical coherence tomography: A review", in *Diabetes and Fundus OCT*, pp. 191–221, 2020. doi: https://doi.org/10.1016/B978-0-12-817440-1.00007-3

[2] K. Elmasry, A.S. Ibrahim, S. Abdulmoneim and M. Al-Shabrawey, "Bioactive lipids and pathological retinal angiogenesis", *British Journal of Pharmacology*, vol. 176, no. 1, pp. 93–109, 2019. doi: https://doi.org/10.1111/bph.14507

[3] B. Potsaid, I. Gorczynska, V.J. Srinivasan, Y. Chen, J. Jiang, A. Cable and J.G. Fujimoto, "Ultrahigh speed spectral/Fourier domain OCT ophthalmic imaging at 70,000 to 312,500 axial scans per second", *Optics Express*, vol. 16, no. 19, pp. 15149–15169, 2008. doi: https://doi.org/10.1364/OE.16.015149

[4] M.E. Velthoven, D.J. Faber, F.D. Verbraak, T.G. Leeuwen and M.D. Smet, "Recent developments in optical coherence tomography for imaging the retina", *Progress in Retinal and Eye Research*, vol. 26, no. 1, pp. 57–77, 2007. doi: https://doi.org/10.1016/j.preteyeres.2006.10.002

[5] S.A. Boppart, G.J. Tearney, B.E. Bouma, J.F. Southern, M.E. Brezinski and J.G. Fujimoto, "Noninvasive assessment of the developing Xenopus cardiovascular system using optical coherence tomography", *SPIE Milestone Series*, vol. 165, pp. 614–619, 2001. doi: https://https://doi.org/10.1073/pnas.94.9.4256

[6] M.J. Suter, S.K. Nadkarni, G. Weisz, A. Tanaka, F.A. Jaffer, B.E. Bouma and G.J. Tearney, "Intravascular optical imaging technology for investigating the coronary artery", *JACC: Cardiovascular Imaging*, vol. 4, no. 9, pp. 1022–1039, 2011. doi: https://doi.org/10.1016/j.jcmg.2011.03.020

[7] L. Chen, Y. Xie, X. Du, W. Qin, L. Huang, J. Dai, K. Qin and J. Huang, "The effect of different doses of ciprofol in patients with painless gastrointestinal endoscopy", *Drug Design, Development and Therapy*, vol. 17, pp. 1733–1740, 2023. doi: https://doi.org/10.2147/DDDT.S414166

[8] G. Santacroce, I. Zammarchi, C.K. Tan, G. Coppola, R. Varley, S. Ghosh and M. Iacucci, "Present and future of endoscopy precision for inflammatory bowel disease", *Digestive Endoscopy*, vol. 36, no. 3, pp. 292–304, 2024. doi: https://doi.org/10.1111/den.14672

[9] E. Farina, J.J. Nabhen, M.I. Dacoregio, F. Batalini and F.Y. Moraes, "An overview of artificial intelligence in oncology", *Future Science OA*, vol. 8, no. 4, p. FSO787, 2022. doi: https://doi.org/10.2144/fsoa-2021-0074

[10] C. Jubelin, J. Muñoz-Garcia, L. Griscom, D. Cochonneau, E. Ollivier, M.-F. Heymann, F.M. Vallette, L. Oliver and D. Heymann, "Three-dimensional in vitro culture models in oncology research", *Cell & Bioscience*, vol. 12, no. 1, p. 155, 2022.

[11] K. Liopyris, S. Gregoriou, J. Dias and A.J. Stratigos, "Artificial intelligence in dermatology: challenges and perspectives", *Dermatology and Therapy*, vol. 12, no. 12, pp. 2637–2651, 2022. doi:

[12] Y. Ming, S. Hao, F. Wang, Y.R. Lewis-Israeli, B.D. Volmert, Z. Xu, A. Goestenkors, A. Aguirre and C. Zhou, "Longitudinal morphological and functional characterization of human heart organoids using optical coherence tomography", *Biosensors and Bioelectronics*, vol. 207, p. 114136, 2022. doi: https://doi.org/10.1016/j.bios.2022.114136

[13] H. Zhang, H. Sun, J. Li, Z. Lv, Y. Tian, and X. Lei, "Gene expression is associated with brain function of insomnia disorder, rather than brain structure", *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 136, p. 111209, 2025. doi: doi.org/10.1016/j.pnpbp.2024.111209

[14] K.M. Eberhardt, M. Treitl, K. Boesenecker, D. Maxien, M. Reiser and J. Rieger, "Prospective evaluation of optical coherence tomography in lower limb arteries compared with intravascular ultrasound", *Journal of Vascular and Interventional Radiology*, vol. 24, no. 10, pp. 1499–1508, 2013. https://doi.org/10.1016/j.jvir.2013.06.015

[15] S. Wang and I.V. Larina, "In vivo imaging of the mouse reproductive organs, embryo transfer, and oviduct cilia dynamics using optical coherence tomography", in *Mouse Embryogenesis: Methods and Protocols*, pp. 53–62, 2018. https://doi.org/10.1007/978-1-4939-7714-7_5

[16] A. Kilic, M.A. Mathier, G.W. Hickey, I. Sultan, V.O. Morell, S.R. Mulukutla and M.E. Keebler, "Evolving trends in adult heart transplant with the 2018 heart allocation policy change", *JAMA Cardiology*, vol. 6, no. 2, pp. 159–167, 2021. doi:10.1001/jamacardio.2020.4909

[17] A.P. Farrell and D.R. Jones, "The heart", in *Fish Physiology*, vol. 40, pp. 91–172, Elsevier, 2023.

[18] I. Dinstein, K. Shanmugam and R. Haralick, "Textural features for image classification", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973. https://doi.org/10.1109/TSMC.1973.4309314

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886–893, 2005. https://doi.org/10.1109/CVPR.2005.177

[20] Z. Guo, L. Zhang and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching", *Pattern Recognition*, vol. 43, no. 3, pp. 706–719, 2010. https://doi.org/10.1016/j.patcog.2009.08.017

[21] D.S. Kermany, M. Goldbaum, W. Cai, C.C. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning", *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018. https://doi.org/10.1016/j.cell.2018.02.010

[22] M. Hirota, S. Ueno, T. Inooka, Y. Ito, H. Takeyama, Y. Inoue, E. Watanabe and A. Mizota, "Automatic screening of the eyes in a deep-learning–based ensemble model using actual eye checkup optical coherence tomography images", *Applied Sciences*, vol. 12, no. 14, p. 6872, 2022. https://doi.org/10.3390/app12146872

[23] A. García-Floriano, A. Ferreira-Santiago, O. Camacho-Nieto and C. Yáñez-Márquez, "A machine learning approach to medical image classification: Detecting age-related macular degeneration in fundus images", *Computers & Electrical Engineering*, vol. 75, pp. 218–229, 2019. https://doi.org/10.1016/j.compeleceng.2017.11.008

[24] A. Wali, Z. Suhail and A. Sipani, "Optical coherence tomography (OCT) image classification for retinal disease using a random forest classifier",2024. https://doi.org/10.21203/rs.3.rs-4551679/v1

[25] A. Wali, Z. Suhail, S. Naz and I. Younas, "An ensemble deep learning model for OCT image detection and classification",2024. https://doi.org/10.21203/rs.3.rs-4923941/v1

[26] K. Bhatia, S. Arora and R. Tomar, "Diagnosis of diabetic retinopathy using machine learning classification algorithm", in *Proceedings of the 2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*,pp. 347–351, 2016. https://doi.org/10.1109/NGCT.2016.7877439

[27] S. Akbar, M.U. Akram, M. Sharif, A. Tariq and U.U. Yasin, "Decision support system for detection of papilledema through fundus retinal images", *Journal of Medical Systems*, vol. 41, pp. 1–16, 2017. https://doi.org/10.1007/s10916-017-0712-9

[28] V. Rajinikanth, S. Kadry, R. Damaševičius, D. Taniar and H.T. Rauf, "Machine learning scheme to detect choroidal neovascularization in retinal OCT image", in *Proceedings of the 2021 Seventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII)*, pp. 1–5, 2021. https://doi.org/10.1109/ICBSII51839.2021.9445134

[29] R. Rasti, A. Mehridehnavi, H. Rabbani and F. Hajizadeh, "Automatic diagnosis of abnormal macula in retinal optical coherence tomography images using wavelet-based convolutional neural network features and random forests classifier", *Journal of Biomedical Optics*, vol. 23, no. 3, p. 035005, 2018. https://doi.org/10.1117/1.JBO.23.3.035005

[30] G. Kalra, H. Cetin, J. Whitney, S. Yordi, Y. Cakir, C. McConville, V. Whitmore, M. Bonnay, L. Lunasco, A. Sassine et al., "Machine learning-based automated detection and quantification of geographic atrophy and hypertransmission defects using spectral domain optical coherence tomography", *Journal of Personalized Medicine*, vol. 13, no. 1, p. 37, 2022. https://doi.org/10.3390/jpm13010037

[31] M.R. Mathews and S.T.M. Anzar, "A lightweight deep learning model for retinal optical coherence tomography image classification", *International Journal of Imaging Systems and Technology*, vol. 33, no. 1, pp. 204–216, 2023. https://doi.org/10.1002/ima.22808

[32] T.M. Aslam, D.C. Hoyle, V. Puri and G. Bento, "Differentiation of diabetic status using statistical and machine learning techniques on optical coherence tomography angiography images", *Translational Vision Science & Technology*, vol. 9, no. 4, p. 2, 2020. https://doi.org/10.1167/tvst.9.4.2

[33] D.N. Phu, A.Q. Tran, N.Q. Tran and T.H. Dang, "The classification of optical coherence tomography images using machine learning for macular hole and diabetic macular edema diagnoses", in *Proceedings of the 2023 12th International Conference on Control, Automation and Information Sciences (ICCAIS)*,pp. 634–638, 2023. https://doi.org/10.1109/ICCAIS59597.2023.10382327

[34] U.S. Khan and S.U.R. Khan, "Boost diagnostic performance in retinal disease classification utilizing deep ensemble classifiers based on OCT", *Multimedia Tools and Applications*, vol. 84, no. 19, pp. 21227–21247, 2025.https://doi.org/10.1007/s11042-024-19922-1

[35] A.J. Prabha, C. Venkatesan, M.S. Fathimal, K. Nithiyanantham and S.A. Kirubha, "RD-OCT Net: hybrid learning system for automated diagnosis of macular diseases from OCT retinal images", *Biomedical Physics & Engineering Express*, vol. 10, no. 2, p. 025033, 2024.https://doi.org/10.1088/2057-1976/ad27ea

[36] S. Agarwal, A.K. Dohare, P. Saxena, J. Singh, I. Singh and U.K. Sahu, "HDL-ACO hybrid deep learning and ant colony optimization for ocular optical coherence tomography image classification", *Scientific Reports*, vol. 15, no. 1, p. 5888, 2025.https://doi.org/10.1038/s41598-025-89961-7

[37] P. Dutta, K.A. Sathi, M.A. Hossain and M.A.A. Dewan, "Conv-ViT: a convolution and vision transformer-based hybrid feature extraction method for retinal disease detection", *Journal of Imaging*, vol. 9, no. 7, p. 140, 2023. https://doi.org/10.3390/jimaging9070140

[38] R.T. Yanagihara, C.S. Lee, D.S.W. Ting and A.Y. Lee, "Methodological challenges of deep learning in optical coherence tomography for retinal diseases: a review", *Translational Vision Science & Technology*, vol. 9, no. 2, p. 11, 2020.https://doi.org/10.1167/tvst.9.2.11

[39] J.F. Arevalo, A.F. Lasave, J.D. Arias, M.A. Serrano and F.A. Arevalo, "Clinical applications of optical coherence tomography in the posterior pole: the 2011 José Manuel Espino lecture–part II", *Clinical Ophthalmology*, pp. 2181–2206, 2013.https://doi.org/10.2147/OPTH.S51158

[40] V.W. Hui, S.K. Szeto, F. Tang, D. Yang, H. Chen, T.Y. Lai, A. Rong, S. Zhang, P. Zhao, P. Ruamviboonsuk et al., "Optical coherence tomography classification systems for diabetic macular edema and their associations with visual outcome and treatment responses–an updated review", *Asia-Pacific Journal of Ophthalmology*, vol. 11, no. 3, pp. 247–257, 2022.https://doi.org/10.1097/APO.0000000000000468

[41] E.J. Kim, W.V. Lin, S.M. Rodriguez, A. Chen, A. Loya and C.Y. Weng, "Treatment of diabetic macular edema", *Current Diabetes Reports*, vol. 19, pp. 1–10, 2019. https://doi.org/10.1007/s11892-019-1188-4

[42] F. De Bats, B. Wolff, M. Mauget-Faÿsse, I. Meunier, P. Denis and L. Kodjikian, "Association of reticular pseudodrusen and early onset drusen", *International Scholarly Research Notices*, vol. 2013, no. 1, p. 273085, 2013.https://doi.org/10.1155/2013/273085

[43] P. Nagaraj, V. Muneeswaran, B. Sunethra, C. Sreeya, U. Dhannushree and S. Nithisaa, "A comparative analysis of retinal disease image classification for OCT using deep learning techniques", in *Proceedings of the 2023 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–10, 2023. https://doi.org/10.1109/ICCCI56745.2023.10128486