



An Optimized DenseNet201–SVM Hybrid Framework for Retinal Disease Classification in Optical Coherence Tomography Image

Shreemat Kumar Dash¹, M.V. Subbaraob², Sudarson Jena¹, Prabira Kumar Sethy^{3*}, Santi Kumari Behera⁴, Aziz Nanthaamornphong⁵

¹ Department of Computer Science Engineering and Applications, SUIIT, Sambalpur University, Burla 768019, India

² Department of CSE, Malla Reddy Engineering College for Women, Secunderabad 500100, India

³ Department of Electronics Engineering, Sambalpur University, Burla 768019, India

⁴ Department of CSE, VSSUT Burla, Sambalpur 768018, India

⁵ College of Computing, Prince of Songkla University, Phuket 83000, Thailand

Corresponding Author Email: prabirasethy@suniv.ac.in

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310209>

ABSTRACT

Received: 26 October 2025

Revised: 5 January 2026

Accepted: 15 February 2026

Available online: 28 February 2026

Keywords:

retinal damage classification, optical coherence tomography images, DenseNet201, support vector machine, Bayesian optimization

Accurate and reliable classification of retinal diseases from optical coherence tomography (OCT) images is critical for early diagnosis and clinical decision-making. This study proposes an Optimized Hybrid Deep Learning Framework that integrates DenseNet201-based feature extraction with a support vector machine (SVM) classifier, enhanced through Bayesian hyperparameter optimization. To ensure an objective model selection process, eighteen transfer learning-based convolutional neural network (CNN) backbones are systematically evaluated and ranked using Duncan's multiple range test, with DenseNet201 identified as the most effective feature extractor. The proposed framework replaces the conventional softmax layer with an SVM classifier to improve decision boundary robustness and classification performance. Experiments conducted on a large-scale OCT dataset comprising 35,168 images across four classes demonstrate that the optimized DenseNet201–SVM model achieves superior performance, with an accuracy of 99.69%, sensitivity of 99.54%, and precision of 99.43%. Additional evaluations on multiple external datasets further confirm the generalizability of the proposed approach. The results indicate that combining deep feature representations with classical machine learning classifiers, along with principled hyperparameter optimization, provides a robust and scalable solution for automated retinal disease classification.

1. INTRODUCTION

Medical image classification and analysis using deep learning (DL) for disease detection and diagnosis have been a huge breakthrough in many aspects of the medical imaging world. The existing methods for the detection of diseases are in general characterized by less reliability, greater error proneness, and generally take a longer time to arrive at any conclusions. However, with recent developments in the field of machine learning, a new and better technique to help with the efficient and effective identification of various diseases is introduced, with a drastic improvement in accuracy. One such problem where this improvement has made a huge change is the early detection of vision loss due to retinal damage.

The retina is a thin, light-sensitive tissue layer present at the back segment of the eye which, if damaged or affected due to any other diseases can lead to significant impairment of vision. Thus, it is important to detect any abnormalities in the retina at the earliest, to avoid any permanent loss of vision. In this paper, we study and report the application of computer-assisted, automated retinal disease detection methods, to detect diseases such as Choroidal neovascularization (CNV),

diabetic macular edema (DME), drusen and normal conditions.

AMD majorly affects people above the age of 50 years and is the leading cause of loss of central vision, as well as the major cause of irreversible blindness across the world [1]. It is a disease that affects the central part of the retina and broadly has two types namely dry AMD and wet AMD. In the dry form of AMD, Drusen are deposits of lipids and proteins that promote inflammation, which is seen as a build-up of waste material between the retinal pigment epithelium (RPE) and the Bruch's membrane, resulting in the RPE elevations [2]. Wet-type AMD shows the presence of CNV which is a condition where abnormal blood vessels grow from the choroid into the sub-retinal pigment epithelium (RPE) space through a break in the Bruch's membrane [3]. DME is caused by neurovascular degeneration in people with diabetes and is the leading cause of vision loss among 20–79 years old people across the world [4]. Therefore, in all such conditions, an early diagnosis and treatment is important [5].

Optical coherence tomography (OCT) is an imaging technique that provides high-resolution, in-vivo, cross-sectional, tomographic imaging of the retina using the principle of light coherence. It is a non-invasive biomedical

imaging technology that has become an essential tool for the early diagnosis and treatment monitoring of retinal diseases such as AMD and DME [6]. It is particularly used in the diagnosis of eye-related issues like AMD, diabetic retinopathy, glaucoma and others. There have been many works where a classification between CNV, DME, drusen and normal retinal conditions have been used, this is a relatively difficult task to be performed as the images that are used vary from the OCT systems used to produce them, features of both conditions can overlap, and the availability of large annotated OCT databases for training and generalization of deep learning models is scarce [7].

Transfer learning is a type of deep learning method where pre-trained neural networks are used. This paper also describes and leverages on some of the benefits of the transfer learning approach to improve the time and performance efficiency for medical image analysis. The use of DL in combination with transfer learning for retinal disease detection can facilitate quicker and more accurate medical diagnoses.

While many deep-learning based works train OCT classifiers, three limitations exist in the literature: (1) few works train many modern CNN backbones to perform a broad/statistically-rigorous evaluation in order to objectively select an ideal feature extractor for OCT; (2) few works replace their final softmax-based classifier layer with a classical classifier (i.e. SVM) and optimize the hyperparameters of said classical classifier layer to improve prediction robustness and visualize learned multiclass decision boundaries; and (3) little work has been done to analyze cross-dataset generalization, along with lightweight/deployment feasibilities. Here we aim to fill each of these voids by (i) testing 18 transfer-learning CNNs and selecting an ideal backbone using Duncan's multirange test, (ii) replacing DenseNet201's softmax layer with an SVM classifier and utilizing Bayesian hyperparameter optimization to optimize SVM hyperparameters, and (iii) cross-validating said optimized hybrid model on public OCT datasets.

Automated retinal disease classification using OCT images as described in this work is a classical machine learning classification problem. The work that this paper does in terms of automating the process is an efficient and effective implementation of a machine learning algorithm to do a four-way retinal classification for retinal diseases such as CNV, drusen, DME and normal retinal conditions.

With the increasing usage and deployment of OCT machines in the medical domain, there are huge improvements to be seen in automatic detection, treatment and diagnosis of retinal issues, and computer assisted diagnosis system can play a huge role in this regard [8-11]. The conventional approach is to use a lot of image pre-processing and then to perform the analysis with a shallow neural network which is a very time-consuming process. Transfer learning approaches have been used to counter this problem, and an extensive range of CNN architectures are explored, compared and the most optimal architecture has been chosen based on a wide range of metrics, and in this case, DenseNet201 has been implemented.

The main contributions of the proposed approach include:

- A hybrid deep learning framework (DenseNet201 in combination with SVM) to classify retinal damage and impairment from OCT images of the eyes.
- Improving the classification performance by removing the softmax layer from the DenseNet201 architecture and combining it with an SVM classifier, to achieve an increased accuracy and precision.

- The application of Bayesian optimization for automatic hyperparameter tuning which has also been used in our work and is the key to the significant performance improvement of our hybrid model.
- The best optimized model with the above-mentioned DenseNet201+SVM model achieved state-of-the-art performance, with 99.6% accuracy, 99.54% sensitivity and 99.43% precision, and can also be used for the early diagnosis and treatment planning in case of retinal disease classification.

The paper is organized as follows: Section 2 gives an overview of past work related to the paper. Section 3 includes the description of the new research and databases used. Section 4 provides a description of the experimental results. Section 5 briefly discusses the proposed work, and Section 6 provides a conclusion and future work.

2. LITERARY REVIEW

2.1 Traditional machine-learning approaches

The first works to classify OCT imagery relied on handcrafted features paired with classical classifiers. These analyses validated the feasibility of OCT image classification, but required manual design of feature extractors and generally suffered from limited scalability. Alsaih et al. [8] trained a linear SVM to perform a binary classification task of healthy versus DME retinas using manually designed features from retinal images. Similarly, Hussain et al. [12] trained a random forest classifier using engineered SD-OCT features to differentiate between healthy retina scans versus diseased scans, reporting >94% accuracy and 0.99 Area Under the Curve (AUC) on select tasks. Another handcrafted feature descriptor uses semivariograms to perform automated diagnosis of AMD from SD-OCT topographic maps and pairs the descriptors with an SVM classifier [13]. These methods are helpful for data-constrained tasks, but ultimately performance hinges on handcrafted features and decisions related to feature-selection.

2.2 Deep-learning and transfer-learning approaches

Following the introduction of deep convolutional neural networks (CNNs), researchers began to apply techniques from transfer learning to automate feature extraction, typically removing the need for engineered features while often improving accuracy. A large body of work adopted pretrained CNN backbones (VGG, ResNet, Inception, DenseNet, MobileNet, etc.) trained on public image benchmarks and fine-tuned them on OCT datasets, regularly reporting high accuracy on task (note that some of these works were limited by the use of small or single datasets and lack consistent validation procedures). Some early works applying CNNs include Tan et al. [14] which used a deep CNN for automated AMD detection and achieved mean testing accuracy between ~91–95% depending on cross-validation scheme. Two works leveraged VGG [15-20] and ResNet-based transfer learning [17]. Li et al. [16] reported 98.6% accuracy using a VGG-16 based transfer-learning approach. Li et al. [17] used four enhanced ResNet50 ensembles and reported 97.3% accuracy and AUC of 0.995. Multi-scale CNNs and fusion networks like DMF-CNN by Das et al. [18] achieved accuracies of 96.03% and 99.60% on the UCSD and NEH datasets, respectively.

DenseNet and other modern architectures have also been evaluated by researchers such as Akinniyi et al. [19], who used a pyramid ensemble of DenseNets (trained from scratch rather than using transfer-learning) with a DenseNet backbone and achieved accuracies of 97.78% and 99.69% on UCSD and another dataset respectively. Using pretrained VGG-19 weights as a transfer-learning method achieved average accuracies of ~99.17% on publicly available large OCT datasets by Choudhury et al. [21] and Choudhary et al. [29]. The recent introduction of transformer models and hybrids has yielded models such as SViT by Hemalakshmi et al. [28], which proposed a hybrid SqueezeNet–ViT model which achieved 99.90% accuracy on OCT2017. Despite high performance in many deep learning works, comparisons between the use of many backbones on a single dataset in a controlled evaluation are rare. Cross-dataset evaluation is also often not performed.

2.3 Hybrid methods and optimization-based methods

Hybrid methods that pair deep feature extractors with classical machine learning classifiers or performing hyperparameter optimization have been used to improve the robustness, interpretability, or efficiency of CNNs. Tuncer et al. used a hybrid CNN+SVM approach as a detection pipeline for CNV/DME/drusen achieving overall accuracies of 98.96% [23]. Subramanian et al. used Bayesian optimization to perform hyperparameter tuning to improve the performance of deep networks for automated diagnosis of retinal diseases [27]. Other studies have focused on applying modules or ensembles specialized for OCT classification tasks and achieving strong AUC/accuracy on studied datasets [15, 26]. Methods that perform comparisons and ensembles of deep learning models, such as MCME and DL-Net by Rasti et al. [15] and Nagamani and Rayachoti [22] respectively have been shown to reach accuracies of ~99.6-99.67% on studied datasets. Hybrid models can provide improvement over individually studied backbones by smoothing decision boundaries and limiting certain modes of error. Systematic approaches to ranking many backbones via statistical testing, then training a targeted hybrid model (typically deep feature extractor + classical classifier) with principled hyperparameter tuning is relatively unexplored in the OCT image classification literature.

2.4 Summary and gap

In summary: classical machine-learning approaches demonstrated feasibility of OCT classification, but rely on handcrafted features [8, 12, 13]; transfer learning with deep CNNs showed significant improvement and are the majority of recent approaches, but direct comparison of many backbones on the same task and cross-dataset validation is limited [14-19, 21, 28, 29]; and hybrid model approaches and optimization-based hyperparameter tuning show promise to systematically improve robustness and model decision boundaries, but have not been applied in a comparative way to many CNN backbones on OCT datasets followed by statistical testing on collected results [15, 22, 23, 26, 27]. Together, these points show there are three specific gaps our work aims to address: systematic comparison of many CNN backbones on OCT classification (this work tests 18 models and uses Duncan’s multirange test to statistically rank them), replacing the softmax classifier with classical classifiers + optimization (we test SVM + Bayesian hyperparameter optimization), and

testing on multiple public OCT datasets to validate cross-dataset accuracy and include discussion of time and memory constraints for deployment.

3. MATERIALS AND METHODOLOGY

3.1 About dataset

For our experiments, we used an open-access dataset of labeled OCT images that is accessible to the public. The original dataset consists of 35,168 OCT images that are divided into four groups: normal, drusen, DME, and CNV. The distribution of images among these classifications is unbalanced, nevertheless. In order to ensure the predictive model’s efficacy, the retinal illness classification problem must be approached with roughly equal data sizes for each class label. 35,168 images in total were used in this study; 28,264 of those images were used for training throughout the four classes, and 6,904 images were used for assessment. Table 1 shows how the original dataset and Figure 1 shows the Sample of OCT Images (a) CNV (b) DME (c) DRUSEN (d) Normal.

Table 1. Distribution of OCT images

OCT Images	No. of Images for Training and Validation	No. of Images for Testing
CNV	6902	1726
DME	7106	1726
DRUSEN	7132	1726
NORMAL	7124	1726
	Subtotal =28,264	Subtotal=6904
		Total=35,168

Note: OCT = optical coherence tomography; CNV = Choroidal neovascularization; DME = diabetic macular edema

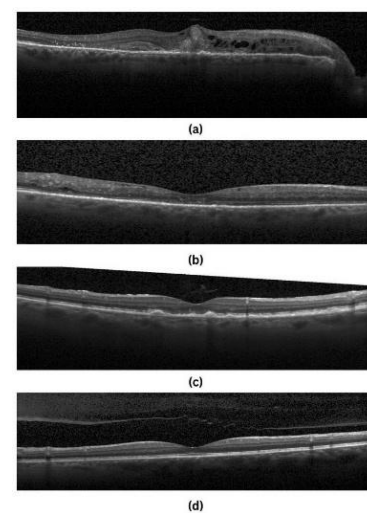


Figure 1. Sample of OCT Images (a) CNV (b) DME (c) DRUSEN (d) Normal

Note: OCT = optical coherence tomography; CNV = Choroidal neovascularization; DME = diabetic macular edema. [source: <https://www.kaggle.com/datasets/paultimothymooney/germany2018>]

The original 35,168 images were partitioned into a held-out test set and a training (development) pool. From the original dataset, 6,904 images (1726 images per class) were reserved for testing and 28,264 images were used for model development. We used the 28,264 development images as follows: For training/fine-tuning CNNs we held out 10% of

the development pool as an internal validation set (this was created by stratified sampling) to monitor training and prevent overfitting. We used this internal validation set for tuning hyperparameters and making early- stopping decisions for the CNN baseline. For Bayesian optimization of SVM hyperparameters used with DenseNet201 features, BOmin evaluated candidate SVM hyperparameters using stratified 5-fold cross-validation within the development pool (i.e., the 28,264 images were split into five folds repeatedly to estimate validation error under each hyperparameter configuration). After selecting the final hyperparameters (using Bayesian optimization) we trained the final classifier on the entire development pool (28,264 images) and reported results on the held-out test set of 6,904 images. All foldings and stratified splits preserved class balance.

3.2 Methodology

Using the public OCT image dataset, a novel Optimized Hybrid Deep Learning Framework is suggested for the multiple classification of retinal disorders. By performing an image pre-processing procedure on the original dataset images, the input dimension of the suggested model is made consistent. For this use, the resolution of the input image is 224×224 . To avoid the over-fitting issue, the OCT image dataset is subjected to image normalization. The proposed method for identifying retinal illness on OCT images consists of four main stages: Preprocessing, Classification, model evaluation via Statistical Analysis, classification and Optimization. Figure 2 shows the system flow diagram for the newly suggested innovative Optimized Hybrid Deep Learning Framework for various retinal disease classifications.

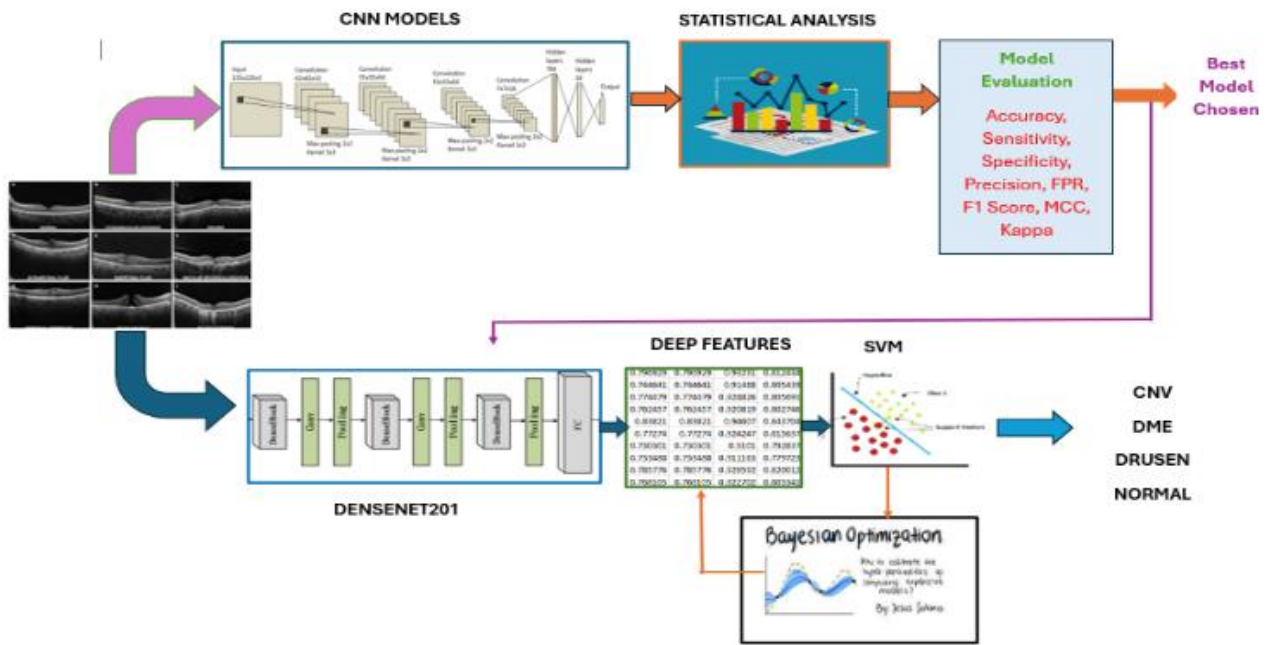


Figure 2. Optimized hybrid deep learning model for oct image classification

3.2.1 Data pre-processing

The retinal layer, retinal structural edge, and image features are improved by a number of image processing approaches. Applying the image enhancement technique reduces background noise. The OCT images are blurry and have low contrast because of speckle noise. Blinking and eye movements are the primary sources of speckle noise during the capturing of images. Random diffuse scattering, pixel value distortion, and camera noise from competing ultrasonic pulses are other perpetrators. In this study, the speckle noise is decreased via median filtering. A median filter provides optimal speed performance and high-performance results during the reduction process. The medium filter determines if a pixel displays continuity with its neighbors by taking into account the values of each neighboring pixel. The noise pixel's value is updated using the medium of the surrounding pixel value.

3.2.2 Model evaluation

We applied all 18 of our CNN models in the Transfer Learning paradigm, where the classification layer (softmax)

remained the same for all models, meaning the only difference between all 18 backbones was the feature extraction. Since DenseNet201 outperformed all other models under this paradigm, we can say that it is the best end-to-end model as well as the best feature extractor. In order to check that conclusion in an independent and objective way, we replaced the softmax classifier with SVM to check the performance of the features. The DenseNet201, still, outperformed all other models in this hybrid model-checking as well, which proves that the former conclusion is true (i.e., Densenet201 is the best feature extractor). Also, since we used SVM in this stage we used Bayesian optimization for further improvement in the classification through parameter tuning of the SVM. Furthermore, by ranking the mean performance scores, we objectively and statistically chose the top CNN model, with less bias to random chance in performance score fluctuations.

3.2.3 DenseNet201 convolutional neural network model

DenseNet201 is a CNN model that can be used for a wide range of computer vision tasks, including medical image analysis. DenseNet201 is part of a family of models called

DenseNets, which are known for their parameter efficiency and feature reuse capabilities. DenseNet201 is a deep model that consists of 201 layers. It uses a dense connectivity pattern where each layer receives feature maps from all its preceding layers as input and passes its feature maps to all the subsequent layers. This connectivity pattern encourages feature reuse throughout the network and makes it easier to train very deep networks by mitigating the vanishing gradient problem. The deep, dense architecture of DenseNet201 allows it to learn rich, hierarchical features from the high-resolution input images. This makes it suitable for processing complex image data such as OCT scans. The model's parameter-efficient architecture, resulting from its dense connectivity, means that the network has fewer parameters than a typical deep CNN of similar depth. This helps to reduce overfitting, especially when training data is limited, which is often the case in medical imaging applications. DenseNet201 is capable of learning a diverse set of features at multiple scales, from low-level textures to high-level semantic information. This multi-scale feature learning can be beneficial for detecting subtle differences in retinal status, such as CNV, drusen, DME, or a normal retina. Transfer learning using DenseNet201 and pre-trained weights on large-scale datasets like ImageNet will result in faster convergence and better performance on the target task after fine-tuning. The architectural benefits of DenseNet201 also help it to generalize better across different datasets, reducing the risk of overfitting and making it more applicable to real-world clinical scenarios.

3.2.4 Classification

We use pre-trained DenseNet201 to extract high-level feature representation from the last layer. This feature representation preserves the fine details and captures complex hierarchical patterns, which is important for medical image analysis. DenseNet201 with 201 densely connected layers has the ability to learn discriminative features to represent the variations among the retinal tissues. The feature extracted from DenseNet201 will be used as an input to the SVM. The SVM can recognize different types of retinal tissues by building an optimal decision boundary. By leveraging the DenseNet201's ability to learn informative and rich feature representation and SVM's excellent classification performance, this combined approach offers a robust and efficient solution for retinal abnormalities detection and vision impairment assessment.

To help ensure reproducibility, features from DenseNet201 were generated by flattening the output of the final global pooling layer (right before original classification/softmax layer) into a 1920 dimensional vector. Input images were median filtered to remove speckle-noise and subsequently resized to 224×224 and preprocessed with ImageNet per-channel mean subtraction and scaling prior to input to DenseNet201. In the hybrid experiments the DenseNet201 network backbone was kept frozen (used as a fixed feature extractor) without softmax, and the resulting 1920-D vector output was provided as input to a multi-class SVM implemented with one-vs-one coding and an RBF (Gaussian) kernel. The SVM box constraint (hyperparameter C) and kernel scale (hyperparameter γ) were tuned using Bayesian optimization; the values provided to the optimizer as the initial starting configuration were $C = 221.4236$, kernel scale = 108.1465. Validation classification error was used as the objective to minimize using a Gaussian-process surrogate function. No additional standardization of inputs was

performed in the SVM models when using DenseNet features (standardize = false). During Bayesian optimization evaluations of candidate configurations were performed with stratified 5-fold cross-validation on the training portion of the data. The final model reported is provided after evaluation on the held-out test set. In the baseline CNN experiments, ImageNet pretrained DenseNet201 weights were used (softmax present, with weights fine-tuned), batch size of 64, 30 epochs, using the Adam optimizer with learning rate of 1e-4. Reported metrics were generated using the aforementioned 5-fold cross-validation procedure and hold-out test split of the data (1,726 images per class) and are averaged across independent runs.

3.2.5 Bayesian optimization

Bayesian optimization is a promising approach to improve the classification of retinal diseases. Instead of using computationally expensive search methods like grid search or random search, Bayesian optimization intelligently explores the hyperparameter space of a given model (usually a probabilistic model such as a Gaussian process) by iteratively selecting hyperparameter values based on a trade-off between exploitation and exploration. This allows it to quickly converge to an optimal hyperparameter configuration with fewer training iterations. This not only saves valuable computational resources but can also help prevent overfitting. When applied to OCT image analysis, for example within deep learning models, Bayesian optimization can be used to tune the architecture of the deep model (e.g., the number and types of layers), as well as learning rate, regularization parameters, and other hyperparameters. The ability of Bayesian optimization to handle noisy and black-box functions can help ensure that the selected hyperparameters generalize well to unseen data, leading to more accurate and reliable retinal disease classification.

4. RESULT AND DISCUSSION

The proposed Optimized Hybrid Deep Learning Model was implemented in MATLAB on a system with a Core i5 processor (2.50 GHz), 16 GB RAM, and an NVIDIA GTX3050 GPU, ensuring sufficient resources for analyzing OCT image classifications. A 5-fold cross-validation was applied for reliable evaluation. The study utilized OCT images divided into training (80%) and validation (20%), with an additional 1,726 test images per class covering CNV, DME, Drusen, and Normal categories. All pre-trained networks are trained with a batch size of 64, over 30 epochs, using the Adam optimizer with a learning rate of 0.0001, ensuring stable convergence. The performance of all pre-trained networks is illustrated in Table 2 and Table 3.

Table 2 and Table 3 show performance measures of different CNN models with different classification model. Each configuration had 30 independent runs to ensure stability. Table 2 included accuracy, sensitivity, specificity, and precision. False Positive Rate (FPR), F1 Score, Matthews correlation coefficient (MCC), and kappa. To get the best model Duncan Multirange test is carried out. Duncan's Multirange Test offers several advantages over ANOVA. While ANOVA identifies the existence of significant differences among groups, it does not specify where these differences lie. Duncan's Test builds on ANOVA by performing detailed pairwise comparisons and grouping

models into statistically significant clusters, providing granular insights into relative performance. It is specifically designed to control the experiment-wise error rate, reducing the likelihood of Type I errors, which is particularly critical when evaluating a large number of models. The test's ability to group models into performance categories (e.g., "d," "m," "a") simplifies interpretation and supports clear decision-making. Furthermore, Duncan's Test is more sensitive to variations within groups compared to other post hoc tests, such as Tukey's HSD, making it highly effective in identifying subtle differences in key performance metrics like MCC, Kappa, and FPR, which are essential for robust classification tasks. The statistical analysis by Duncan multirange test reveal the following points:

- DenseNet201 has the highest performance across accuracy, sensitivity, specificity, and precision, as

reflected by its group "d."

- ResNet50 and MobileNetV2 are also strong performers in groups "m" and "i" across all metrics.
- Models like AlexNet and NASNetMobile are at the lower end in group "a" or "k," indicating significantly weaker performance compared to higher-grouped models.
- DenseNet201 has the lowest FPR, indicating fewer false positives, and the highest MCC and Kappa, showing excellent classification balance.
- ResNet50 and MobileNetV2 show strong performance, but not as strong as DenseNet201 in these categories.
- VGG16, VGG19, and NASNetMobile perform poorly across the board, belonging to the lowest groups.

Table 2. Performance of CNN models in terms of Accuracy, Sensitivity, Specificity, and Precision

Model	Accuracy	Sensitivity	Specificity	Precision
Alexnet	.73091 ^a	.73091 ^a	.91030 ^a	.77816 ^a
Darknet19	.74576 ^b	.74576 ^b	.91525 ^b	.78098 ^b
Darknet53	.759603 ^c	.75960 ^c	.91986 ^c	.79317 ^c
Densenet201	.820640 ^d	.82064 ^d	.94021 ^d	.84141 ^d
Efficientnetb0	.748102 ^e	.74810 ^e	.91603 ^e	.78439 ^e
Googlenet	.737891 ^f	.73789 ^f	.91263 ^f	.78408 ^f
Inceptionresnetv2	.775946 ^g	.77594 ^g	.92531 ^g	.80930 ^g
Inceptionv3	.795138 ^h	.79513 ^h	.93171 ^h	.81584 ^h
Mobilenetv2	.810312 ⁱ	.81031 ⁱ	.93677 ⁱ	.83034 ⁱ
Nasnetlarge	.743530 ^j	.74353 ^j	.91451 ^j	.78004 ^j
Nasnetmobile	.724145 ^k	.72414 ^k	.90804 ^k	.76536 ^k
Resnet18	.756677 ^l	.75667 ^l	.91889 ^l	.79273 ^l
Resnet50	.815295 ^m	.81529 ^m	.93843 ^m	.84208 ^m
Resnet101	.804673 ⁿ	.80467 ⁿ	.93489 ⁿ	.83688 ⁿ
Shufflenet	.784559 ^o	.78455 ^o	.92818 ^o	.81240 ^o
Vgg16	.726052 ^p	.72605 ^p	.90868 ^p	.78059 ^p
Vgg19	.723387 ^q	.72338 ^q	.90779 ^q	.76668 ^q
xception	.778312 ^r	.77831 ^r	.92610 ^r	.81263 ^r

**** Results are in mean value 30 independent runs**

Note: CNN = Convolutional Neural Network.

Table 3. Performance Evaluation of CNN Models in terms of FPR, F1 Score, MCC and Kappa

Model	FPR	F1 Score	MCC	Kappa
Alexnet	.089695 ^a	.72064 ^a	.65897 ^a	.30875 ^a
Darknet19	.084746 ^b	.74257 ^b	.67596 ^b	.32390 ^b
Darknet53	.080132 ^c	.75664 ^c	.69369 ^c	.36082 ^c
Densenet201	.059786 ^d	.81945 ^d	.76984 ^d	.52170 ^d
Efficientnetb0	.083965 ^e	.74438 ^e	.67939 ^e	.33015 ^e
Googlenet	.087369 ^f	.73204 ^f	.66970 ^f	.30302 ^f
Inceptionresnetv2	.074684 ^g	.77337 ^g	.71535 ^g	.40252 ^g
Inceptionv3	.068287 ^h	.79537 ^h	.73607 ^h	.45370 ^h
Mobilenetv2	.063229 ⁱ	.81049 ⁱ	.75602 ⁱ	.49416 ⁱ
Nasnetlarge	.085489 ^j	.73993 ^j	.67344 ^j	.31795 ^j
Nasnetmobile	.091951 ^k	.72103 ^k	.64971 ^k	.27852 ^k
Resnet18	.081107 ^l	.75018 ^l	.68958 ^l	.35285 ^l
Resnet50	.061568 ^m	.81156 ^m	.76471 ^m	.50745 ^m
Resnet101	.065108 ⁿ	.80427 ⁿ	.75382 ⁿ	.48375 ⁿ
Shufflenet	.071813 ^o	.77860 ^o	.72344 ^o	.42549 ^o
Vgg16	.091315 ^p	.72230 ^p	.65862 ^p	.28024 ^p
Vgg19	.092204 ^q	.71766 ^q	.64928 ^q	.26236 ^q
Xception	.073895 ^r	.77799 ^r	.71944 ^r	.40883 ^r

**** Results are in mean value 30 independent runs**

Note: CNN = Convolutional Neural Network; FPR = False Positive Rate; MCC = Matthews Correlation Coefficient.

Using Duncan's test grouping, we can compare the statistical significance of the models based on their performance metrics. Here are the general patterns:

- DenseNet201 consistently ranks in the highest-performing groups across all metrics, demonstrating that it is statistically superior to many other models.
- ResNet50, MobileNetV2, and InceptionV3 also perform strongly but fall into lower groups than DenseNet201, indicating statistically significant differences.
- AlexNet, NASNetMobile, and VGG models generally perform poorly and are grouped at the lowest levels, showing their significant underperformance.

Hence, DenseNet201 is the clear top performer across nearly all metrics.

Hence, to improve the performance, a hybrid model is introduced, where denset201 is used for feature extraction, SVM for classification and Bayesian optimizer for tuning the hyperparameters of the model. The performance of hybrid model is illustrated in Figure 3 and Table 4. Figure 4 illustrated Bayesian optimization process for tuning SVM hyperparameters for classification of Retinal damage and vision impairments. The confusion matrix (Figure 3)

illustrates the performance of the hybrid model in classifying retinal damage and vision impairments, demonstrating its effectiveness across the four categories: CNV, DME, Drusen, and Normal. The optimization process for tuning the SVM hyperparameters using Bayesian optimization is depicted in Figure 4, providing insights into the steps taken to enhance the model's performance. The hybrid framework integrates DenseNet201 for feature extraction, SVM for classification, and Bayesian optimization for fine-tuning the hyperparameters, as shown in Figure 3 and detailed in Table 4. This combination ensures improved accuracy and robustness, with the figures serving as visual aids to better understand the workflow and results.

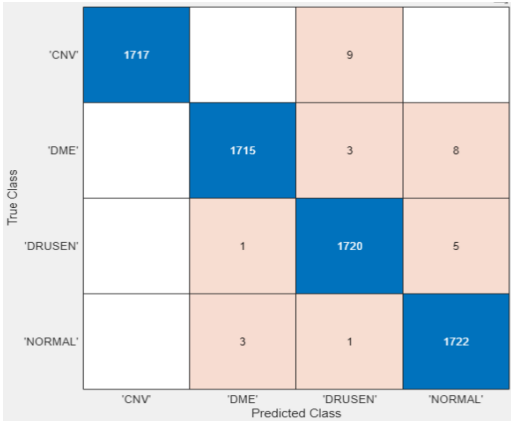


Figure 3. Confusion matrix of optimized hybrid deep learning model for classification of retinal damage and vision impairments

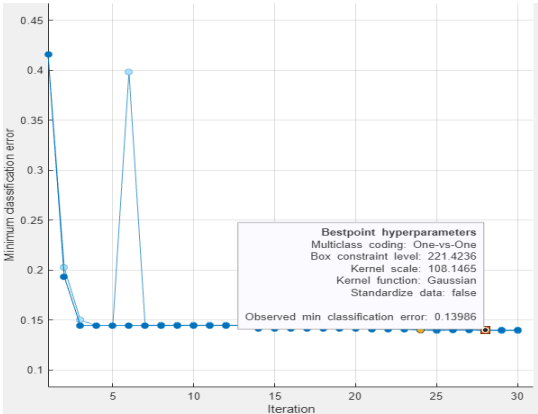


Figure 4. Bayesian optimization process for tuning SVM hyperparameters for classification of Retinal damage and vision impairments
Note: SVM = support vector machine.

The proposed model achieved exceptional performance with 99.6% accuracy, 99.43% precision, 99.54% sensitivity, 0.012 false positive rate, 98.89% MCC, 98.78% kappa coefficient, and an F1 score of 0.99. These metrics, computed from the confusion matrix and supported by ROC-AUC analysis, confirm the robustness of the framework in differentiating retinal diseases. Overall, the integration of DenseNet201 for deep feature extraction with SVM classification demonstrates strong reliability and generalizability, highlighting its potential for clinical application in retinal disease diagnosis.

Table 4 shows the Performance of Optimized Hybrid Deep Learning model of each class in terms of TPR, FNR, PPV and FDR.

Table 4. Performance of optimized hybrid deep learning model of each class in terms of TPR, FNR, PPV and FDR

	CNV	DME	DRUSEN	NORMAL
TPR	99.5%	99.4%	99.7%	99.8%
FNR	0.5%	0.6%	0.3%	0.2%
PPV	100%	99.8%	99.2%	99.3%
FDR	0%	0.2%	0.8%	0.7%

Note: CNV = Choroidal neovascularization; DME = Diabetic Macular Edema; TPR = True Positive Rate; FNR = False Negative Rate; PPV = Positive Predictive Value; FDR = False Discovery Rate.

In this experiment, we employ the proposed Bayesian optimization in order to further improve the classification error in the retinal disease detection problem. For this, we apply the SVM algorithm in a transfer learning framework, as a replacement to the softmax layer in our network. The reason behind this modification is two-fold: better robustness and multiclassification abilities of SVM with the one-vs-one algorithmic technique. As for the hyperparameters of SVM box constraint and kernel scale, we set their initial values to be 221.4236 and 108.1465, respectively, and then further optimize them with the Bayesian optimization approach equipped with a Gaussian kernel. With this, we could expect better decision boundaries for the improved retinal disease classification model. The initial error is 0.42, and after two iterations, it was reduced to 0.14, with further decrease to 0.13986 after a total of 28 iterations, as is illustrated in Figure 4. To validate our approach, we used a 10% test set, with 5-fold cross validation to ensure reliability of our results. The SVM classifier was used in tandem with the set of features from DenseNet-201 in a transfer learning framework. Table 5 presents the results, with the proposed method outperforming the other, in terms of decreasing the error and increasing the accuracy of retinal disease classification, which is also illustrated in Figure 4.

Table 5. The outcomes and specifics of the suggested and current approaches

Reference	Approach	Accuracy (%)	Sensitivity (%)	Specificity (%)
Tuncer et al., 2021 [23]	CNN-SVM	98.96	97.90	97.43
Upadhyay et al., 2022 [24]	CNN	97.16	96.30	98.50
Ara et al., 2022 [25]	CNN	99.00	-	-
Sotoudeh-Paima et al., 2022 [26]	FPN-VGG16	94.50	94.82	97.23
Subramanian et al., 2022 [27]	DenseNet201	93.01	-	-
Choudhury et al., 2023 [29]	Vgg19	99.17	99.00	99.5
Zhang., 2025 [22]	Multi-modal Interactive Projection Module	86.90	72.50	n/a
Proposed framework	DenseNet201 with SVM and Bayesian optimizer	99.69	99.54	99.09

Note: CNN = Convolutional Neural Network; SVM = support vector machine; FPN = Feature Pyramid Nrtwork; DL = deep learning.

The method achieves high accuracy and sensitivity, minimizing false negatives and ensuring reliability in clinical diagnostics. Its lightweight design and efficient computational requirements make it scalable and suitable for deployment in resource-constrained clinical settings. Furthermore, the use of SVM for classification provides clearer decision boundaries, enhancing result interpretability and aiding clinicians in making informed decisions.

Regarding the performance improvement, while the numerical difference may appear marginal, even small gains in accuracy, sensitivity, and specificity are crucial in medical image analysis, where high precision is required for clinical decision-making. The proposed hybrid approach achieves an impressive accuracy of 99.69%, sensitivity of 99.54%, and precision of 99.43%, ensuring a statistically reliable enhancement over previous methods. By systematically integrating feature extraction, classification, and optimization, our approach reduces the risk of overfitting and improves model robustness.

While the proposed Optimized Hybrid Deep Learning Framework demonstrates high classification performance, certain limitations must be acknowledged to provide a balanced evaluation. Although the study initially relied on a single publicly available dataset, the model has also been evaluated on three additional publicly available datasets: "OCTDL: Retinal OCT Images Dataset" [30], "OCTID: OCTImage Database" [31], and "Retinal OCT Image Classification - C8" [32]. This broader validation enhances the model's generalizability to diverse populations and imaging conditions, illustrated in Table 6.

Table 6. Performance of proposed model on different datasets

Dataset	Labels	Performance
OCTDL [30]	AMD, DME, ERM, NO, RAO, RVO, VID	Accuracy 95.73%, Sensitivity 95.4%, specificity 95.71%
OCTID [31]	AMD, CSR, DR, MH, NO	Accuracy 99.89%, Sensitivity 99.44%, Specificity 99.58%
Retinal OCT Image Classification - C8 [32]	AMD, CNV, CSR, DME, DR, DRUSEN, MH, NORMAL	Accuracy 89.59%, Sensitivity 84.46%, Specificity 87.58%

Note: OCTDL= Optical Coherence Tomography Dataset for Image-Based Deep Learning Methods; OCTID= Optical coherence tomography image database.

One potential criticism of our 99.69% accuracy is that it is high. There are several reasons we were able to achieve such high performance: (a) we trained on a large amount of labeled data (35168 images) and used sufficiently large balanced test splits (1726 images/class) to reduce sampling variance, (b) we objectively selected the backbone through both evaluation of 18 pretrained CNNs and Duncan's multirange test prior to hybridization in order to determine DenseNet201 was the best performing feature extractor, (c) we used DenseNet201 final global-pooled features (1920-D) and paired it with an RBF SVM while also using Bayesian optimization to search over box-constraint and kernel-scale to improve decision boundaries, and (d) we used stratified 5-fold CV during hyperparameter search in addition to keeping a fully held-out test set (6904 images) that was used for final evaluation only. Finally, to help ensure generalizability and alleviate concerns over overfitting we (i) tested the optimized hybrid on three

other public OCT datasets (OCTDL, OCTID, and C8) and report these dataset-wise results in Table 6, (ii) report the confusion matrix and show per-class TPR/PPV (Table 4 / Figure 3) to ensure performance is balanced across all classes, and (iii) report a number of metrics in addition to accuracy (sensitivity, precision, F1, FPR, MCC, Kappa) to show high accuracy is supported by other metrics.

We evaluated our final DenseNet201+SVM model on three external OCT datasets using two testing protocols: (1) directly transfer -- use the model trained on OCT2017 as is to establish zero shot generalization and (2) domain adapted -- fine-tune just the SVM on each dataset's development split (DenseNet weights frozen) using the same Bayesian hyperparameter tuning procedure as above. The direct transfer results indicate out-of-domain generalization; domain adapted results (in Table 6) indicate what can be gained with minimal training effort. Keep in mind that these results should be interpreted carefully as the datasets vary in image acquisition properties, labels used, and split protocols; strong results on the domain adapted splits suggest the DenseNet features learned are generalizable to other datasets, while poor direct transfer results suggest there is further work needed to align domains before model deployment.

A key limitation of this work is that clinical deployment considerations were not addressed. We did not evaluate inference latency or the hardware and compute constraints typical of real-world settings, nor did we assess integration with PACS/workflows or the regulatory pathways required for clinical validation. Clinician acceptance remains untested, particularly with respect to explainability and prospective clinical testing. In addition, interpretability analyses and expert review were limited: future work should include systematic clinician evaluation of attribution maps and lesion-level validation to verify that model attention aligns with clinically relevant findings before any clinical deployment.

5. CONCLUSION

The objective of this research was to optimize a hybrid deep learning architecture to yield a refined classification of retinal disease from OCT images. The proposed model demonstrated a highly efficient and optimized approach by merging the DenseNet201 architecture and SVM classifier for disease classification. The use of Bayesian optimization for hyperparameter tuning resulted in a significant improvement in model performance, with an accuracy of 99.69%, sensitivity of 99.54%, and F1 score of 0.991. The optimization process itself contributed significantly to the enhancement of the model's ability to predict retinal diseases such as Drusen, DME, and CNV correctly. This simplified approach has the potential to provide a valid and reliable diagnostic tool that can be employed to aid in the early detection of retinal disorders and, in general, improve patient outcomes and speed up treatment decisions.

However, there are certain limitations of the current study that must be addressed. First, the proposed model was built on a single dataset, which limited its generalization power. The model's performance on more diverse datasets must be assessed and verified to evaluate its robustness and reliability in different populations and imaging conditions. Second, although the model exhibited excellent performance on the test metrics, it may face computational efficiency and hardware

constraints when applied in real-world clinical settings. The scalability of the model must be assessed, and any performance bottlenecks must be eliminated to make it practically usable in clinical environments. Third, the model could be further improved in terms of inference speed. Obtaining prediction times that are less than real-time while maintaining accuracy is critical for using the model in clinical decision support systems and real-time applications. In the future, the dataset can be expanded to cover more types of retinal diseases and explore the fusion of multimodal data sources to further improve diagnostic accuracy. The use of explainable AI (XAI) methods may also help understand how the model made its predictions, providing transparency and trust in its output. Finally, the hybrid deep learning model could be extended to other medical imaging modalities besides retinal disease diagnosis, which could open up further applications in multiple diagnostic areas.

AUTHOR CONTRIBUTION STATEMENT

"Shreemat Kumar Dash and Sudarson Jena contributed to the conception, design and implementation of the deep learning architecture for classification of retinal injury in OCT images as well as data analysis and interpretation. Ashoka Kumar Ratha and Prabira Kumar Sethy provided important inputs to model optimization and experimental design.". Prabira Kumar Sethy also provided research guidance and coordinated the research team to integrate the machine learning methods. Aziz Nanthaamornphong assisted in analyzing the model's performance measures and reviewed the manuscript for strengthening its technical rigor. Santi Kumari Behera assisted in data analysis and helped revise the manuscript and provided critical comments to enhance the clinical significance of the study. M.V. Subbarao adopted the reviewer's comments and improved the quality of the manuscript.

AVAILABILITY OF DATA AND MATERIALS

The dataset is available on: <https://www.kaggle.com/datasets/paultimothymooney/kerman2018>.

REFERENCES

- [1] Liu, L., Li, C., Yu, H., Yang, X. (2022). A critical review on air pollutant exposure and age-related macular degeneration. *Science of The Total Environment*, 840: 156717. <https://doi.org/10.1016/j.scitotenv.2022.156717>
- [2] Carozza, G., Zerti, D., Tisi, A., Ciancaglini, M., Maccarrone, M., Maccarrone, R. (2024). An overview of retinal light damage models for preclinical studies on age-related macular degeneration: Identifying molecular hallmarks and therapeutic targets. *Reviews in the Neurosciences*, 35(3), 303-330.
- [3] Monis, M.D., Ali, S.M., Bhutto, I.A., Mahar, P.S., Ali, S., Mahar, P.S. (2023). Idiopathic choroidal neovascularization in pregnancy: A case report. *Cureus*, 15(2): e34611. <https://doi.org/10.7759/cureus.34611>
- [4] Madjedi, K., Pereira, A., Ballios, B.G., Arjmand, P., Kertes, P.J., Brent, M., Yan, P. (2022). Switching between anti-VEGF agents in the management of refractory diabetic macular EDEMA: A systematic review. *Survey of Ophthalmology*, 67(5): 1364-1372. <https://doi.org/10.1016/j.survophthal.2022.04.001>
- [5] Karabaş, V.L., Tokuç, E.Ö., Şermet, F. (2022). Survey of intravitreal injection preferences for the treatment of age-related macular degeneration and macular EDEMA among members of the Turkish Ophthalmological Association. *Turkish Journal of Ophthalmology*, 52(3): 179. <https://doi.org/10.4274/tjo.galenos.2021.37075>
- [6] Diao, S., Su, J., Yang, C., Zhu, W., Xiang, D., Chen, X., Peng, Q., Shi, F. (2023). Classification and segmentation of OCT images for age-related macular degeneration based on dual guidance networks. *Biomedical Signal Processing and Control*, 84: 104810. <https://doi.org/10.1016/j.bspc.2023.104810>
- [7] Schmitt, J.M. (1999). Optical coherence tomography (OCT): A review. *IEEE Journal of Selected Topics in Quantum Electronics*, 5(4): 1205-1215. <https://doi.org/10.1109/2944.796348>
- [8] Alsaih, K., Lemaitre, G., Rastgoo, M., Massich, J., Sidibé, D., Meriaudeau, F. (2017). Machine learning techniques for diabetic macular EDEMA (DME) classification on SD-OCT images. *Biomedical Engineering Online*, 16(1): 68. <https://doi.org/10.1186/s12938-017-0352-9>
- [9] Nandy Pal, M., Roy, S., Banerjee, M. (2021). Content based retrieval of retinal OCT scans using twin CNN. *Sādhanā*, 46(3): 174. <https://doi.org/10.1007/s12046-021-01701-5>
- [10] Daanouni, O., Cherradi, B., Tmiri, A. (2020). Automatic detection of diabetic retinopathy using custom CNN and grad-cam. In: Saeed, F., Al-Hadhrani, T., Mohammed, F., Mohammed, E. (eds) *Advances on Smart and Soft Computing. Advances in Intelligent Systems and Computing*, Springer, Singapore. https://doi.org/10.1007/978-981-15-6048-4_2
- [11] Kepp, T., Ehrhardt, J., Heinrich, M.P., Hüttmann, G., Handels, H. (2019). Topology-preserving shape-based regression of retinal layers in OCT image data using convolutional neural networks. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, pp. 1437-1440. <https://doi.org/10.1109/ISBI.2019.8759261>
- [12] Hussain, M.A., Bhuiyan, A.D., Luu, C., Theodore Smith, R.H., Guymer, R., Ishikawa, H., Schuman, J.S., Ramamohanarao, K. (2018). Classification of healthy and diseased retina using SD-OCT imaging and random forest algorithm. *PloS One*, 13(6): e0198281. <https://doi.org/10.1371/journal.pone.0198281>
- [13] Santos, A.M., Paiva, A.C., Santos, A.P., Mpinda, S.A., Gomes Jr, D.L., Silva, A.C., Braz Jr., G., de Almeida, J.D.S., Gattass, M. (2018). Semivariogram and Semimadogram functions as descriptors for AMD diagnosis on SD-OCT topographic maps using Support Vector Machine. *Biomedical Engineering Online*, 17(1): 160. <https://doi.org/10.1186/s12938-018-0592-3>
- [14] Tan, J.H., Bhandary, S.V., Sivaprasad, S., Hagiwara, Y., Bagchi, A., Raghavendra, U., Rao, A.K., Raju, B., Shetty, N.S., Gertych, A., Chua, K.C., Acharya, U.R. (2018). Age-related macular degeneration detection using deep convolutional neural network. *Future Generation Computer Systems*, 87: 127-135. <https://doi.org/10.1016/j.future.2018.05.001>
- [15] Rasti, R., Rabbani, H., Mehridehnavi, A., Hajizadeh, F.

- (2017). Macular OCT classification using a multi-scale convolutional neural network ensemble. *IEEE Transactions on Medical Imaging*, 37(4): 1024-1034. <https://doi.org/10.1109/TMI.2017.2780115>
- [16] Li, F., Chen, H., Liu, Z., Zhang, X., Wu, Z. (2019). Fully automated detection of retinal disorders by image-based deep learning. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 257(3): 495-505. <https://doi.org/10.1007/s00417-018-04224-8>
- [17] Li, F., Chen, H., Liu, Z., Zhang, X.D., Jiang, M.S., Wu, Z.Z., Zhou, K.Q. (2019). Deep learning-based automated detection of retinal diseases using optical coherence tomography images. *Biomedical Optics Express*, 10(12): 6204-6226. <https://doi.org/10.1364/BOE.10.006204>
- [18] Das, V., Dandapat, S., Bora, P.K. (2021). Automated classification of retinal OCT images using a deep multi-scale fusion CNN. *IEEE Sensors Journal*, 21(20): 23256-23265. <https://doi.org/10.1109/JSEN.2021.3108642>
- [19] Akinniyi, O., Rahman, M.M., Sandhu, H.S., El-Baz, A., Khalifa, F. (2023). Multi-stage classification of retinal OCT using multi-scale ensemble deep architecture. *Bioengineering*, 10(7): 823. <https://doi.org/10.3390/bioengineering10070823>
- [20] Hassan, E., Elmougy, S., Ibraheem, M.R., Hossain, M.S., AlMutib, K., Ghoneim, A., AlQahtani, S.A., Talaat, F.M. (2023). Enhanced deep learning model for classification of retinal optical coherence tomography images. *Sensors*, 23(12): 5393. <https://doi.org/10.3390/s23125393>
- [21] Shakor, M.Y., Khaleel, M.I. (2024). Recent advances in big medical image data analysis through deep learning and cloud computing. *Electronics*, 13(24): 4860. <https://doi.org/10.3390/electronics13244860>
- [22] Zhang, H., Bai, X., Hou, G., Quan, X. (2025). A multi-step interaction network for multi-class classification based on OCT and OCTA images. *Information Fusion*, 120: 103041. <https://doi.org/10.1016/j.inffus.2025.103041>
- [23] Tuncer, S.A., Çınar, A., Firat, M. (2021). Hybrid CNN based computer-aided diagnosis system for choroidal neovascularization, diabetic macular edema, drusen disease detection from OCT images. *Traitement du Signal*, 38(3): 673-679. <https://doi.org/10.18280/ts.380314>
- [24] Upadhyay, P.K., Rastogi, S., Kumar, K.V. (2022). Coherent convolution neural network based retinal disease detection using optical coherence tomographic images. *Journal of King Saud University-Computer and Information Sciences*, 34(10): 9688-9695. <https://doi.org/10.1016/j.jksuci.2021.12.002>
- [25] Ara, R.K., Matiolański, A., Dziech, A., Baran, R., Domin, P., Wiczorkiewicz, A. (2022). Fast and efficient method for optical coherence tomography images classification using deep learning approach. *Sensors*, 22(13): 4675. <https://doi.org/10.3390/s22134675>
- [26] Sotoudeh-Paima, S., Jodeiri, A., Hajizadeh, F., Soltanian-Zadeh, H. (2022). Multi-scale convolutional neural network for automated AMD classification using retinal OCT images. *Computers in Biology and Medicine*, 144: 105368. <https://doi.org/10.1016/j.compbmed.2022.105368>
- [27] Subramanian, M., Kumar, M.S., Sathishkumar, V.E., Prabhu, J., Karthick, A., Ganesh, S.S., Meem, M.A. (2022). Diagnosis of retinal diseases based on Bayesian optimization deep learning network using optical coherence tomography images. *Computational Intelligence and Neuroscience*, 2022(1): 8014979. <https://doi.org/10.1155/2022/8014979>
- [28] Hemalakshmi, G.R., Murugappan, M., Sikkandar, M.Y., Begum, S.S., Prakash, N.B. (2024). Automated retinal disease classification using hybrid transformer model (SViT) using optical coherence tomography images. *Neural Computing and Applications*, 36(16): 9171-9188. <https://doi.org/10.1007/s00521-024-09564-7>
- [29] Choudhary, A., Ahlawat, S., Urooj, S., Pathak, N., Lay-Ekuakille, A., Sharma, N. (2023). A deep learning-based framework for retinal disease classification. *Healthcare*, 11(2): 212. <https://doi.org/10.3390/healthcare11020212>
- [30] Kulyabin, M., Zhdanov, A., Nikiforova, A., Stepichev, A., Kuznetsova, A., Ronkin, M., Borisov, V., Bogachev, A., Korotkich, S., Constable, P.A., Maier, A. (2024). Octdl: Optical coherence tomography dataset for image-based deep learning methods. *Scientific Data*, 11(1): 365. <https://doi.org/10.1038/s41597-024-03182-7>
- [31] Gholami, P., Roy, P., Parthasarathy, M.K., Lakshminarayanan, V. (2020). OCTID: Optical coherence tomography image database. *Computers & Electrical Engineering*, 81: 106532. <https://doi.org/10.1016/j.compeleceng.2019.106532>
- [32] Subramanian, M., Shanmugavadivel, K., Naren, O.S., Premkumar, K., Rankish, K. (2022). Classification of retinal oct images using deep learning. In *2022 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, pp. 1-7. <https://doi.org/10.1109/ICCCI54379.2022.9740985>
- [33] Zhang, X., Xiao, Z., Yang, B., Wu, X., Higashita, R., Liu, J. (2024). Regional context-based recalibration network for cataract recognition in AS-OCT. *Pattern Recognition*, 147: 110069. <https://doi.org/10.1016/j.patcog.2023.110069>
- [34] Xiao, Z., Zhang, X., Zheng, B., Guo, Y., Higashita, R., Liu, J. (2024). Multi-style spatial attention module for cortical cataract classification in AS-OCT image with supervised contrastive learning. *Computer Methods and Programs in Biomedicine*, 244: 107958. <https://doi.org/10.1016/j.cmpb.2023.107958>