

Impact of Optimizer Algorithm on NasNetMobile Model for Eight-class Retinal Disease Classification from OCT Images

Madhumithaa. S^{ORCID}, and Masoodhu Banu N. M^{ORCID}.

Department of Biomedical Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India

Corresponding author: Madhumithaa.S (e-mail: vtd1152@veltech.edu.in), **Author(s) Email:** Masoodhu Banu N.M. (e-mail: dr.masoodhubanu@veltech.edu.in)

Abstract Artificial intelligence (AI) is an emerging technology that plays a vital role in various fields, including the medical field. Ophthalmology is the earliest field to adopt AI for diagnosing several retinal diseases. Many imaging techniques are available, but Optical Coherence Tomography (OCT) is particularly useful for early-stage diagnosis. OCT is a non-invasive imaging method that offers high-resolution visualization of the retinal structure, aiding the ophthalmologist in differentiating between normal and abnormal retina. Automated OCT-based retinal disease classification using deep learning (DL) is important for early disease detection. Most DL models achieved high performance, but the influence of the optimizer on model behaviour, convergence, and explainability remains a challenge. To bridge the gap, this study evaluates the performance and convergence of five optimizers, such as RMSprop, AdamW, Adam, Nadam, and SGD, on the NasNetMobile model. The model was trained on the OCT-8 dataset, which comprises seven diseased retinal classes and one normal class of Optical Coherence Tomography (OCT) images. The seven diseases are Age-related Macular Degeneration (AMD), choroidal neovascularization (CNV), Central Serous retinopathy (CSR), diabetic macular edema (DME), diabetic retinopathy (DR), DRUSEN, and Macular Hole (MH). The study also analyzes convergence behaviour and explainability through early stopping regularization technique and GradCAM XAI, respectively. The model achieved 71%, 93%, 96%, 97%, and 97% of accuracy, respectively. Compared with other optimizers, the SGD optimizer achieved high accuracy in 22 epochs, which indicates better generalization. GradCAM XAI highlights the disease-relevant region across different retinal diseases. This framework emphasizes the significance of selecting an appropriate optimizer for robust retinal disease classification using a DL model trained on OCT images.

Keywords Retinal diseases; NasNetMobile model; Deep learning; SGD optimizer; Nadam optimizer; AdamW optimizer; GradCAM XAI;

1. Introduction

Optical Coherence Tomography (OCT) is an important imaging modality for the early detection of retinal diseases. This provides a cross-sectional image of the retinal layers and clearly enables the structural abnormalities in the retinal and macular regions. Retinal diseases can cause irreversible vision loss if not detected early. Due to high screen time, stress, age, lack of exercise, etc., which increase the prevalence of retinal diseases and also require rapid, accurate, and reproducible evaluation [1-3]. For these reasons, automated detection of retinal disease from OCT images has increased attention among researchers. The Convolutional Neural Network (CNN) models, such as ResNet, DenseNet, AlexNet, Inception, and transformer models, as well as lightweight models like MobileNet and NasNetMobile, have been used for retinal disease classification and

have achieved better performance [4]. Many Machine learning (ML) and Deep Learning (DL) models have been proposed to enhance OCT-based retinal disease classification. Kermany et al. [5] adapted the inception V3 model for four-class classification of retinal disease using the Adam optimizer and obtained 93.4% accuracy. Nithya et al. [6] developed a five-layer CNN model with SGD optimizer and achieved 97% test accuracy. Madhumithaa et al. [7] trained the AlexNet model with the SGD optimizer, achieving 97% accuracy, whereas they also trained two other models, such as InceptionV3 and ResNet50, with the Adam optimizer, achieving 95% and 79.4% accuracy, respectively. Chen et al. [8], Kaur et al. [9] and Imram et al. [10], compared the SGD optimizer and Adam optimizer with different model combinations, such as Retina CNN, DenseNet121, and MobileNetV3, respectively. Hassan et al. [11] proposed enhanced DL model (modified ResNet50 + random forest) with dual

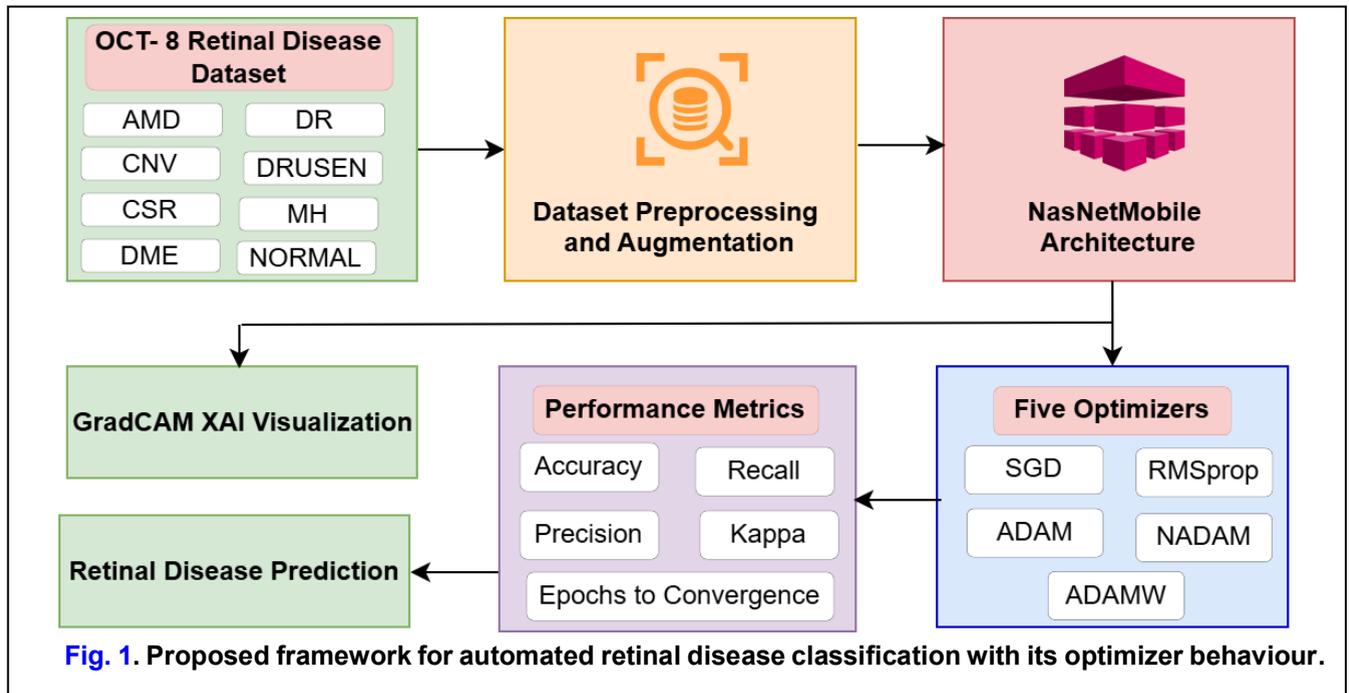
Table 1. Summary of the existing work, highlighting Optimizer choice and reported accuracies

Author	Dataset	Model	Optimizer	Accuray
Ma et al.[30] (2022)	3-class & 4-class	HCT-Net		86.18% & 91.56%
Malliga et al.[21] (2022)	8-class	VGG19	Adam	97%
Dutta et al.[31] (2023)	4-class	Conv-ViT		94%
Rabbi et al.[32] (2023)	4-class	CNN + CBAM		96%
Wang et al.[33] (2023)	Private	ViT	AdamW	95%
Laouarem et al. [34] (2024)	4-class,3-class, 8-class	HTC-retina		99%, 97% and 99%
Novely et al.[35] (2024)	8-class	DenseNet + CBAM		96.28%
Khalil et al.[36] (2024)	4-class & 5- class	Inception + dual attention	SGD	99.5% & 99.6%
Alenezi et al.[37] (2024)	4-class	EfficientNetB7 + attention module		
Shahzadi et al.[38] (2024)	4-class	Modified VGG19 & OCTNet	Adam	97% & 99%
Jannat et al.[39] (2025)	2-class	OCT-SelfNet		
Pan et al.[40] (2025)	4-class	MobileViT + CBAM	AdamW	99.8%

SGD and Adam optimizer for retinal disease classification, and achieved 95% accuracy. Stanojevic et al. have trained four models, such as AlexNet, VGG, Inception, and residual network, with the RMSprop optimizer and the Adam optimizer. The Inception model with the RMSprop optimizer obtained 95% classification accuracy. The summary of the existing literature with its optimizers is given in Table 1. Even though advancements in CNN architectures have been made, the function of the optimizer during training has not yet been thoroughly examined for multiclass datasets. Moreover, choosing an optimizer for CNN models is highly sensitive, especially for models trained on medical images, due to noisy gradients, small datasets, and class imbalance. However, this study utilizes the OCT-8 dataset, which is balanced, allowing us to analyze the optimizer's behaviour specifically. The role of the optimizer is not only to control the model weight update during backpropagation, but also to have an influence on stability, convergence speed, sensitivity to overfitting, and generalization. Most prior OCT retinal disease classification studies used traditional optimizers such as Stochastic Gradient Descent (SGD) or Adam, and achieved high classification accuracy. These studies focused solely

on final performance, without analyzing convergence behaviour, generalization characteristics, or computational efficiency across various optimizers. Especially, researchers believed that faster convergence must have better generalization, but it's still not fully investigated in medical imaging applications. Additionally, the impact of optimizer selection on model interpretability, particularly in highlighting the disease-relevant area through gradient-based visualization methods, remains unexplored. Therefore, there is a lack of systematic comparative analysis to evaluate optimizers across training stability, convergence speed, generalization performance, and explainability, particularly in the field of OCT retinal disease classification.

To bridge this gap, the study compares the five commonly used optimizers, like RMSprop, SGD, Adam, Nadam, and AdamW, for the NasNetMobile model for eight-class retinal disease classification from OCT images. We evaluated the optimizer performance in terms of validation stability, computation capability, convergence behaviour, generalization, and class-wise performance. The study also examined interpretability variation using GradCAM, an explainable artificial intelligence (XAI) method, to understand how the



optimizer guides the model, focusing on disease-relevant areas. The NasNetMobile model was trained individually for each optimizer and its training stability was assessed using performance metrics. The key contribution of this study is as follows.

1. The proposed work presents the systematic comparison of five optimizers for the classification of retinal diseases from OCT images. To the best of our knowledge, no prior work has been reported.
2. The study utilized the NasNetMobile model because it can capture deeper and higher-level features from the images and also maintains low computational complexity.
3. The performance of the optimizers was evaluated using standard performance metrics, Cohen's kappa. Learning stability and convergence behaviour were analyzed through the early stopping regularization technique.
4. Model interpretability was examined using GradCAM XAI, which highlights the model's decision-making focus region.

II. Methods

A detailed proposed framework for OCT-based retinal disease classification using different optimization algorithms is shown in Fig. 1. The proposed methodology consists of five modules. First, seven retinal disease and normal retinal OCT images were collected; then the images were preprocessed, and augmentation techniques were employed. The

Table 2. Image distribution in the OCT-8 Dataset

Class	Train	Validation	Test
AMD	2,300	350	350
CNV	2,300	350	350
CSR	2,300	350	350
DME	2,300	350	350
DR	2,300	350	350
DRUSEN	2,300	350	350
MH	2,300	350	350
NORMAL	2,300	350	350

preprocessed images were fed into the NasNetMobile model for training. In this stage, we selected and evaluated five optimizers, such as SGD, RMSprop, Adam, Nadam, and AdamW functions. Finally, the performance metrics of five models were examined, and GradCAM XAI was used for interpretability.

A. Dataset

The brief explanation of the retinal diseases is as follows. Age-related macular degeneration (AMD) affects people above 50 age [12], [13]. It has two stages: DRUSEN (the initial stage) contains yellow deposits under the retina [14], the next stage, choroidal neovascularization (CNV), characterized by abnormal blood vessel growth beneath the retina [15], [16]. A person with diabetes is affected by two retinal diseases, such as diabetic retinopathy (DR) [17] and diabetic macular edema (DME) [18], [19], which cause damage to blood vessels and fluid accumulation due to elevated glucose levels. Central Serous retinopathy (CSR), accumulates subretinal fluid [20], whereas a macular hole (MH) affects the macular region. The

OCT-8 dataset [21] consists of these seven retinal diseases and normal OCT retinal images. This dataset consists of 24,000 images, of which 18,400 are for training, and 2,800 each for validation and testing. The distribution of images per class is shown in Table 2

B. Data Preprocessing and Augmentation

The OCT images were resized into 224 x 224 pixels, and they were normalized to range between 0 and 1 using Eq. (1) [22]. In addition, the study utilized “KerasDataImageGenerator”, which generates real-time augmented images during model training and enhances the generalizability [23]. The augmentations, such as rotation ($\pm 10^\circ$), zoom (0.1), horizontal flip (True), height shift (0.1), and width shifts (0.1), were chosen for this study.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where X' represents the normalized image, X denotes the original intensity of the image, X_{min} refers to the minimum pixel value in the image (0), and X_{max} indicates the maximum pixel value in the image (255).

A. NasNetMobile Architecture

Neural architecture search Network (NASNet) is a

The study selected the NasNetMobile model because it offers several advantages over other lightweight models, such as MobileNet and EfficientNet. First, its neural architecture search nature extracts cell structures, generating repeatable building blocks that respond differently under various optimization techniques compared with manually designed models. Next, with 4.2M trainable parameters, the model achieved an optimal trade-off between performance and computational efficiency, which makes it suitable for analyzing various optimization behaviours. Next, its skip pathway and deeper connection produce a more complex loss landscape, which acts as an ideal framework for assessing how various optimizers direct gradient updates. Finally, it integrates medical image classification into its architecture, enabling us to analyze the interaction between the optimizer and the DL model, particularly for retinal disease classification from OCT images.

B. Optimizer Selection

Most retinal diseases show structural similarities on OCT images. The selection of the optimizer plays a vital role because it helps the model to learn these fine

Table 3. Summarize the selected Optimizers, their mechanism, and the motivation to include in this study

Optimizer	Key Mechanism	Motivation for use
SGD (Non-Adaptive)	Uniform updates, fixed LR with momentum for gradient smoothing.	Strong generalization, stable convergence.
RMSprop (Adaptive)	Grading based LR scaling	Faster convergence for noisy gradient
Adam (Adam+ Momentum)	Combine momentum and adaptive rates with bias correction	Widely used
Nadam (Adaptive+ Nesterov momentum)	Look-ahead for accelerated convergence	Complex gradient
AdamW (Adaptive+ Regularization)	Decoupled weight decay	Improved regularization

lightweight CNN that has the ability to capture hierarchical and complex features from OCT retinal images. The architecture is built from two basic components: the normal cell and the reduction cell. First, the model contains a stem convolution and is then stacked with multiple cell blocks. These layers systematically extract disease-relevant structural changes, such as neovascular membranes, drusen deposits, fluid pockets, edge patterns, and retinal layer textures. Then, the model contains depthwise separable convolution, which helps to maintain low computational complexity. Finally, the extracted features are passed to the global pooling layer (GAP) and then to the fully connected and softmax layer, which predicts the retinal disease.

features efficiently. In this study, five optimization algorithms were employed to evaluate their effects on NasNetMobile performance for OCT-based retinal disease classification, and the rationale for selecting these optimizers is provided in Table 3. Each and every optimizer updates the model parameters by reducing the categorical cross-entropy loss as shown in Eq. (2) [24].

$$L = - \sum_{i=1}^j y_j \log(\hat{y}_j) \quad (2)$$

where L represents cross entropy loss, i denotes class index, j specify number of classes, y_j indicates the ground truth label and \hat{y}_j represent predicted probability.

1. Stochastic Gradient Descent (SGD)

SGD utilized the current mini-batch gradient for calculating gradients, as shown in Eq. (3) [25], [26] and Eq. (4) [25], [26],

$$v_{t+1} = \mu v_t + \alpha g_t \quad (3)$$

$$\theta_{t+1} = \theta_t - v_{t+1} \quad (4)$$

where v_{t+1} denotes updated velocity, μ signifies the momentum factor, v_t refers to the velocity, α indicates the learning rate, g_t denotes the gradient, θ_{t+1} is a new weight value, and θ_t represents the model weight at time step t . It provides stable generalization for OCT images by avoiding excessive parameter adaptation and reducing overfitting to fine details.

2. RMSprop

RMSprop used a running average of squared gradients for normalizing the gradient as Eq. (5) [27] and Eq. (6) [27]. In OCT images, the gradients varied across retinal layers due to differences in structure.

$$v_t = \beta v_{t-1} + (1 - \beta) g_t^2 \quad (5)$$

$$\theta_{t+1} = \theta_t - \alpha \frac{g_t}{\sqrt{v_t + \epsilon}} \quad (6)$$

where v_t describes the updated second moment estimate, β is a decay rate, v_{t-1} denotes the prior second moment estimate, g_t^2 signifies the square of the gradient, θ_{t+1} shows the updated model weight, θ_t represent present model weight, α indicates the learning rate, g_t is the gradient, and ϵ is a constant term added to prevent division by zero. It allows different convergence rates across model layers by adapting the learning rate for each parameter based on the recent gradient magnitude. This factor is particularly significant for OCT images because the retinal layers exhibit complex features and contrast variations.

3. Adam

Adam incorporates momentum with an adaptive per-parameter learning rate and simultaneously tracks the first and second moments of the gradients, and is widely used in medical imaging, especially in retinal disease classification from OCT images. It can handle the high variability in lesion patterns and retinal texture, and also converges quickly. Parameter updates are computed by Eq. (7) [24], Eq. (8) [24], and Eq. (9) [24],

$$m_t = \beta m_{t-1} + (1 - \beta_1) g_t \quad (7)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (8)$$

$$\theta_{t+1} = \theta_t - \alpha \frac{m_t}{\sqrt{v_t + \epsilon}} \quad (9)$$

where m_t represent first moment estimate, β denotes the decay rate, m_{t-1} is the previous first moment estimate, β_1 and β_2 describe the exponential decay rates, v_t refers to the second moment estimate, v_{t-1} indicates the previous second moment estimate, g_t^2 denotes the square of the gradient, θ_{t+1} is a new weight vector, θ_t indicates current model weight, α is a learning rate, ϵ represents constant which prevent division by zero and g_t denotes the gradient.

4. Nesterov-accelerated Adam (Nadam)

Nadam combines Nesterov momentum with Adam, providing a look-ahead mechanism that can accelerate convergence of gradient dynamics. This mechanism potentially helps the optimizer better direct gradient directions towards the ideal solution, especially in OCT retinal images with different pathological features. This is the first study to utilize Nadam optimizer for retinal disease classification from OCT images. Parameter updates are calculated by Eq. (10) [28],

$$\theta_{t+1} = \theta_t - \alpha \left(\frac{\beta_1 m_t}{\sqrt{v_t + \epsilon}} + \frac{(1 - \beta_2) g_t}{\sqrt{v_t + \epsilon}} \right) \quad (10)$$

where θ_{t+1} indicates adjusted weight values, θ_t is the present weight network weight, α denotes the learning rate, β_1 and β_2 demonstrate the exponential decay rates, m_t denotes the first moment estimate, v_t signifies the velocity, and ϵ represents a small constant to avoid division by zero.

5. AdamW

AdamW splits weight decay from gradient updates as shown in Eq. (11) [29], unlike Adam, where weight decay interacts with the gradient updates. This feature offers effective regularization and prevents overfitting.

Table 4. Hyperparameter setting used in this work

Components	Specification
Input size	224 x 224
Stride	2
Kernal size	3 x 3
Activation function	ReLu, softmax
Batch size	16
Number of epochs	100
Early stopping patience	10

$$\theta_{t+1} = \theta_t - \left(\alpha \frac{m_t}{\sqrt{v_t + \epsilon}} + \lambda \theta_t \right) \quad (11)$$

where θ_{t+1} is the updated network weight value, θ describes the current model weight, α represents the learning rate, m_t is the first moment estimate, v_t indicates the second moment estimate, ϵ is a constant, and λ denotes the weight decay coefficient.

C. Impact of OCT Image Characteristics on Optimizer Behaviour

Many natural OCT imaging properties lead to unique challenges for optimization algorithms. The main challenge is speckle noise, which arises from coherent light interference during OCT image acquisition. This noise produces random intensity fluctuations that generate noisy gradients during backpropagation, particularly in the initial convolutional layers, where the model learns low-level features. The next challenge is the structural similarities among retinal diseases, which exhibit small intensity variations. These features produce complex training dynamics that different optimizers can direct differently. Next, the retina is a layered structure that produces different contrast levels

Table 5. Performance Metrics for the NasNetMobile model on the OCT-8 dataset for five optimizers.

Optimizer	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Macro F1 (%)	Cohen's Kappa value (%)	Epoch to convergence
RMSprop	70.50 ± 2.12	75.50 ± 3.54	70.50 ± 2.12	70.50 ± 2.12	68.00 ± 1.41	66.000 ± 1.414	11
AdamW	89.00 ± 5.66	90.50 ± 4.95	89.00 ± 5.66	89.00 ± 5.66	89.00 ± 5.66	87.000 ± 7.071	32
Adam	96.00 ± 0.00	96.00 ± 0.00	96.00 ± 0.00	96.00 ± 0.00	96.00 ± 0.00	96.000 ± 1.414	38
Nadam	97.00 ± 0.00	97.00 ± 0.00	97.00 ± 0.00	97.00 ± 0.00	97.00 ± 0.00	95.500 ± 0.707	52
SGD	97.00 ± 0.00	97.00 ± 0.00	97.00 ± 0.00	97.00 ± 0.00	97.00 ± 0.00	97.00 ± 0.00	22

across anatomical areas, which can benefit from various learning rate adaptations across model parameters. These properties have precise significance for optimization performance. SGD with consistent learning rates offers gradient smoothing, potentially decreasing the influence of noisy gradients from speckle noise. On the other hand, the adaptive optimizer utilizes per-parameter learning rates that adapt to varying contrast levels but it risks over-adaptation to noisy gradients due to aggressive parameter updates. The RMSprop optimizer's learning rate handles the non-stationarity of gradients from different pathological features. The capabilities of each optimizer algorithm in relation to OCT-specific challenges provide an empirical foundation for this comparative study.

D. Evaluation Metrics and XAI

To assess the performance of the NasNetMobile model with various optimizers, this study employed many commonly used performance metrics from Eq.12 [34], Eq.13 [34], Eq.14 [34], and Eq.15 [34], respectively. In addition to this, early stopping and convergence analyses, and the Cohen's kappa value were computed. Cohen's kappa was utilized to evaluate the agreement between the actual and predicted classes. Although the OCT-8 dataset used in this study maintains a balanced class distribution, kappa remains valuable for OCT retinal image classification because some retinal diseases share similar structural features, leading to uncertain predictions. These metrics not only differentiate true model learning from chance-level prediction, particularly for structurally similar retinal diseases, but also provide confidence in model reliability for practical deployment. GradCAM XAI was used to interpret where the model focuses during decision-making. This shows a difference in the focusing region.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (13)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (14)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

where TP refers to true Positive, TN represents True Negative, FP denotes False Positive, and FN corresponds to false Negative.

GradCAM XAI was applied to the model trained across all optimizers to compare whether optimization strategies affect feature localization and explainability. It computed the gradient of the output class-probability score with respect to the feature maps of the final convolutional layer, which produced a heatmap. This heatmap highlighted the region that the model influences for classification.

E. Implementation and Hyperparameter Setting

The present work was performed on an NVIDIA GPU. The programming was executed using Python 3.11 and TensorFlow in Jupyter Notebook. The whole work was conducted in the same environment, and the details are given in Table 4. These specifications were selected through empirical validation and standard practices for medical image classification. The input size of 224 x 224 pixels adheres to the standard setup of the NasNetMobile model, offering a good balance between performance and computational efficiency. A batch size of 16 was selected due to GPU memory constraints and training stability. Training was limited to 100 epochs with early stopping, which prevents overfitting and unnecessary computation. The early stopping technique was utilized with patience 10, which stops training when there is no improvement in the validation loss, and the minimum mode stops training.

III. Result

The study evaluated the performance of five optimizers on the NasNetMobile model, and the detailed performance metrics on the OCT-8 dataset are presented in Table 5 for two multiple runs.

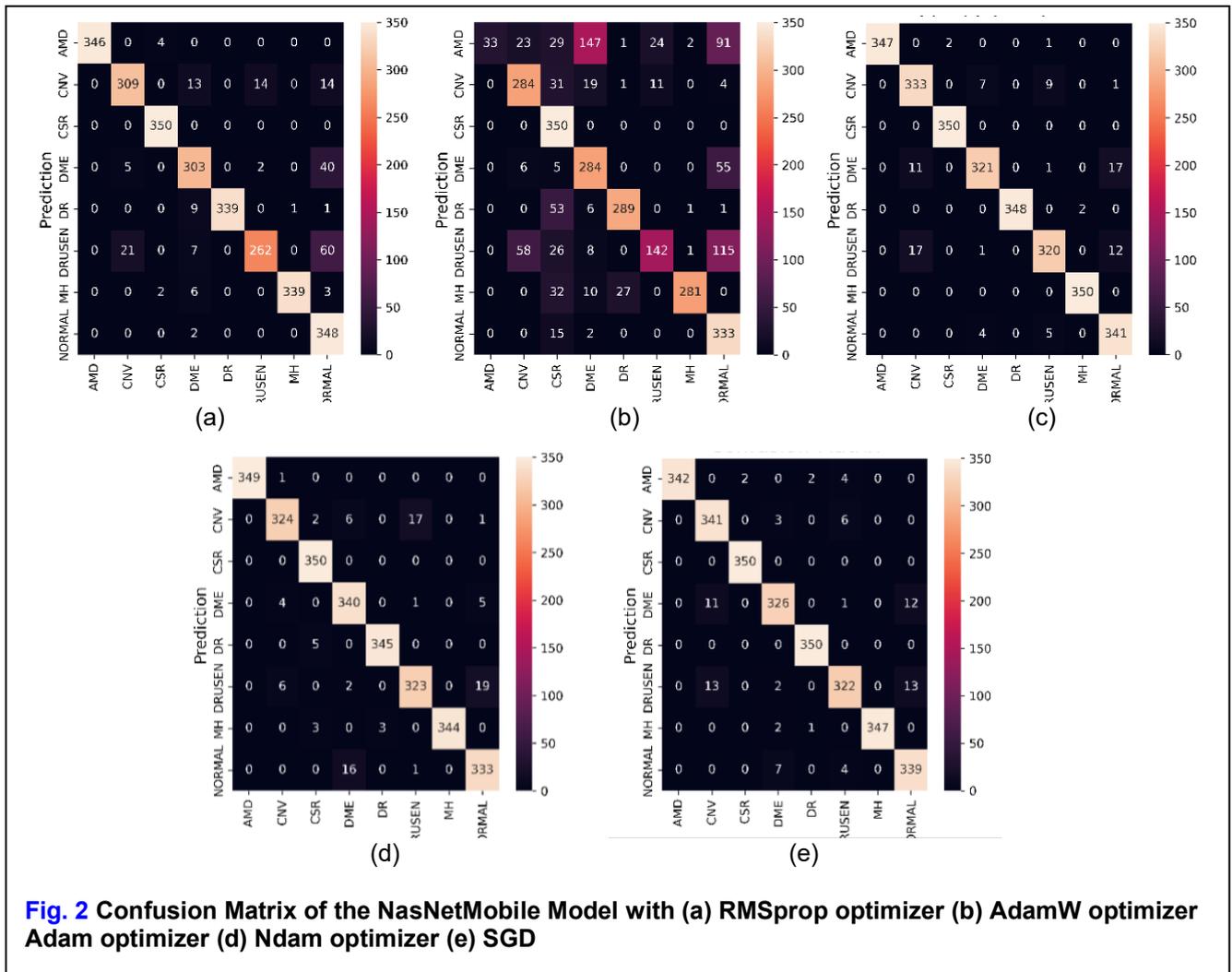


Fig. 2 Confusion Matrix of the NasNetMobile Model with (a) RMSprop optimizer (b) AdamW optimizer Adam optimizer (d) Ndam optimizer (e) SGD

The model with SGD and Nadam optimizers achieved the highest accuracy of 97% at 22 and 52 epochs, respectively, whereas the model with the RMSprop optimizer achieved the lowest accuracy of 71% at 11 epochs. The class-wise performance is shown in Fig. 2 confusion matrix, and the best three optimizer convergence is shown in Fig. 3 and Fig. 4. In Fig. 3, the SGD optimizer achieved stable convergence around 22 epochs and illustrated a smooth, consistent decrease in training loss with fluctuations across the training process, which indicates uniform gradient updates. Unlike the adaptive optimizer, which exhibits initial faster convergence within the first 15 epochs. The RMSprop optimizer exhibits intermediate behaviour, with fluctuating validation performance, which weakens training progress. In Fig. 4, the SGD optimizer maintains a small, constant gap between the validation and training losses, indicating better generalisation and avoiding overfitting. In contrast, adaptive optimizers are unable to maintain a consistent gap between the training and validation losses, indicating overfitting. The SGD optimizer achieved stable validation loss after

convergence, whereas adaptive optimizers continue to vary even in later epochs. The comparison of validation accuracy with the number of epochs for each optimizer is shown in Fig. 5. To understand the model decision, GradCAM XAI generates the heatmap. It highlighted the most influential area in red colour within the image, as shown in Fig. 6. In the CNV case, the GradCAM XAI heatmap focused on the subretinal hyperreflective area, a diagnostic feature of CNV. In the DME case, the highlighted region indicates fluid leakage in the macula, while for the DRUSEN case, the model focuses on lipid deposits between the retinal pigment epithelium (RPE) and Bruch's membrane. This observation demonstrates that the model learned meaningful pathological features rather than image artifacts.

IV. Discussion

The result presented in Table 5 shows an assessable and statistically significant difference in the optimizer performance when applied to NasNetMobile for OCT-based eight-class retinal disease classification.

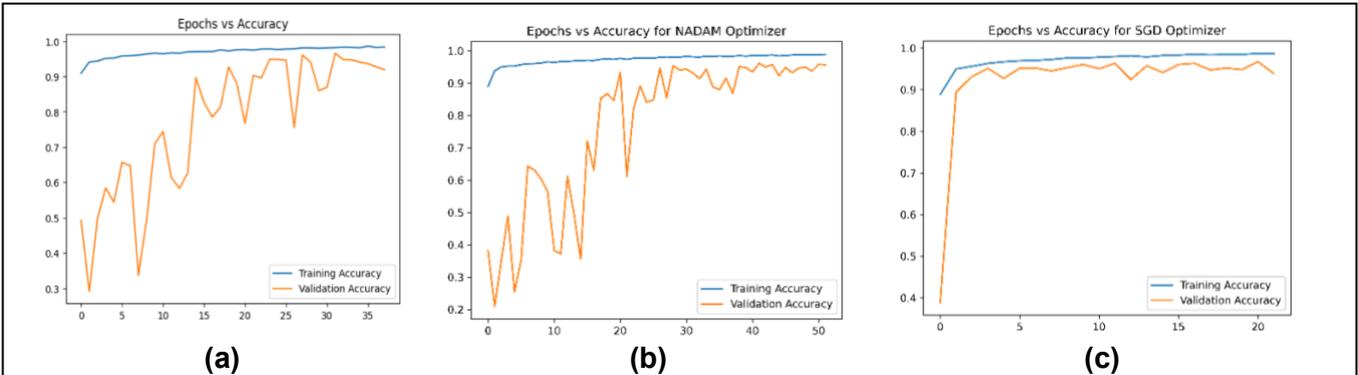


Fig. 3 Training- validation accuracy of the NasNetMobile Model with (a) Adam optimizer (b) Nadam optimizer (c) SGD optimizer.

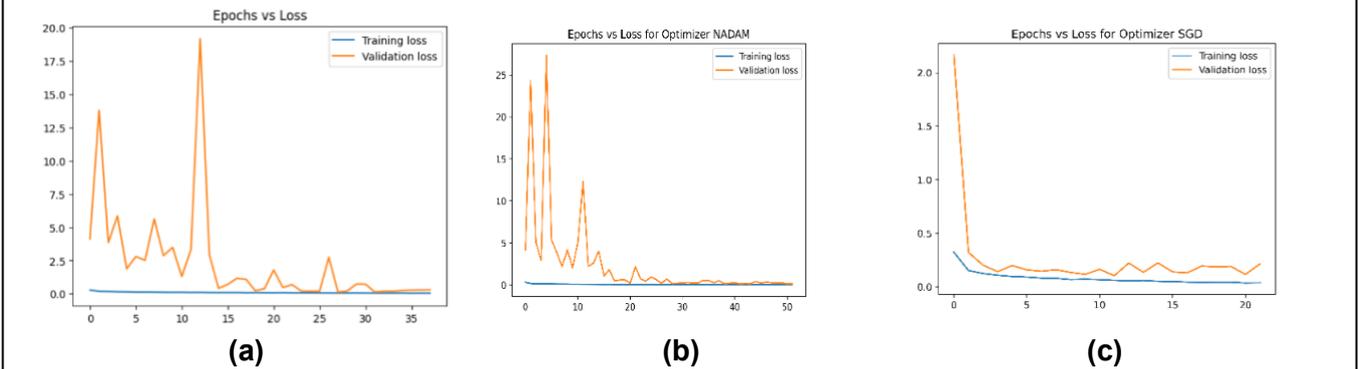


Fig. 4 Training- validation loss of the NasNetMobile Model with (a) Adam optimizer (b) Nadam optimizer (c) SGD optimizer.

The SGD optimizer exhibits superior performance across all performance metrics, with an accuracy of $97\% \pm 0.00$, recall of $97\% \pm 0.00$, precision of $97\% \pm 0.00$, macro F1 of $97\% \pm 0.00$, F1 score of $97\% \pm 0.00$, and Cohen Kappa value of $97\% \pm 0.00$, converging in 22 epochs. All SGD metrics that achieved a uniform zero deviation are a critical indicator for training stability, suggesting that the model yields similar performance across multiple runs; this consistency cannot be achieved by any other optimizer in this study. The Nadam optimizer obtained the same performance metrics of 97% across all metrics, but its Cohen Kappa value of $95.5\% \pm 0.707$ was 1.5 points lower than SGD and also required 52 epochs for convergence, which is 2.36 times more epochs than SGD. This variation is clinically important; a higher kappa value specifies stronger agreement beyond chance for each class prediction. The observed difference in the number of epochs between SGD and Nadam indicates that SGD achieves better generalization with a reduced computational complexity.

The Adam optimizer achieved consistently $96\% \pm 0.00$ across all metrics, with a Cohen kappa of 96 ± 1.414 , converging at 38 epochs. Compared with SGD, its metrics have decreased by 1%, and the kappa standard deviation is ± 1.414 , indicating a minor

inconsistency in per-class agreement. The AdamW optimizer achieved an accuracy of $89\% \pm 5.66$, which is considerably low, and a kappa of $87\% \pm 7.071$ at 32 epochs, indicating inconsistent stability during training. RMSprop showed low performance, achieving $70\% \pm 2.12$ on the performance metrics and a kappa of $66\% \pm 1.414$, with convergence at 11 epochs (the fastest among five optimizers). The fastest convergence, combined with the lowest accuracy,

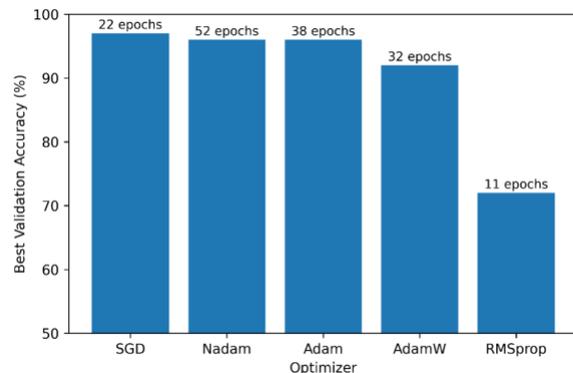


Fig 5. Comparative analysis for five optimizers in terms of validation accuracy, with the number of epochs required to reach convergence

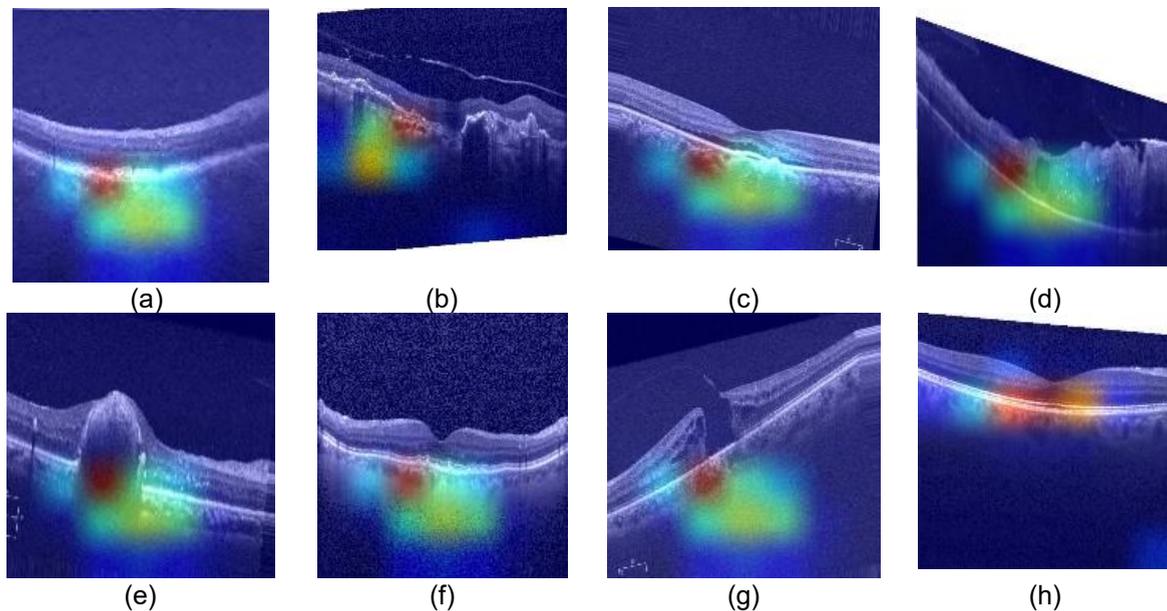


Fig 6. GradCAM Visualization for the NasNetMobile model with SGD optimizer (a) AMD, (b) CNV, (c) CSR, (d) DME, (e) DR, (f) DRUSEN, (g) MH, (h) NORMAL.

indicates that premature convergence, rather than computational efficiency. The optimizer settled into a local minimum of 70% accuracy and stopped improving further. It may be due to the influence of the irregular, high-frequency texture patterns in the OCT images in the optimizer's adaptive gradient scaling mechanism. The bar chart in Fig. 5 illustrates the clear evidence of these results, where SGD and Nadam optimizers achieved the superior validation accuracy, while RMSprop attained low performance with fewer epoch counts. The highest SGD performance in retinal disease classification can be attributed to the primary characteristics of the OCT imaging field. OCT images contain fine structural features like subretinal fluid pockets in CNV and DME, DRUSEN deposits beneath the RPE layer, which are differentiated by detailed, layered texture gradients instead of significant anatomical patterns. These features need consistent, gradient updates to learn constant, distinct boundaries between retinal classes. The constant learning rate technique of SGD, guided by early stopping, ensures consistent parameter updates that neither decelerate prematurely near local minima, as in AdamW, nor overshoot, as in RMSprop. However, adaptive optimizers like Adam, Nadam, and AdamW converge faster at the start due to their adaptive learning rate mechanisms. These optimizers needed a larger number of epochs to reach peak validation accuracy because the adaptive technique often demonstrates accelerated behaviour, which slows down the later training phase, especially when it is near a local minima. The study incorporated the early

stopping technique across all optimizers, strengthening SGD, as it provides uniform learning progression, and attained better generalization at 22 epochs compared to adaptive optimizers. The GradCAM visualization for SGD is illustrated in Fig. 6 for all eight classes. The heatmap confirms that the SGD focuses only on the disease-specific region rather than the background noise. A comparison of recent studies on retinal disease classification from OCT images with our proposed study is presented in Table 6. Based on the optimizer choice, three prior studies [31], [32], and [33] utilized Adam and AdamW and achieved accuracy of 94%, 96%, and 95%, respectively. In this study, Adam achieved 96% while AdamW achieved 89%, but SGD overperformed these two optimizers with 97%, confirming that the prevalent use of Adam can ignore potential accuracy gain. The most directly comparable work is [35], which combines DenseNet and CBAM and selects SGD as optimizer for OCT eight-class retinal classification. This study achieved 96.28% accuracy, but the present study outperforms this by 0.72% using the lightweight NasNetMobile model. A significant gap in Table 6 is the absence of estimates of Cohen's kappa in previous studies. However, our study evaluates a kappa value for all optimizers, which gives a complete picture of diagnostic robustness beyond accuracy alone and emphasizes the need for standardized metrics in future OCT-classification studies. Even though this study achieved promising results, limitations remain. The study used only a single dataset and a backbone model, comparing the optimizer only with its default learning rates. Moreover,

Table 6. Comparison with related work

Author	Model	Optimizer	Class	Accuracy
Dutta et al.[31] (2023)	Conv-ViT	Adam	4	94%
Rabbi et al.[32] (2023)	CNN + CBAM	Adam	4	96%
Wang et al.[33] (2023)	ViT	AdamW	Private	95%
Novely et al.[35] (2024)	DenseNet + CBAM	SGD	8	96.28%
Our Study	NasNetMobile	SGD	8	97%

optimization performance is largely determined by dataset characteristics and model complexity; therefore, the results of this study are generally not generalizable and cannot be directly applied to other imaging modalities or architectures. The study utilized only five optimizers; the latest optimizers like AdaGrad, Lion, and so on, were not included, and it is still uncertain whether these alternatives will overperform the SGD optimizer in retinal disease classification. The study achieved 97% accuracy with zero standard deviation and without external validation, which limits clinical translation claims. The computational complexity metrics, such as inference time, model FLOPs, training time per epoch, and memory usage, were not estimated; these parameters are important for evaluating feasibility for clinical deployment.

The results of the study have immediate implications for the deployment of DL-based systems for retinal disease diagnosis. The study highlights the SGD as an optimal optimizer for OCT-based retinal disease classification on the NasNetMobile model. The model achieved 97% across all metrics, with a kappa value of $97\% \pm 0.00$, in just 22 epochs. This provides evidence that the light-weight NasNetMobile is recommended for AI -assisted in OCT retinal disease classification, which should be trained with the SGD optimizer, especially when the training image consists of detailed layered texture patterns. Based on the 58% reduction in computational requirement by SGD compared with Nadam convergence, SGD can be recommended for scenarios involving frequent model retraining with dynamically updated clinical datasets.

V. Conclusion

The main aim of the study is to systematically evaluate the influence of the optimizer on the NasNetMobile model for eight-class retinal disease classification. The result demonstrates that the selection of the optimizer plays a vital role in enhancing accuracy and convergence behaviour. The result showed that the SGD optimizer achieved 97% accuracy in 22 epochs, outperforming adaptive optimizers while maintaining superior performance and fast convergence. Nadam also achieved similar accuracy, but it requires 52

epochs for convergence, whereas Adam achieved comparable accuracy 96% in 38 epochs, followed by AdamW, which achieved 89% in 32 epochs, and RMSprop, which achieved 70% in 11 epochs, showing the lowest performance. The study not only analyzes the accuracy but also examines the Cohen's kappa value and convergence behaviour, which also indicates that SGD shows better generalization across the retinal disease. The integration of early stopping not only reduced training time without affecting performance but also prevented overfitting. The implementation of GradCAM XAI provided significant insights into the model's decision. Overall, the observation emphasized that a suitable optimizer can improve the model performance without increasing computational resources. These results provide practical insights into selecting appropriate optimizers for developing an OCT-based retinal disease classification task in reduced time. In the future, the study will focus on evaluating more datasets, multiple architectures, a real-time clinical environment, and exploring a hybrid optimizer to further enhance robustness and generalization.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

Publicly available, and it is referred to in the manuscript.

Author Contribution

Madhumithaa S conceptualized and designed the study, conducted data collection, participated in data analysis and interpretation, and wrote the main draft. Masoodhu Banu conducted supervision and provided critical feedback on the manuscript. All authors reviewed and approved the final version of the manuscript and agreed to be responsible for all aspects of the work, ensuring integrity and accuracy.

Declarations

Ethical Approval

Not applicable.

Consent for Publication Participants.

Consent for publication was given by all participants.

Competing Interests

The authors declare no competing interests.

References

- [1] B. Das, "Sustainable global vision care," *Kerala Journal of Ophthalmology*, vol. 34, no. 2, p. 92, 2022, doi: 10.4103/kjo.kjo_64_22.
- [2] S. G. Honavar, "Self-sustainable and inclusive eye care – Where equity meets excellence," *Indian J. Ophthalmol.*, vol. 70, no. 5, pp. 1441–1442, May 2022, doi: 10.4103/ijo.IJO_1008_22.
- [3] J. Mai and U. Schmidt-Erfurth, "Rolle der künstlichen Intelligenz bei verschiedenen retinalen Erkrankungen," *Klin. Monbl. Augenheilkd.*, vol. 241, no. 09, pp. 1023–1031, Sep. 2024, doi: 10.1055/a-2378-6138.
- [4] A. Choudhary, S. Ahlawat, S. Urooj, N. Pathak, A. Lay-Ekuakille, and N. Sharma, "A Deep Learning-Based Framework for Retinal Disease Classification," *Healthcare*, vol. 11, no. 2, p. 212, Jan. 2023, doi: 10.3390/healthcare11020212.
- [5] D. S. Kermany *et al.*, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.
- [6] N. Rajagopalan, V. N., A. N. Josephraj, and S. E., "Diagnosis of retinal disorders from Optical Coherence Tomography images using CNN," *PLoS One*, vol. 16, no. 7, p. e0254180, Jul. 2021, doi: 10.1371/journal.pone.0254180.
- [7] M. S, M. B. N.M, J. M. M, and A. P. Sarma, "Comparison of DL model for retinal diseases classification using OCT images," in *2024 Tenth International Conference on Bio Signals, Images, and Instrumentation (ICBSII)*, IEEE, Mar. 2024, pp. 1–5. doi: 10.1109/ICBSII61384.2024.10562391.
- [8] A. Naik, B. S. Pavana, and K. Sooda, "Retinal Disease Classification from Retinal-OCT Images Using Deep Learning Methods," 2022, pp. 95–104. doi: 10.1007/978-981-19-2347-0_8.
- [9] A. Kaur, V. Kukreja, M. Kumar, A. Choudhary, and R. Sharma, "RetDense: A Fine-tuned DenseNet121 Model for Retinal Eye Disease Detection," in *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, IEEE, Mar. 2024, pp. 1–5. doi: 10.1109/IATMSI60426.2024.10502833.
- [10] S. Imran, A. L U, D. M S, T. Sharma, and T. Khanum, "Deep Learning-Based Classification of Retinal Diseases from OCT Images with LLM-Powered Patient Query Support," *Int. J. Health Sci. Res.*, vol. 15, no. 7, pp. 314–323, Jul. 2025, doi: 10.52403/ijhsr.20250738.
- [11] E. Hassan *et al.*, "Enhanced Deep Learning Model for Classification of Retinal Optical Coherence Tomography Images," *Sensors*, vol. 23, no. 12, p. 5393, Jun. 2023, doi: 10.3390/s23125393.
- [12] M. E. Subasi, S. Patnaik, and A. Subasi, "Optical coherence tomography image classification for retinal disease detection using artificial intelligence," in *Applications of Artificial Intelligence in Healthcare and Biomedicine*, Elsevier, 2024, pp. 289–323. doi: 10.1016/B978-0-443-22308-2.00009-3.
- [13] J. Subhedar and A. Mahajan, "A Review on Recent Work On OCT Image Classification for Disease Detection," in *2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON)*, IEEE, Feb. 2023, pp. 1–6. doi: 10.1109/OTCON56053.2023.10114003.
- [14] Q. Chen *et al.*, "Automated drusen segmentation and quantification in SD-OCT images," *Med. Image Anal.*, vol. 17, no. 8, pp. 1058–1072, Dec. 2013, doi: 10.1016/j.media.2013.06.003.
- [15] S. Saha *et al.*, "Automated detection and classification of early AMD biomarkers using deep learning," *Sci. Rep.*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-47390-3.
- [16] S. M. Waldstein, W. D. Vogl, H. Bogunovic, A. Sadeghipour, S. Riedl, and U. Schmidt-Erfurth, "Characterization of Drusen and Hyperreflective Foci as Biomarkers for Disease Progression in Age-Related Macular Degeneration Using Artificial Intelligence in Optical Coherence Tomography," *JAMA Ophthalmol.*, vol. 138, no. 7, pp. 740–747, Jul. 2020, doi: 10.1001/jamaophthalmol.2020.1376.
- [17] S. Akhtar *et al.*, "A deep learning based model for diabetic retinopathy grading," *Sci. Rep.*, vol. 15, no. 1, p. 3763, Jan. 2025, doi: 10.1038/s41598-025-87171-9.
- [18] T. A. Ciulla, A. G. Amador, and B. Zinman, "Diabetic Retinopathy and Diabetic Macular Edema Pathophysiology, screening, and novel therapies," *Diabetes Care*, vol. 26, no. 9, pp. 2653–2664, Sep. 2003, doi: 10.2337/DIACARE.26.9.2653.
- [19] A. Markan, A. Agarwal, A. Arora, K. Bazgain, V. Rana, and V. Gupta, "Novel imaging biomarkers in diabetic retinopathy and diabetic macular edema," *Ther. Adv. Ophthalmol.*, vol. 12, Jan. 2020, doi: 10.1177/2515841420950513.

- [20] S. A. Hassan, S. Akbar, A. Rehman, T. Saba, H. Kolivand, and S. A. Bahaj, "Recent Developments in Detection of Central Serous Retinopathy Through Imaging and Artificial Intelligence Techniques&x2013;A Review," 2021, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2021.3108395.
- [21] M. Subramanian, K. Shanmugavadivel, O. S. Naren, K. Premkumar, and K. Rankish, "Classification of Retinal OCT Images Using Deep Learning," in *2022 International Conference on Computer Communication and Informatics, ICCCI 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICCCI54379.2022.9740985.
- [22] H. Benhar, A. Idri, and J. L. Fernández-Alemán, "Data preprocessing for heart disease classification: A systematic literature review," *Comput. Methods Programs Biomed.*, vol. 195, p. 105635, Oct. 2020, doi: 10.1016/j.cmpb.2020.105635.
- [23] A. Soliman and J. Terstriep, "Keras Spatial," in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, New York, NY, USA: ACM, Nov. 2019, pp. 69–76. doi: 10.1145/3356471.3365240.
- [24] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
- [25] L. Bottou, "Large-Scale Machine Learning with Stochastic Gradient Descent," in *Proceedings of COMPSTAT'2010*, Heidelberg: Physica-Verlag HD, 2010, pp. 177–186. doi: 10.1007/978-3-7908-2604-3_16.
- [26] H. Robbins and S. Monro, "A Stochastic Approximation Method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, Sep. 1951, doi: 10.1214/aoms/1177729586.
- [27] R. Elshamy, O. Abu-Elnasr, M. Elhoseny, and S. Elmougy, "Improving the efficiency of RMSProp optimizer by utilizing Nestrovo in deep learning," *Sci. Rep.*, vol. 13, no. 1, p. 8814, May 2023, doi: 10.1038/s41598-023-35663-x.
- [28] A. N. T. Kissiedu, G. K. Aggrey, M. G. Asante-Mensah, and A. Asante, "Development of Pneumonia Identification System: A Comparative Analysis of Some Selected CNN Architectures Using Adam, Nadam, and RAdam Optimizers," in *2024 IEEE SmartBlock4Africa*, IEEE, Sep. 2024, pp. 1–12. doi: 10.1109/SmartBlock4Africa61928.2024.10779552.
- [29] H. Azis, M. S. Andini, Herman, L. Syafie, A. R. Manga', and L. B. Ilmawan, "A Comparative Study of SGD, Adam, AdamW, and RMSprop Optimizers for VGG19-Based Rupiah Banknote Classification," in *2025 9th International Conference On Electrical, Electronics And Information Engineering (ICEEIE)*, IEEE, Sep. 2025, pp. 1–6. doi: 10.1109/ICEEIE66203.2025.11251942.
- [30] I. Khalil, A. Mehmood, H. Kim, and J. Kim, "OCTNet: A Modified Multi-Scale Attention Feature Fusion Network with InceptionV3 for Retinal OCT Image Classification," *Mathematics*, vol. 12, no. 19, Oct. 2024, doi: 10.3390/math12193003.
- [31] A. M. Alenezi *et al.*, "Multiscale attention-over-attention network for retinal disease recognition in OCT radiology images," *Front. Med. (Lausanne)*, vol. 11, 2024, doi: 10.3389/fmed.2024.1499393.
- [32] N. Novelty, S. Mahmud Shuvo, and Md. F. Faruk, "Improving Pre-Trained CNNs with CBAM and Skip Connections for Multi-Class Retinal Diseases Classification using OCT Images," Association for Computing Machinery (ACM), Oct. 2024, pp. 946–953. doi: 10.1145/3723178.3723304.
- [33] Z. Ma, Q. Xie, P. Xie, F. Fan, X. Gao, and J. Zhu, "HCTNet: A Hybrid ConvNet-Transformer Network for Retinal Optical Coherence Tomography Image Classification," *Biosensors (Basel)*, vol. 12, no. 7, Jul. 2022, doi: 10.3390/bios12070542.
- [34] 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," *J. Imaging*, vol. 9, no. 7, Jul. 2023, doi: 10.3390/jimaging9070140.
- [35] S. F. Rabbi, M. Al Mamun, M. F. Faruk, S. M. Mahedy Hasan, and A. Y. Srizon, "A Multi-branch and Attention Based CNN Architecture for the Classification of Retinal Diseases from OCT Images," in *2023 International Conference on Information and Communication Technology for Sustainable Development, ICICT4SD 2023 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 36–40. doi: 10.1109/ICICT4SD59951.2023.10303384.
- [36] Q. Wang, K. Chen, W. Dou, and Y. Ma, "Cross-Attention Based Multi-Resolution Feature Fusion Model for Self-Supervised Cervical OCT Image Classification," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 20, no. 4, pp. 2541–2554, Jul. 2023, doi: 10.1109/TCBB.2023.3246979.

- [37] A. Laouarem, C. Kara-Mohamed, E. B. Bourennane, and A. Hamdi-Cherif, "HTC-retina: A hybrid retinal diseases classification model using transformer-Convolutional Neural Network from optical coherence tomography images," *Comput. Biol. Med.*, vol. 178, Aug. 2024, doi: 10.1016/j.compbimed.2024.108726.
- [38] Z. Shahzadi and M. Zubair, "Multiclass Classification of Retinal Disorders Using Optical Coherence Tomography Images," in *2024 Horizons of Information Technology and Engineering (HITE)*, IEEE, Oct. 2024, pp. 1–6. doi: 10.1109/HITE63532.2024.10777173.
- [39] F.-E. Jannat, S. Gholami, M. N. Alam, and H. Tabkhi, "OCT-SelfNet: a self-supervised framework with multi-source datasets for generalized retinal disease detection," *Front. Big Data*, vol. 8, Jul. 2025, doi: 10.3389/fdata.2025.1609124.
- [40] H. Pan *et al.*, "A lightweight model for the retinal disease classification using optical coherence tomography," *Biomed. Signal Process. Control*, vol. 101, p. 107146, Mar. 2025, doi: 10.1016/j.bspc.2024.107146.

industry experience, specializing in digital signal processing and embedded programming. Notably, she has also worked as a Scientist Grade C at ISRO, where she contributed to significant projects in space technology. Her recent publications reflect her research interests and contributions to Brain-Computer Interfaces, Image Processing, and Artificial Intelligence. Through her academic and research efforts, she plays a significant role in advancing knowledge and fostering innovation in these areas.

Author Biography



Madhumithaa S. is an esteemed research scholar at Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology in Avadi, Chennai. In her current position, she focuses on pioneering research that advances both technology and science.

Her academic journey includes a previous role as an Assistant Professor at an engineering college, where she spent 3.5 years educating and mentoring students across various engineering disciplines. Through her scholarly endeavors, Madhumithaa has made notable contributions to the fields of Artificial Intelligence, Image Processing, and Signal Processing, resulting in several impactful publications. Her work reflects her deep commitment to fostering knowledge and innovation in these dynamic areas, underscoring her dedication to advancing the frontiers of technology and education.



Masoodhu Banu N.M serves as a professor at the Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology in Avadi, Chennai. With two decades of teaching experience, she possesses profound expertise in her field. In addition to her academic role, she has ten years of

