

# Automated Retinal Disease Classification from OCT Images Using a Lightweight CNN–SE Architecture

Parth R. Dave<sup>1,2</sup>, Nikunj H. Domadiya<sup>2</sup>, and Jaiwal Patel<sup>3</sup>

<sup>1</sup> Department of Computer Engineering, Gujarat Technological University, Gujarat, India

<sup>2</sup> Department of Computer Engineering, L. D. College of Engineering, Gujarat, India

<sup>3</sup> Science Stream, GEMS New Millennium School, Dubai, UAE

**Corresponding author:** Nikunj H. Domadiya (e-mail: domadiyanikunj002@gmail.com), **Author(s) Email:** Parth R. Dave (e-mail: parthdave.ce@ldce.ac.in) Jaiwal Patel (e-mail: ujaiwal@outlook.com)

**Abstract** Automated classification of retinal diseases from Optical Coherence Tomography (OCT) images can support timely and consistent identification of retinal abnormalities; however, many existing approaches rely on computationally intensive architectures or complex preprocessing procedures. This study proposes a lightweight Convolutional Neural Network with Squeeze-and-Excitation blocks (CNN–SE) for automated multi-class retinal disease classification while maintaining a compact computational structure. The proposed architecture consists of convolutional layers with SE-based channel recalibration, global average pooling, a fully connected layer, and dropout. Unlike approaches that depend on pretrained backbones or multi-stage segmentation, the proposed framework integrates channel-wise feature recalibration directly into a compact CNN, allowing the model to learn discriminative retinal representations with limited architectural complexity. Data augmentation was applied only to the training images, and the network was optimized using the Adam optimizer. The model was evaluated on the eight-class OCT-C8 dataset and the four-class OCT2017 dataset using accuracy, precision, recall, F1-score, and macro-ROC-AUC. The proposed model achieved 97.00% accuracy on OCT-C8 and 99.00% on OCT2017. To assess domain generalization, the model was trained exclusively on OCT-C8 and directly evaluated on the independent OCT2017 test set without retraining or fine-tuning. Across five random seeds, the cross-dataset evaluation achieved  $98.74 \pm 0.11\%$  accuracy, with consistently high precision, recall, F1-score, and macro-ROC-AUC. The proposed network maintains a compact computational footprint and enables efficient image inference compared with the evaluated lightweight architectures. These results demonstrate that the proposed CNN–SE architecture provides strong classification performance and cross-dataset generalization while maintaining computational efficiency, making it a promising lightweight framework for automated retinal OCT image analysis.

**Keywords** Convolutional Neural Network; Artificial Intelligence; Machine Learning; Deep Learning; Optical Coherence Tomography

## 1. Introduction

Optical Coherence Tomography (OCT) is a valuable imaging technique used for the early detection of retinal diseases. It generates high-resolution tomographic images of retinal tissue that enable accurate detection of structural abnormalities within the retina and macula parts. Retinal diseases may sometimes result in irreversible vision loss and therefore the early detection of abnormalities is essential. Retinal OCT imaging is widely used in ophthalmology to assist in treatment planning and to monitor the success of treatments such as laser therapy, eye injections, and surgical interventions. There are several factors that can lead to retinal disorders, such as genetic defects, stress, aging, lack of physical activity, excessive screen

exposure, and unhealthy eating habits. As a result, retinal diseases are becoming more common, making it essential to carry out fast, precise, and consistent evaluations. Automated retinal OCT image classification techniques can be helpful as a suggestive system for doctors, providing impartial, reliable, and consistent diagnostic support for identifying abnormalities. Conventional machine learning algorithms use texture-based detailing retrieved from images to represent local or global information. Various feature extraction techniques such as Local Binary Pattern (LBP) [1], Histogram of Oriented Gradients (HOG) [2], Gray-Level Co-Occurrence Matrix (GLCM) [3], edge detection, and shape detection are applied to obtain a discriminative feature vector. The extracted

feature representation is then fed into supervised learning classifiers, including Artificial Neural Networks (ANN) [4] and Support Vector Machines (SVM) [5], or other classifiers to perform the final image classification. However, one major limitation of classical machine learning approaches is that they depend on manually extracted features instead of learning the features directly from the data. Since the image representation relies only on predefined feature descriptors, some important information from the original image may be lost during the feature extraction process. This loss of information can affect how well the model understands the data and may lead to lower accuracy and reduced overall performance of the classification system.

Deep learning has increasingly become the preferred approach for automated OCT image classification because CNNs [6] can learn discriminative representations directly from retinal image data. Recent studies have explored layer-guided CNNs, segmentation-assisted architectures, transfer learning, ensemble models, and other specialized frameworks to improve classification performance. Although these approaches have demonstrated strong results, their increased architectural depth, dependence on additional preprocessing stages, or reliance on multiple models can introduce greater computational and implementation complexity.

CNNs learn important spatial characteristics and structural relationships in input images by employing successive convolutional and pooling layers. As the data passes through these consecutive layers, increasingly relevant features are extracted and subsequently forwarded to the dense layers where the ultimate classification decision is generated. The fully connected part may consist of multiple hidden units just before reaching the final output stage. The last layer consists of output nodes representing the different categories defined in the dataset for classification, which enables the model to accurately assign the input sample to its corresponding category. Appropriate activation functions, model parameters, and hyperparameters are chosen to enhance the learning performance of the network and enhance its generalization capability and robustness on unseen test dataset. In addition, several techniques are employed to manage problems such as underfitting and overfitting. These methods enable the model to learn efficiently from the training data while ensuring consistent performance when tested on the test dataset.

Within retinal OCT analysis, pretrained and transfer-learning-based architectures such as NASNetMobile [7], AlexNet [8], VGG16 [9], VGG19 [9], ResNet50 [10], and GoogLeNet [11], DenseNet121 [12], Inception-ResNet-v2 [13], Xception [14], MobileNetV2 [15], EfficientNet-B0

[16] have been explored for learning discriminative image representations. Although these architectures have demonstrated their effectiveness in image classification, their comparatively large model structures and computational requirements may limit their suitability for resource-constrained deployment. Other OCT classification methods incorporate retinal-layer segmentation, region-of-interest extraction, image alignment, or ensemble learning to improve feature discrimination. While these strategies can enhance classification performance, they may also introduce additional processing stages and implementation overhead. These limitations motivate the development of a compact architecture that can learn disease-relevant representations directly from OCT images while avoiding dependence on complex preprocessing or large pretrained backbones.

Automated OCT image analysis plays a crucial role as a suggestive system between a patient and a doctor to identify anomalies in a grater detailed manner. Many efforts have been carried out over the years to achieve this by applying conventional and contemporary machine learning algorithms with various techniques. P. P. Srinivasan et al. [17] performed image reshaping, noise reduction, and cropping of the retinal images to extract Histogram of Oriented Gradients (HOG) descriptors [2]. For multi-class classification, a Support Vector Machine (SVM) classifier [5] was used with a one-versus-one strategy, resulting in three linear SVM classifiers corresponding to the three dataset classes: AMD, DME, and Normal. The retinal image dataset used in the study was locally developed using OCT images collected from patients with different retinal disorders. Y.-Y. Liu et al. [18] proposed a traditional OCT image classification method using multi-resolution and PCA-based [19] LBP features. An SVM classifier was employed to classify MH, ME, and AMD from a locally collected dataset. Alsaih K et al. [20] proposed a conventional OCT classification approach using Histogram of Oriented Gradients (HOG) [2] and Local Binary Patterns (LBP) [1] features with PCA for feature reduction. The optimized features were classified using a Support Vector Machine (SVM) [5] on a locally collected dataset.

On the other hand, several advanced techniques have been proposed for classifying retinal OCT images. L. Huang et al. [21] developed a layer-based CNN that learned features from individual retinal layers for OCT image classification. The extracted features were combined using transfer learning, achieving improved performance on the OCT2017 [22] and HUCM [21] OCT image datasets. J. Kim and L. Tran [23] proposed a two-stage deep learning model for OCT image classification. U-Net-based ROI extraction and image alignment enhanced feature learning, leading to better accuracy than L. Huang et al. [21] on

the OCT2017 [22] dataset. Shengyong Diao et al. [24] proposed CM-CNN and CAM-UNet for AMD classification from OCT images. Their approach improved region localization and achieved promising

performance. Hassan E. et al. [25] proposed the EOCT framework, combining ResNet-50 [10] feature extraction with a Random Forest classifier for OCT image classification.

Table 1. Overview of Existing Approaches for Retinal OCT Image Classification

| Reference               | Category               | Dataset Used  | Key Methodology                                                                     | Limitation                                                                                         |
|-------------------------|------------------------|---------------|-------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------|
| Liu et al. [18]         | Machine Learning       | Local dataset | SVM classifier using both global and local image features                           | Classification performance is highly dependent on accurate feature extraction and image alignment. |
| Alsaih et al. [20]      | Machine Learning       | Local dataset | HOG, LBP, and PCA features combined with an SVM classifier                          | Relies on handcrafted feature engineering, which may reduce adaptability to diverse datasets.      |
| Huang et al. [21]       | Deep Learning          | OCT2017, HUCM | Layer-guided CNN developed using transfer learning                                  | Requires retinal layer segmentation before classification, increasing preprocessing complexity.    |
| Kim and Tran [23]       | Deep Learning          | OCT2017       | U-Net-based preprocessing with image rotation followed by binary CNN classifiers    | Multiple preprocessing stages increase computational complexity and execution time.                |
| Diao et al. [24]        | Deep Learning          | Local dataset | CM-CNN integrated with CAM-Unet for retinal image analysis                          | Separate segmentation and classification networks make the framework more complex.                 |
| Hassan et al. [25]      | Hybrid Learning        | OCT2017       | ResNet-50 used for feature extraction followed by a Random Forest classifier (EOCT) | Hybrid architecture requires additional training effort and higher implementation complexity.      |
| Paul et al. [27]        | Ensemble Deep Learning | OCTx, OCT2017 | Ensemble of multiple deep learning models for retinal OCT classification            | Limited training data may restrict the model's ability to generalize to unseen clinical images.    |
| Yang et al. [28]        | CNN Ensemble           | OCT2017       | Ensemble learning strategy combining multiple CNN models                            | Increased memory usage and computational requirements during training and inference.               |
| Stanojević et al. [29]  | Deep Learning          | OCT2017       | Comparative evaluation of CNN models with InceptionV1 showing the best performance  | High computational demand limits its suitability for real-time applications.                       |
| Madhumithaa et al. [30] | Deep Learning          | OCT-C8        | NasNetMobile architecture optimized using the SGD optimizer                         | Computationally demanding, making deployment on resource-limited systems challenging.              |
| Dave and Domadiya [33]  | Deep Learning          | OCT2017       | CNN architecture incorporating hybrid pooling with an optimized training strategy   | Validation was limited to four retinal disease categories, reducing its broader applicability.     |

The model, trained using SGD and Adam, achieved state-of-the-art performance. On the OCT2017 dataset [22]. For the automatic categorization of normal and AMD retinal scans, Lee [26] suggested a 21-layer VGG16-based convolutional neural network trained on 2.6 million EMR-linked OCT pictures. By combining

scan-level predictions, the model attained 87.63% image-level, 88.98% OCT-level, and 93.45% patient-level accuracy (AUROC up to 97.46%). Nevertheless, the framework's clinical generalizability was constrained by its binary classification capability, lack

of external validation, and training using data from a single institution and OCT instrument.

In [27], the authors proposed OCTx, an ensemble deep learning model for four-class OCT image classification. It achieved strong performance but was evaluated on a limited dataset. In [28], the authors presented a CNN model using ImageNet pre-trained weights for three-class retinal OCT image classification. Fine-tuning and ensemble learning increased the classification accuracy from 68.17% to 92.89%. In [29], the authors evaluated AlexNet, VGG, Inception, and ResNet for retinal OCT image classification. The Inception model with RMSprop achieved the best results, reaching 95.53% accuracy and an F1-score of 0.93687. In [30], the authors applied NasNetMobile [7] to classify eight retinal diseases using the OCT-C8[31] dataset. Among the tested optimizers, SGD delivered the best results with 97% accuracy and stable convergence within 22 epochs.

Jannat et al. [32] presented OCT-SelfNet as a SwinV2-based self-supervised framework that combines masked autoencoder pre-training with cross-dataset data fusion, achieving 99% classification accuracy with high AUC-ROC, AUC-PR, and F1-score. However, its evaluation is limited to binary classification (AMD vs. Normal), restricting its applicability to multi-class retinal disease diagnosis. Dave and Domadiya et al. [33] formulated a CNN-based model for retinal OCT image classification. Using hybrid pooling, data augmentation, and learning rate decay, the model achieved 98.75% accuracy on the OCT2017 [22] dataset. A thorough analysis of earlier machine learning and deep learning approaches used for OCT image classification is summarized in Table 1. Despite the progress reported in OCT image classification, an important trade-off remains between predictive performance and model complexity. Layer-guided and segmentation-based approaches require additional anatomical processing, transfer-learning methods depend on comparatively large pretrained backbones, and ensemble strategies increase computational and memory requirements. Moreover, reported evaluations are often limited to a single dataset or a restricted set of disease categories, making it difficult to assess whether high within-dataset performance translates to robustness across different data distributions. Therefore, there remains a need for a compact CNN architecture that combines effective feature selection with efficient computation while being evaluated across multiple OCT datasets. To address this research gap, this study proposes a lightweight CNN-based model for retinal OCT image classification using the OCT-C8 [31] and OCT2017 [22] datasets. The proposed architecture integrates Squeeze-and-Excitation (SE) blocks within a compact 13-layer CNN to enhance channel-wise feature representation, together with global average

pooling and dropout for efficient feature learning and improved generalization. Unlike approaches that rely on pretrained backbones with additional SE modules, the proposed network incorporates channel-wise feature recalibration directly into the CNN architecture, while data augmentation and the Adam optimizer support robust and stable training. The model contains approximately 375,000 trainable parameters, maintaining a compact structure while retaining the capacity to learn diagnostically relevant retinal features. Its effectiveness is assessed through within-dataset classification, cross-dataset validation, ablation experiments, and computational-efficiency analysis, providing a comprehensive evaluation of both predictive performance and model efficiency.

The remainder of this manuscript is organized as follows: Section II describes the proposed methodology, Section III presents the experimental results, Section IV discusses the findings and limitations, and Section V concludes the study and outlines future research directions.

## II. Method

The detailed architecture of the proposed framework is shown in Fig. 1. The proposed framework mainly has five working modules. The first one is the input datasets; secondly, the preprocessing and augmentation; thirdly, a CNN architecture with SE blocks; the second-to-last module is selecting a proper optimizer for the multiclass classification; and the last one is the evaluation of the model using two different datasets.

### A. Data Preparation

The OCT-C8 dataset [31] contains 8 classes: Drusen, MH, AMD, CNV, CSR, DME, DR, and Normal. There are 2,800 validation images (350 per class), 2800 test images (350 per class), and 18,400 training images (2,300 per class). Feature extraction is more difficult because each image has a resolution of 96 dpi with a 24-bit depth and contains noise and different intensity patterns. The OCT2017 dataset [22] comprises 83,484 training images, 968 test images (242 per class), and 32 validation images, divided into four classes: CNV, DME, Drusen, and Normal. As example samples, the training images vary in size and contain noise. The publicly available OCT dataset was used as provided by its original authors, with images collected from different patients. Before model training, the images were resized to 200 × 200 pixels using bilinear interpolation, converted to RGB format, and normalized to the range [0,1]. No additional quality screening or duplicate removal was performed, as the dataset was used in its original curated form.

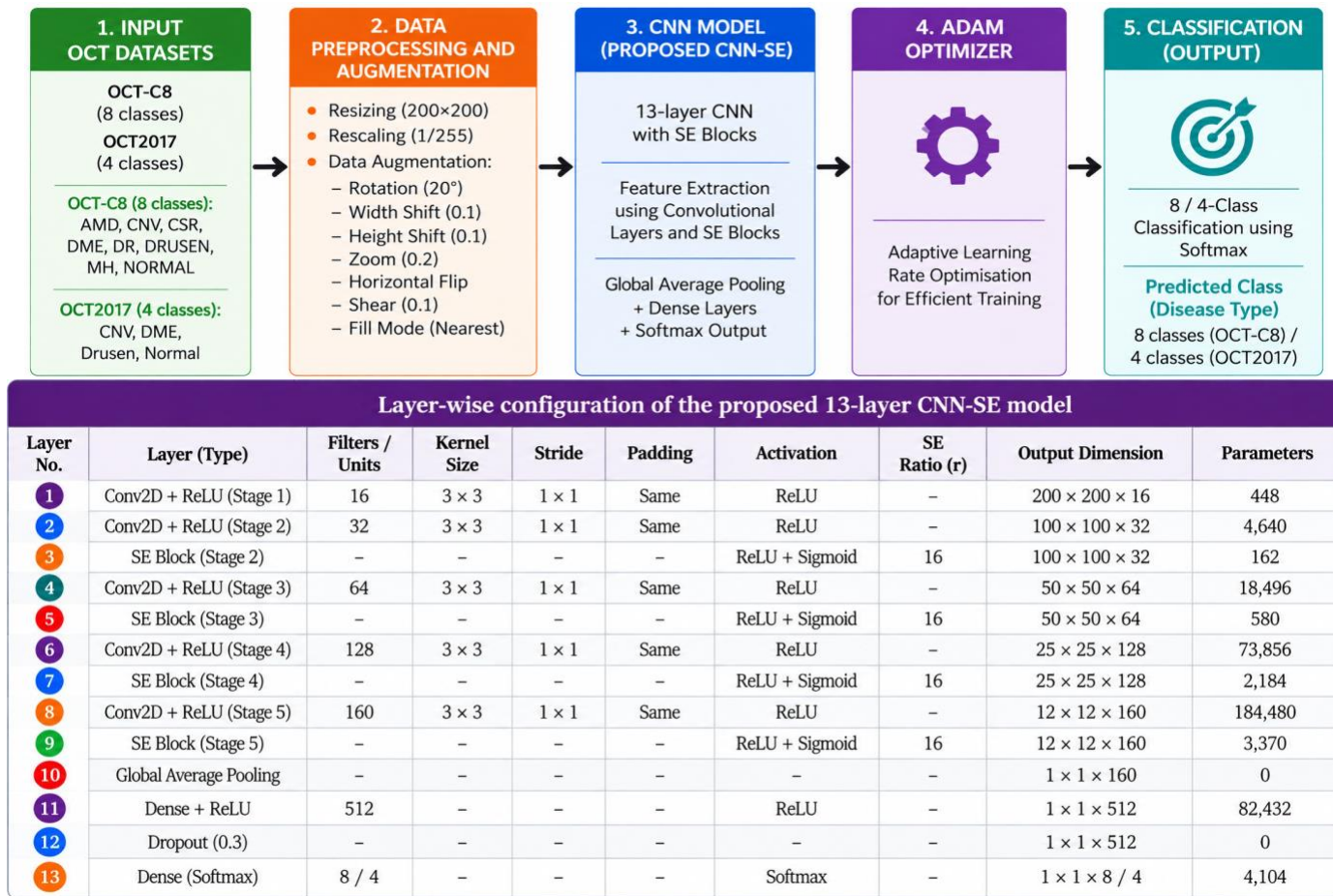


Fig. 1 Proposed CNN Model (OCT-C8: 8 Class, OCT2017: 4 Class)

## A. Data Augmentation

By transforming the source images, image augmentation creates a variety of training examples that aid CNNs in learning minor differences across comparable classes [34]. This lessens overfitting, increases generalization, and improves feature representation. Table 2 lists the augmentation settings utilized in the suggested strategy, which creates augmented images during training without storing them permanently. To improve the model's ability to generalize while preserving the anatomical characteristics of retinal OCT images, augmentation was applied exclusively to the training set. The training images were subjected to controlled transformations, including horizontal flipping, small-angle rotation, limited horizontal and vertical translation, moderate zooming, and mild shearing, to introduce realistic variations in image appearance and spatial alignment. These transformations increase the diversity of the training distribution and encourage the network to learn stable, discriminative retinal features rather than relying on specific image orientations or local patterns. Vertical flipping and excessive geometric transformations were

deliberately avoided because they may alter the natural arrangement of retinal layers and produce anatomically unrealistic pathological patterns. The augmented samples were generated online during training, so the original images were not permanently duplicated, and no augmentation was applied to the validation or test sets. Representative transformed images were also visually inspected to verify that retinal-layer continuity, tissue boundaries, and disease-related morphological structures remained identifiable after augmentation.

## B. Squeeze-and-Excitation Blocks

The Squeeze-and-Excitation (SE) block [35] by finding the most informative feature channels and simulating channel-wise dependencies, it enhances CNN feature learning. It produces more selective representations of features by enhancing significant traits while suppressing less important ones. The overall architecture for the SE block is depicted in Fig. 2. The operation of the SE block is divided into three key stages: squeeze, excitation, and recalibration. In the squeeze stage, global average pooling is applied to the input feature map  $X \in R^{H \times W \times C}$  to produce a channel-wise descriptor. In this stage, each channel's spatial

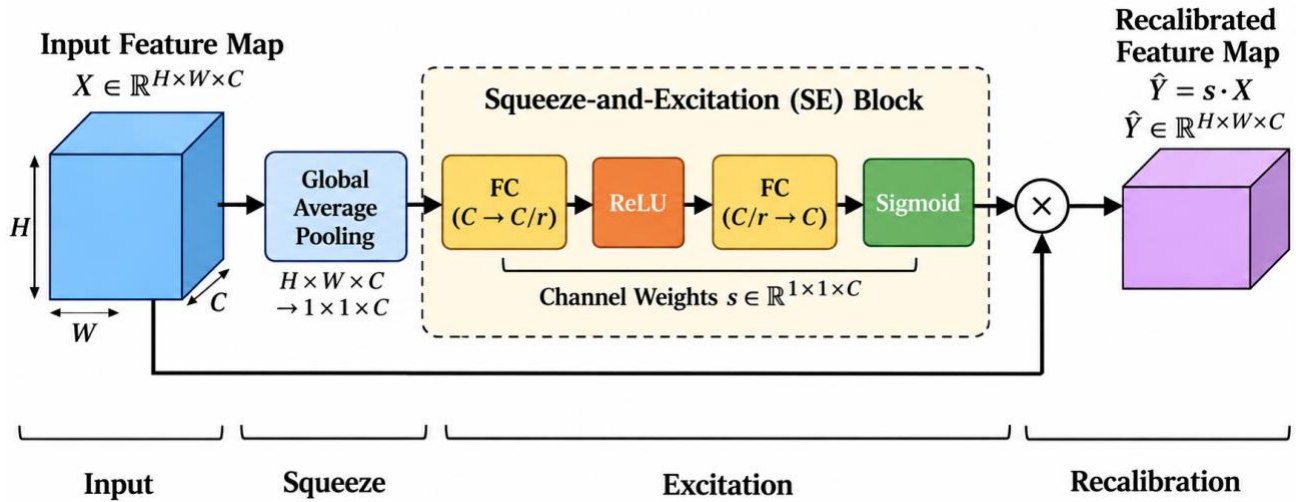


Fig. 2. Architecture of SE block (Reproduced from [35])

dimensions  $H \times W$  are reduced into one representative value, effectively capturing the global distribution of feature responses. Mathematically, the output of the squeeze operation for the  $C^{th}$  channel is computed as Eq. (1)[35] where  $Z_c$  represents the aggregated statistic of the  $C^{th}$  channel. This descriptor encodes the global contextual information and reflects the overall importance of each channel.

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (1)$$

After the squeeze stage, the excitation module learns the importance of each channel using two dense layers with a bottleneck structure. ReLU [36] and sigmoid activations generate normalized channel-wise weights that emphasize informative features while reducing computational cost. This process can be expressed as Eq. (2)[35]

$$S = \sigma(W_2 * \delta(W_1 * z)) \quad (2)$$

where “ $z$ ” is the squeezed vector,  $W_1$  and  $W_2$  are the learnable parameters of the fully connected layers,  $\delta$  denotes the ReLU [36] activation, and  $\sigma$  represents the sigmoid function. In the final feature refinement step, the learned channel parameters are applied to the actual feature map. Each feature channel is scaled based on its assigned importance value, enabling the network to highlight the most meaningful features while minimizing the influence of less important ones. This is given by Eq. (3).

$$\tilde{X} = S_c * X_c \quad (3)$$

where  $S_c$  is the learned weight for the  $C^{th}$  channel. This feature refinement process improves the network's ability to learn meaningful channel-wise information by giving major emphasis to the most pertinent attributes while simultaneously diminishing the impact of less beneficial ones. As a result, the model achieves better classification performance with only a small increase in

computational cost. The global average pooling of SE block can be depicted by Eq. (4)[35]

$$G_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (4)$$

### C. Adam Optimizer

The Adam optimizer is a widely used method for training deep learning models due to its ability to perform well in a variety of scenarios. It integrates the concepts of adaptive learning and momentum, allowing the model to adjust how much each parameter is

Table 2. Configuration Details of Image Transformation Techniques

| Parameter Name            | Parameter Value |
|---------------------------|-----------------|
| Pixel value normalization | 1./255          |
| Rotation angle            | 20°             |
| Horizontal shift          | 0.1             |
| Vertical shift            | 0.1             |
| Zoom factor               | 0.2             |
| Horizontal flipping       | True            |
| Shear intensity           | 0.1             |
| Fill mode                 | nearest         |

updated during training. Instead of using the same learning rate for all parameters, Adam adapts it based on the behavior of the gradients, which helps the model learn more efficiently and with better stability. Eq. (5)[37] computes the first moment vector using the first hyperparameter (typically 0.9) to maintain an exponentially decaying average of past gradients that tracks directional momentum.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (5)$$

Eq. (6)[37] calculates the uncentered second moment vector using the second hyperparameter (typically 0.999) to track gradient variance by maintaining an exponentially decaying average of squared gradients.

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (6)$$

Eq. (7)[37] yields the bias-corrected first moment estimate by dynamically adjusting the time step, which compensates for the moving average's initial bias toward zero.

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \quad (7)$$

Eq. (8) [37] derives the bias-corrected second moment estimate by adjusting for the time step, correcting the initial zero-bias in variance tracking during early iterations.

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \quad (8)$$

Eq. (9)[37] updates the model parameters  $\theta$  by adjusting the learning rate with the ratio of these bias-corrected moments, providing adaptive step sizes stabilized by a small constant.

$$\theta = \theta - \alpha * \frac{m}{(\sqrt{v}) + (\epsilon)} \quad (9)$$

where  $m_t$  and  $v_t$  denote the first- and second-moment estimates,  $\hat{m}_t$  and  $\hat{v}_t$  are their bias-corrected forms,  $g_t$  represents the gradient at iteration  $t$ ,  $\beta_1$  and  $\beta_2$  are the decay rates,  $\alpha$  is the learning rate, and  $\epsilon$  epsilon is a small constant for numerical stability.  $m$  captures the average of past gradients, and  $v$  represents the average of their squared values. These estimates are continuously updated during training, helping the optimizer adapt to the shape of the loss function.

#### D. CNN Architecture

The proposed CNN processes 200×200×3 OCT images through successive convolution and max-pooling layers to learn hierarchical image features. SE blocks are incorporated after each stage to emphasize important feature channels while suppressing less relevant ones. The gradual increase in filters enables effective feature extraction, leading to improved retinal OCT image classification. The extracted feature maps are condensed using global average pooling and passed to a 512-neuron dense layer with 0.3 dropout to improve generalization. A softmax layer then produces the final class predictions for the OCT-C8 and OCT2017 datasets. A 2x2 max-pooling procedure

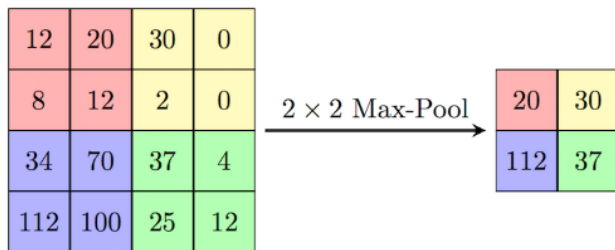


Fig. 3. Illustration of 2x2 max pooling

follows each convolutional layer to preserve the most informative features while reducing the size of the feature map by keeping just the strongest activations. As seen in Fig. 3 and Eq. (10)[11], this increases the

computing efficiency and aids in the model's learning of more robust features.

$$\text{MaxPool}(x) = \max(x_{ij})_{1 \leq i \leq k, 1 \leq j \leq k} \quad (10)$$

The ReLU [36] activation function is used throughout the CNN to introduce non-linearity by retaining positive values and suppressing negative ones. Its computational efficiency and ability to mitigate the vanishing gradient problem enable faster and more stable network convergence. The mathematical equation is depicted by Eq. (11) [36]

$$F(x) = \max(x, 0) \quad (11)$$

The 6×6×160 feature maps are compressed by global average pooling into a 1×1×160 feature vector following the final convolution and max-pooling stage, and then they are sent to a 512-neuron dense layer for discriminative feature learning. To lessen overfitting and enhance model generalisation, a dropout rate of 0.3 is used during training. Eight neurons are used in the last output layer for the OCT-C8 [31] dataset and four neurons for the OCT2017[22] dataset, with each neuron representing a distinct class to generate the final classification results. The last layer uses the softmax activation function to calculate the likelihood of each output class. Using these probabilities, the framework assigns the input picture to the class with the greatest confidence score. The suggested model minimizes the discrepancy between the true and predicted class labels during training by using categorical cross-entropy loss for multi-class classification.

The equation, Eq. (12)[8][10], represents a 2D convolutional layer where a filter matrix with a bias value slides across an input image to calculate a weighted sum at each overlapping region. This mathematical operation extracts localized spatial features from the input data to produce a specific coordinate pixel value within the output feature map, as detailed in Eq. (12)[8][10].

$Y(i,j) = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} X(i+m,j+n) * W(m,n) + b$  (12)  
The spatial dimensions of the output feature map after convolution are calculated in Eq. (13)[8], which determines the final spatial output size of a feature map by adding twice the padding (P) to the input dimension (N) and subtracting the kernel size (K). This result is then divided by the stride rate (S), rounded down to the nearest integer using the floor function, and increased by one to account for the final filter placement, as outlined in Eq. (13)[8].

$$O = \left\lfloor \frac{N-K+2P}{S} \right\rfloor + 1 \quad (13)$$

The mathematical formulation of the softmax function is presented in Eq. (14)[11]. The function transforms the output values of the network into normalized probability scores ranging from 0 to 1, with the total probability across all classes equal to one. In Eq. (14)[11], Z

represents the input vector received by the output layer, and  $j$  refers to an individual output neuron, where  $j = 1, 2, \dots, K$ .

$$\sigma(z)_j = \frac{e^{(z)_j}}{\sum_{k=1}^K e^{(z)_k}} \quad (14)$$

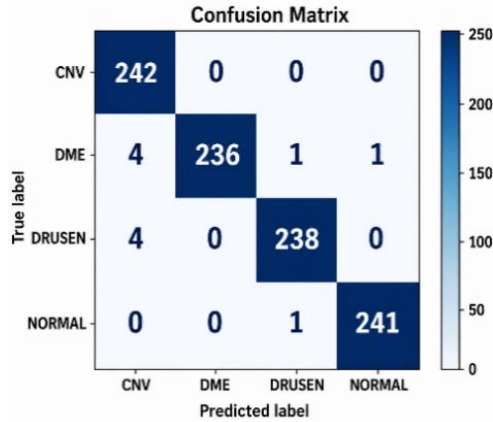
The dense layer transforms the extracted features into a higher-level representation as shown in Eq. (15)[9], where  $(W)$  denotes the weight matrix,  $(x)$

represents the input feature vector,  $(b)$  is the bias, and  $(f(\cdot))$  denotes the ReLU activation function. This transformation enables the network to learn complex feature representations for accurate classification

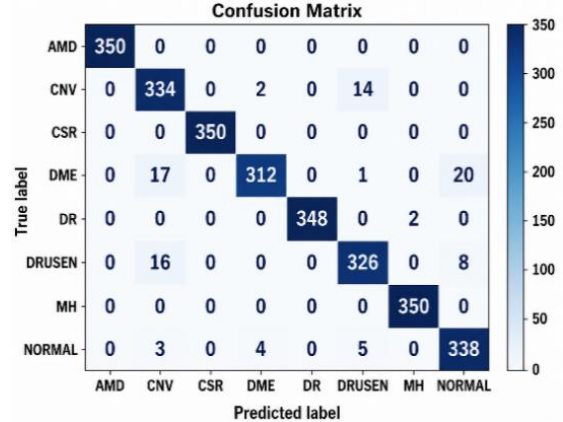
$$h = f(Wx + b) \quad (15)$$

As discussed earlier, dropout can be denoted by Eq. (16)[11], which reduces overfitting to make the model more generalizable.

$$h_{drop} = r \odot h, \quad r \sim \text{Bernoulli}(p) \quad (16)$$



(a)



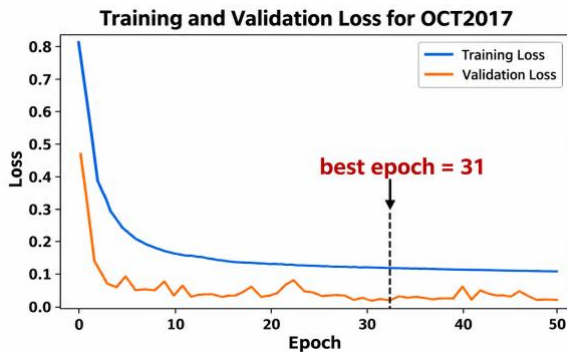
(b)

**Fig. 4.** Confusion matrix for (a) OCT2017 and (b) OCT-C8 test datasets, showing raw class-wise prediction counts.

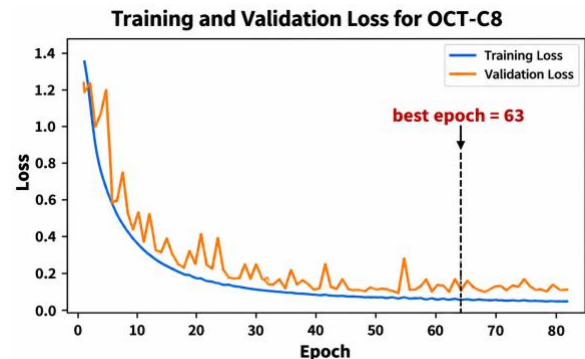
### III. Results

The proposed CNN-SE model was implemented in Python 3.12.13 using TensorFlow 2.21.0, Keras, scikit-learn, and Matplotlib on an AMD Ryzen 7 5700G (3.80 GHz) CPU with 16 GB RAM. The model used  $200 \times 200 \times 3$  input images, a batch size of 32, Adam optimization with a learning rate of 0.001, categorical cross-entropy loss, an SE reduction ratio of 16, and a dropout rate of 0.30. Training was performed for a maximum of 100 epochs with early stopping (patience = 20), retaining the best-performing weights;

convergence occurred at epoch 31 for OCT2017 [22] and epoch 63 for OCT-C8 [31]. The same hyperparameter configuration was applied to both datasets without test-set-specific tuning. OCT-C8 [31] contained 18,400 training and 2,800 validation/testing images, whereas OCT2017 [22] contained 83,484 training, 32 validation, and 968 testing images. The lightweight CNN-SE model required approximately 13 h/109 ms for training/testing on OCT2017 and 6 h/144 ms on OCT-C8. Performance was comprehensively evaluated using confusion matrices, training-validation curves, and class-wise evaluation metrics.



(a)



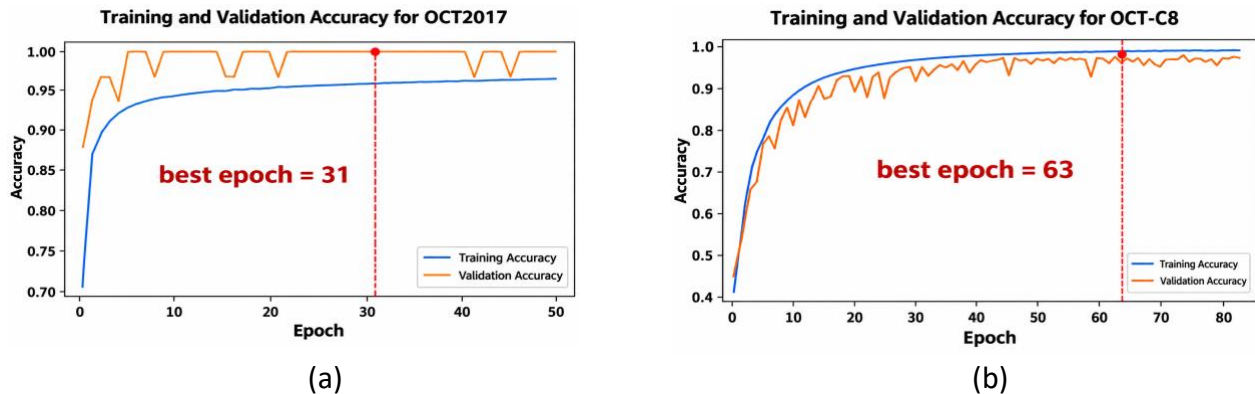
(b)

**Fig. 5.** Training and validation loss curves for the proposed CNN-SE model: (a) OCT2017 dataset and (b) OCT-C8 dataset.

The confusion matrices for the OCT2017 and OCT-C8 test datasets are presented in Fig. 4(a) and Fig. 4(b) respectively. Most predictions are concentrated along the diagonal, indicating effective discrimination across the evaluated retinal categories, although a small number of errors were observed, particularly for DME and Drusen in OCT2017. The performance difference between the datasets can be partly attributed to the larger number of classes and higher degree of inter-class morphological similarity in OCT-C8. In this dataset, the remaining errors were mainly observed among CNV, DME, Drusen, and Normal, where overlapping OCT characteristics such as retinal thickening, intraretinal or subretinal fluid, and variations in retinal reflectivity can make class boundaries less distinct. Subtle or localized abnormalities may also cause some diseased scans to

resemble normal retinal structures. In contrast, AMD, CSR, DR, and MH showed comparatively fewer misclassifications, suggesting that their structural patterns were more readily distinguished by the learned feature representations. Overall, the observed errors were primarily associated with visually similar or less pronounced retinal abnormalities, while the strong diagonal pattern across both confusion matrices indicates that these cases had limited impact on the overall classification performance.

The training and validation loss curves for the OCT2017 and OCT-C8 datasets are shown in Fig. 5(a) and Fig. 5(b). Smooth decreases in both curves with a small training-validation gap indicate stable convergence without substantial overfitting. Owing to its larger training set, fewer classes, and lower inter-class



**Fig. 6.** Training and validation accuracy curves for the proposed CNN-SE model: (a) OCT2017 dataset and (b) OCT-C8 dataset.

complexity, OCT2017 converges faster than OCT-C8. The corresponding training and validation accuracy curves are presented in Fig. 6(a) and Fig. 6(b), showing a consistent increase with closely aligned curves, which indicates stable learning and good generalization. Although OCT-C8 converges more slowly because of its higher-class complexity and inter-class similarity, the model maintains strong classification performance on both datasets.

Table 3 summarizes the class-wise classification performance of the proposed CNN-SE model on the OCT2017 and OCT-C8 datasets. On OCT2017, the model achieved an overall accuracy of 99%, with precision, recall, and F1-score values close to 1.00 across CNV, DME, Drusen, and Norma. For OCT-C8, the model obtained an overall accuracy of 97%, with the highest class-wise performance observed for CNV, DME, Drusen, and Normal, while AMD, CSR, DR, and MH also showed strong classification results. The macro- and weighted-average metrics were consistently 0.99 for OCT2017 and 0.97 for OCT-C8, indicating balanced performance across the evaluated

classes despite the higher-class diversity and morphological similarity in OCT-C8.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (17)$$

$$Precision = \frac{TP}{TP+FP} \quad (18)$$

$$Recall = \frac{TP}{TP+FN} \quad (19)$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (20)$$

The model was evaluated using Eqs. (17), (18), (19) and (20)[11].

#### A. Cross-Dataset Validation

To further examine domain robustness, the CNN-SE model was trained exclusively on OCT-C8 and evaluated directly on the OCT2017 dataset without retraining or fine-tuning, using the four shared diagnostic classes, namely CNV, DME, Drusen, and Normal. Across five repeated experiments with random seeds 11, 22, 33, 44, and 55, the model achieved an overall accuracy of  $98.74 \pm 0.11\%$ , with weighted precision, recall, and F1-score of  $98.75 \pm 0.12\%$ ,  $98.74 \pm 0.11\%$ , and  $98.74 \pm 0.11\%$ , respectively. The corresponding macro-ROC-AUC of  $0.9995 \pm 0.0002$

Table 3. Class-wise Classification Results of CNN-SE Model for OCT2017 and OCT-C8 datasets

|              | OCT2017   |        |          |         | OCT-C8    |        |          |         |
|--------------|-----------|--------|----------|---------|-----------|--------|----------|---------|
|              | Precision | Recall | F1-score | Support | Precision | Recall | F1-score | Support |
| Class        |           |        |          |         |           |        |          |         |
| AMD          | NA        |        |          |         | 1.00      | 1.00   | 1.00     | 350     |
| CNV          | 0.97      | 1.00   | 0.98     | 242     | 0.90      | 0.95   | 0.93     | 350     |
| CSR          | NA        |        |          |         | 1.00      | 1.00   | 1.00     | 350     |
| DME          | 1.00      | 0.98   | 0.99     | 242     | 0.98      | 0.89   | 0.93     | 350     |
| DR           | NA        |        |          |         | 1.00      | 0.99   | 1.00     | 350     |
| Drusen       | 0.99      | 0.98   | 0.99     | 242     | 0.94      | 0.93   | 0.94     | 350     |
| MH           | NA        |        |          |         | 0.99      | 1.00   | 1.00     | 350     |
| Normal       | 1.00      | 1.00   | 1.00     | 242     | 0.92      | 0.97   | 0.94     | 350     |
| Accuracy     | 0.99      |        |          |         | 0.97      |        |          |         |
| Macro Avg    | 0.99      | 0.99   | 0.99     | 968     | 0.97      | 0.97   | 0.97     | 2800    |
| Weighted Avg | 0.99      | 0.99   | 0.99     | 968     | 0.97      | 0.97   | 0.97     | 2800    |

Table 4. Cross-Dataset Testing Performance of the Proposed CNN-SE Model Across Five Random Seeds.

| Metric                     | Mean     | SD       | Mean ± SD (%)   |
|----------------------------|----------|----------|-----------------|
| Accuracy                   | 0.987397 | 0.001132 | 98.74 ± 0.11    |
| Weighted Precision         | 0.987460 | 0.001153 | 98.75 ± 0.12    |
| Weighted Recall            | 0.987397 | 0.001132 | 98.74 ± 0.11    |
| Weighted F1-score          | 0.987397 | 0.001141 | 98.74 ± 0.11    |
| Macro F1-score             | 0.987397 | 0.001141 | 98.74 ± 0.11    |
| Macro ROC-AUC              | 0.999545 | 0.000236 | 0.9995 ± 0.0002 |
| Compatible Prediction Rate | 1.000000 | 0.000000 | 100.00 ± 0.00   |

indicates consistent discrimination across the evaluated classes, while the 100.00 ± 0.00% compatible prediction rate confirms that all evaluated samples were mapped to the intended diagnostic categories across the repeated experiments. These findings provide evidence of stable cross-dataset robustness and generalization, summary statistics in Table 4.

B. Computational Efficiency and Inference Performance

The computational characteristics of the proposed CNN-SE were further examined against MobileNetV2, EfficientNetB0, and NASNetMobile using a consistent evaluation procedure as described in Table 5. With only 0.3748 M trainable parameters, 0.3670 G FLOPs, and 183.52 M MACs, the proposed network occupied just 1.54 MB and required an average of 9.40 ms to process an image. It also recorded the highest inference throughput among the evaluated models, reaching

106.4 images/s. These measurements show that the proposed architecture achieves its feature-recalibration capability without introducing the computational burden typically associated with larger networks. The resulting combination of a small model footprint and rapid inference indicates a favorable efficiency profile for retinal OCT classification, particularly in settings where computational resources may be limited. FLOPs were calculated using TensorFlow’s graph profiler with a fixed input size of 200 × 200 × 3, using the convention that one MAC corresponds to two FLOPs. Inference latency was measured at batch size 1 after 20 warm-up runs, followed by 100 timed runs, using the same evaluation procedure for all models.

In terms of diagnostic performance, the proposed CNN–SE model achieved 97% accuracy on the OCT-C8 test set, with a macro-AUROC of 0.9985, indicating strong discriminative capability. In the independent

Table 5. Computational Efficiency Comparison of CNN Models under Consistent Conditions

| Model            | Parameters<br>(Millions) | FLOPs<br>(G) | MACs<br>(M) | Model<br>Size<br>(MB) | Peak<br>Memory<br>(MB) | Latency<br>(ms/image) | Throughput<br>(images/s) |
|------------------|--------------------------|--------------|-------------|-----------------------|------------------------|-----------------------|--------------------------|
| Proposed CNN- SE | 0.3748                   | 0.3670       | 183.52      | 1.54                  | 0.0012                 | 9.40                  | 106.40                   |
| MobileNetV2      | 2.2682                   | 0.5342       | 267.09      | 9.21                  | 0.0039                 | 42.13                 | 23.74                    |
| EfficientNetB0   | 4.0598                   | 0.7083       | 354.16      | 16.29                 | 0.0312                 | 62.01                 | 16.13                    |
| NASNetMobile     | 4.2782                   | 0.9983       | 499.15      | 18.95                 | 0.0703                 | 82.93                 | 12.06                    |

cross-dataset evaluation, where the model was trained on OCT-C8 and tested on OCT2017 across five random seeds as stated earlier, it achieved  $98.74 \pm 0.11\%$  accuracy and  $0.9995 \pm 0.0002$  macro-AUROC. The corresponding MCC ( $0.9832 \pm 0.0015$ ) and Cohen's  $\kappa$  ( $0.9832 \pm 0.0015$ ) further demonstrate strong agreement and stable classification performance on the external dataset as presented in Table 6.

C. Ablation Study

To systematically assess the individual contributions of the SE blocks, data augmentation, and optimizer

guaranteeing stable optimization, the suggested Model E which combined SE blocks, data augmentation, and the Adam optimizer achieved the best overall performance. The ablation findings verify that each element adds to the model's efficacy, and when combined, they produce the most robust and well-balanced retinal OCT image. The results of the ablation study are briefly presented in Table 7

. To examine the contribution of the main components, the ablation results were analyzed through controlled pairwise comparisons using the configurations

Table 6. Within-Dataset and Cross-Dataset Validation Results with Four Shared Classes

| Train Dataset: Test Dataset | Test Set (#) | Accuracy (%)     | Macro-AUROC         | MCC                 | Cohen's $\kappa$    |
|-----------------------------|--------------|------------------|---------------------|---------------------|---------------------|
| OCT-C8 → OCT-C8             | 2,800        | 97               | 0.9985              | 0.9626              | 0.9624              |
| OCT-C8 → OCT2017            | 968          | $98.74 \pm 0.11$ | $0.9995 \pm 0.0002$ | $0.9832 \pm 0.0015$ | $0.9832 \pm 0.0015$ |

selection under the same training and testing settings, ablation research was carried out using the OCT2017 and OCT-C8 datasets. At the beginning, a basic CNN model was trained without including SE blocks and without applying data augmentation. We call it Model A for the denotation which is shown in Fig. 1, in that SE blocks are not included. This architecture showed lower classification performance, especially for classes like CNV, DME, and Drusen where the features are quite similar. Since no augmentation was used, the training data lacked variation, which made it harder for the model to generalize well on unseen test dataset images. In the second model implementation, Model B, data augmentation was added while keeping the model architecture the same as Model A. With more varied training samples, the model was able to learn better patterns, which improved overall accuracy and reduced overfitting. There was also a clear improvement in recall, especially for minority classes and those with

similar visual features. While Model D indicated that substituting SGD for Adam led to slower convergence and somewhat reduced accuracy, Model C showed that adding SE blocks enhanced feature representation and classification performance, especially for structurally related retinal disorders. By strengthening feature learning, boosting generalization, and

reported in the study presented by Table 8. Removing the SE blocks reduced OCT2017 accuracy from 97.9% to 96.2%, while adding augmentation improved the baseline from 96.2% to 97.8%; similarly, replacing SGD with Adam increased OCT2017 accuracy from 95.3% to 99.0% and OCT-C8 accuracy from 94.1% to 97.0%. These comparisons indicate that SE blocks, data augmentation, and optimizer selection contribute to the overall performance of the proposed model. Among the tested factors, the optimizer changes from SGD to Adam produced the largest observed improvement. Alternative SE reduction ratios, pooling strategies, and dropout rates were not included in the reported experiments; therefore, no unsupported conclusions are drawn for these components.

IV. Discussion

A. Comparison with Existing Approaches

Table 9 presents a comprehensive comparison of the proposed model against existing state-of-the-art techniques evaluated on the OCT2017 and OCT-C8 datasets. This comparative analysis was conducted to contextualize the performance of the proposed CNN-SE architecture relative to previously published approaches,

taking into account differences in network depth, parameter count, and optimization strategies employed across the literature. Such benchmarking is essential not only to validate the effectiveness of the proposed method but also to identify the specific architectural choices that contribute to performance gains over conventional CNN-based classifiers. The proposed CNN-SE achieves a competitive accuracy of 99.00% on OCT2017, exceeding the 88.40–98.75% range of the compared approaches as shown in Table 9. This performance is achieved with only 0.375 M trainable parameters, substantially fewer than the reported

Table 7. Ablation study of different model implementations

| Model                       | SE Blocks | Data Augmentation | Optimizer | Accuracy (OCT2017) | Accuracy (OCT-C8) |
|-----------------------------|-----------|-------------------|-----------|--------------------|-------------------|
| Model A                     | No        | No                | Adam      | 96.2%              | 92.8%             |
| Model B                     | No        | Yes               | Adam      | 97.8%              | 94.6%             |
| Model C                     | Yes       | No                | Adam      | 97.9%              | 94.87%            |
| Model D                     | Yes       | Yes               | SGD       | 95.3%              | 94.1%             |
| Model E<br>(Proposed Model) | Yes       | Yes               | Adam      | 99.0%              | 97.0%             |

Table 8. Component wise ablation study analysis (pp = percentage point)

| Comparison        | Component isolated     | OCT2017 Accuracy Change | OCT-C8 Accuracy Change |
|-------------------|------------------------|-------------------------|------------------------|
| Model A → Model B | Data augmentation      | +1.60 pp                | +1.80 pp               |
| Model A → Model C | SE blocks              | +1.70 pp                | +2.07 pp               |
| Model D → Model E | Optimizer (SGD → Adam) | +3.70 pp                | +2.90 pp               |
| Model C → Model E | Data augmentation      | +1.10 pp                | +2.13 pp               |

models such as Huang et al. [21]. (24.88 M), Kim et al. [23]. (235.6/452.1 M), Hassan et al. [25] (25.64 M), Opoku et al.[38] (8.2 M), Stanojević et al.[29] (6.6 M), and Yang et al. [28] (21.4 M). Unlike these studies, the

Table 9. Comparison with existing approaches

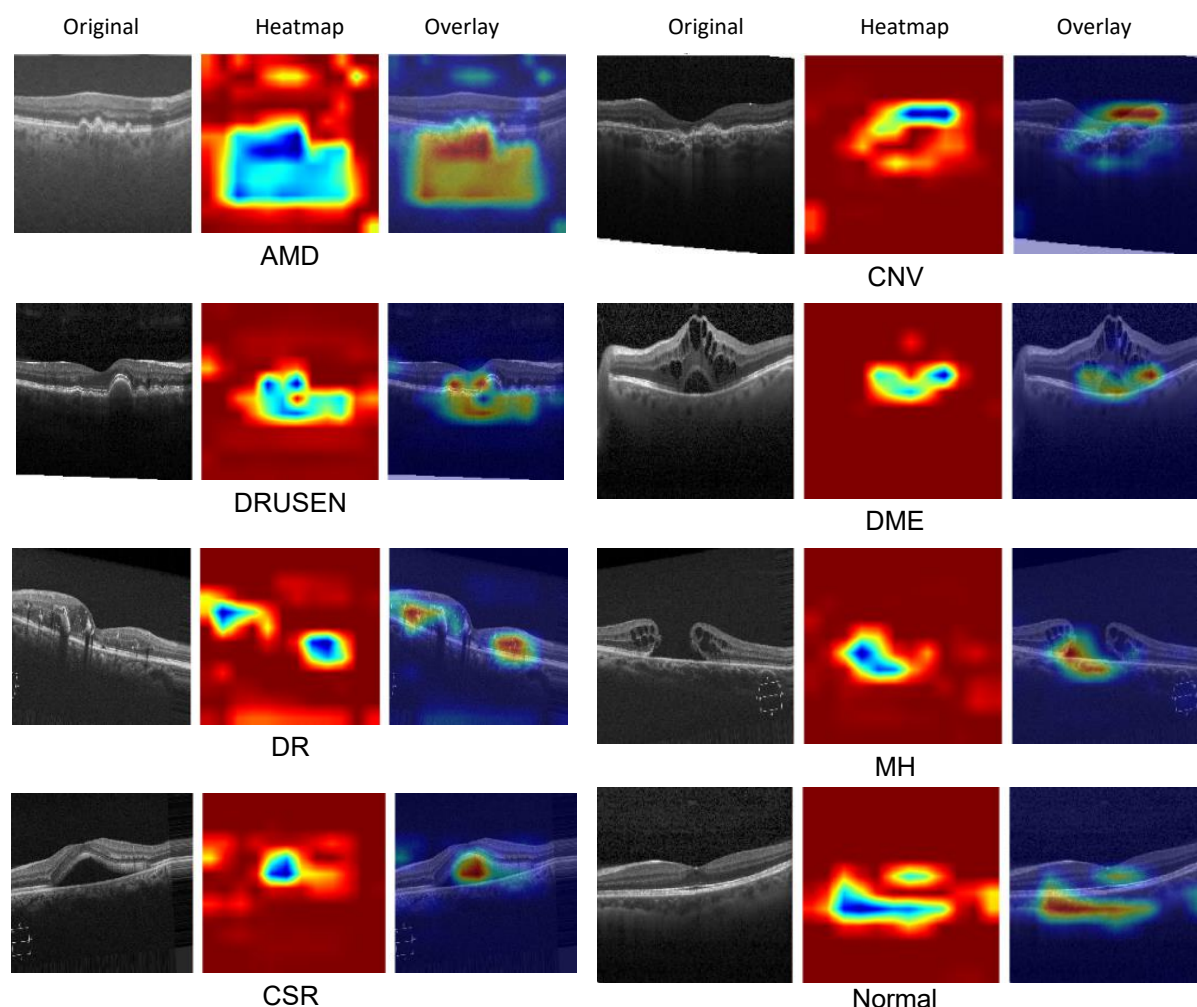
| Publication Year | References                    | Approach                                                                                    | Acc (OCT2017 Dataset) | Acc (OCT-C8 Dataset) |
|------------------|-------------------------------|---------------------------------------------------------------------------------------------|-----------------------|----------------------|
| 2019             | L.Huang et al. [21]           | Layer-guided convolutional neural network                                                   | 88.4                  | NA                   |
| 2020             | Le et al.[39]                 | Transfer learning                                                                           | 87.20                 |                      |
| 2020             | Saraiva et al. [40]           | Deep CNN architecture                                                                       | 93.3                  |                      |
| 2021             | J Kim et al. [23]             | Two binary CNN variants (Model 1 / Model 2 with high computational complexity)              | 98.1/98.7             |                      |
| 2022             | Stanojević, M et al. [29]     | Deep convolutional neural network model                                                     | 95.55                 |                      |
| 2023             | Hassan E et al [25]           | Enhanced Optical Coherence Tomography (EOCT)-based framework                                | 97.47                 |                      |
| 2023             | Diao et al. [24]              | CNN-based classification framework                                                          | 96.93                 |                      |
| 2023             | Michael Opoku et al. [38]     | Capsule network integrated with contrast-limited adaptive histogram equalization            | 97.7                  |                      |
| 2024             | Yang J et al. [28]            | CNN, EfficientNet-V2, and ResNet-based ensemble architecture                                | 97.89                 |                      |
| 2025             | Dave and Domadiya et al. [33] | CNN with data augmentation, exponential learning rate decay, and hybrid max–average pooling | 98.75                 |                      |
| 2026             | Proposed Model                | CNN enhanced with SE blocks, Adam optimization, and global average pooling                  | 99.00                 | 97.00                |

proposed model is evaluated on both four-class OCT2017 and eight-class OCT-C8 datasets, achieving 99.00% and 97.00% accuracy, respectively. To further assess generalization across dataset distributions, cross-dataset validation was performed by training exclusively on the OCT-C8 training set and evaluating on the independent OCT2017 test set without additional training or fine-tuning. Across repeated runs, this evaluation achieved  $98.74 \pm 0.11\%$  accuracy,  $98.75 \pm 0.12\%$  weighted precision,  $98.74 \pm 0.11\%$  weighted recall and F1-score, and a macro-ROC-AUC of  $0.9995 \pm 0.0002$ . These results provide additional evidence of consistent generalization beyond a single dataset distribution.

### A. Explainable AI

Grad-CAM analysis was performed to examine the spatial features contributing to the predictions of the proposed CNN-SE model across the eight OCT-C8 classes, as illustrated in Fig. 7. The activation maps predominantly highlighted localized retinal regions associated with class-discriminative structural patterns,

with stronger responses observed around abnormal retinal morphology in AMD, CNV, CSR, DME, DR, Drusen, and MH, whereas Normal samples generally exhibited more localized activation within the retinal layers. The concentration of activations around relevant retinal structures suggests that the learned convolutional representations capture spatially meaningful features for disease discrimination, particularly in cases involving subtle or overlapping morphological characteristics. Grad-CAM visualizations of representative misclassified samples further illustrate regions associated with the model's incorrect predictions and provide additional insight into challenging class boundaries. Since Grad-CAM reflects the contribution of image regions to a model prediction rather than causal evidence of a specific pathological feature, these visualizations are interpreted as an explanatory analysis of the learned feature representations rather than a clinical localization method. Despite these encouraging findings, several limitations should be considered. The study relies on publicly available curated datasets, and



**Fig. 7. Grad-CAM visualizations of representative OCT-C8 dataset**

no external hospital-based prospective validation was performed. Differences in OCT scanners, acquisition protocols, patient populations, and clinical settings may therefore affect model performance under real-world conditions. In addition, patient-level prospective testing was not available in the present study. Future work will

## V. Conclusion

This study aimed to develop a lightweight and computationally efficient framework for automated retinal OCT image classification. The proposed CNN-SE model integrates SE blocks, data augmentation, and the Adam optimizer and achieved 99.00% and 97.00% accuracy on the OCT2017 and OCT-C8 datasets, respectively. Cross-dataset evaluation across five random seeds further achieved  $98.74 \pm 0.11\%$  accuracy and  $0.9995 \pm 0.0002$  macro-ROC-AUC when trained on OCT-C8 and directly tested on OCT2017, indicating robust generalization across dataset distributions. The model contains approximately 0.3748 M trainable parameters and achieves an average inference latency of 9.40 ms/image, supporting its computational efficiency. Grad-CAM analysis provided additional visual insight into the regions contributing to model predictions. Overall, the findings demonstrate that the proposed CNN-SE framework provides an effective balance of classification performance, generalization, and computational efficiency. Future work will focus on multicenter and prospective clinical validation to further assess its practical applicability.

## Acknowledgment

The authors have no acknowledgments to declare.

## Funding

This research received no external funding.

## Data Availability

The OCT2017 and OCT-C8 datasets used in this study are publicly available and are cited in References [22] and [31], respectively. No new dataset was generated.

## Author Contribution

Parth R. Dave: Conceptualization, Methodology, Data Curation, Software, Formal Analysis, Writing Original Draft, Writing, Review & Editing. Nikunj H. Domadiya: Conceptualization, Methodology, Supervision, Validation, Formal Analysis, Visualization, Writing, Review & Editing. Jaiwal Patel: Formal Analysis, Mathematical Analysis, Validation, Writing Review & Editing. All authors reviewed and approved the final manuscript.

## Code Availability

focus on evaluating the proposed model using multicenter, hospital-based OCT datasets with diverse acquisition conditions, patient-level evaluation, and prospective clinical validation to establish its robustness and clinical applicability.

The code used for model implementation and analysis is available from the corresponding author upon reasonable request.

## Declarations

### Ethics Approval

This study used publicly available, de-identified datasets and did not involve direct recruitment of human participants. Additional institutional ethics approval was therefore not required for the present secondary data analysis.

### Consent for Publication

Not applicable. No identifiable personal information is reported.

### Conflict of Interest

The authors declare no competing financial or non-financial interests.

### Informed Consent

Not applicable.

### Generative AI Declaration

During the preparation of this manuscript, the authors used generative AI-assisted tools solely for language refinement, grammar correction, and improvement of readability. The authors subsequently reviewed and edited the content and take full responsibility for the accuracy, originality, interpretation, and integrity of the manuscript.

### Declaration of Originality

The authors declare that this manuscript is original, has not been published previously, and is not under consideration for publication elsewhere.

### Declaration of Research Integrity

The authors confirm that the reported results are based on the described experiments and analyses and that no data, figures, or results have been fabricated, falsified, or inappropriately manipulated.

## References

- [1] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.

- [2] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 2005, pp. 886-893 vol. 1. doi: [10.1109/CVPR.2005.177](https://doi.org/10.1109/CVPR.2005.177).
- [3] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973, doi: [10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314).
- [4] G. Li, G. Wang, Q. Wang, F. Fei, S. Lü, and D. Guo, "ANN: a heuristic search algorithm based on artificial neural networks," in *Proceedings of the 1st International Conference on Intelligent Information Processing, ICIP-16*, 2016, Art. no. 51, pp. 1-9, doi: [10.1145/3028842.3028893](https://doi.org/10.1145/3028842.3028893).
- [5] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18-28, Jul. 1998, doi: [10.1109/5254.708428](https://doi.org/10.1109/5254.708428).
- [6] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 2018, pp. 278-282. doi: [10.1109/ICSCCC.2018.8703316](https://doi.org/10.1109/ICSCCC.2018.8703316).
- [7] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 8697-8710. doi: [10.1109/CVPR.2018.00907](https://doi.org/10.1109/CVPR.2018.00907).
- [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84-90, May 2017, doi: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [9] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, Apr. 2015, doi: [10.48550/arXiv.1409.1556](https://doi.org/10.48550/arXiv.1409.1556).
- [10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770-778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [11] C. Szegedy et al., "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1-9. doi: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
- [12] G. Huang, Z. Liu, L. van der Maaten, and K. Weinberger, "Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 2261-2269. doi: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
- [13] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, Aug. 2017, doi: [10.1609/aaai.v31i1.11231](https://doi.org/10.1609/aaai.v31i1.11231).
- [14] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 1800-1807. doi: [10.1109/CVPR.2017.195](https://doi.org/10.1109/CVPR.2017.195).
- [15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 4510-4520. doi: [10.1109/CVPR.2018.00474](https://doi.org/10.1109/CVPR.2018.00474).
- [16] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv preprint arXiv:1905.11946*, 2019. doi: [10.48550/arXiv.1905.11946](https://doi.org/10.48550/arXiv.1905.11946).
- [17] P. P. Srinivasan et al., "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed. Opt. Express*, vol. 5, no. 10, p. 3568, Oct. 2014, doi: [10.1364/BOE.5.003568](https://doi.org/10.1364/BOE.5.003568).
- [18] Y.-Y. Liu et al., "Computerized Macular Pathology Diagnosis in Spectral Domain Optical Coherence Tomography Scans Based on Multiscale Texture and Shape Features," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 11, p. 8316, Oct. 2011, doi: [10.1167/iovs.10-7012](https://doi.org/10.1167/iovs.10-7012).
- [19] A. Maćkiewicz and W. Ratajczak, "Principal components analysis (PCA)," *Comput. Geosci.*, vol. 19, no. 3, pp. 303-342, 1993, doi: [10.1016/0098-3004\(93\)90090-R](https://doi.org/10.1016/0098-3004(93)90090-R).
- [20] K. Alsaih et al., "Classification of SD-OCT volumes with multi pyramids, LBP and HOG descriptors: Application to DME detections," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, Aug. 2016, pp. 1344-1347. doi: [10.1109/EMBC.2016.7590956](https://doi.org/10.1109/EMBC.2016.7590956).
- [21] L. Huang, X. He, L. Fang, H. Rabbani, and X. Chen, "Automatic Classification of Retinal Optical Coherence Tomography Images With Layer Guided Convolutional Neural Network," *IEEE Signal Process. Lett.*, vol. 26, no. 7, pp. 1026-1030, Jul. 2019, doi: [10.1109/LSP.2019.2917779](https://doi.org/10.1109/LSP.2019.2917779).
- [22] 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](https://doi.org/10.1016/j.cell.2018.02.010).
- [23] J. Kim and L. Tran, "Retinal Disease Classification from OCT Images Using Deep Learning Algorithms," in *2021 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, IEEE, Oct. 2021, pp. 1–6, doi: [10.1109/CIBCB49929.2021.9562919](https://doi.org/10.1109/CIBCB49929.2021.9562919).
- [24] S. Diao *et al.*, "Classification and segmentation of OCT images for age-related macular degeneration based on dual guidance networks," *Biomed. Signal Process. Control*, vol. 84, p. 104810, Jul. 2023, doi: [10.1016/j.bspc.2023.104810](https://doi.org/10.1016/j.bspc.2023.104810).
- [25] 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](https://doi.org/10.3390/s23125393).
- [26] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images," *Ophthalmol. Retina*, vol. 1, no. 4, pp. 322–327, 2017, doi: [10.1016/j.oret.2016.12.009](https://doi.org/10.1016/j.oret.2016.12.009).
- [27] D. Paul, A. Tewari, S. Ghosh, and K. C. Santosh, "OCTx: Ensembled Deep Learning Model to Detect Retinal Disorders," in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, Jul. 2020, pp. 526–531, doi: [10.1109/CBMS49503.2020.00105](https://doi.org/10.1109/CBMS49503.2020.00105).
- [28] J. Yang, G. Wang, X. Xiao, M. Bao, and G. Tian, "Explainable ensemble learning method for OCT detection with transfer learning," *PLoS One*, vol. 19, no. 3, p. e0296175, Mar. 2024, doi: [10.1371/journal.pone.0296175](https://doi.org/10.1371/journal.pone.0296175).
- [29] M. Stanojević, D. Drašković, and B. Nikolić, "Retinal disease classification based on optical coherence tomography images using convolutional neural networks," *Journal of Electronic Imaging*, vol. 32, no. 03, Nov. 2022, doi: [10.1117/1.JEI.32.3.032004](https://doi.org/10.1117/1.JEI.32.3.032004).
- [30] M. Selvarajan and M. B. N. M., "Impact of Optimizer Algorithm on NasNetMobile Model for Eight-class Retinal Disease Classification from OCT Images," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 8, no. 2, pp. 490–503, Mar. 2026, doi: [10.35882/jeeemi.v8i2.1464](https://doi.org/10.35882/jeeemi.v8i2.1464).
- [31] Obuli Sai Naren, "Retinal OCT Image Classification - C8" Kaggle, 2021. [Online]. Available: Accessed: Jul. 20, 2026. doi: [10.34740/KAGGLE/DSV/2736749](https://doi.org/10.34740/KAGGLE/DSV/2736749)
- [32] 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," *Frontiers in Big Data*, vol. 8, Art. no. 1609124, 2025, doi: [10.3389/fdata.2025.1609124](https://doi.org/10.3389/fdata.2025.1609124).
- [33] P. R. Dave and N. H. Domadiya, "Lightweight deep learning approach for retinal OCT image classification: A CNN with hybrid pooling and optimized learning," *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 15, no. 1, p. 414, Mar. 2026, doi: [10.11591/ijict.v15i1.pp414-427](https://doi.org/10.11591/ijict.v15i1.pp414-427).
- [34] Z. Yang, R. O. Sinnott, J. Bailey, and Q. Ke, "A survey of automated data augmentation algorithms for deep learning-based image classification tasks," *Knowl. Inf. Syst.*, vol. 65, no. 7, pp. 2805–2861, 2023, doi: [10.1007/s10115-023-01853-2](https://doi.org/10.1007/s10115-023-01853-2).
- [35] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 7132–7141, doi: [10.1109/CVPR.2018.00745](https://doi.org/10.1109/CVPR.2018.00745).
- [36] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proc. 27th International Conference on Machine Learning (ICML)*, 2010, pp. 807–814, doi: [10.5555/3104322.3104425](https://doi.org/10.5555/3104322.3104425).
- [37] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint arXiv:1412.6980*, 2017, doi: [10.48550/arXiv.1412.6980](https://doi.org/10.48550/arXiv.1412.6980).
- [38] M. Opoku, B. A. Weyori, A. F. Adekoya, and K. Adu, "CLAHE-CapsNet: Efficient retina optical coherence tomography classification using capsule networks with contrast limited adaptive histogram equalization," *PLoS One*, vol. 18, no. 11, p. e0288663, Nov. 2023, doi: [10.1371/journal.pone.0288663](https://doi.org/10.1371/journal.pone.0288663).
- [39] D. Le *et al.*, "Transfer Learning for Automated OCTA Detection of Diabetic Retinopathy," *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 35, Jul. 2020, doi: [10.1167/tvst.9.2.35](https://doi.org/10.1167/tvst.9.2.35).
- [40] A. A. Saraiva, D. B. S. B. S. Santos, P. Pimentel, J. V. M. Sousa, N. M. F. Ferreira, J. E. S. Batista Neto, S. Soares, and A. Valente, "Classification of optical coherence tomography using convolutional neural networks," in *Proc. 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC-2020)*, BIOINFORMATICS, 2020, pp. 168–175, doi: [10.5220/0009091001680175](https://doi.org/10.5220/0009091001680175).

## Authors' Biography



**Prof. Parth R. Dave**    was born at Halol in Gujarat, India on 7th August, 1989. Prof. Dave completed B.E. and M. Tech with specialization in Computer Engineering from Dharmsinh Desai University, Nadiad, Gujarat, India in the years 2011 and 2013, respectively.

The author's major field of study covers machine learning, deep learning, image processing, and multimedia information retrieval. He had worked as an Assistant Professor in the Computer Engineering Department at Dharmsinh Desai University, Nadiad, Gujarat, India from 2013 to 2020. He is currently working as an Assistant Professor in the Computer Engineering Department at L. D. College of Engineering, Ahmedabad, Gujarat, India since 2020. He delivers expert talks and serves as a reviewer for international conferences and journals. He has also presented papers in International Journals, IEEE Conferences. He can be contacted at email: [prd7889@gmail.com](mailto:prd7889@gmail.com).

Django-based student book-exchange application. Beyond research and programming, Jaiwal is the founder of EcoRevive Dubai, a student-led e-waste initiative. He hopes to pursue undergraduate study in computer science and applied artificial intelligence.



**Nikunj H. Domadiya**    completed M. Tech and Ph.D. from NIT-Surat, Gujarat, India, in 2013 and 2020, respectively. The author's major field of research covers Data Privacy and Security, Machine learning, Deep learning, and multimedia information

retrieval. He is currently working as an Assistant Professor in the Computer Engineering Department at L. D. College of Engineering, Ahmedabad, Gujarat, India since 2020. He can be contacted at email: [domadiyanikunj002@gmail.com](mailto:domadiyanikunj002@gmail.com)



**Jaiwal Patel** is a Grade 12 student at GEMS New Millennium School in Dubai, with interests in computer science, artificial intelligence, mathematics, and socially responsible technology. His academic work includes independent study in advanced STEM subjects and

has also developed software projects, including EcoRevive OS, a platform designed to support community e-waste collection operations, and a