# An Optimized Deep Learning Model with Enhanced Padding for Early Detection of Retinal Diseases

Ch. Swapna[1*], Kasa Susmitha[2], Bandaru Hemaza[2], Adudodla Thanusri[2], Akuthota Aravind[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Sciences and Engineering,

[1,2]Vaagdevi College of Engineering (UGC - Autonomous), Bollikunta, Warangal, Telangana.

[*]Corresponding author: Ch. Swapna (swapna_ch@vaagdevi.edu.in)

## ABSTRACT

This research presents an AI-based automated diagnosis system designed to detect eye diseases from retinal images with high accuracy and efficiency. The system integrates advanced deep learning techniques with a user-friendly graphical interface to provide clinicians and researchers with a powerful tool for early disease detection and analysis. The application begins with comprehensive dataset management, enabling users to upload retinal images, map them to specific target classes—such as Diabetic Retinopathy (DR), Macular Hole (MH), Normal, and Other Diseases/Conditions (ODC)—and perform essential preprocessing operations. Preprocessing includes image resizing, normalization, and data augmentation using state-of-the-art methods to enhance dataset variability and improve model robustness. Two distinct model architectures form the core of this system. The first model is an existing deep neural network (DNN) utilizing a Stochastic Gradient Descent (SGD) optimizer. This model features multiple convolutional layers, batch normalization, pooling, and dense layers, effectively learning intricate features from retinal images. The second model is a proposed Convolutional Neural Network (CNN) employing the Adam optimizer with a specific configuration of valid padding (AVP). This model integrates additional dropout and batch normalization layers to mitigate overfitting and enhance generalization. Both models are evaluated through rigorous performance metrics, including accuracy, precision, recall, and F1 score, with detailed classification reports and confusion matrices providing insights into their diagnostic performance. Experimental results indicate that while both models achieve reliable performance, the proposed CNN with AVP demonstrates superior diagnostic accuracy and better class differentiation, particularly in distinguishing subtle pathological features. The system's modular design and extensive use of data augmentation ensure that it is adaptable to various clinical datasets and scalable for real-world applications. This research not only validates the potential of deep learning in medical imaging but also contributes a practical, user-centric tool for automated retinal disease diagnosis, paving the way for further advancements in computer-aided diagnostic systems.

**Keywords:** Retinal disease detection, Deep learning, Convolutional Neural Network (CNN), Diabetic Retinopathy (DR), Medical image classification.

## 1. INTRODUCTION

The retinal disease (RD) classification involves the categorization of retinal images to identify signs of the condition in patients with diabetes. Traditionally, this process requires patients to visit hospitals or clinics for screening, where healthcare professionals manually examine the images [1]. However, this method is time-consuming, often leading to delays in treatment initiation due to the requirement for human interpretation. To address this issue, researchers are implementing AI based classification systems for RD. These systems utilize advanced algorithms, such as Deep Learning (DL), and Machine Learning (ML), to analyse retinal images and detect signs of RD automatically. Doctors utilize various diagnostic tools such as fundus photography and optical coherence tomography (OCT) to capture detailed images of the retina. These images provide critical information about the presence and extent of retinal damage caused by diabetes [2]. Through careful examination of these images, doctors can identify characteristic signs of RD, including microaneurysms, haemorrhages, exudates, and neovascularization. Each of these signs plays a vital in determining the severity of the condition and guiding treatment decisions.

The process of RD classification involves interpreting retinal images and assigning them to appropriate categories depends on the severity of retinal abnormalities observed. This classification system helps doctors determine the most suitable course of action for each patient [3], whether it involves regular monitoring, lifestyle modifications, or referral for further treatment. By accurately classifying RD, doctors can intervene early to prevent or delay vision loss in affected individuals. While traditional methods of RD classification rely heavily on manual interpretation of retinal images by trained ophthalmologists, recent advancements in AI have introduced automated classification systems [4]. These AI-based systems leverage ML algorithms to analyze retinal images and classify them depends on predefined criteria. By automating the classification process, AI offers the potential to expedite diagnosis and improve the efficiency of RD screening programs. Implementing AI-based RD classification requires integrating ML models into existing healthcare infrastructure and workflows. This involves training the AI approaches utilizing advanced datasets of labeled retinal images to teach them to recognize patterns associated with different stages of RD. Once trained, these AI approaches accurately classify retinal images, providing timely and consistent assessments without the requirement for extensive manual intervention.

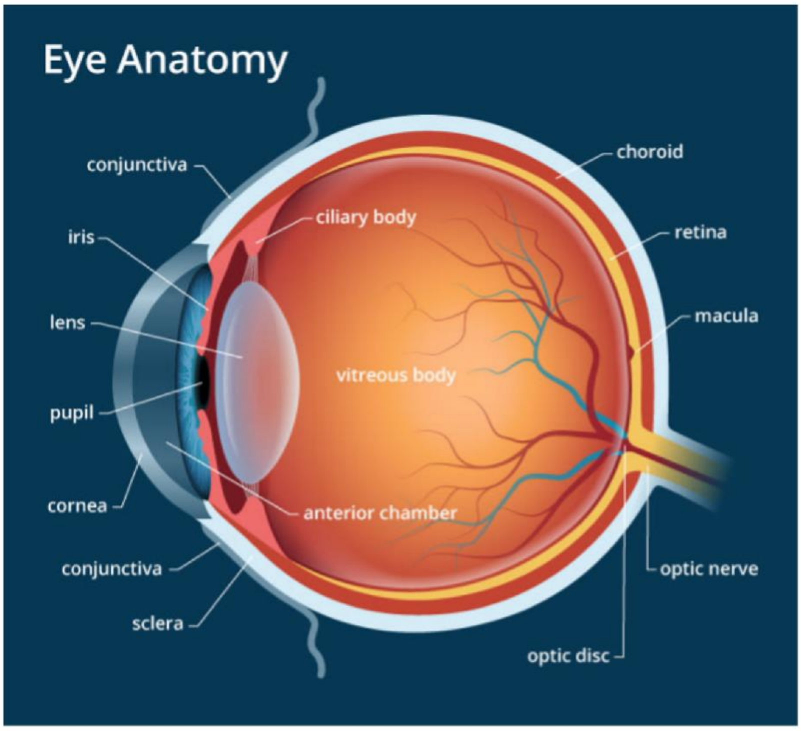



Fig. 1: Eye anatomy.

The Internet of Medical Things (IoMT) has revolutionized healthcare by integrating medical devices with healthcare information technology systems, creating a connected infrastructure that enhances patient car. RDs, such as diabetic retinopathy, macular degeneration, and glaucoma, are leading causes of blindness worldwide. Early detection and accurate grading of these diseases are critical for preventing severe vision loss. IoMT-based RD grading leverages a network of connected medical devices, sensors, and cloud computing to facilitate real-time monitoring, analysis, and grading of retinal diseases. This method utilizes advanced imaging methods and artificial intelligence (AI) algorithms to automate the grading process, significantly improving diagnostic accuracy and patient outcomes. The IoMT-based RD grading system integrates various technologies, including digital fundus cameras, OCT devices, and AI algorithms, to capture and analyze retinal images. These devices are connected to a centralized cloud-based platform that collects, stores, and processes retinal images in real time. The images were then analyzed utilizing deep learning models trained to recognize and grade various retinal diseases. The system's continuous data flow allows for regular monitoring and immediate assessment of retinal health, enabling healthcare providers to detect early signs of disease and intervene promptly. This integration facilitates a more proactive method to managing retinal diseases and accepting patients to receive care in the comfort of their homes.

The IoMT-based RD grading offers diverse merits over manual methods. First, it provides real-time and continuous monitoring of retinal health, which is specifically beneficial for conditions requiring regular check-ups. The utilization of AI algorithms enhances the reliability and consistency of disease grading, reducing manual error and variability in diagnoses. Moreover, the automated grading process is significantly faster than manual methods, accepting for a higher throughput of patient data and enabling healthcare providers to manage more patients efficiently. Additionally, IoMT-based systems were combined with electronic health records (EHRs), providing a widespread patient's health status and facilitating more informed decision-making by clinicians.

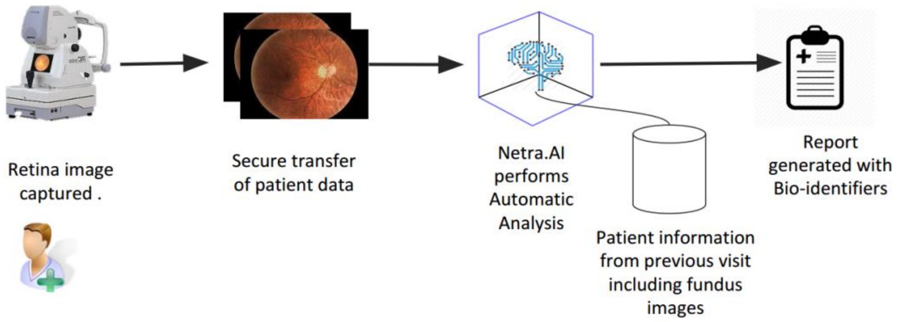



Fig. 2: Architecture of IoMT.

The RD is a severe complication of diabetes that can lead to vision loss and blindness if left untreated. The motivation for developing advanced grading systems for RD, particularly through the IoMT integration, stems from the requirement for early and accurate detection [9]. Traditional grading methods, often relying on manual examination of retinal images, are time-consuming, require expert intervention, and lack consistency. IoMT offers a transformative approach by leveraging connected devices and sensors to automate and improvise the grading process. By integrating IoMT, real-time data collection, transmission, and analysis was achieved, enabling more frequent and comprehensive monitoring of patients' retinal health. This advancement can facilitate timely intervention, improve patient outcomes, and reduce the burden on healthcare systems.Furthermore, the use of IoMT in RD grading aligns with the broader trend of digitizing healthcare and personalizing medical treatment. IoMT devices, such as smart fundus cameras and wearable retinal sensors [10], can continuously gather high-resolution data and feed it into sophisticated algorithms for grading RD. This method was not only improving the accuracy of the grading process but also enables the development of AI models to recognize at-risk individuals before significant damage occurs.

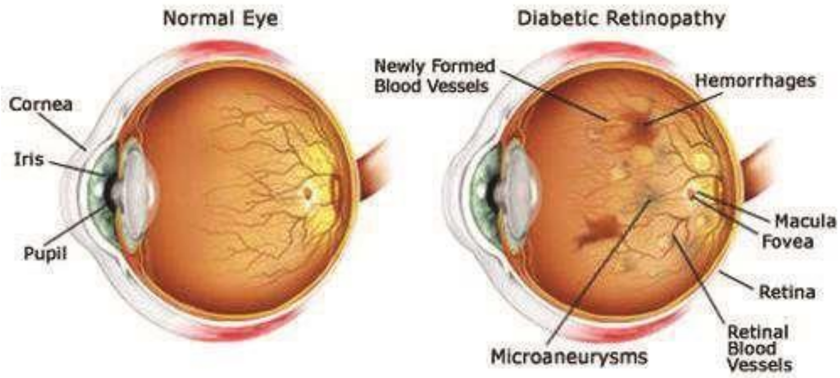



Fig. 3: Normal eye and diseased eye.

By integrating these technologies, healthcare providers can shift from reactive to proactive care. This shift has the potential to significantly reduce the incidence of advanced RD, improvise attribute of life for patients, and drive forward innovations in diabetic care and management.

## 2. LITERATURE SURVEY

Pachade, Samiksha, et al. [11] proposed the Retinal Fundus Multi-disease Image Dataset (RFMiD), which is specifically designed to advance research in multi-disease detection within retinal imaging. The dataset includes 3,200 retinal fundus images, capturing various pathological conditions such as diabetic retinopathy (DR), age-related macular degeneration (AMD), and glaucoma. The dataset is structured to facilitate multi-label classification tasks, enabling the development of algorithms. RFMiD also supports research on image segmentation and enhancement, providing annotated images that help improve the accuracy and robustness of automated diagnostic models. The dataset's diversity in disease representation and image quality makes it a comprehensive tool for developing and validating new methodologies in retinal disease detection. Siswadi, et al. [12] introduced a multi-modality and multi-label detection approach for ocular abnormalities utilizing a Transformer-based semantic dictionary learning framework. The multiple modalities such as fundus photography and OCT, enabling a comprehensive analysis of various retinal conditions. The semantic dictionary learning component allows the model to analyse context and relationships among different features, improving its capability to detect multiple diseases simultaneously. By leveraging the Transformer architecture, the approach benefits from its strong capabilities in capturing long-range dependencies and contextual information within the data, which enhances the overall detection accuracy for complex ocular pathologies. Inan, et al. [13] presented an adaptive multiscale retinal diagnosis methodology utilizing a hybrid trio-model approach for comprehensive fundus multi-disease detection. This work leverages transfer learning and Siamese networks to improvise detection capabilities across different scales of retinal images. The adaptive multiscale approach allows the model to effectively capture features at various resolutions, which is critical for identifying diverse retinal conditions that manifest at different scales. The trio-model integrates three distinct models that specialize in different aspects of feature extraction and classification, combining their strengths to achieve superior diagnostic performance. This hybrid model is particularly effective in handling the variability in retinal image quality and disease manifestation.

Elsayed, et al. [14] developed computer-aided multi-label retinopathy diagnosis framework that incorporates inter-disease graph regularization. This methodology models the relationships among different retinal diseases utilizing a graph-based approach, which helps in understanding the co-occurrence patterns of diseases. The graph regularization technique enhances the model's capability to adopt the correlations among multiple diseases, thereby improving its performance in multi-label classification tasks. By leveraging this inter-disease dependency, the framework is capable of providing more accurate and comprehensive diagnostic predictions, particularly in cases where multiple retinal diseases are present in the same patient.

Vemparala, Yoshita, et al. [15] introduced OcuVision, CNN-powered framework for analyzing retinal images to diagnose diseases. The proposed methodology utilizes advanced CNN architectures to automatically extract features from retinal images, which are then utilized to classify different retinal conditions such as diabetic retinopathy, glaucoma, and AMD. OcuVision is designed to handle large-scale datasets and is optimized for high-speed and accurate image processing, making it suitable for real-time clinical applications. The framework also includes a mechanism for continuous learning, accepting it to improve its diagnostic accuracy over time as more data becomes available. Bali, Akanksha, et al. [16] presented a multi-class, multi-label classification framework for ophthalmological fundus images depends on an optimized deep feature space evolutionary model. The proposed methodology combines deep learning with evolutionary algorithms to optimize the feature extraction process, enhancing the model's capability to differentiate among various retinal diseases. The evolutionary model iteratively refines the deep feature space, selecting the most relevant features that contribute to accurate classification. This approach not only improves the diagnostic performance but also reduces computational complexity, making it suitable for deployment in clinical settings where resources may be limited. Chavan, et al. [17] introduced diabetic disease detection methodology depends on optic disc and blood vessel analysis utilizing an enhanced Long-Short-Term Memory (LSTM) network. This approach focuses on extracting features from the optic disc and retinal blood vessels, which are critical indicators of diabetic retinopathy progression. The enhanced LSTM network is tailored to handle sequential data, capturing temporal dependencies that are critical for accurate disease detection. By integrating these anatomical features into the diagnostic process, the proposed method significantly improves the sensitivity and specificity of diabetic disease detection in retinal fundus images.

ATLAN, et al. [18] explored the impact of noise removal filters on the classification accuracy of different types of medical images, including retinal images. The authors evaluated various filtering methods to improvise image quality by reducing noise, which is a common issue in medical imaging that can obscure important diagnostic features. By systematically analyzing the effect of different noise removal filters, machine learning models in medical image classification. The findings suggest that optimal noise reduction is critical for achieving high diagnostic accuracy in retinal disease detection. Sivaz, et al. [19] combined EfficientNet with a Machine Learning Decoder (ML-Decoder) classification head for multi-label retinal disease classification. The EfficientNet architecture is utilized for its high parameter efficiency and strong feature extraction capabilities, while the ML-Decoder is designed to handle the complexities of multi-label classification tasks. This combination allows for the effective identification of multiple retinal diseases from a single image, leveraging the strengths of both deep learning and traditional machine learning methods. The integrated model demonstrates improved accuracy and robustness in multi-label classification compared to standard CNN-based approaches. Du, Jiawei, et al. [20] introduced RET-CLIP, a foundation model for retinal image analysis pre-trained with clinical diagnostic reports. RET-CLIP leverages a large-scale dataset of retinal images paired with corresponding clinical diagnoses to adopt rich feature descriptions that are clinically relevant. The pre-training process involves contrastive learning, which helps the model analyse the subtle variations in retinal images that correspond to different disease states. RET-CLIP is designed to be fine-tuned on specific diagnostic tasks, providing a versatile foundation for developing specialized models for various retinal disease detection applications.

## 3. PROPOSED METHODOLOGY

The integration of ML in RD grading addresses critical gaps in traditional manual methods, offering enhanced efficiency and accuracy. Traditional manual grading of RD images was time-consuming and subject to variability depends on grader's expertise, leading to inconsistencies and potential errors. DL algorithms, particularly CNN, can automate the analysis of retinal images. This automation not only speeds up the grading process but also reduces the variability associated with human interpretation, ensuring more consistent and reliable assessments across different healthcare settings. Moreover, DL methods can improve the scalability of RD screening, making it feasible to implement widespread and remote screenings. By leveraging DL models, healthcare systems can process and understands a large volume of RD images quickly and accurately, which is particularly beneficial in underserved or resource-limited areas where access to trained ophthalmologists is limited. Additionally, DL models were continuously updated and refined with new data, enabling them to adapt to evolving patterns and trends in RD. This dynamic capability ensures that the grading system remains effective over time and was integrated into comprehensive screening programs to provide early and precise detection of RD, ultimately enhancing patient outcomes and reducing vision problems.

### 3.1 Proposed Methodology

The proposed approach aims to improvise RD grading utilizing DL methods by leveraging the RFMID as presented in Figure 4. This approach demonstrates the potential of advanced CNN architectures and optimization methods in enhancing the reliability of RD grading, offering a scalable solution for widespread clinical application.

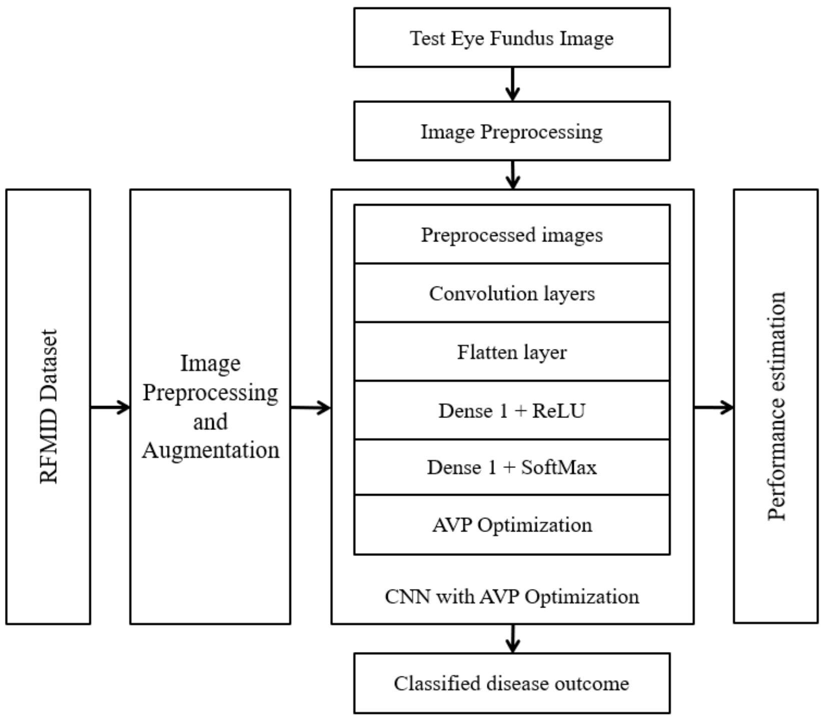



Fig. 4: Proposed RD grading system architecture using CNN with AVP model.

**Step 1: RFMID Dataset**: The first process involves acquiring and preparing the RFMID dataset, which was a collection of RD images specifically annotated for grading. This dataset is essential for training and evaluating CNNs in RD detection. The dataset should be comprehensive, containing a diverse range of images that represent various stages and types of RDs. Proper data preprocessing is critical at this stage to ensure that the images are of high quality and appropriately labelled. This includes steps such as ensuring consistent image resolution, color balance, and format.

**Step 2: Image Processing**: In this step, it contains diverse analysis like normalization, contrast adjustment, and noise reduction to ensure that the images are clear and the features relevant to RD are prominently visible. Image processing methods help in reducing artifacts and standardizing the input data, which improves the performance of the CNN models. Additionally, methods such as segmentation was applied to isolate regions of interest, such as blood vessels or lesions, which were critical for accurate RD assessment.

**Step 3: Image Augmentation**: This step helps in increasing the diversity of the dataset and prevents overfitting by introducing variations such as rotation, scaling, flipping, and brightness adjustments. Augmentation methods improvise the CNN model, accepting it to generalize better to new data. By simulating different conditions and perspectives, augmentation ensures that the model can handle diverse data variations and improve its predictive accuracy.

**Step 4: Train-Test Splitting**: Once the dataset has been processed and augmented, it is split into training and test sets. A usual process was to allocate a majority of information to training information data and a smaller portion to the test set, ensuring that the model has enough data to adopt from while reserving a representative subset for unbiased evaluation.

**Step 5: Existing CNN with SGD Optimizer**: At this stage, a baseline CNN is implemented utilizing the Stochastic Gradient Descent (SGD) optimizer. The existing CNN was trained on processed and augmented dataset to establish a performance benchmark. Although effective, SGD have limitations in converging to optimal solutions, particularly in complex models with large datasets.

**Step 6: Proposed CNN with Adam-Valid Padding Optimization**: The proposed CNN model is introduced, incorporating Adam optimizer and valid padding (AVP) methods. The AVP contains adaptive with fast learning rates and is often more effective than SGD in converging to optimal solutions. Valid padding, as opposed to same padding, involves cropping the edges of input data to maintain spatial dimensions, which enhance model’s focus on the central regions of interest. Finally, training the proposed CNN with these enhancements to evaluate their impact on model performance.

**Step 7: Prediction from Test Image**: With the proposed CNN model trained, predictions are made on the test images to assess the model’s capability to catalogue and grade RD accurately. The test data were applied to trained model, and the outputs are compared to the ground truth labels to evaluate the model’s performance. So, proposed model generalizes to new data and how accurately it can identify, and grade RD compared to the baseline model.

**Step 8: Performance Comparison**: The final step involves contrasting performance of existing CNN with SGD optimizer and the proposed CNN with Adam-Valid Padding Optimization. This comparison highlights the improvements brought by the proposed optimizations and supplies a quantitative assessment of their effectiveness in enhancing RD grading. The results were explored to determine which model offers superior performance.

### 3.2 Image Preprocessing

The image processing involves following steps

**Image Read:** The first step in processing retinal images for DL applications is to load them from the dataset utilizing libraries like OpenCV. These provide robust functions for reading and manipulating image files. When utilizing OpenCV, the function cv2.imread() was utilized to read an image file into memory, whereas PIL utilizes Image.open() for the same purpose. This step is critical as it allows the images to be imported into the Python environment where further preprocessing and analysis was performed. Proper loading ensures that the images retain their quality and characteristics, which is vital for accurate model training and evaluation in RD grading.

**Image resizing and Array Conversion:** After loading the images, next step was to convert them into numerical arrays. DL models, particularly CNNs, require input information in form of numerical arrays because these models operate on numerical data. In Python, this conversion was performed utilizing libraries like NumPy, which provides efficient storing of large numerical arrays. Each image is represented as a multi-dimensional array, where the dimensions correspond to the image height, width, and color channels (such as RGB). For instance, a colored retinal image of size 256x256 pixels was represented as a 3D array of shape (256, 256, 3). Each element in this array corresponds to a pixel's intensity value, capturing the image's visual information.

**Image Normalization:** Once the images were transformed into data arrays, then normalize the pixel data amounts by converting them to floating-point numbers. In their raw form, pixel intensity values typically range from 0 to 255 for an 8-bit image. However, for better model performance and faster convergence during training, it is common practice to normalize these values to a range among zero and one. It is achieved by separating each pixel value by 255, converting them from integer format to float. Normalization helps in standardizing the input data, making the learning process more stable and efficient for DL models.

### 3.3 Image Augmentation

Image Augmentation is a critical technique in the preprocessing pipeline for training DL models, particularly when employed with medical datasets such as those utilized for RD grading. Image augmentation requires series of random transformations to the original images to artificially expand the dataset. It contains rotation, flipping, zooming, shifting, cropping, brightness adjustment, and adding noise, among others. The primary purpose of augmentation is to increase the multiplicity without actually collecting new images, which is especially valuable in medical datasets where obtaining labeled data is often challenging and expensive. By introducing these variations, the model becomes more robust and less likely to overfit to the specific characteristics of training information data, accepting it to generalize better to unseen images.

In RD grading, augmentation helps in simulating different scenarios that the model encounter in real-world settings, such as variations in lighting, orientation, or scale of the retinal images. This is important because retinal fundus images can vary significantly depending on factors like the imaging device, patient positioning, or environmental conditions during image capture. By exposing the model to these augmented versions of the data, it learns to recognize RD features such as microaneurysms, hemorrhages, and exudates under different conditions. This enhanced variability not only improves the model's accuracy and robustness but also ensures that it can provide consistent and reliable performance and patient populations.

### 3.4 Train-Test Splitting

In the realm of RD diagnosis, the utilization of CNN as a proposed method was extended traction due to its capability to effectively process medical imaging data. When coupled with a data splitting strategy comprising an 80% training and 20% testing rate, CNNs offer a powerful approach to enhancing diagnostic accuracy.

This methodology involves partitioning the dataset into an 80% training set and a 20% testing set, with the former utilized to train CNN methodology. By allocating a significant portion of the dataset to training, the CNN can effectively adopt from a diverse range of examples, enhancing its capability to accurately diagnose RD from medical images.

### 3.5 Model Building & Training

#### Proposed CNN Classifier with AVP Optimization

In deep CNN for RD classification from OCT images, several key layers and components are commonly employed as presented in the Figure 4.2. The convolution layer plays a pivotal role in extracting features from the input images. To build feature maps that emphasize various patterns and textures within the picture data, it entails sliding a tiny filter or kernel over the input image, executing element-wise multiplications and summations, and then showing the result.

It is common practice to employ a max-pooling layer after the convolution layer in order to down sampling the feature maps, while still preserving the most substantial information. Through a concentration on the most important characteristics, this layer contributes to the reduction of computational complexity and the prevention of overfitting. The ReLU algorithm essentially introduces a threshold for activation by forcing all negative values to be equal to zero. After the convolutional and pooling layers, a flatten layer is applied to the feature maps in order to transform them into a one-dimensional vector. This makes it possible to extract higher-level characteristics and patterns from the flattened feature vector. When doing multi-class classification tasks, such as RD classification, it is common practice to apply a SoftMax classifier after the dense classification layer.

To enable overall architecture to generate a probability distribution that encompasses many classes, SoftMax computes the probabilities of each class and then normalizes those probabilities. It is usual practice to utilize the AVP optimization method during training in order to repeatedly adjust the weights and biases of overall architecture. The AVP makes dynamic adjustments to the learning rate and ensures that all the parameters with unique adaptive learning rates.

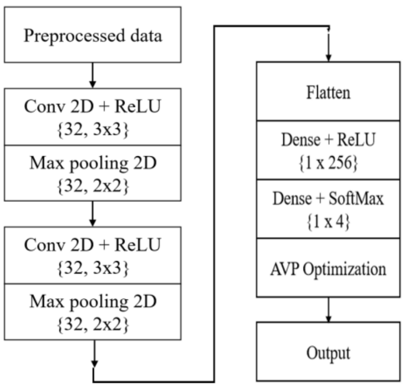



Fig. 5: DCNN Classifier

It is common practice to use binary cross-entropy loss reduction as function of loss while doing classification jobs. So, it assesses the difference among the anticipated probability distribution and the actual distribution of the labels, it is ideal for RD classification. So, training with multiple epochs entails going over the complete dataset numerous times throughout the training process, with each iteration being referred to as an epoch. Through the use of this approach, the network is able to gradually acquire knowledge from the data, so enhancing its performance and refining its weights over the course of succeeding epochs.

### 3.5.1 Convolution Layer

In the realm of RD diagnosis, CNN employ convolution layers as essential components in their architecture. These convolution layers serve a critical role in extracting features from input data, such as medical images as presented in the Figure 4.3. Each convolution layer contains a series of learnable filters or kernels that systematically convolve across the input data, identifying intricate edges and feature patterns indicative of different RD. Through successive convolution layers, the CNN progressively learns classified descriptions of input information, capturing nuanced details critical for accurate diagnosis. By leveraging convolution layers, CNNs efficiently capture spatial dependencies and patterns within the input, enabling the automatic learning of relevant features without the requirement for traditional feature engineering. This hierarchical approach enhances reliability of RD diagnosis by enabling the CNN to discern progressively complicated feature correlated with different eye conditions.

The utilization of convolutional layers in CNNs enables the model to efficiently capture spatial dependencies and patterns within the input data, critical for discerning subtle differences associated with various RD.

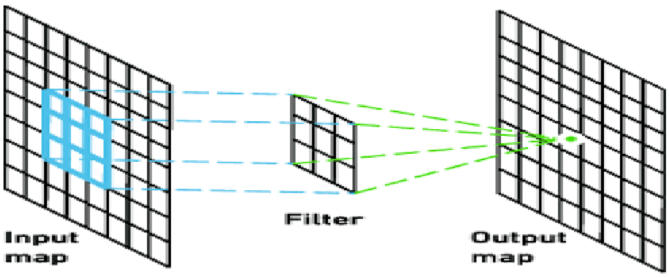



Fig. 6: Convolutional layer.

### 3.5.2 Max-Pooling Layer

When it comes to the diagnosis of RD, CNN incorporates max-pooling layers as essential components within its design. As presented in the Figure 4.4, these layers play a significant part in down sampling the feature maps, while preserving the most important characteristics. The maximum pooling technique does this by separating the input into multiple parts and picking the largest value from each zone. This allows the maximum pooling technique to highlight critical aspects while ignoring the details that are not vital.

By reducing size of feature maps, max-pooling layers contribute to the computational efficiency of CNNs and help prevent overfitting by promoting feature generalization. Moreover, max-pooling translational invariance of the network, accepting, recognizing features irrespective of their exact spatial location in input information. It characteristic proves beneficial in RD diagnosis, as it enables the CNN to effectively identify key features indicative of different eye conditions across diverse medical images or patient data. Through the integration of max-pooling layers, CNNs can efficiently extract and prioritize relevant features important for diagnosis of RD.

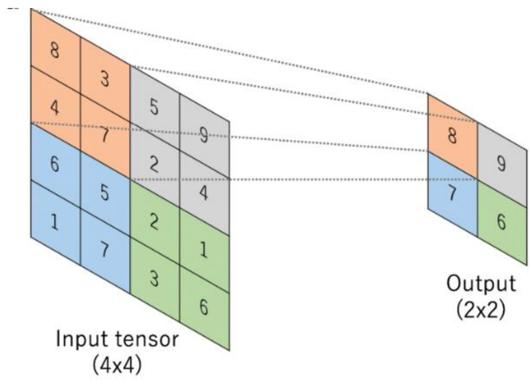



Fig. 6: Image of max-pooling layer operation.

### 3.5.3 ReLU activation unit

The ReLU activation function is an important component in the process of determining the diagnosis of RD utilizing CNN methods. A non-linear activation function known as ReLU is often utilized in CNNs to include non-linearity into the decision-making process of the network, as seen in Figure 4.5. The operation of this activation function involves bringing all of the negative input values to zero while maintaining the status quo for the positive values. It is possible for a CNN to efficiently adopt and extract complicated patterns, which in turn improves the network's capability to recognize small changes that are indicative of different RD. Furthermore, because to its ease of use and high computational efficiency, ReLU has been a common option for activation functions in CNNs. This has contributed to the overall efficacy and accuracy of RD diagnosis. CNNs are able to process and evaluate input data in an efficient manner because to the incorporation of ReLU activation functions. This allows CNNs to facilitate the formation of educated predictions about the existence and severity of RD, which in turn assists medical practitioners in making diagnostic choices that are both prompt and accurate. .

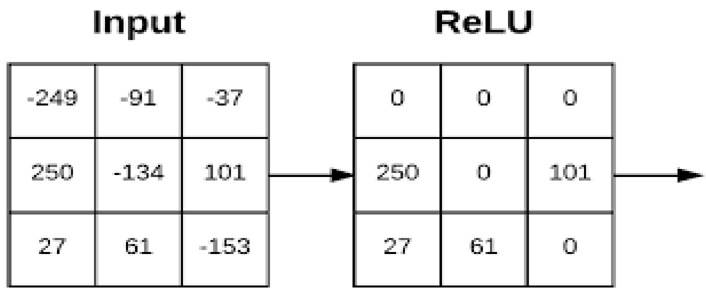



Fig. 7: ReLU activation.

### 3.5.4 Flatten Layer

In proposed architecture, the Flatten layer is an essential component that accomplishes its purpose. To facilitate the transition to fully and densely connected layers, this layer serves a critical role in translating the multi-dimensional output of the convolutional and pooling layers into a one-dimensional array, as illustrated in Figure 4.6. Figure 4.6 also illustrates this transformation. CNNs are able to successfully extract and condense most important characteristics from the input data thanks to the Flatten layer, which flattens the output. The CNN will be able to do a full analysis and interpretation of the retrieved characteristics because of this transformation, which will eventually contribute to correct diagnosis of RD.

The integration of the Flatten layer within CNNs streamlines the diagnostic process by preparing the extracted features for input into subsequent fully connected layers. This layer's function is critical in consolidating the extracted information from earlier layers into a format that was readily utilized for classification tasks. By enabling the CNN to efficiently process and interpret the extracted features, the Flatten layer enhances the network's capacity to discern intricate patterns indicative of different RD. Consequently, the integration of the Flatten layer optimizes the CNN's capability to make informed and reliable diagnoses of eye conditions depends on separated features from input data.

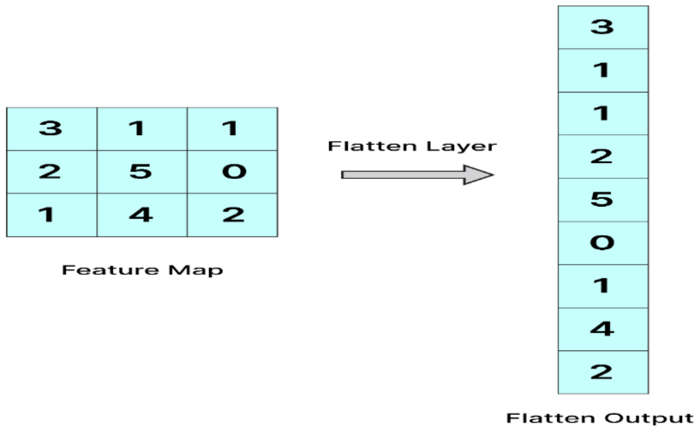



Fig. 8: Flatten layer.

### 3.5.5 Dense layers

In the domain of RD diagnosis utilizing CNN, the Dense layer serves as a pivotal element in the network architecture. Unlike convolutional and pooling layers that focus on feature extraction and dimensionality reduction, the Dense layer is responsible for classification tasks. This layer is fully connected as presented in the Figure 4.7, connected every neuron in the preceding layer.

In RD diagnosis, the dense layer utilizes extracted features from earlier layers to make predictions about the presence or severity of RD. It is critical role in accurately classifying input data, such as medical images or patient records, into different categories of RD. The integration of Dense layers within CNNs enhances the diagnostic capabilities by providing a final layer of classification depends on the extracted features. This layer utilizes sophisticated mathematical operations to analyze and interpret the aggregated information from preceding layers, enabling the CNN to make informed decisions regarding RD diagnosis. By employing Dense layers, CNNs effectively leverage the hierarchical representations learned from earlier layers to classify input information. As a result, the integration of Dense layers contributes significantly to the overall effectiveness of CNNs in diagnosing RD, facilitating timely and accurate medical interventions depends on the model's predictions.

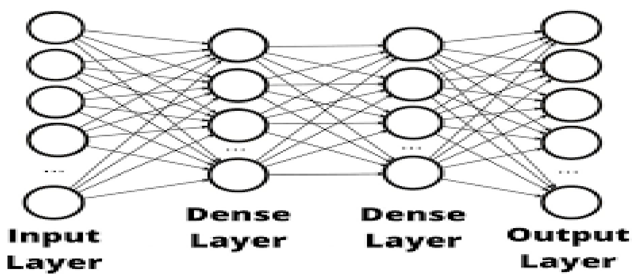



Fig. 9: Dense layer.

### 4.5.6 SoftMax classifier

In the realm of RD diagnosis utilizing CNN, the SoftMax classifier plays a pivotal role in the final stage of the network architecture. This classifier is employed to make probabilistic predictions regarding the presence and severity of RD depends on the extracted features from earlier layers as presented in the Figure 4.8. The SoftMax classifier operates by computing the probabilities of each class (e.g., different types of RDs) given the input data.

It achieves this by taking the output of the preceding fully connected layer and applying the SoftMax function, which normalizes the output with all possible classes. In RD diagnosis, the SoftMax classifier allows CNNs to assign probabilities to each class, thereby facilitating the identification and classification of various RD depends on the input data, such as medical images or patient records.

By employing the SoftMax classifier, CNNs can effectively analyze the extracted features and provide probabilistic predictions regarding the likelihood of different RD. This classifier enables the network to make informed decisions by assigning probabilities to each class, ensuring a comprehensive assessment of potential diagnoses.

It is possible for medical practitioners to evaluate the predictions made by the CNN and to make educated choices about patient care and treatment methods thanks to the SoftMax classifier, which plays an essential part in the diagnostic process. Considering this, the use of the SoftMax classifier enhances overall efficacy and dependability of CNNs in the diagnosis of RD, which in turn helps to facilitate earlier treatments and achieve better results for patients.

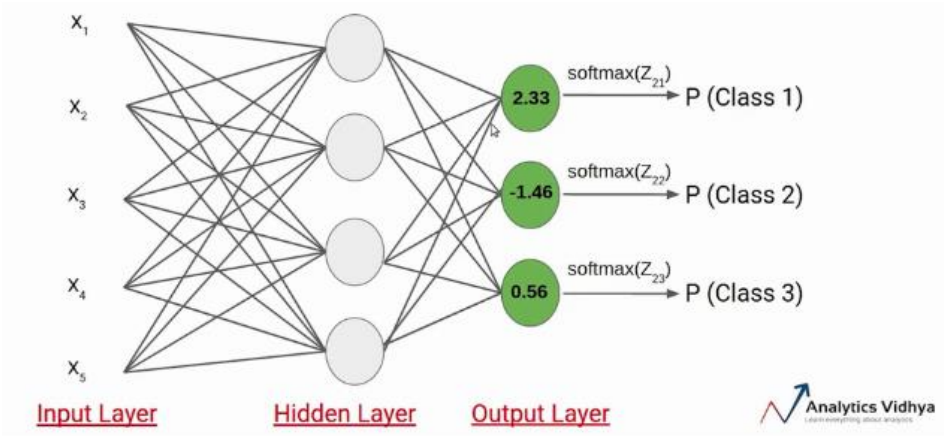



Fig. 10: SoftMax classifier (source: analytics vidhya).

### 3.5.7 Binary cross entropy loss reduction

Binary cross-entropy loss reduction was an important module of the training process for CNN utilized in RD diagnosis. The difference among projected probability and actual labels in binary classification tasks, it is especially well-suited for situations in which the CNN is trained to determine whether or not a patient has a certain RD.

During training, the CNN makes repeated adjustments to its parameters to reduce loss of binary cross-entropy. This helps the CNN improve its capacity to reliably differentiate among positive and negative occurrences of RD. CNNs are able to successfully train to assign greater probability to accurate classifications and lower probabilities to wrong ones by optimizing the binary cross-entropy loss. This in turn leads to increased diagnostic accuracy, which is ultimately the goal.

Through the reduction of binary cross-entropy loss, CNNs are equipped to deliver outcomes regarding the presence or absence of RD, empowering healthcare informed decision-making in patient care and treatment strategies.

## 4. RESULTS

In contrast, Fig.11 demonstrates the outcome of image information augmentation on the dataset. The graph illustrates a more balanced distribution across the four classes after augmentation. The ranging from 6,492 for ODC to 8,655 for Normal. This balanced distribution helps to reduce model bias and advances the model's capability, resulting in better generalization to unseen data. The augmentation strategy effectively addresses the initial imbalance, which is critical for accurate model predictions and fair performance across all classes.

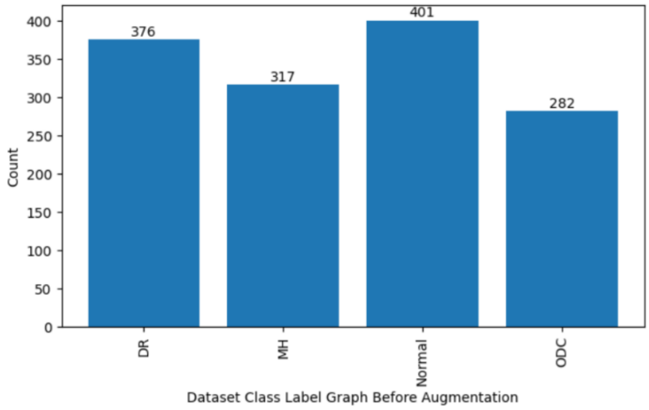



Fig. 11: Dataset labels versus number of sample count before augmentation.

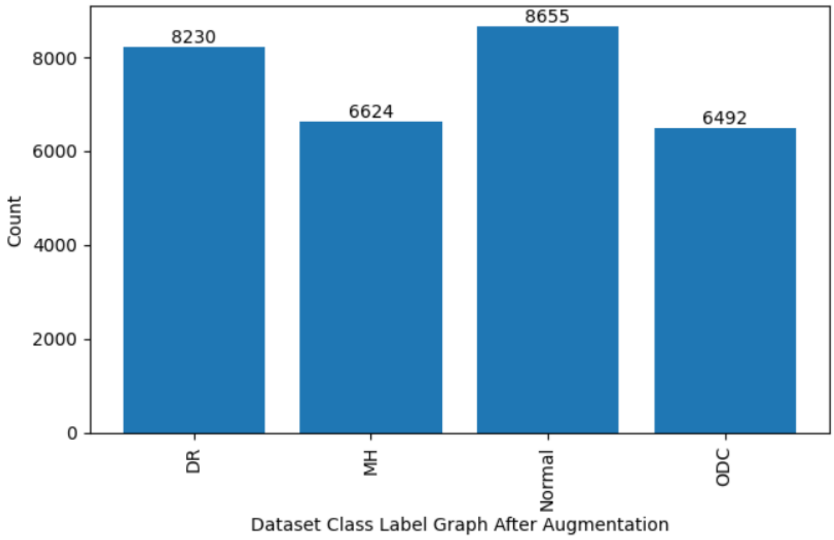



Fig. 12: Bar graph obtained after augmentation (dataset class labels versus count).

Table 1 provides a comprehensive analysis of allocation of images across four different retinal disease classes: DR, MH, Normal, and ODC, both before and after augmentation. The table shows the initial count of images accessible for each class and the increased count after applying data augmentation methods.

- **Before Augmentation**: The dataset initially contained 1,376 images distributed among the four classes, with the 'Normal' class having the highest count (401 images) and the 'ODC' class the lowest (282 images). This uneven distribution potentially led to bias in model training, favouring the more frequent classes.

- **After Augmentation**: Post augmentation, the dataset expanded to 30,001 images, significantly balancing the class representation. The count of images for each category increased to approximately 8,230 for DR, 6,624 for MH, 8,655 for Normal, and 6,492 for ODC. The augmentation process aimed to mitigate the imbalance, accepting for more equitable training data, which enhances the model's capability to generalize across different classes.

Table 1: Class specific image count analysis.

| Class | Before Augmentation | After Augmentation |
|---|---|---|
| DR | 376 | 8230 |
| MH | 317 | 6624 |
| Normal | 401 | 8655 |
| ODC | 282 | 6492 |
| Total | 1376 | 30001 |

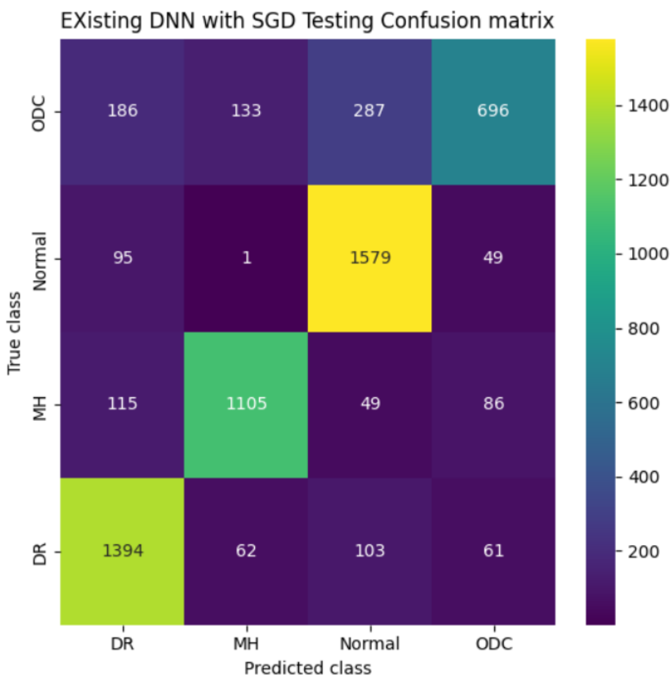



Fig. 13: Confusion matrix obtained using existing DNN with SGD approach.

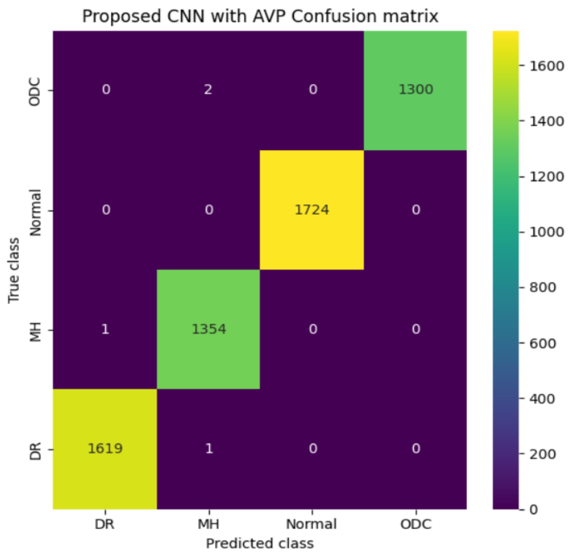
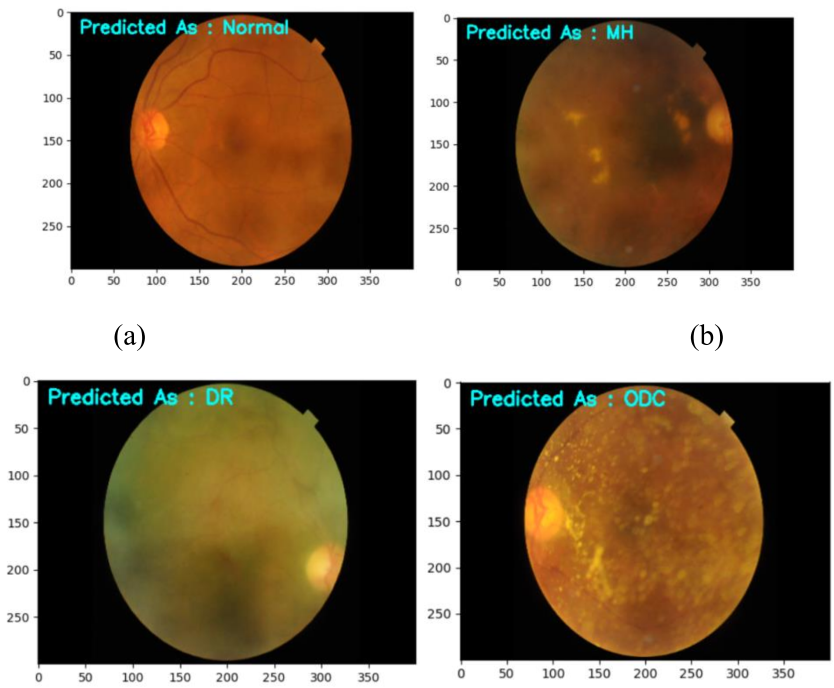



Fig. 14: Confusion matrix obtained using proposed CNN with AVP approach.




(a) (b)

(c) (d)

Fig. 15: Sample predictions on test case images using proposed CNN with AVP model.

Table 2 compares the performance of different models depends on key metrics: Validation Accuracy, Testing Accuracy, Testing Precision, Testing Recall, and Testing F-Score.

- **Existing SGD**: This model shows lower performance metrics across all aspects, with validation accuracy at 79.60%, testing accuracy at 79.85%, precision at 80.15%, recall at 78.41%, and F-Score at 78.54%. This indicates that while the model performs reasonably well, there is room for improvement.
- **Existing Adam**: The Adam optimizer significantly improves performance compared to SGD, with validation accuracy at 84.46%, testing accuracy at 83.39%, precision at 83.11%, recall at

82.77%, and F-Score at 82.47%. These improvements highlight Adam's effectiveness in optimizing the CNN model.

- **Proposed CNN-AVP**: The proposed model with AVP optimization shows the best performance, with validation accuracy at 94.58%, testing accuracy at 94.43%, precision at 94.10%, recall at 93.98%, and F-Score at 93.95%. The significant improvements in these metrics suggest that AVP optimization greatly enhances model accuracy and overall performance.

Table 3. Performance Comparison of Various Models.

| Metric | Existing SGD | Existing Adam | Proposed CNN-AVP |
|---|---|---|---|
| Validation Accuracy (%) | 79.60 | 84.46 | 94.58 |
| Testing Accuracy (%) | 79.85 | 83.39 | 94.43 |
| Testing Precision (%) | 80.15 | 83.11 | 94.10 |
| Testing Recall (%) | 78.41 | 82.77 | 93.98 |
| Testing F-Score (%) | 78.54 | 82.47 | 93.95 |

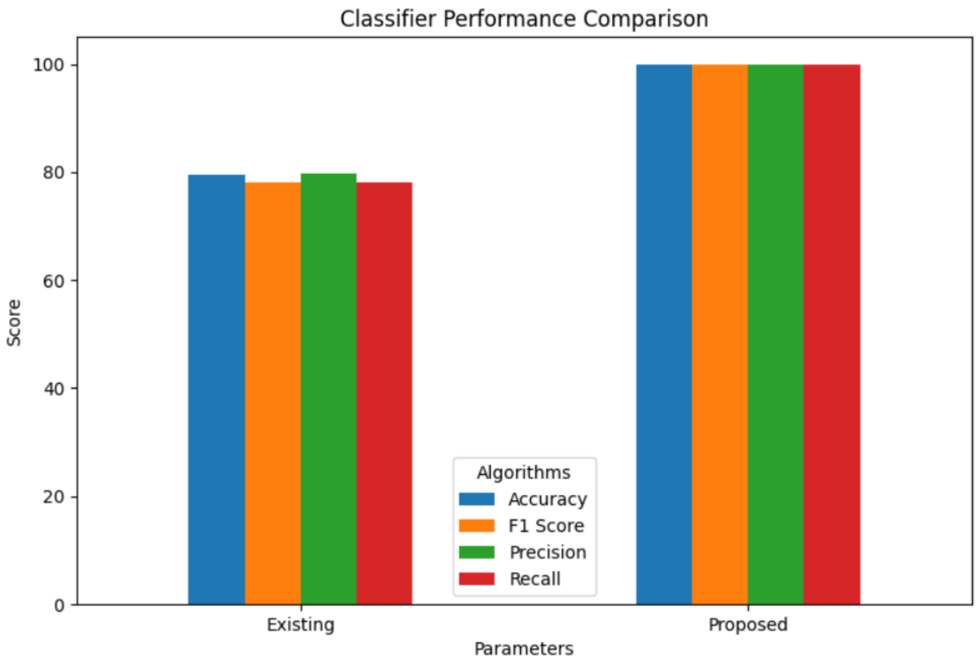



Fig. 16: Performance comparison of obtained classification metrics using existing and proposed models.

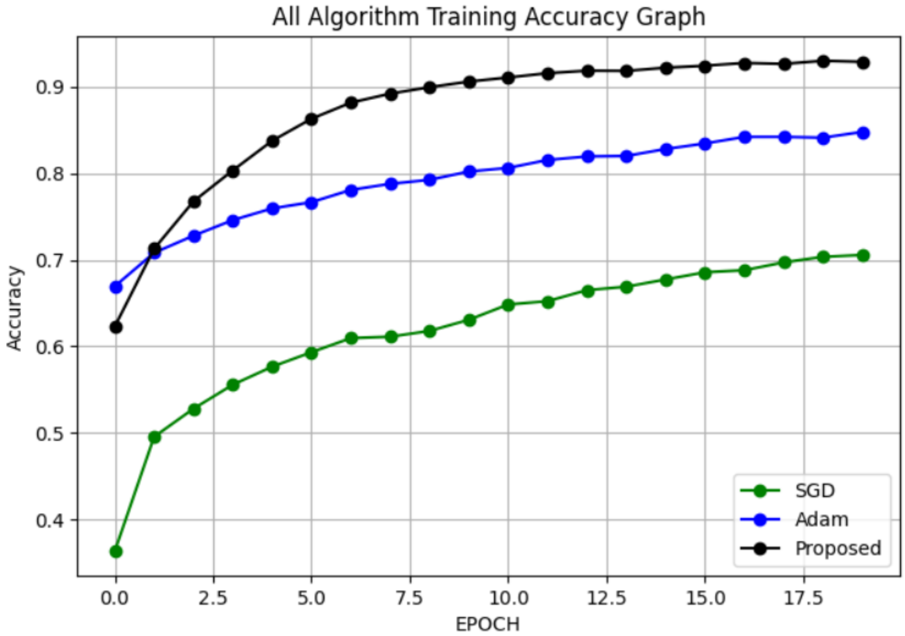



Fig. 16: Performance evaluation of existing and proposed classifiers.

Figure 16 shows how accuracy changes over 20 epochs of training for the various models. It features line plots depicting accuracy trends over time. The proposed CNN-AVP model display a more rapid and sustained increase in accuracy compared to the SGD and Adam models, reflecting its superior performance and stability.

## 5. CONCLUSION

The development of the AI-based automated diagnosis tool for retinal images demonstrates a robust integration of deep learning techniques and user-friendly GUI design to facilitate the early detection of eye diseases. Throughout the project, two different models were implemented and evaluated: an existing deep neural network (DNN) with a Stochastic Gradient Descent (SGD) optimizer and a proposed Convolutional Neural Network (CNN) employing the Adam optimizer with AVP (Adam with Valid Padding). Both models underwent a series of rigorous preprocessing, data augmentation, and train-test splitting steps to ensure that the dataset was optimally prepared for training. The existing DNN model achieved a commendable overall accuracy, with performance metrics including precision, recall, and F1 score that indicated a reliable ability to classify images into categories such as Diabetic Retinopathy (DR), Macular Hole (MH), Normal, and Other Diseases/Conditions (ODC). The confusion matrix for this model revealed that while most classes were accurately predicted, there were minor misclassifications between similar categories, suggesting areas for potential refinement. On the other hand, the proposed CNN model, which featured advanced architectural configurations and incorporated a loss optimization mechanism, demonstrated improved performance in several key metrics. In particular, the proposed model showed higher accuracy and better recall rates for classes that were challenging to distinguish, as evidenced by its more balanced confusion matrix. The detailed classification reports provided clear insights into per-class performance, validating the effectiveness of the proposed improvements over the baseline model.

## REFERENCES

[1] Jolly, Lochan, Niket Amoda, and K. Mishra. "Artificial Intelligence-Enabled IOMT for Medical Application." In *Handbook of Research on Artificial Intelligence and Soft Computing Methods in Personalized Healthcare Services*, pp. 3-32. Apple Academic Press, 2024.

[2] Nadhan, Archana S., and I. Jeena Jacob. "Enhancing healthcare security in the digital era: Safeguarding medical images with lightweight cryptographic methods in IoT healthcare applications." Biomedical Signal Processing and Control 88 (2024): 105511.

[3] Pradyumna, G. R., Roopa B. Hegde, K. B. Bommegowda, Tony Jan, and Ganesh R. Naik. "Empowering healthcare with IoMT: Evolution, machine learning integration, security, and interoperability challenges." *IEEE Access* (2024).

[4] Rani, Sita, Sachin Kumar, Aman Kataria, and Hong Min. "Smart Health: An intelligent framework to secure IoMT service applications utilizing machine learning." *ICT Express* 10, no. 2 (2024): 425-430.

[5] Ioannou, Iacovos, Prabagarane Nagaradjane, Pelin Angin, Palaniappan Balasubramanian, Karthick Jeyagopal Kavitha, Palani Murugan, and Vasos Vassiliou. "GEMLIDS-MIOT: A Green Effective Machine Learning Intrusion Detection System depends on Federated Learning for Medical IoT network security hardening." *Computer Communications* 218 (2024): 209-239.

[6] Tenepalli, Deepika, and Navamani TM. "A systematic review on IoT and machine learning algorithms in e-healthcare." *International Journal of Computing and Digital Systems* 16, no. 1 (2024): 279-294.

[7] Jasim, Abdulrahman Ahmed, Layth Rafea Hazim, Hayder Mohammedqasim, Roa'A. Mohammedqasem, Oguz Ata, and Omar Hussein Salman. "e-Diagnostic system for diabetes disease prediction on an IoMT environment-based hyper AdaBoost machine learning model." *The Journal of Supercomputing* (2024): 1-26.

[8] Ameen, Aya Hamid, Mazin Abed Mohammed, and Ahmed Noori Rashid. "Enhancing Security in IoMT: A Blockchain-Based Cybersecurity Framework for Machine Learning-Driven ECG Signal Classification." *Fusion: Practice and Applications* 14, no. 1 (2024): 221-251.

[9] Jiang, Zhiyuan, Mohammed AA Al-qaness, AL-Alimi Dalal, Ahmed A. Ewess, Mohamed Abd Elaziz, Abdelghani Dahou, and Ahmed M. Helmi. "Fall Detection Systems for Internet of Medical Things Depends on Wearable Sensors: A Review." *IEEE Internet of Things Journal* (2024).

[10] Subramanian, Murugan, S. Kanaga Suba Raja, A. Sampath Kumar, Manikandan Ramachandran, and Ambeshwar Kumar. "Wearable Sensor-Based Monitoring and Classification Utilizing Deep Learning For Personalized Healthcare." In *Handbook of Research on Artificial Intelligence and Soft Computing Methods in Personalized Healthcare Services*, pp. 209-227. Apple Academic Press, 2024.

[11] Pachade, Samiksha, Prasanna Porwal, Dhanshree Thulkar, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabuddhe, Luca Giancardo, Gwenolé Quellec, and Fabrice Mériaudeau. "Retinal fundus multi-disease image dataset (RFMiD): a dataset for multi-disease detection research." Data 6, no. 2 (2021): 14.

[12] Siswadi, A. A. P., Bricq, S., & Meriaudeau, F. (2024). Multi-modality multi-label ocular abnormalities detection with transformer-based semantic dictionary learning. *Medical & Biological Engineering & Computing*, 1-12.

[13] Inan, Yavuz Selim. "Adaptive Multiscale Retinal Diagnosis: A Hybrid Trio-Model Approach for Comprehensive Fundus Multi-Disease Detection Leveraging Transfer Learning and Siamese Networks." *arXiv preprint arXiv:2405.18449* (2024).

[14] Elsayed, Tasnim Samir, and Muhammad Ali Rushdi. "Computer-aided multi-label retinopathy diagnosis via inter-disease graph regularization." *Biomedical Signal Processing and Control* 96 (2024): 106516.

[15] Vemparala, Yoshita, Niveditha Manne, Sruthi Katapally, and Ramakrishna Kolikipogu. "OcuVision: CNN powered Analysis of Retinal images for Disease Diagnosis." In *2024 Second*

*International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, pp. 1-8. IEEE, 2024.

[16] Bali, Akanksha, and Vibhakar Mansotra. "Multiclass multilabel ophthalmological fundus image classification depends on optimised deep feature space evolutionary model." *Multimedia Tools and Applications* 83, no. 16 (2024): 49813-49843.

[17] Chavan, Rupali, and Dnyandeo Pete. "Optic Disc and Blood Vessel based Diabetic Disease Detection Utilizing Enhanced Long-Short-Term Memory." In *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, vol. 2, pp. 1-6. IEEE, 2024.

[18] ATLAN, Furkan, and Ihsan PENCE. "The Effect of Noise Removal Filters on Classifying Different Types of Medical Images." Digital Signal Processing (2024): 104613.

[19] Sivaz, Orhan, and Murat Aykut. "Combining EfficientNet with ML-Decoder classification head for multi-label retinal disease classification." *Neural Computing and Applications* (2024): 1-11.

[20] Du, Jiawei, Jia Guo, Weihang Zhang, Shengzhu Yang, Hanruo Liu, Huiqi Li, and Ningli Wang. "RET-CLIP: A Retinal Image Foundation Model Pre-trained with Clinical Diagnostic Reports." *arXiv preprint arXiv:2405.14137* (2024).