# A Novel Fusion Deep Learning Approach for Retinal Disease Diagnosis Enhanced by Web Application Predictive Tool

Nani Gopal Barai[1], Subrata Banik[2], F M Javed Mehedi Shamrat[3]
Bangladesh Japan Information Technology Limited (BJIT Limited), Dhaka, Bangladesh[1, 2]
Department of Computer System and Technology, University of Malaya, Kuala Lumpur, Malaysia[3]

***Abstract*—Retinal disorders such as age-related macular degeneration and diabetic macular edema can lead to permanent blindness. Optical coherence tomography (OCT) enables professionals to observe cross-sections of the retina, which aids in diagnosis. Manually analyzing images is time-consuming, difficult, and prone to mistakes. In the dynamic and constantly evolving domain of artificial intelligence (AI) and medical imaging, our research represents a significant development in the field of retinal diagnostics. In this study, we introduced "RetiNet", an advanced hybrid model that is derived from the best features of ResNet50 and DenseNet121. To the model, we utilized an open-source retinal dataset that underwent a meticulous refinement process using a series of preprocessing techniques. The techniques involved Histogram Equalization for the purpose of achieving optimal contrast, Gaussian blur to mitigate noise, morphological operations to facilitate precise feature extraction, and Data Balancing to ensure impartial model training. These operations led to the attainment of a test accuracy of 98.50% by RetiNet, surpassing the performance standard set by existing models. A web application has been developed with the purpose of disease prediction, providing doctors with assistance in their diagnostic procedures. Through the development of RetiNet, our research not only transforms the accuracy of retinal diagnostics but also introduces an innovative combination of deep learning and application-oriented solutions. This innovation brings in a novel era characterized by improving reliability and efficiency in the field of medical diagnostics.**

***Keywords*—Retinal disease; RetiNet; hybrid model; learning; Web application; gaussian blur; histogram equalization**

## I. INTRODUCTION

This Retinal disorders commonly result from a combination of predispositions and environmental factors, mostly affecting the structures within the eyes, such as the membrane, lens, and nerve systems within eye. These conditions exert a significant impact on individuals' quality of life. Notably, approximately 7% of those aged 65 and above report experiencing a form of visual impairment [1]. The importance of timely and accurate diagnosis cannot be overstated in mitigating the severity of these conditions. The usage of inappropriate diagnostic approaches may exacerbate the issue. Diabetic Eye Disease (DED), comprising conditions such as diabetic retinopathy, glaucoma, and cataracts, is a significant cluster of visual impairment that impacts individuals with diabetes. Extended periods of diabetics can result in a decline in visual acuity and, in severe cases, substantial visual loss. As stated by the World Health Organization (WHO), among the global population of 2.2 billion people affected by visual impairments, it is estimated that approximately one billion cases may have been preventable with proper diagnostics and treatment [2].

Artificial Intelligence has become an essential resource in assisting healthcare professionals in the early diagnosis of diseases [3, 4]. Currently, numerous AI-driven systems integrate medical test results with domain-specific knowledge to detect and categorize diseases. Furthermore, deep learning (DL) has been employed in various practical contexts, showcasing its promise. Specifically, researchers have utilized DL techniques to identify retinal disorders by analyzing retinal fundus images. Although DL approaches in the field of machine learning (ML) succeed at distinguishing between healthy and diseased retinal images, the challenge of classifying varied retinal disorders into multiple classes remains complex and unresolved.

Numerous studies have attempted to predict or characterize retinal health by analyzing eye images. For instance, the authors in [5] proposed an enhanced technique that employs a novel CenterNet model in conjunction with a DenseNet-100 feature extractor. The study aimed to identify and characterize lesions associated with diabetic retinopathy and macular edema. The approach demonstrated exceptional accuracies of 97.93% and 98.10% when assessed on the APTOS-2019 and IDRiD benchmark datasets. In a comparable way, the study conducted by [6] tested the performance of three distinct classification algorithms across various classes. These algorithms included the Convolutional Neural Network (CNN), Visual Geometry Group 16 (VGG16), and InceptionV3. The analysis and comparison of each approach were performed using the confusion matrix.

The prevalence of retinal disease in ophthalmology has been a significant concern due to its diverse manifestations and rising frequency. Traditional diagnostic approaches, although valuable, may prove ineffective due to inherited human fallibility, namely inaccurate assessment of the intricate details of retinal vision. In order to address challenges, we devised "RetiNet". RetiNet has been developed to surpass the existing limitations in the field of retinal image diagnosis by harnessing the potential of DL to address the associated challenges. The primary focus of RetiNet is the enhancement of diagnostic accuracy. Our study presents a comprehensive approach that presents cutting-edge AI techniques aiming to improve the

standards of retinal disease detection. Fig. 1 represents the overall workflow of the proposed system. The primary findings of the study are as follows:

- This study introduced "RetiNet", a novel hybrid model generated from the fusion of ResNet50 and DenseNet121.
- A comprehensive set of preprocessing techniques were utilized, including Histogram Equalization, Gaussian Blur, Morphological Operations, and Data Balancing. The methods guaranteed optimal image quality and consistency, hence facilitating efficient model training and accurate diagnosis.
- The performance of RetiNet surpasses that of other popular models. The outcome emphasizes the efficiency of our hybrid methodology, demonstrating its potential for real-world implementation and reliability.
- This study incorporates a web application in a novel way, providing an efficient instrument for the diagnosis of retinal diseases.

The paper's structure is outlined in the following manner: Section II thoroughly examines previous research. Section III, which focuses on technique, addresses several issues such as data acquisition, data preparation, use of pre-trained models, proposed models, and their corresponding hyperparameters. Section IV provides an in-depth analysis of the results obtained from the research. Section V is on the development and implications of a functional web application. Section VI involves a discourse and comparative examination with prior research. Section VII eventually concludes the article by presenting essential findings and valuable perspectives.

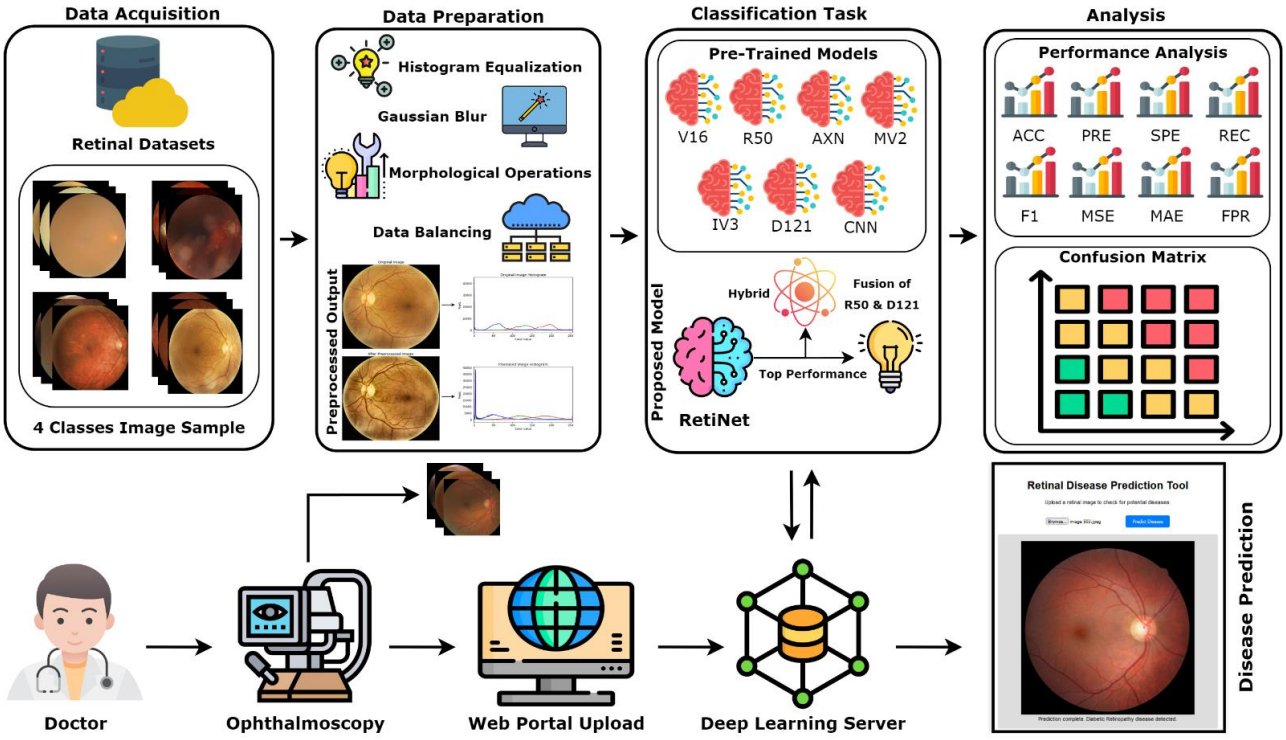


Fig. 1. The overall architecture of the proposed hybrid model RetiNet.

## II. Literature Review

Throughout several decades, the scientific community has diligently advocated for the advancement of automated diagnostic systems, with the ultimate goal of revolutionizing the field of medical diagnostics. Traditional expert systems, albeit groundbreaking, operated based on meticulously defined rules, occasionally encountering difficulties when confronted with complex classification challenges. Nevertheless, introducing machine and deep learning has brought forth a renewed vigor in this sector. Through the use of data training, these algorithms have made a substantial contribution to the improvement of research endeavors. Significantly, within the scope of medical research, machine learning distinguishes itself due to its remarkable versatility, demonstrating proficiency in detecting and classifying a multitude of disorders across various domains.

In a recent study by Arslan et al. [7], various CNN architectures were employed to evaluate a 2748 Retinal Fundus images dataset. This dataset comprised 1374 images from healthy individuals and 1374 images from diverse diseased groups. The CNN models underwent thorough evaluation using the 10-fold cross-validation methodology. Remarkably, the EfficientNet architecture demonstrated superior performance with an impressive accuracy and recall rate of 94.88% in measurements. Similarly, Malik et al. [8] developed a rapid diagnosis approach for eye diseases by utilizing a diverse range of machine learning models using Neural Networks. This array consisted of algorithms such as the Decision Tree, Random

Forest, and Naïve Bayes. The ICD-10 codes were used for specific disease diagnoses. Among the classifiers, the Random Forest model had superior performance, achieving an accuracy rate of 86.63%. The Neural Network model closely trailed after, achieving an accuracy rate of 85.98%. In a separate investigation centered on Optical Coherence Tomography, Metin and Karasulu [9] employed two CNN models, ResNet50 and MobileNetV2, as the foundation for their research. When the aforementioned models were employed to analyze data pertaining to several retinal diseases, ResNet50 exhibited an accuracy of 94% whilst MobileNetV2 dropped to 81%.

Sarki et al. [10] introduced an automated technique for the detection of Diabetic Eye Diseases (DEDs) by the analysis of retinal fundus images. The classifier chosen for their study was the CNN model, which was further optimized by fine-tuning using the RMSprop optimizer. The dataset, comprised of images from multiple sources, was specifically curated to facilitate multi-class classification. Remarkably, their classifier achieved an accuracy rate of 81.33% in retinal disease classification. Hussain et al. [11] conducted another study utilizing a dataset consisting of 251 spectral-domain optical coherence tomography (SD-OCT) images. Among the sample, a total of 192 cases were indicative of retinal diseases, while the remaining 59 cases were of healthy eyes. The researchers utilized the Random Forest classification technique to differentiate between data indicative of illness and normal conditions. To ensure the model's robustness, a 15-fold cross-validation process was utilized, yielding a remarkable classification accuracy of 96.89%.

Additionally, Almansour et al. [12] introduced a CNN architecture that drew inspiration from the VGG16 model, with a specific focus on glaucoma detection. The researchers acquired data from a total of seven distinct databases. The localization of the Region of Interest (ROI) was given particular emphasis, resulting in the allocation of a total of 2084 samples for classification subsequent to ROI analysis. To enhance the performance of the VGG16 model, two more layers were incorporated, ending with the Softmax activating function. Upon evaluating the combined data from all the datasets, the model demonstrated a noteworthy accuracy rate of 78%. In an independent study, Seker et al. [13] developed a CNN model using Keras framework for the purpose of glaucoma classification. The fundus images were first subjected to preprocessing using the Irfanview graphic schemes prior to being inputted into the classification pipeline. The suggested framework was robust, consisting of 49 layers and employing the Adam optimizer with binary cross-entropy loss function. This demonstrates a classification accuracy of 85%.

The limitations in the current retinal disease identification using conventional imaging techniques have been evident through various research. Numerous contemporary models encounter challenges pertaining to the accuracy rate of predictions and the complexity involved in augmenting retinal image datasets. In order to address these challenges and enhance the quality of retinal diagnostics, we introduced "RetiNet". The primary objective with RetiNet is to utilize its hybrid architecture to achieve exceptional accuracy in the diagnosis and classification of retinal diseases. This attempt highlights our commitment to enhancing diagnostic skills and redefining the parameters of retinal imaging diagnostics.

## III. Methodology

### A. Data Acquisition

In the context of retinal disease diagnosis investigation, we employed an extensive dataset of retinal images collected from the renowned open-source platform Kaggle [14]. The dataset is crucial to our study as it contains many images corresponding to various eye conditions. This comprehensive collection allows for a balanced and intricate approach to our analysis. The dataset consists of 1074 images depicting Healthy Eyes, 1038 images of Cataract, 1098 images implying Diabetic Retinopathy, and 1007 images displaying Glaucoma. Such heterogeneity not only contributes to the robust classification of various eye diseases but also provides a dimension of authenticity to our investigation. For an understanding of the range of diversity present in the dataset, Fig. 2 may serve as a useful reference, as it displays sample images from each category. The utilization of this dataset has played a pivotal role in our research, facilitating the establishment of evaluations and findings based on empirical evidence from the real world.

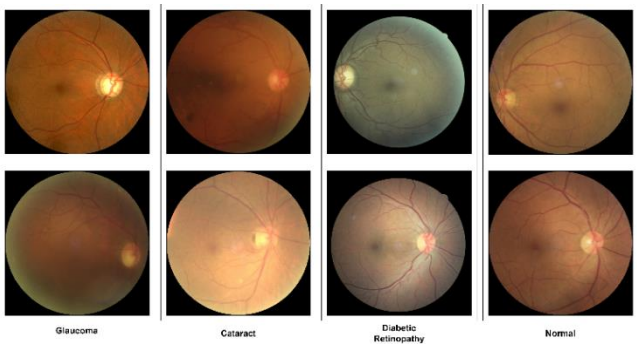


Fig. 2. Sample images of the four classes from the selected retinal datasets.

### B. Data-preparation

*1) Histogram equalization:* The study utilized a specialized histogram equalization technique to manage variations in contrasts across color images effectively. The first step involved the conversion of each image from the RGB color system to the YCbCr color space. The application of histogram equalization was restricted solely to the Y channel, representing luminance, to mitigate any potential color distortion. The equalization procedure commenced by calculating the histogram for the Y channel, which signifies the distribution of pixel intensities. The resulting Cumulative Distribution function (CDF) was utilized to remap each pixel intensity, r in the Y channel according to the following formula:

$$s = \frac{(CDF(r) - CDF_{min})}{(1 - CDF_{min})} (L - 1) \quad (1)$$

Here, L is the grayscale level, and CDFmin is the smallest non-zero value in the CDF.

This process facilitates the even dispersion of luminance intensities, consequently enhancing regions with low contrast. After equalization, the modified Y channel was reintegrated into the YCbCr image, which was subsequently transformed back to RGB. This approach resulted in a standardized contrast through the entire dataset, generating a consistent foundation for subsequent analysis. Algorithm 1 outlines the operational steps involved in the implementation of histogram equalization.

**Algorithm 1:** Histogram Equalization for Retinal Images

**Procedure** ColorHistEqu (Image $I$, GrayscaleLevels $L$)
  $I_{YCcCr} \leftarrow$ Convert $I$ to YCbCr color space
  Extract $Y$ channel as $I_Y$
  Initialize histogram array $H$ [0….$L$-1] to zeros
  Initialize CDF array $CDF$ [0….$L$-1] to zeros
  **for** each pixel $p$ in $I_Y$ **do**
    $H\ [intensity\ of\ P] \leftarrow H\ [intensity\ of\ P] + 1$
  **end for**
  $CDF\ [0] \leftarrow H\ [0]$
  **for** $i = 1$ to $L$-1 **do**
    $CDF\ [i] \leftarrow CDF\ [i-1] + H[i]$
  **end for**
  $cdf_{min} \leftarrow$ minimum non-zero value of $CDF$
  **for** $i = 0$ to $L-1$ **do**
    $normalized_{cdf}\ [i] \leftarrow \frac{(CDF[i]-cdf_{min})}{(1-cdf_{min})}\ (L-1)$
  **end for**
  **for** each pixel $p$ in $I_Y$ **do**
    $P_{newintensity} \leftarrow normalized_{cdf}[intensity\ of\ P]$
  **end for**
  Replace Y channel in $I_{YCbCr}$ with $I_Y$
  Convert $I_{YCbCr}$ back to RGB color spcae
  Save $I_{YCbCr}$ as ‘Output.jpg’
  **return** $I_{YCbCr}$
**end procedure**

*2) Gaussian blur:* The Gaussian blur is a widely used convolutional technique where Gaussian Kernel G is convolved with an image I. In the domain of image processing, a technique akin to a weighted average of pixel values is employed, wherein the weights progressively diminish as the distance from the center pixel increases. The Gaussian function in two dimensions is mathematically defined in Eq. (2).

$$G(x,y) = \frac{1}{2\pi\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

Here, the variables x and y represent spatial coordinates, whereas σ is the standard deviation that governs the extent of the Gaussian kernel’s distribution. The kernel radius is commonly selected as μ =3 σ to account for more than 99% of the Gaussian distribution. To implement the Gaussian Blur on an image, a convolution operation is executed between the image and the Gaussian kernel.

$$(I * G)(u,v) = \sum_{x=-\mu}^{\mu} \sum_{y=-\mu}^{\mu} I(u-x, v-y) \cdot G(x,y) \quad (3)$$

The variable $(u, v)$ represent the pixel coordinates in image $I$, whereas the variables $(x, y)$ iterate over the dimensions of the Gaussian kernel. The outcome, denoted as $(I * G)(u, v)$, represents a blurred image.

The convolution process is employed to amplify the influence of the central pixels relative to the distant ones, resulting in a visible smoothing effect. As σ increases, the level of blurring becomes more pronounced, resulting in a greater impact on the surrounding pixels. Algorithm 2 is an illustration of the Gaussian blur technique.

**Algorithm 2:** Gaussian Blur for Image Smoothing

**Procedure** GaussianBlur(Image $I$, StandardDeviation $\sigma$)
  Compute Gaussian kernel radius $\mu \leftarrow 3\times\sigma$
  Initialize Gaussian Kernel G with size $(2\mu + 1) \times (2\mu + 1)$
  **for** $x = -\mu to \mu, y = -\mu to \mu$ **do**
    $G[x+\mu][y+\mu] \leftarrow \frac{1}{2\pi\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}}$
  **end for**
  Normalize kernel $G$ such that its sum is 1
  **for** each pixel $(u, v)$ in $I$ **do**
    $I_{blurred}[u][v] \leftarrow \sum_{x=-\mu}^{\mu} \sum_{y=-\mu}^{\mu} I(u-x, v-y) \times G[x+y][y+\mu]$
  **end for**
  **return** $I_{blurred}$
**end procedure**

*3) Morphological operations:* The assessment of eye diseases using retinal images requires meticulous attention to image quality and clarity. Morphological methods, which are fundamentally non-linear in nature, have been integral to our efforts in enhancing these images based on their form. For instance, the dilation technique is used to amplify white regions of the foreground of images. The process of enlarging retinal images improves recognizable characteristics, particularly, blood vessels, to make them easier to identify when applied on images. The concept can be articulated as,

$$f \oplus s\ (x,y) = max_{(a,b)\in s}\{f(x-a, y-b)\} \quad (4)$$

In contrast, erosion functions as an inverse of dilation as it reduces the white regions in the images. This contraction is highly advantageous for the purpose of severing associated objects or eliminating minor noise components, as observed in the context.

$$f \ominus s\ (x,y) = min_{(a,b)\in s}\{f(x+a, y+b)\} \quad (5)$$

The improvement of images can be achieved by implementing the opening operation, which involves a sequential process of erosion followed by dilation. This technique effectively eliminates minor protrusions or objects, which is a necessary tool for reducing noise and artifacts. This operation can be represented as,

$$f \circ s = (f \ominus s) \oplus s \quad (6)$$

The final component of the morphological approaches is the closure procedure. Starting with dilating, followed by degrading an image, it adeptly closes tiny holes or breaches in

the foreground, required to restore the discontinuities in blood vessels. The aforementioned procedure can be depicted as,

$$f \bullet s = (f \oplus s) \ominus s \quad (7)$$

Following the series of preprocessing techniques, such as histogram Equalization, Gaussian Blur, and Morphological Operations, the retinal images underwent processing to achieve optimal clarity and enhancement of features. The images generated are prepared for training and subsequently fed into our neural network model, as illustrated in Fig. 3. The histogram of both images (original and preprocessed) shows the changes after applying all the preprocessing techniques (Histogram Equalization, Gaussian blur, and morphological operations).

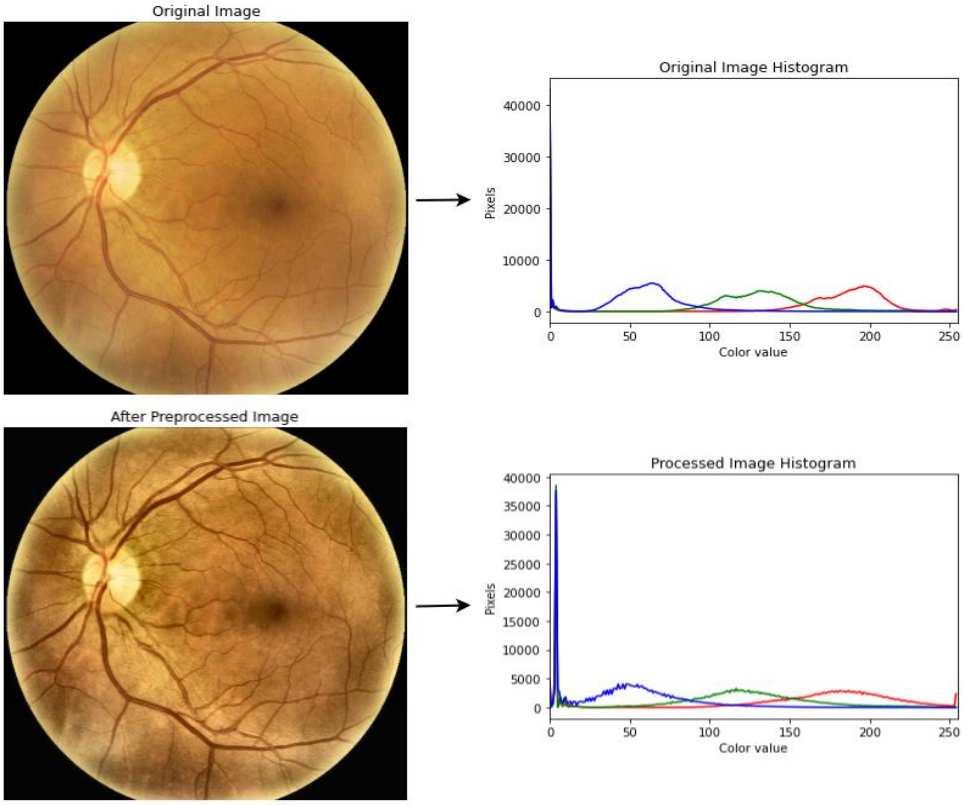


Fig. 3. After all preprocessed steps histogram showing the changes of image enhancement ratio.

*4) Data balancing:* In the domain of machine learning and data analysis, the presence of an unbalanced dataset has the potential to introduce bias into the model predictions, particularly when one class is disproportionately represented in comparison to the others. The disparity between the classes was detected in our study by analyzing retinal images. The data under-sampling method was employed to resolve this issue. Strategic undersampling involves reducing the number of instances in the overrepresented or majority classes in order to align with the size of the minority class rather than artificially increasing the number of instances in the underrepresented class. This approach equalized 1000 images for each class covering Normal, Diabetic Retinopathy, Cataract, and Glaucoma. While implementing this strategy decreased the overall amount of data for the majority classes, it played a vital role in achieving a balanced distribution of data. A balanced dataset reduces the potential for biased predictions while simultaneously improving the model's capacity for generalization. Algorithm 3 illustrates the general under-sampling technique.

**Algorithm 3:** Data Undersampling for Class Balance

**Procedure** Undersample (Dataset $D$)
  Determine the size of the smallest class, *minSize*
  **for** each class $c$ in $D$ **do**
    $data_c \leftarrow$ Data instances of class $c$
    **if** size of $data_c >$ *minSize* **then**
      Randomly select *minSize* instances from $data_c$ to form $newdata_c$
    **else**
      $newdata_c$
    **end if**
  **end for**
  Merge all $newdata_c$ to form the balanced dataset $D'$
  **return** $D'$
**end procedure**

After establishing a balanced dataset, we further split our data into training, test, and validation subsets to assist the modeling stage. The division adopted a ratio of 70:20:10. The training set consists of 70% of the images from each class, serving as the fundamental data for model learning. The test set comprised 20% of the data and was used to evaluate the performance of the training model on previously unseen data.

The remaining 10% of the data was allocated as the validation set, which serves as an essential component for repeated modification and fine-tuning of the model during the training process. Table I presents a comprehensive analysis detailing the distribution of images across each set and class. The table presented visually demonstrates the rigorous distribution method adopted to achieve effective model training and validation.

TABLE I. FINAL DATA DISTRIBUTION FOR EACH CLASS AND THE ALLOCATION OF DATA INTO TRAINING, TESTING, AND VALIDATION SETS

| Class | Original | Balance |
|---|---|---|
| Normal | 1074 | 1000 |
| Diabetic Retinopathy | 1098 | 1000 |
| Cataract | 1038 | 1000 |
| Glucoma | 1007 | 1000 |
| Total | 4217 | 4000 |
| Train | 2800 Images (70%) | |
| Test | 800 Images (20%) | |
| Validation | 400 Images (10%) | |

### C. Model Evaluation and Classification

The primary objective of our proposed approach is to automatically identify eye disorder while delivering an improved level of classification accuracy. To address the challenges of classifying retinal conditions from retinal datasets, we have developed a novel and robust hybrid CNN model RetiNet. This model is characterized by its distinctiveness, reliability and resilience. In the pursuit of identifying the most effective transfer learning approach for the classification task at hand, we tested seven recognized pre-trained models: VGG16 [15], ResNet50 [16], AlexNet [17], MobileNetV2 [18], InceptionV3 [19], DenseNet121 [20] and a CNN [21]. Table II provides an overview of the important characteristics and fundamental elements of the chosen deep CNN models. The subsequent sections provide a comprehensive evaluation of the architecture and performance of RetiNet, as well as the primary focus of the study.

TABLE II. FEATURES AND ATTRIBUTES OF EVALUATED DEEP LEARNING MODELS

| Model | Input Shape | Custom Input Shape | Parameters | Size (MB) |
|---|---|---|---|---|
| VGG16 | 224×224 | 224×224 | 138,35,7544 | 528 |
| ResNet50 | 224×224 | 224×224 | 25,636,712 | 98 |
| AlexNet | 227×227 | 224×224 | 62,378,344 | 233 |
| MobileNetV2 | 224×224 | 224×224 | 35,38,984 | 14 |
| InceptionV3 | 229×229 | 224×224 | 23,851,784 | 92 |
| DenseNet121 | 224×224 | 224×224 | 80,62,504 | 33 |
| CNN | 224×224 | 224×224 | 78,81,365 | 39 |
| RetiNet | 224×224 | 224×224 | 35,100,000 | 136 |

*1) Proposed Models (RetiNet):* In the context of eye disease classification from retinal images, a sensitive field of study, the deliberate and evidence-driven decision to combine the capabilities of ResNet50 and DenseNet121. ResNet50 is renowned for its innovative skip or "residual connections" which efficiently counteracts the vanishing gradient issues that often arise in deep neural networks. This ensures that, regardless of the depth of the network, gradients flow smoothly, hence facilitating rapid learning. This ResNet has consistently exhibited its efficacy in many image classification challenges, successfully detecting both broad structures, such as the overarching form of blood vessels, as well as the intricate details, such as subtle deviations that may indicate potential disorders. In contrast, DenseNet121 exhibits an exceptionally thick architecture, where each layer establishes intimate interconnections with all other layers, facilitating a seamless transmission of information.

Since minor details in retinal images might function as early indicators of disorders such as diabetic retinopathy or glaucoma, the dense linkage encourages feature reuse. Additionally, it functions as a built-in regularization mechanism, safeguarding against overfitting, a critical consideration when dealing with constrained datasets. While architectures such as VGG and Inception possess their own merits, our hybrid model is particularly well-suited for the complex demands of retinal image classification due to the combination of the depth and skip connections of ResNet50, along with the extensive feature extraction capabilities of DenseNet-121. The proposed RetiNet model is outlined as follows:

When presented with an input image I, both the ResNet50 and DenseNet121 architectures execute a series of convolutional operations to extract feature maps.

$$F_{resnet}(I) = ResNet50\ (I) \tag{10}$$

$$F_{densenet}(I) = DenseNet121\ (I) \tag{11}$$

where, $F_{resnet}$ and $F_{densenet}$ are the feature maps from ResNet50 and DenseNet121, respectively.

Furthermore, we apply global average pooling (GAP) to these feature maps to get a fixed-size feature vector. For a given feature map F, the GAP operation can be expressed as,

$$GAP\ (F) = \frac{1}{W \times \mathrm{H}} \sum_{i=1}^{W} \sum_{j=1}^{H} F(i,j) \tag{12}$$

where, W and H are the width and height of the feature maps, respectively.

However, the feature vectors obtained from the GAP operation on both networks are concatenated.

$$C = Concatenate\ (GAP(F_{resnet}), GAP(F_{densenet})) \tag{13}$$

The concatenated feature vector $C$ is passed through a dense layer with a ReLU activation function,

$$D = \sigma(W_d \times \mathrm{C} + b_d) \tag{14}$$

where, $\sigma$ is the ReLU activation function, $W_d$ represents the weight of the dense layer and $b_d$is the bias.

Finally, the feature vector from the dense layer is passed through another dense layer with softmax activation to classify the image into the four classes of eye disease.

$$O = Softmax(W_o \times \mathrm{D} + b_o) \tag{15}$$

where, $W_o$ represents the weights of the output layer and $b_o$ is the bias.

The output $O$ will be a vector with four values, each representing the probability of the image belonging to the corresponding class (Normal, Diabetic Retinopathy, Cataract, and Glaucoma). Algorithm 4 demonstrates the procedural steps of the proposed RetiNet model. Furthermore, Fig. 4 illustrates the architecture of the hybrid model.

**Algorithm 4:** RetiNet: Hybrid Model for Eye Disease Classification

**Procedure** RetiNet (Image $I$)
  $I_{norm} \leftarrow$ Normalize $I$
  $F_R \leftarrow$ ResNet50_Extract ($I_{norm}$)
  $F_D \leftarrow$ DenseNet121_Extract ($I_{norm}$)
  $G_R \leftarrow$ GAP ($F_R$)
  $G_D \leftarrow$ GAP ($F_D$)
  $C \leftarrow Concat\ (G_R, G_D)$
  $D \leftarrow Dense\_ReLU\ (C)$
  $O \leftarrow Softmax\ (D)$
  ***return*** $O$
**end procedure**

### D. Hyperparameters Optimization

Hyperparameters are significant variables that might have an influence on the training dynamics and overall performance of the model [22], [23]. The variables include the number of epochs, batch size, image dimensions, optimizer options, activation functions, learning rate, decay rate, dropout rate, and regularization parameters. During the experiment, we repeatedly modified several parameters, including batch size, learning rate, and regularization variables, resulting in improvements in the model's accuracy and efficiency. RetiNet was subjected to benchmarking against several prominent architectures, namely VGG16, ResNet50, AlexNet, MobileNetV2, InceptionV3, DenseNet121, and a customized CNN. The training process for each model consisted of 300 epochs, during which various optimizers were employed to facilitate a comprehensive evaluation. To fine-tune our model and to determine the optimal hyperparameter configuration, we apply the keras-tune tool, followed by an extensive grid search technique. Table III presents the final set of hyperparameters post-tuning.

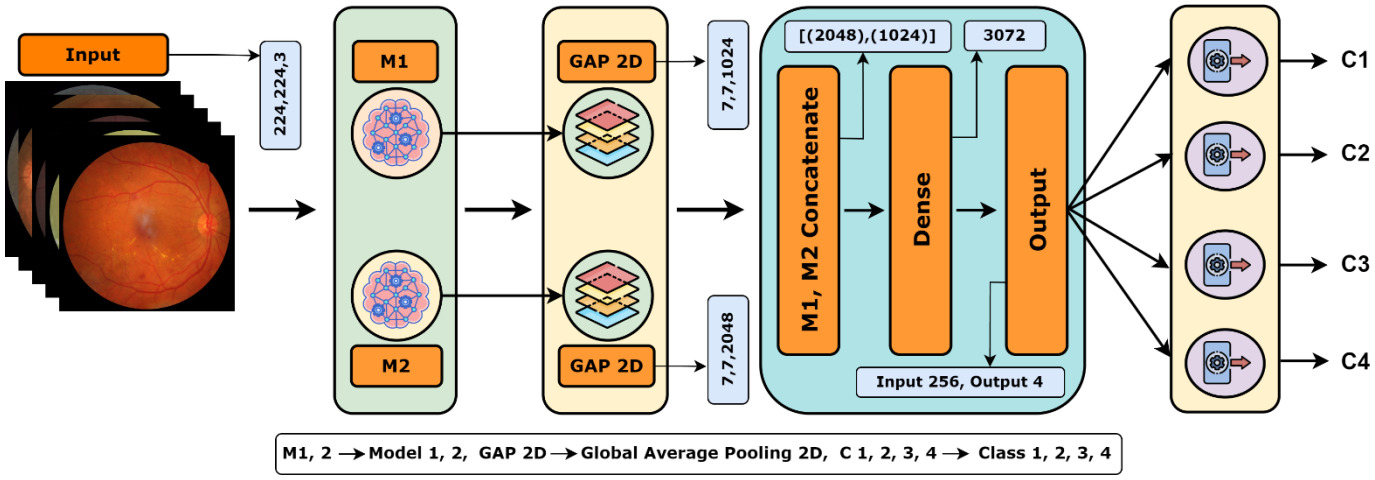


Fig. 4. The architecture of the proposed model RetiNet.

TABLE III. OPTIMIZED HYPERPARAMETERS FOR MODEL TRAINING

| Model | No. of Epochs | Batch Size | Image Size | Optimizers | Activation Function | Learning Rate | Decay Rate | Dropout Rate | Regularizer |
|---|---|---|---|---|---|---|---|---|---|
| VGG16 | 300 | 64 | 224×224 | Adam | Softmax | 0.000001 | 1e-3 | 0.5 | 5e-4 |
| ResNet50 | 300 | 64 | 224×224 | SGD | ReLU | 0.0001 | 1e-4 | - | 1e-4 |
| AlexNet | 300 | 64 | 224×224 | Adagrad | ReLU | 0.00001 | 1e-2 | 0.2 | 5e-4 |
| MobileNetV2 | 300 | 64 | 224×224 | SGD | Softmax | 0.1 | 1e-4 | 0.5 | 1e-5 |
| InceptionV3 | 300 | 64 | 224×224 | Adam | Sigmoid | 0.001 | 1e-3 | - | 1e-4 |
| DenseNet121 | 300 | 64 | 224×224 | Adam | ReLU | 0.0000001 | 1e-2 | - | 1e-4 |
| CNN | 300 | 64 | 224×224 | Adam | ReLU | 0.001 | 1e-2 | 0.5 | 1e-4 |
| RetiNet | 300 | 128 | 224×224 | RMSProp | Sigmoid | 0.001 | 1e-6 | 0.9 | 1e-4 |

## IV. ANALYSIS OF EXPERIMENTAL OUTCOMES

### A. Environmental Setup and Tools

We implemented all the deep learning models using Keras (version 2.10.0) and TensorFlow (version 2.0) frameworks using Python 3.7. For data visualization, we deployed the Seaborn and Matplotlib packages. The evaluation was carried out on a device powered by an AMD Ryzen 7 CPU clocked at 3.90 GHz, 32 GB RAM, and an AMD Radeon RX 580 series GPU, functioning on a Windows 10 operating system.

### B. Assessment Metrics

To comprehensively evaluate the performance of our model in the classification of retinal images, we adopted a set of statistical metrics that involves Accuracy, Specificity, Recall,

Precision, False Positive Rate (FPR), F1-score, Mean Squared Error (MSE), and Mean Absolute Error (MAE).

**Accuracy**: Represents the proportion of correct predictions made by the model over the total predictions.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (16)$$

Here, $TP$ = True Positives, $TN$= True Negatives, $FP$= False Positives and $FN$ = False Negatives.

Precision: Indicates the fraction of relevant instances among the instances that the model predicted as positive. It provides insight onto the correctness of positive predictions.

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

Recall: Signifies the proportion of real positive occurrences that the model managed to predict accurately, demonstrating its ability to detect positive cases.

$$Recall = \frac{TP}{TP+FN} \quad (18)$$

Specificity: Measures the actual negative rate, indicating the model's effectiveness in accurately recognizing the negative class amongst all the classes.

$$Specificty = \frac{TN}{TN+FP} \quad (19)$$

FPR: Proportion of negative instances that are incorrectly classified as positive.

$$FPR = \frac{FP}{FP+TN} \quad (20)$$

F1-Score: Provides a balance between Precision and Recall by combining them into a single measure. This metric also shows how well a model can be used to identify both positive and negative data.

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (21)$$

MSE: Determines the average squared variance between the predicted outcomes and the actual values, delivering a sense of prediction error size.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (22)$$

Here $y_i$ is the actual value, $\hat{y}_i$ is the predicted value, and $n$ is the number of observations.

MAE: Provides the mean of the absolute variance between predictions and actual observations, exhibiting the model's accuracy.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (23)$$

### C. Study Outcomes

In this study, eight algorithms were utilized to classify retinal image data. These algorithms comprised seven transfer learning models and the novel RetiNet model. These models aim to assist in identifying eye conditions by accurately detecting abnormalities in retinal images. In order to conduct an extensive evaluation, each model underwent training for 300 epochs, with the results being recorded at each iteration. The performance measures of each model can be calculated by utilizing Eq. (16) to Eq. (23). This thorough methodology provides a profound understanding of the functioning of each model when exposed to retinal data.

This study systematically evaluated the performance of eight distinct models on their ability to classify retinal images with high consistency. Fig. 5 illustrates the performance of all the models. Among the employed transfer learning models, VGG16 exhibited commendable performance, achieving an accuracy of 91.58% and a precision of 92.1%. These results serve as evidence of its proficiency in accurately classifying positive instances. The model, ResNet50, achieved an accuracy of 86.36% and a specificity of 86.7%, indicating its ability to classify negative samples effectively. The AlexNet model, with its accuracy and recall rates of 92.76% and 92.9%, respectively, demonstrated its adeptness in identifying actual positive rates.

The MobileNetV2 and InceptioV3 models exhibited distinct performance characteristics, achieving accuracies of 93.31% and 93.5%, respectively. Both models displayed impressive F1 scores, representing the harmonic mean of accuracy and recall, indicating a well-balanced performance across both metrics. InceptionV3 demonstrated a balanced detection capability, as evidenced by the achievement of an F1-score of 93.35%. When comparing the performance of DenseNet121 and CNN, it is observed that both models achieved satisfactory results, with accuracies of 90.86% and 88.34%, respectively. These findings suggest that there may be potential for further enhancement in the model's performance. The accuracy rate of the improved model, RetiNet is 98.50%. Another notable statistical measure that evaluates the average squared variations between predicts and actual observations, known as the Mean Squared Error (MSE), was found to be exceptionally low for RetiNet at 0.015. This indicates that the predictions made by RetiNet are highly accurate and closely aligned with the actual results.

Nevertheless, the focal point of the assessment was the cutting-edge RetiNet model. The system exhibits a remarkable level of accuracy of 98.50% and F1-score of 98.65%. Additionally, it demonstrates a precision of 98.7% and an impressively low FPR of 1.7%. This accomplishment showcases the proficiency of RetiNet in accurately identifying and distinguishing retinal abnormalities with high precision.

The combination of all these measurements offers an exhaustive overview of the capabilities of each model. The detailed analysis depicted in Fig. 5 not only highlights the potential of models such as RetiNet but also paves the path for future advancements in the field of retinal image diagnostics.

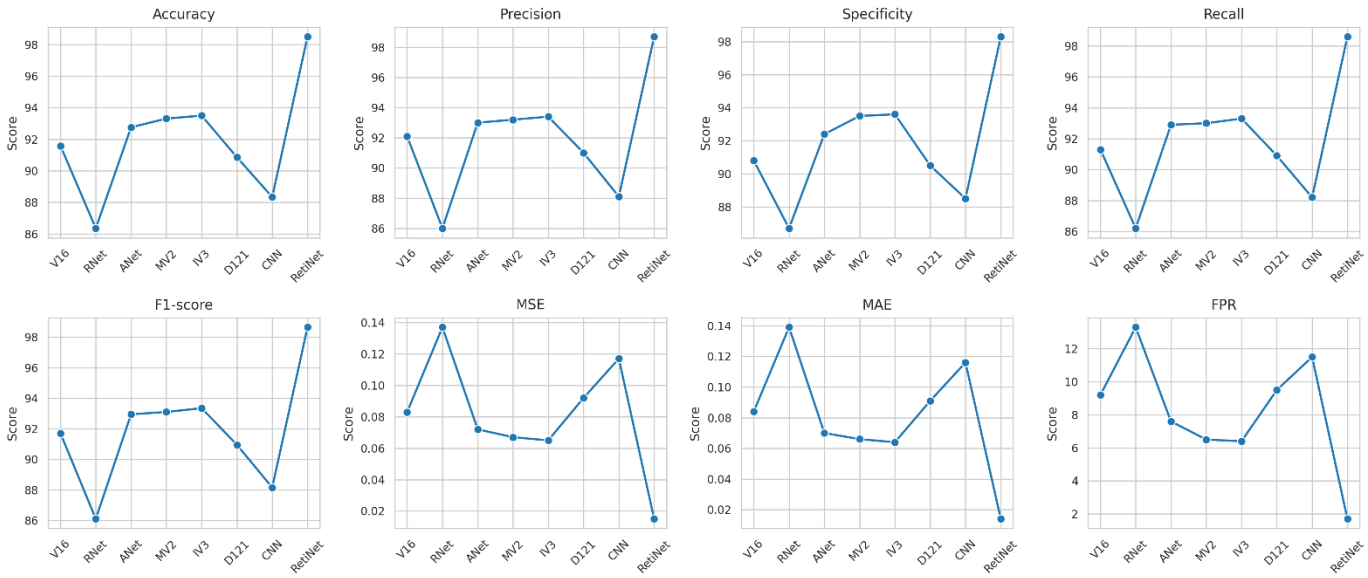


Fig. 5. Comparative performance matrics of retinal image classification models. Here 'V16' indicates 'VGG16', 'RNet' indicates 'ResNet50', 'Anet' indicates 'AlexNet', 'MV2' indicates 'MobileNetV2', 'IV3' indicates 'InceptionV3', 'D121' indicates 'DenseNet121'.

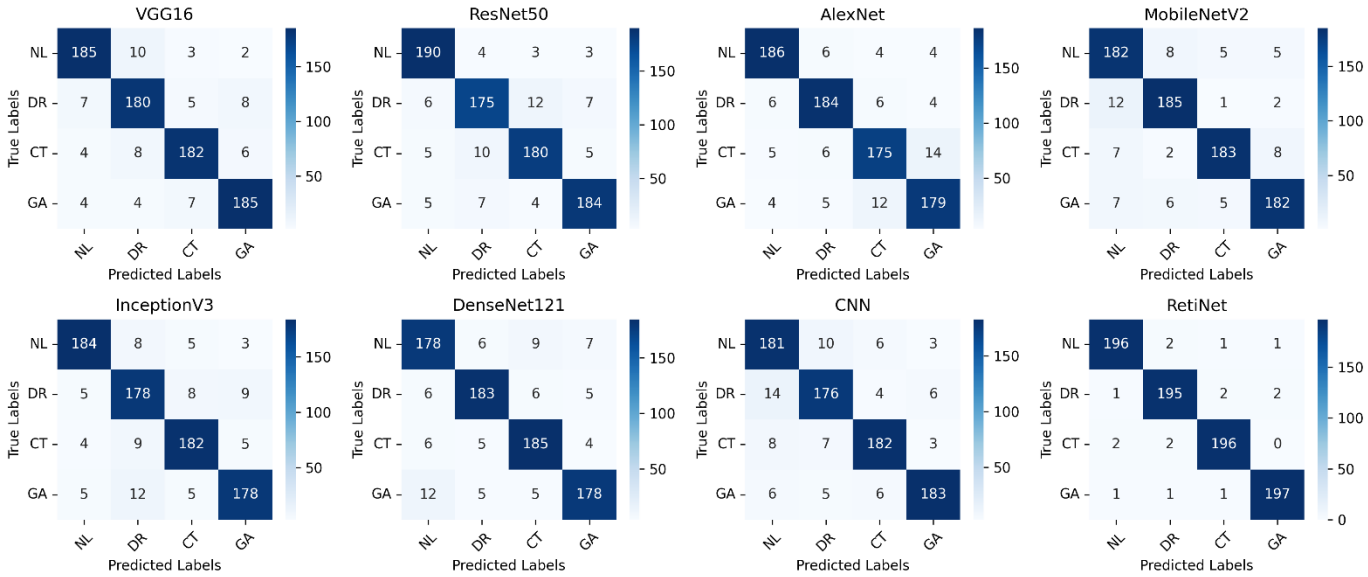


Fig. 6. Confusion matrix of all classification models. Here 'NL' indicates 'Normal' class, 'DR' indicates 'Diabetic Retinopathy' class, 'CT' indicates 'Cataract' class, GA indicates 'Glaucoma' class.

During the testing phase, the model's performance is evaluated using confusion matrices and observing significant patterns of classification and misclassification illustrated in Fig. 6. Out of the 200 Normal data, the VGG16 model accurately classified 185 data while misclassifying 15 data. More specifically, 10 of the data was misclassified as Diabetic Retinopathy. In the Diabetic Retinopathy class, 180 images were correctly identified, whereas five images were incorrectly categorized as Cataract. On the other hand, 182 images depicting Cataract were accurately identified, whereas eight images were erroneously interpreted as Diabetic Retinopathy. Lastly, 185 images were accurately classified in the Glaucoma class, yet seven were mislabeled as Cataracts.

The ResNet50 model accurately classified 190 images as normal and misclassified 10 images, of which four were incorrectly labeled as Diabetic Retinopathy. In the class focused on Diabetic Retinopathy, 175 images were accurately recognized, whereas 12 were mislabeled as Cataract. Among the images in the Cataract class, 180 were found to be correct, while five were classified as Glaucoma. Finally, for the Glaucoma class, 184 accurate classifications were made, and four images were misclassified as Cataract.

AlexNet properly classified 186 images as Normal, while six images as misdiagnosed as Diabetic Retinopathy. In the case of Diabetic Retinopathy, 184 images were correctly classified, except six were misclassified as Cataract. From the

Cataract class, a total of 175 images were correctly identified, and 14 were wrongly identified as Glaucoma. A total of 179 images from the Glaucoma class were classified, and 12 were misclassified as Cataract. The model MobileNetV2 demonstrated proficiency in correctly classifying 182 images as Normal but misclassified eight as Diabetic Retinopathy. The condition known as Diabetic Retinopathy found 185 accurate classifications, but 12 images were mistakenly classified as Normal. The model correctly classified 183 Cataract images but labeled 8 as Glaucoma. Similarly, 182 Glaucoma images could be correctly predicted but seven images were misinterpreted as Normal.

In the case of the model InceptionV3, 184 Normal images were correctly classified and eight were classified as Diabetic Retinopathy. The model identified 178 images of Diabetic Retinopathy but incorrectly identified eight images as Cataract. Within the Cataract class, 182 images were diagnosed, but 5 were misdiagnosed as Glaucoma. From the Glaucoma class, 178 instances were identified, while 12 instances were misidentified as Diabetic Retinopathy. The DenseNet121 properly identified 178 Normal images while incorrectly identifying six as Diabetic Retinopathy. The model accurately classified 183 Diabetic Retinopathy images yet six images wrongly fell into the Cataract category. Within the Cataract class, a total of 185 images were correctly recognized, while four images were erroneously classified as Glaucoma. In the Glaucoma class, there were 178 accurate predictions and 12 instances where the classification was incorrectly assigned as Normal.

In the CNN model, 200 Normal images were tested, of which 181 were correctly classified, while 10 were inaccurately classified as Diabetic Retinopathy. The model properly recognized 176 Diabetic Retinopathy images while incorrectly reporting 14 as the Normal group. For cataract, 182 were accurately recognized, with three misinterpreted for Glaucoma. In the Glaucoma category, 183 were correctly classified, with five misclassified as Diabetic Retinopathy.

Finally, the RetiNet model, the standout performer, successfully identified an outstanding 196 out of 200 Normal images, with only two misclassifications into the Diabetic Retinopathy group. It also properly detected 195 Diabetic Retinopathy images, with misinterpretation of two images into the Cataract class. In the cataract group, 196 interpretations were correctly made, with only two misclassifications into the Glaucoma class. The Glaucoma class witnessed 197 valid classifications with just a single misclassification into the Cataract class. The performance of all the eight models employed in the study is presented in Table IV.

TABLE IV. PERFORMANCE SCORES OF ALL EMPLOYED MODELS FOR RETINAL DISEASE DIAGNOSIS

| Model | VGG16 | ResNet50 | AlexNet | MobileNetV2 | InceptionV3 | DenseNet121 | CNN | RetiNet |
|---|---|---|---|---|---|---|---|---|
| Accuracy | 91.58% | 86.36% | 92.76% | 93.31% | 93.50% | 90.86% | 88.34% | 98.50% |
| Precision | 92.10% | 86.00% | 93.00% | 93.20% | 93.40% | 91.00% | 88.10% | 98.70% |
| Specificity | 90.80% | 86.70% | 92.40% | 93.50% | 93.60% | 90.50% | 88.50% | 98.30% |
| Recall | 91.30% | 86.20% | 92.90% | 93.00% | 93.30% | 90.90% | 88.20% | 98.60% |
| F1-score | 91.70% | 86.10% | 92.95% | 93.10% | 93.35% | 90.95% | 88.15% | 98.65% |
| MSE | 0.083 | 0.137 | 0.072 | 0.067 | 0.065 | 0.092 | 0.117 | 0.015 |
| MAE | 0.084 | 0.139 | 0.070 | 0.066 | 0.064 | 0.091 | 0.116 | 0.014 |
| FPR | 9.2% | 13.3% | 7.6% | 6.5% | 6.4% | 9.5% | 11.5% | 1.7% |

## V. WEB INTERFACE

A precisely designed digital interface has been built exclusively for medical professionals, with a focus on facilitating their use of medical imaging for diagnosing different retinal diseases. The implementation of this modern interface plays a crucial role in facilitating healthcare professionals, particularly doctors, and experts, in effectively and accurately identifying the medical issues impacting their patients. Fig. 7 displays the general interface of the web application.

Doctors initiate the procedure by entering medical images of the patient's retina into the system. The images are instantly uploaded to the server and merged into the proposed RetiNet diagnostic model. Concurrently, the diagnostic form that goes with it is filled out with essential patient data and sent to the server again.

During this phase, the server conducts an essential verification process evaluating the quality of the retinal images that have been submitted to ensure they satisfy the requisite criteria for precise analysis. After conducting this verification process the server utilizes proposed RetiNet technology to examine the images and generate an advanced AI supported estimation of the stage of the disease. This process entails employing of an advanced algorithmic interpretation of the retinal images to determine the condition's progression and severity.

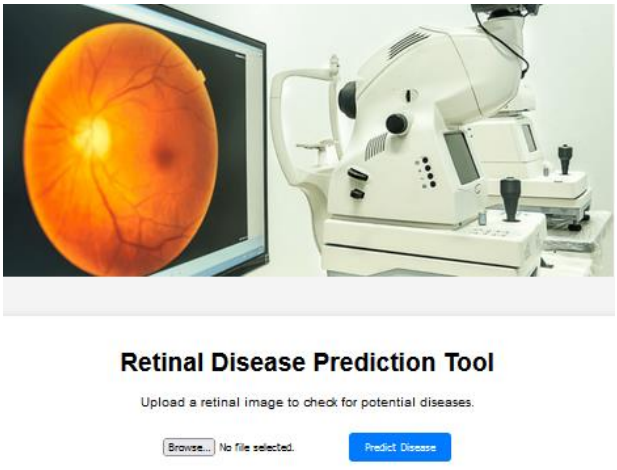


Fig. 7. Web application interface.

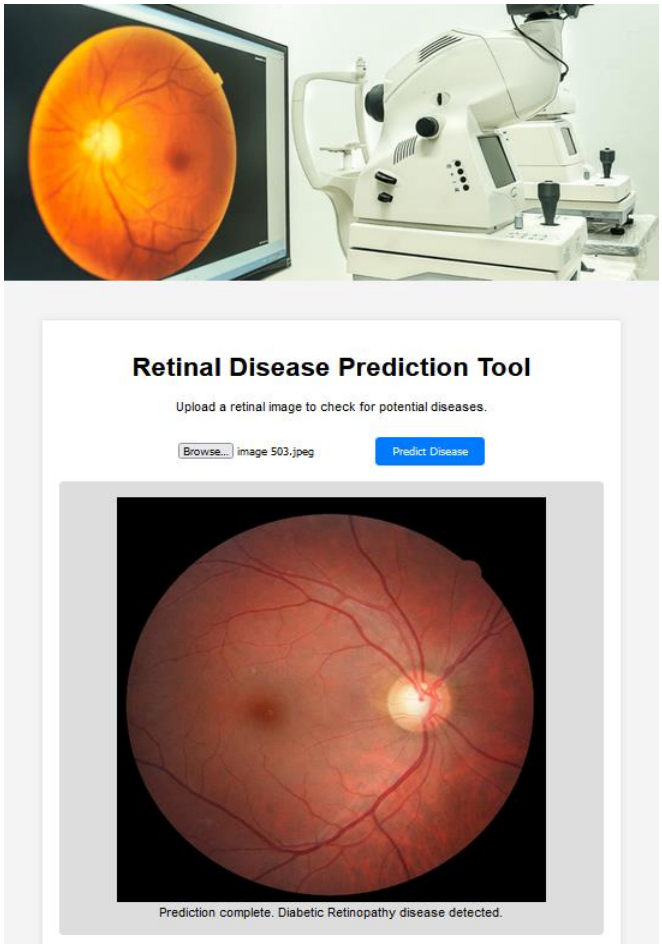


Fig. 8. Disease diagnosis interface.

After the analysis is accomplished, the server aggregates the findings, which include the AI-generated estimate of disease's stage. The doctor is then instantly informed of these findings. Fig. 8 shows the prediction of AI technology help doctors diagnose patients more accurately with more expertise.

## VI. Comparison of Existing Studies

This study presents the "RetiNet" model, an advanced CNN model specifically developed for accurately classifying retinal images. One of the primary objectives is to solve the significant challenge of accurately diagnosing retinal disease within the imaging datasets. Prior to training, the dataset was subjected to thorough preprocessing techniques. These techniques involve Histogram Equalization for optimum image contrast, Gaussian Blur for noise reduction, and Morphological Operations for enhanced feature extraction and Data Balancing to offset class imbalances. During the testing phase, RetiNet demonstrates a remarkable accuracy of 98.50%, surpassing other existing models and effectively enabling disease detection. Moreover, A web-based tool has been created to assist medical professionals in detecting and diagnosing retinal disorders. This initiative set the benchmark for retinal image-based diagnosis and provides medical practitioners with a robust diagnostic tool. Table V provides a detailed analysis of the performance of RetiNet in comparison to other prominent models.

TABLE V. Proposed Model Comparison of Existing Studies

| Authors | Methods | Accuracy |
|---|---|---|
| Arslan et al. [7] | EfficientNet | 94.88% |
| Malik et al. [8] | Random Forest | 86.63% |
| Metin and Karasulu [9] | ResNet50 | 94.00% |
| Sarki et al. [10] | CNN | 81.33% |
| Hussain et al. [11] | Random Forest | 96.89% |
| Almansour et al. [12] | Fine-tuned VGG16 | 78.00% |
| Seker et al. [13] | Keras-based CNN | 85.00% |
| Barai et al. | Proposed Model (RetiNet) | 98.50% |

## VII. Conclusion

This study discusses the significant advantages of employing retinal images for accurately identifying retinal diseases through utilizing our cutting-edge "RetiNet" model. The study subjects each image to a comprehensive preprocessing procedure that includes Histogram Equalization, Gaussian Blur, Morphological Operations, and Data Balancing. We have obtained the highest degree of classification for images that are well-suited for the purpose of extracting features efficiently. A web application has been designed and developed to assist medical professionals in identifying retinal diseases. Although RetiNet's ability did detect some misclassification, its overriding performance spotlighted the transformational potential of CNN models in retinal imaging. Further studies will prioritize the improvement of these methods in order to increase preciseness, with the ultimate goal of transforming the diagnosis of retinal diseases worldwide and significantly influencing ophthalmic healthcare.

## References

[1] Pelletier, A. L., Rojas-Roldan, L., & Coffin, J. (2016). Vision loss in older adults. American family physician, 94(3), 219-226.

[2] World Health Organization (2019), https://www.who.int/newsroom/newsletters, (Last Access: 02.04.2023).

[3] Shamrat, F. J. M., Azam, S., Karim, A., Ahmed, K., Bui, F. M., & De Boer, F. (2023). High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. Computers in Biology and Medicine, 155, 106646.

[4] Shamrat, F. J. M., Akter, S., Azam, S., Karim, A., Ghosh, P., Tasnim, Z., ... & Ahmed, K. (2023). AlzheimerNet: An effective deep learning based proposition for alzheimer's disease stages classification from functional brain changes in magnetic resonance images. IEEE Access, 11, 16376-16395.

[5] Nazir, T., Nawaz, M., Rashid, J., Mahum, R., Masood, M., Mehmood, A., Ali, F., Kim, J., Kwon, H. & Hussain, A. Detection of Diabetic Eye Disease from Retinal Images Using a Deep Learning Based CenterNet Model. Sensors. 21 (2021), https://www.mdpi.com/1424-8220/21/16/5283.

[6] Smaida, M. & Serhii, Y. Comparative Study of Image Classification Algorithms for Eyes Diseases Diagnostic. International Journal Of Innovative Science And Research Technology. 4 (2019).

[7] ARSLAN, G., & Erdaş, Ç. B. (2023). Detection of Cataract, Diabetic Retinopathy and Glaucoma Eye Diseases with Deep Learning Approach. Intelligent Methods In Engineering Sciences, 2(2), 42-47.

[8] Malik, S., Kanwal, N., Asghar, M. N., Sadiq, M. A. A., Karamat, I., & Fleury, M. (2019). Data driven approach for eye disease classification with machine learning. Applied Sciences, 9(14), 2789.

[9] METİN, B., & KARASULU, B. (2022). Derin Öğrenme Modellerini Kullanarak İnsan Retinasının Optik Koherans Tomografi Görüntülerinden Hastalık Tespiti. Veri Bilimi, 5(2), 9-19.

[10] Sarki, R., Ahmed, K., Wang, H., Zhang, Y., & Wang, K. (2021). Convolutional neural network for multi-class classification of diabetic

eye disease. EAI Endorsed Transactions on Scalable Information Systems, 9(4).

[11] Hussain, M. A., Bhuiyan, A., D. Luu, C., Theodore Smith, R., H. Guymer, R., Ishikawa, H., ... & Ramamohanarao, K. (2018). Classification of healthy and diseased retina using SD-OCT imaging and Random Forest algorithm. PloS one, 13(6), e0198281.

[12] Almansour, A., Alawad, M., Aljouie, A., Almatar, H., Qureshi, W., Alabdulkader, B., ... & Almazroa, A. (2022). Peripapillary atrophy classification using CNN deep learning for glaucoma screening. Plos one, 17(10), e0275446.

[13] Seker, M. E., Koyluoglu, Y. O., Celebi, A. R. C., & Bayram, B. (2022). Effects of Open-Source Image Preprocessing on Glaucoma and Glaucoma Suspect Fundus Image Differentiation with CNN.

[14] Kaggle. Avialble Online: https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification (Accessed on 10 August 2023).

[15] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[17] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

[18] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[19] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[20] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[21] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[22] Akter, S., Shamrat, F. J. M., Chakraborty, S., Karim, A., & Azam, S. (2021). COVID-19 detection using deep learning algorithm on chest X-ray images. Biology, 10(11), 1174.

[23] Sutradhar, A., Al Rafi, M., Ghosh, P., Shamrat, F. J. M., Moniruzzaman, M., Ahmed, K., ... & Moni, M. A. (2023). An Intelligent Thyroid Diagnosis System Utilising Multiple Ensemble and Explainable Algorithms with Medical Supported Attributes. IEEE Transactions on Artificial Intelligence.