Received: March 21, 2024. Revised: April 29, 2024.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

A Lightweight Deep Learning-Based Ocular Disease Prediction Model Using Squeeze-and-Excitation Network Architecture with MobileNet Feature Extraction

Amera W. Al-funjan[1] Hanaa M. Al Abboodi[2]* Najlaa Abd Hamza[3]
Wafaa M. Salih Abedi[4] Alaa H. Abdullah[2]

[1]*College of Education for Pure Sciences, University of Babylon, Babylon, Iraq*
[2]*Department of Electrical Engineering, University of Babylon, Iraq*
[3]*College of Nursing, University of Baghdad, Baghdad, Iraq*
[4]*Management Information System Department, City University Ajman, UAE*
* Corresponding author's Email: hanaa.ali@uobabylon.edu.iq

Abstract: The field of ophthalmology offers great promise for improving patient care and outcomes via automated diagnosis of eye illnesses. Using the Squeeze-and-Excitation Network (SENet) with a Mobile-Net backbone, we describe a unique deep-learning method for automatic illness categorization from retinal images in this paper. Using a lightweight transfer learning model with dramatically decreased parameters, we aim to achieve high accuracy and robust performance in binary classification tasks, such as normal vs. cataract and normal vs. other disease classes. We create and assess a lightweight model using a sizable dataset of retinal pictures from the Ocular Disease Intelligent Recognition (ODIR) database. We achieve remarkable accuracy values that surpass 99.9% on both training and validation datasets for all classification tasks through extensive testing and validation. Our models demonstrate constant performance metrics and small loss values, highlighting their efficacy and dependability in automated illness detection. We also review our models' clinical applicability and possible influence in helping medical professionals with early illness diagnosis, treatment planning, and patient management. Our research constitutes a noteworthy progression in AI-based ocular illness diagnosis, providing a dependable and practically applicable structure for automated disease categorization from retinal images.

Keywords: Ocular disease classification, Squeeze-and-excitation network, Mobile-net, Feature extraction, Transfer learning.

1. Introduction

The eye is one of the most vital organs for daily tasks [1, 2]. Retinal damage can be irreversibly damaged by eye illnesses [3]. Eye problems can lead to blindness or vision impairment, which makes it difficult for people to read, drive, recognize faces, and navigate their surroundings [3]. Motor, linguistic, emotional, social, and cognitive development impairments may occur in children with significant visual impairment [4]. Low academic attainment is a typical issue for visually impaired school-aged children. Reducing symptoms and stopping disease development are treatment goals for these illnesses. Vision rehabilitation has been shown to be extremely helpful in restoring functioning for individuals with irreversible vision loss resulting from a range of eye disorders, including glaucoma and trauma-related consequences [5].

Medical imaging creates visual representations of the internal architecture of the human body using a variety of methods [6]. For non-invasive assessments of physiological processes and anatomical characteristics, it is essential [7, 8]. Medical practitioners use these images to track, identify, and treat illnesses. AI-based techniques can be used by healthcare facilities to treat eye issues effectively [9, 10]. These algorithms can recognize early-stage cases

of eye disorders by evaluating the medical images including optical coherence tomography (OCT) and retinal images [9].

Advanced artificial intelligence (AI) methods proved significant ability in evaluating patient pictures and conducting remote consultations [11]. Furthermore, the AI application enables remote monitoring of patients' ocular issues, decreasing the requirements for frequent in-person examinations [12]. AI algorithms can classify patients, and individuals, critical groups depending on the severity of their conditions [13]. A range of patient data, such as genetic information, imaging results, and medical history are evaluated by machine learning and its algorithm to provide suggestions for individualized therapy [14]. Also, these procedures ensure treatments are tailored to meet patients' needs and characteristics.

Recently, image-based Convolutional Neural Network (CNN) models have been employed to use the retinal, fundus, and OCT scans to categorize eye illnesses in different stages [15]. The first step of this process requires collecting a large scale of medical images that cover various eye disorders. Visual representations of eye disorders were obtained using medical imaging equipment or databases. A labeled dataset has been trained using the CNN model and supervised learning techniques [16]. In training, the model can identify features, textures, and patterns in the pictures that function as biomarkers for different types of ocular disorders [17]. After optimization, the model is tested on a separate dataset to evaluate its real-time performance [18].

In our work, we proposed a unique lightweight deep learning model that uses cutting-edge convolutional neural network (CNN) architectures that integrate the features of MobileNet as a deep feature extraction with a light structure with Squeeze-and-Excitation Network based on channel-wise feature recalibration concept to classify ocular diseases automatically from retinal pictures. Our work presented sophisticated image processing techniques, including Sobel edge detection and principal component analyses of PCA to analyze retinal images. This work investigates binary classification tasks: normal vs. cataract images; normal against various disease classes: glaucoma and mild non-proliferative retinopathy. Obtain data from the Ocular Disease Intelligent Recognition (ODIR) database, carefully pre-process retinal images to guarantee consistency and model architecture compatibility, and choose suitable deep learning models for ocular disease classification. A thorough assessment of model performance has been made using evaluation measurements such as the area under the receiver operating characteristic curve (AUC-ROC), accuracy, recall, F_1 score, and precision. The rest of this paper is structured as follows: Section 2 introduces the related works of eye disease prediction using advanced machine learning models. The methodology includes many steps, including pre-processing and statistical feature extraction, and our transfer learning model, which combines deep features of Mobile-Net and Squeeze-and-Excitation Network (SENet), is explained In Section 3. Model evaluation and performance analysis are described in Section 4. We provide the experimental results and analysis in section 5. Comparative analysis is explained in section 6. We conclude this paper with experimental findings and analysis in section 7.

2. Literature review

Ophthalmologists and optometrists evaluate symptoms, make clinical observations, and conduct diagnostic tests to classify eye ailments conventionally [18]. A comprehensive clinical examination of the patient's ocular condition is performed before the treatment begins. Medical specialists with expertise in optometry or ophthalmology evaluate and analyze visible signs and symptoms of visual problems during this evaluation [19]. Fundus photography provides researchers in this field with excellent-resolution retinal pictures that are vital for better diagnosis [20].

Deep learning models (DL) extract pertinent information from medical pictures, such as OCT and retinal scans, and then classify these images. When classifying eye diseases, these characteristics might include anomalies in the structure of the retina, blood vessels, or lesions. Many types of eye diseases have been detected using DL models [21, 22]. These models can be constructive when patients arrive with many coexisting ocular diseases. DL models may attain consistency in diagnosis and classification [23].

The current investigations have proposed a dataset and classification methods for ocular diseases. Li and colleagues [24] released a fundus imaging collection. They used the dataset to assess cutting-edge DL methods. They stressed the importance of using a successful DL approach when categorizing the fundus pictures. Blood vessel extraction from the photographs proposed by Sundaram et al. [25]. For categorization, they used the hybrid segmentation technique. Their study focused on extracting blood vessels in Fundus Images of the retina through a hybrid segmentation model. This work's main disadvantage was that the authors neglected the feature extraction stage, which may have significantly enhanced the model's accuracy and

complexity reduction. He and colleagues (2021) [26] introduced a multi-label disease classification methodology. They used a deep network and a spatial correlation module SCM to categorize the eye conditions. The main limitation of this work was the model's computational complexities due to the significant number of layers for its structure and the required 100 epochs for network training to achieve an accuracy of around 0.94, which required 74.2 million trainable parameters. Also, the model data was heavily unbalanced.

Gour and Khanna presented a multiclass image classification technique [27]. They used two models based on transfer learning to find anomalous patterns. The best results of around 0.96 as an achieved accuracy were obtained from the first transfer learning model that consists of ResNet, InceptionV3, MobileNet, and VGG16, with total trainable parameters for their suggested model of around 63 million as a main limitation for this model that refer to its computation complexity requirements. Junayed et al. [28] used the CNN model with hyper-parameter optimization to identify only cataracts. Similar to Xu et al. [29] suggested a DL-based approach to identify glaucoma. A limited number of training examples were used to train these two models when each work focused on predicting only a single class that refers to its major disadvantage.

Wahab Sait [30] presented an eye disease classification based on deep learning methods with ODIR and EDC datasets. He achieved validation accuracy with both datasets ODIR and EDC 99.1 and 99.4, respectively. One of the critical drawbacks of this model was its high computation complexity, which resulted from the high trainable parameters of his proposed model, around 24 million with the ODIR dataset. Moreover, limited generalizability may be seen when applying models trained on data from a particular population to a distinct group. DL models have quick image interpretation speed. However, its real-time diagnosis is limited by regulatory clearances, processing resources, and medical device integration. It is essential to develop methods for implementing research findings from academic studies in practical contexts. Consistent benchmark datasets are required for DL models for various eye conditions that would enable more trustworthy evaluations and comparisons of distinct DL models.

Al-Fahdawi and his co-authors [31] presented and enhanced a deep neural model that employed SENet as a primary classifier and tested it with two phases: first when successive pre-processing methods and the second phase applied SENet block directly on OIA-ODIR images without any pre-processing procedures, then the authors compared the classification results for both strategies. One of this work's limitations is that it did not consider the importance of applying the feature extraction method, which is a critical step in enhancing the classification accuracy and decreasing the model computation complexity that aims to reduce the model implementation requirements.

On the other hand, our suggested model employed two feature extraction steps, including engineering feature extraction represented by applying PCA, and deep features have been extracted by the MobileNet pre-trained model. These steps integrated with the SENet block achieved our optimal research goal to construct a general lightweight model that can predict eye diseases from any input image. Our novel model structure that integrated the PCA features and deep features with the binary classification employed to solve the multi-class classification challenges in order to significantly reduce the model complexity, and implementation requirements while maintaining the model's high performance.

3. Methodology

The proposed deep learning structure that combines the Squeeze-and-Excitation Network (SENet) as a deep learning architecture designed to enhance convolutional neural networks' (CNNs') performance based on channel-wise feature recalibration and Mobile-Net as pre-trained models for feature extraction as shown in Fig. 2 as a block diagram. It aims to identify three ocular illnesses in ODIR dataset images. This research seeks to categorize three eye illnesses based on ODIR imaging. Three primary components make up our proposed system. First, the input image is pre-processed to increase contrast, decrease noise, and boost the deep-learning models' learning ability. Second, an efficient framework for deep feature extraction and fusion is used to derive discriminative processing, model architecture selection, training, assessment, performance analysis, comparison with baseline models, and discussion of findings are all included in the technique.

3.1 Data acquisition and preparation

An organized ophthalmic database called Ocular Disease Intelligent Recognition (ODIR) has five thousand patient records [24]. ODIR database gathered extensive retinal images covering several disease classifications, such as mild non-proliferative retinopathy, glaucoma, and cataracts, to ensure enough data for both model training and assessment.

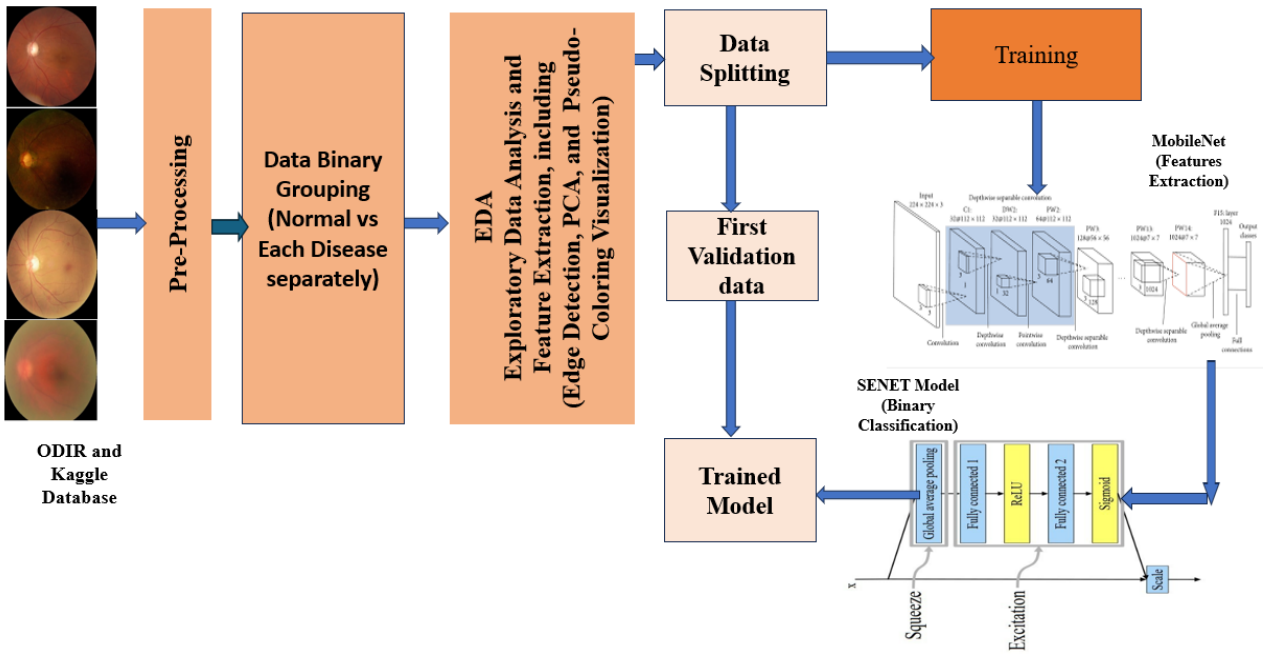

Figure. 1 Methodology for Ocular Disease Classification deep learning model based on transfer learning between Mobil-Net and SENet structure with PCA features with binary classification

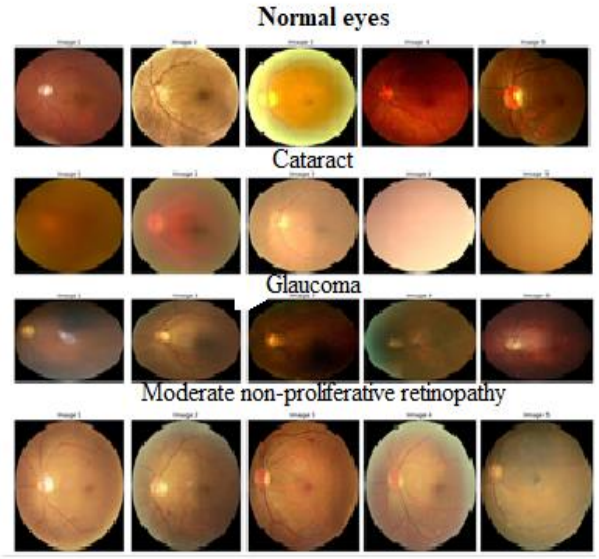

Figure. 2 The sample of our dataset included four classes: Normal, Cataract, Glaucoma, and Moderate non-proliferative retinopathy

In our study, we utilized this database and performed extensive EDA. For classification, we collected and prepared data for four selected eye diseases: cataract, normal glaucoma, and moderate non-proliferative retinopathy.

The training and testing set for these four classes of images have been prepared. The distribution of the training and testing set is 1353 images in the training set and 372 images in the testing set. Now, each training and testing set contains four more classes of diseases. We selected 500 images in the training set (250 for the left eye and 250 for the right eye) for each class and 100 images in the testing set for each class, but glaucoma has a limited number of images, so there we selected 190 images for the left eye and 190 images for right eyes.

Even though the cataract has 518 images in the directory, many images are missing in the root images directory, so we have around 262 cataract images, which we get in the training folder. The glaucoma class also has 190 images available for each eye in the directory, but in the root images directory, we can get up to 200 images only. The same goes for PR eye diseases. In the testing set, there is no such class imbalance.

The class imbalance issue has been sorted out through the image augmentation algorithm in the pre-processing step of our proposed model by artificially increasing the size of images to obtain higher accuracy for each disease. The sample images for each class are explained in Fig. 2.

3.2 Pre-processing

The retinal pictures underwent pre-processing. Many steps have been applied, such as pre-processing, which involves image downsizing to a standard size (224x224 pixels), colour mode conversion, and optional data augmentation methods like rotation and flipping. The pre-processing procedures employed in this work enhance the data resolution for better deep

learning performance. Data Augmentation adds changes to the training data, and data augmentation techniques like flipping, rotating, and zooming can improve the models' resilience and generalizability. We perform data augmentation during training by creating augmented pictures on the fly using tools like ImageDataGenerator. Normalizing the picture pixel values to a similar scale facilitates faster convergence and training process stabilization. Usually, we divide the pixel values by 255 (for the [0, 255] scale) or by mean and the standard deviation to normalize them. The photographs can be cropped to match their aspect ratio, and unnecessary background information can be removed to focus the model's attention on the relevant areas of Interest. This strategy offers two advantages: it reduces computational costs and increases model efficiency. These pre-processing steps ensure the input data is appropriately organized, standardized, and improved to facilitate practical training of deep learning models for classifying ocular disorders from retinal pictures.

3.2.1 Edge detection on ocular disease retinal images

The Sobel operator was applied to extract edges from the grayscale retinal images, emphasizing regions with significant variations in intensity.

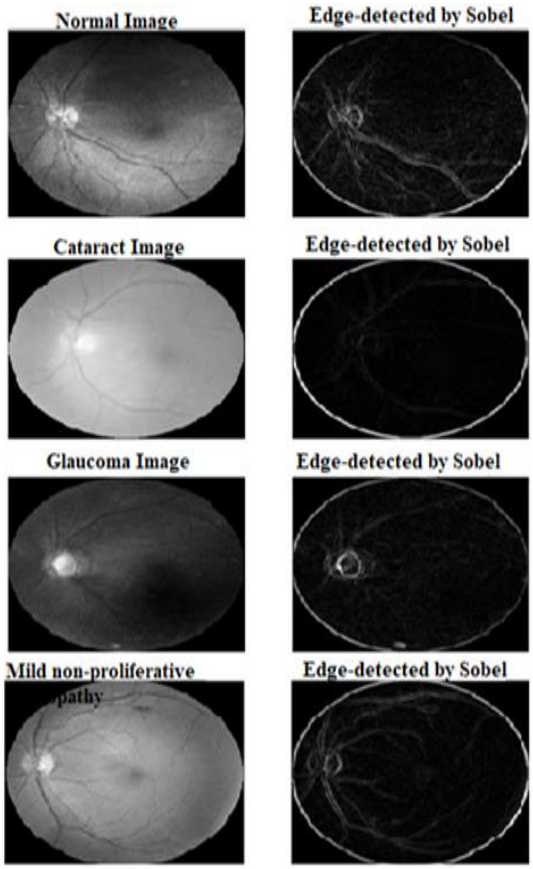

Figure. 3 Edge detection for Normal, Cataract, Glaucoma, and moderate non-proliferative retinopathy

This tool is applied to the input images as a vital pre-processed step to prepare our data for successive processing, including feature extracting and classifier. This method is crucial for identifying the boundaries of structures in images, enabling subsequent processing and extraction of features. The classes from the ODIR database have been analyzed using the Sobel edge detection method, and the results are shown in Fig. 3. We can depict particular boundaries within the cataract image that indicate structural inconsistencies demonstrative of the infection. The edges are repressed within the normal retinal image, suggesting the nonappearance of recognizable neurotic surrenders. This stark differentiation highlights the viability of the Sobel edge location in recognizing sound and unhealthy retinal components.

Additionally, similar trends in edge detection were observed when we expanded our study to include images of moderate non-proliferative retinopathy and glaucoma. Image regions corresponding to pathological abnormalities associated with these diseases have clear and visible edges. This finding confirms the effectiveness of Sobel edge detection in capturing relevant anatomical information important for disease diagnosis and classification. Applying this method improves the accuracy and reliability of eye disease detection algorithms by assisting in extracting valuable details and their subsequent classification.

3.3 Data binary grouping based on image labels

Our work suggests solving the multiclass classification challenge by applying a binary classification technique for classifying eye diseases. Once the data has been pre-processed, it is categorized into three binary classifications based on data labeling: Normal against Cataract, Normal versus Glaucoma, and Moderate non-proliferative Retinopathy versus Normal. The binary classification method was chosen because it can easily accommodate extraneous features in datasets of any size.

3.4 Principal component analysis (PCA) for feature extraction

Principle component analyses PCA is an established method that is widely applied to machine learning and image processing. By transforming high-dimensional data sets to high-dimensional space reduction, it tries to extract the most significant sources of variance while ignoring pertinent information.

We employ PCA in our work to minimize the dimensionality of big data sets while maintaining the

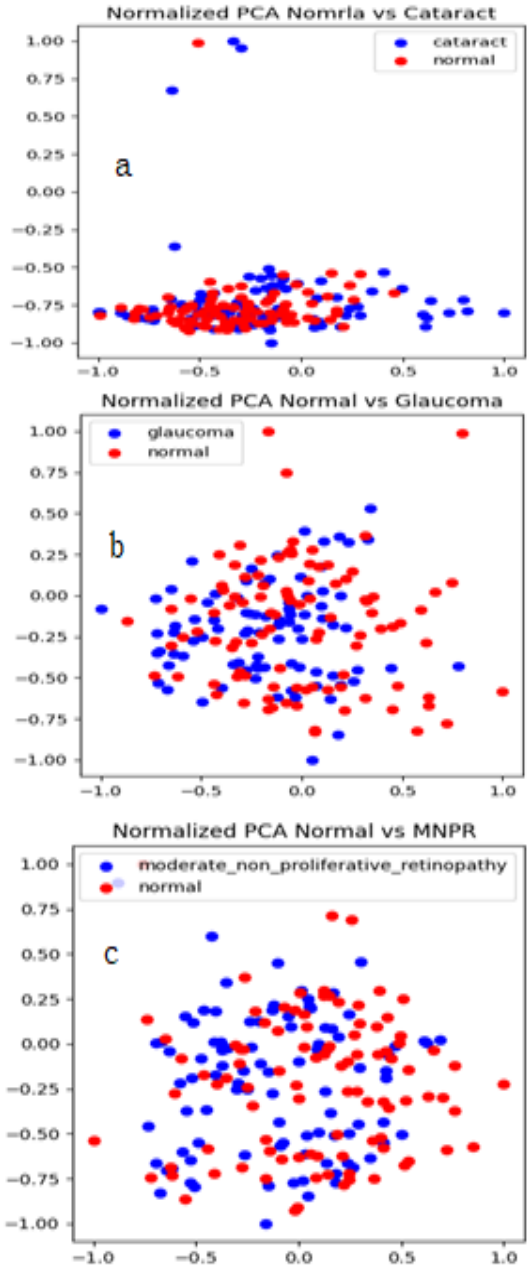

Figure. 4 PCA was applied individually for Normal vs. Cataract, Glaucoma, and moderate non-proliferative retinopathy

highest degree of information from the original data. In this part, PCA is used to minimize the dimensionality of retinal pictures obtained by the Eye Disease Intelligent.

Identification (ODIR) database and make it possible to display details regarding the underlying data distribution. By finding and measuring pertinent data connections, altering the data to preserve the most significant associations, and eliminating the remainder, this technique lowers dimensionality. The association between features is analyzed using the covariance matrix C, as shown below [32].

$$C = \frac{1}{M}\sum_{N=1}^{M} \Phi n \Phi n^{T} = AA \quad (1)$$

$$A = [\Phi 1, \Phi 2, \ldots\ldots, \Phi M] \quad (2)$$

$$\Phi_i = x_i - \tilde{x} \quad (3)$$

$$\tilde{x} = \frac{1}{M}\sum_{i=1}^{M} x_i \quad (4)$$

Where x_i is the $N\times1$ vector representing the input retinal dataset, this vector contains the input eye images. Eigenvectors and eigenvalues have been obtained by performing a linear transformation or eigen decomposition on the Covariance Matrix. Then, our data was converted into principal components using Eigenvectors. The significance of these correlations has been assessed by Fig. 4, which explains the visual representation of applying PCA on the binary grouping data that demonstrated the specific patterns between the Normal Images vs Cataract, Glaucoma, and Mild non-proliferative retinopathy individually. After PCA dimensionality reduction, the distribution of images in a two-dimensional space defined by main components is shown in scatter plots.

Interestingly, discrete clusters arise for every class, suggesting that the PCA treatment identified identifiable patterns. A coherent clustering of cataract photos indicates that the illness has standard structural features. In the limited feature space, normal pictures reflect their unique visual features and create their cluster. This distinction highlights the PCA's effectiveness in discriminating between retinal pictures with cataracts and those without, providing the foundation for further classification tasks. After the dimensionality reduction of PCA, discrete clusters are obtained for every class, which outlines traits inherent in the individual illness states. Images with moderate non-proliferative retinopathy show a coherent clustering, indicative of common structural abnormalities linked to the disease by employing eigenvalues and retaining the crucial components.

3.5 Model architecture

We have chosen a cutting-edge deep learning architecture to classify eye diseases using the Squeeze-and-Excitation Network (SENet) on a Mobile-Net foundation. The architecture was selected because of its effectiveness in eye disease prediction tasks and its capacity to compromise computational efficiency and model complexity.

3.5.1 Mobile-Net deep feature extraction

Mobile-Net is a collection of efficient neural network architectures for embedded and mobile vision applications. Google developed it in response to the need for high-performance, low-weight models that could be used with low-resource devices, such as Internet of Things gadgets, mobile phones, and embedded systems. Specifically, Mobile-Net models have gained widespread recognition for their effectiveness, compact size, and efficiency in various computer vision applications. The advent of Mobile-Net pre-trained models has led to a substantial advancement in constructing neural network architecture. Nowadays, building efficient deep-learning models for embedded and mobile vision applications relies essentially on transfer learning concepts. Mobile-Net models, which offer a careful trade-off between accuracy and efficiency, are essential for enabling artificial intelligence on devices with limited resources because they are crucial in the current context of edge computing and computer vision.

As shown in Fig. 5, MobileNet, the pre-trained model, has been employed as an intelligent feature extraction network due to its key characteristics, including many features that decrease the model size significantly while keeping the model performance efficient. Mobile-Net is based on depth-wise separable convolutions, which splits the convolution process into depth-wise and pointwise convolutions. This characteristic reduces the computation demands since the spatial and channel dimensions are treated independently. The MobileNet model employed a width multiplier parameter to control the network's width, which can balance model performance and size, one of the essential features in MobileNet that decreases the model complexity. A resolution multiplier is an extra parameter of MobileNet that controls the input resolution of the model. Lower resolutions result in smaller models and faster inference. Inverted residual blocks are used in several MobileNet versions to reduce computing costs while maintaining model integrity.

Pre-trained MobileNet as a transfer learning model was trained on large-scale datasets such as ImageNet. These pre-trained models may be modified using smaller, task-specific datasets for various computer vision applications. Depth-wise separable convolutions DSC: A factorized convolution, the MobileNet model uses depth-wise separable convolutions to break down a standard convolution into a depth-wise and one-dimensional convolution known as a pointwise convolution. The depth-wise convolution uses a single filter to train the input channels individually. Then, these single filters are combined using dimension pointwise convolution. A conventional convolution involves just one step to filter and combine inputs into a new set of outputs divided into two layers by the depth-wise separable convolution. These two layers are one for filtering and another for connecting. A significant reduction in processing and model size is achieved with this factorization. A typical convolution layer generates two square feature maps, one for input F with size $D_F \times D_F \times N$ and the second for output G with size $D_G \times D_G \times M$. D_F and D_G represent the input and output feature map spatial dimensions (width and height). CNN layer parameters are described by the convolution kernel K of size $D_K \times D_K \times N \times M$, where D_K refers to the spatial width and height of the kernel and output channels, respectively. For standard convolution, the output feature map is calculated as follows:

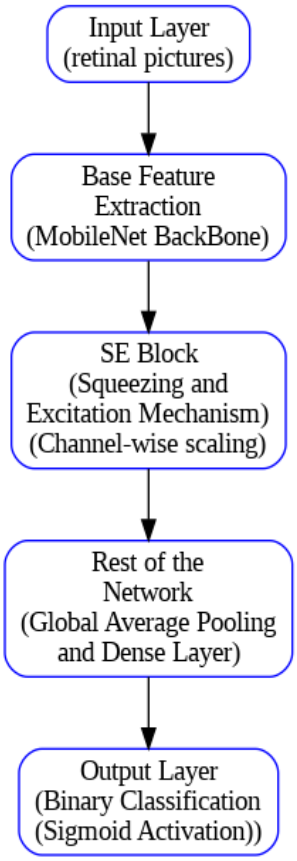

Figure. 5 The proposed architecture of transfer learning by integrating Mobil-Net and SENet Model features for Retina image prediction

$$G_{k,l,m} = \sum_{i,j,n} K_{i,j,m,n} \cdot F_{k+i-1,l+j-1,n} \quad (5)$$

There is a multiplicative relationship between the computational cost, the kernel size (D_k), the feature

map size (D_F), and the input and output channels (N and M, respectively). All of these variables and how they interact are covered by MobileNet models. Depth-wise separable convolutions were initially applied to separate the relationship between the size of the kernel and output channel numbers.

Two layers comprise Depth-wise Separable Convolutions (DSC): pointwise and depth-wise convolutions. Each input channel (input depth) gets its filter via depth-wise convolutions. Afterward, the output of the depth-wise layer is linearly combined using pointwise convolution, a primary 1×1 convolution. MobileNets employ either the batch norm or the ReLU nonlinearity for each layer. One way to express depth-wise convolution with an input depth of one filter per channel is as follows:

$$G\hat{}_{k,l,n} = \sum_{i,j} K\hat{}_{i,j,n} \cdot F_{k+i-1,l+j-1,n} \quad (6)$$

K^ is the kernel of the depth-wise convolution applied to the m^{th} channels in F to generate the mth channels of G^, the output feature map. Depth-wise Separable Convolution (DSC) cost can be computed by combining the dimensional cost of its component (depth-wise convolution and pointwise convolution) by summation process as explained in the following:

$$DSC_{COst} = D_K \times D_K \times N \times D_F D_F + N \times M \times D_F \times D_F \quad (7)$$

The computation cost can be reduced by emphasizing convolution as a two-phase procedure filtering and combining that involving dividing the DSC cost on the standard convolution cost expressed as below:

$$\frac{D_K \times D_K \times N \times D_F \times D_F + N \times M \times D_F \times D_F}{D_K \times D_K \times N \times M \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (8)$$

Width Multiplier: The first hyper-parameter to decrease the CNN computational cost is a fundamental parameter called width multiplier, which is added to construct the lightweight model. In MobileNet networks, the width multiplier is responsible for evenly transforming each layer. Mathematically, the width multiplier will be multiplied by the input and output channels to become αN and αM, respectively. The width multiplier significantly affects the MobileNets due to the reduced computational cost of DSC and the number of parameters fourfold. The width multiplier takes values between 0 and 1. The computational cost of applying a width optimizer on DSC is described as follows:

$$DSC_{COst} = D_K \times D_K \times \alpha N \times D_F \times D_F + \alpha N \times \alpha M \times D_F \times D_F \quad (9)$$

Resolution Multiplier: Resolution Multiplier ρ is another hyperparameter related to the input image resolution employed in the MobileNet to decrease the internal representation of each layer after applying it to the input image. The computational cost of using Resolution Multiplier on depth-wise separable convolution is described as follows:

$$DSC_{COst} = D_K \times D_K \times \alpha N \times \rho D_F \times \rho D_F + \alpha N \times \alpha M \times \rho D_F \times \rho D_F \quad (10)$$

With only a little loss in accuracy, MobileNet employs 3×3 depth-wise separable convolutions, which caused a considerable reduction in the computation of standard convolutions. MobileNet is used in our suggested model structure for deep feature extraction from ODIR pre-processed images due to the low computational cost of depth-wise convolutions. Then, the reduction in MobileNet resolution has been sorted out by applying SENet as a classifier to deal with the MobileNet extracted feature map.

3.5.2 SENet model

The Squeeze-and-Excitation Network (SENet) is a deep learning architecture designed to enhance the performance of convolutional neural networks CNNs by channel-wise feature recalibration. SENet was first described in "Squeeze-and-Excitation Networks" research by Jie Hu, Li Shen, and Gang Sun, published in 2017 [33]. SENet thoroughly describes the interaction among several feature vectors to enhance the representative capacity of neural networks. CNNs have significantly transformed various computer vision applications, such as segmentation, object detection, and image categorization. While CNNs have achieved considerable success, they frequently encounter challenges such as overfitting, vanishing gradients, and extracting significant features from the input data.

A unique design structure called the "Squeeze-and-Excitation block" has been introduced to overcome the challenges of CNNs. SE block aims to eliminate unnecessary features and enhance the ability to collect and emphasize crucial ones, optimizing the network's overall performance. Since SENet is presented as a drop-in module to several CNN architectures, it is a versatile and potent tool for improving model efficiency and accuracy. The construction of the SE building block can be expressed mathematically by the following:

$$F_{tr} = X \to U, X \epsilon R^{\acute{H}\times\acute{W}\times\acute{C}}. U \epsilon R^{H\times W\times C} \quad (11)$$

F_{tr} represents any transformation such as convolution or set of convolutions. To do feature recalibration for these transformations that translates the input X to the feature mappings U where $U \in R^{H\times W\times C}$, such a convolution, for instance, we might construct a similar SE block. A squeeze operation is first done to create a channel descriptor to the features U, aggregating feature maps over their spatial dimensions (H×W). This descriptor's purpose is to encapsulate the worldwide distribution of channel-wise feature responses, enabling data from the global receptive field to be used by all levels of the network. After the aggregation, the excitation process works as a self-gating mechanism that handles the embedding as input and produces a set of per-channel modulation weights. These weights are then applied to the feature maps U to create the output of the SE block, which may then be directly fed into the network layers above.

By translating an input superscript feature mappings $U \in R^{H\times W\times C}$, a transformation F_{tr} may be used to build a computational unit called a squeeze-and-excitation block. For the filter kernels, V refers to the learned vector, which has the parameters of the *c-th* filter as explained below:

$$V = [v1, v2 \ldots, vc] \quad (12)$$

Where *vc* represents the parameters of the filter when the output U of $F_{tr,}$ which is assumed to be a convolutional operator, can be written as the following:

$$U = [u1, u2 \ldots, uC] \quad (13)$$

Where u_C is computed from the convolution relationship denoted by $*$ between a single channel of v_C and corresponding channel X *as* explained in the following:

$$u_c = v_c * X = \sum_{s=1}^{\acute{c}} v_c^s * X^s \quad (14)$$

$$v_c = \left[v_c^1, v_c^2, v_c^3, \ldots\ldots, v_c^{\acute{C}}\right] \quad (15)$$

$$X = \left[x^1, x^2, x^3, \ldots\ldots, x^{\acute{C}}\right] \quad (16)$$

A two-dimensional spatial kernel is denoted by v_c^s Representing a single channel v_c that matches channel X. Because the output results from adding all the channels, v_c inherently incorporates the channel dependencies, yet these dependencies overlap with the spatial correlation of the filter records. Our objective is to ensure the network can tune in more to informative characteristics, which can be used in future transformations while ignoring less relevant ones. To do this, we consider explicitly describing the interconnections of the channels such that the filter responses are adjusted in two stages, squeezing and excitation, before input into the subsequent transformation.

Squeeze operator: Our approach to employing channel dependencies includes examining the signal to each channel in the output attributes. Due to the local receptive field of each learned filter, each unit of the transformation output U cannot use contextual information outside of this area. This issue is especially extreme at lower network layers with narrow receptive fields. The global squeezing of information added spatially into channel descriptors to address local receptive field issues. Using global average pooling generates channel-wise statistics. To get a statistic $z \in R^C$, reduce U on spatial dimensions $H \times W$. The *c-th* element of squeeze transformations z is computed as follows:

$$z_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \quad (17)$$

Where z_c is squeeze transformations of output feature map ($F_{sq}(u_c)$), the outcome of the transformation U may be considered a set of local descriptors with statistics representative of all of the data. Utilizing such knowledge is expected in feature engineering tasks. We use the primary global average pooling method, acknowledging that more advanced aggregating techniques might also be used.

Excitation Operator: After aggregating information in the squeeze process, we perform a second operation to capture channel-wise dependencies comprehensively. The function needs to satisfy two conditions: it must be adaptable, especially in its ability to understand a nonlinear connection between channels, and it should recognize a non-exclusive relationship to allow multiple channels to be emphasized rather than just one-hot activation. A sigmoid activation function with a basic gating procedure has been applied to achieve these requirements as explained in the following:

$$s = \sigma\left(W_2\delta(W_1 z)\right) \quad (18)$$

Where s is the Excitation Operator of the squeeze transformation output with dimensional reduction, σ, and δ are sigmoid and ReLU activation functions.

$W_1 \in R^{\frac{C}{r} \times C}$ and $W_2 \in R^{C \times \frac{C}{r}}$. The structure of the gating procedure is created as a bottleneck using two fully connected (FC) nonlinear layers to control model complexity and improve generalization. These layers include a dimensionality-reduction layer with parameters W_1 and reduction ratio r, a ReLU activation function, and a dimensionality-increasing layer with parameters W_2. The block's final output is achieved by adjusting the transformation output U using the activations.

$$F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (19)$$

The multiplication process between the feature map u_c and s_c represents a channel-wise multiplication to obtain the final output of the SE block.

3.5.3 Proposed model layers architecture

Our research aims to classify eye disorders from retinal pictures by designing a robust, lightweight, deep-learning model. The Squeeze-and-Excitation Network (SENet), a cutting-edge convolutional neural network architecture renowned for collecting necessary information and improving model performance, is incorporated into the suggested design. Table 1 explains our model layers' structural properties. The input layer of the model is made to take retinal images. The dimensions of input images should be (224×224). The MobileNet model is the foundation for feature extraction to use pre-trained weights and leverage transfer learning. Then, MobileNet was chosen because it effectively extracts high-level deep features from ODIR images and its lightweight design. SE Block (Squeezing and Excitation Mechanism) included:

Global Average Pooling: generates channel-wise feature vectors by averaging feature mappings across spatial dimensions.

Squeeze Operation: reduces the size of the feature map using a 1x1 convolutional layer, then adds nonlinearity using a ReLU activation function.

Excitation Operation: creates channel-wise scaling factors using sigmoid in a second 1x1 convolutional layer. These variables adjust each feature channel's weight according to its relevance to the classification job.

Multiply Operation: Feature maps are reweighted by performing element-wise multiplication with the calculated scaling factors, allowing the model to emphasize important traits while attenuating less significant ones. Global Average Pooling Creates an international representation of the input image by combining spatial data from many feature maps. Then, the Dense Layer was chosen as a bottleneck layer, a fully linked layer with 128 units, and ReLU activation allowed for nonlinear transformations and feature abstraction. Finally, the Output Layer involves one unit with a sigmoid activation function that makes up the last layer. This layer produces a probability score that indicates the likelihood of the input image belonging to a specific class (binary classification).

Table 1. Model Layers Properties

Layer	Output Shape	Parameters	Connected to
input_1 (Input Layer)	[(None, 224, 224, 3)]	0	[]
mobilenet_1.00_224 (Functional)	None, 7, 7, 1024)	3228864	['inpute_1[0][0]']
global_average_pooling2d	(None, 1024)	0	['mobilenet_1.00_224[0][0]']
reshape (Reshape)	(None, 1, 1, 1024)	0	['global_average_pooling2d [0][0]']
conv2d (Conv2D)	(None, 1, 1, 64)	65600	['reshape [0][0]']
conv2d_1 (Conv2D)	(None, 1, 1, 1024)	66560	['conv2d [0][0]']
multiply (Multiply)	(None, 7, 7, 1024)	0	['mobilenet_1.00_224[0][0]', 'conv2d_1[0][0]']
global_average_pooling2d_1	(None, 1024)	0	['multiply [0][0]']
dense (Dense)	(None, 128)	131200	['global_average_pooling2d_1[0] [0]']
dense_1 (Dense)	(None, 1)	129	['dense [0][0]']

The Adam Optimiser, a popular choice for training deep neural networks, is utilized to construct the model to meet the need for model compilation.

3.5.4 Model training and hyperparameters

The training dataset, which had been pre-processed, was utilized to train the selected model architecture by stochastic gradient descent optimization. During training, the model acquired the capacity to discern relevant features from the retinal images and classify them into the appropriate disease categories. The model's performance was monitored during its iterative training over multiple epochs using metrics such as accuracy and loss. The model hyperparameters were made thoroughly until our ultimate performance goal was reached, as shown in Table 2. The model is configured for binary classification, using binary cross entropy as the loss function. This loss function distinguishes normal photos from regular patients. In our approach, we have yet to conduct joint training for all classes. Instead, each class is individually classified against normal images. This approach offers a significant advantage, resulting in high model accuracy. Given the critical nature of medical applications, we prioritize performance and do not make any compromises on underfitting or overfitting. To achieve better performance, we employ adequate datasets and fine-tuned models to ensure a balanced distribution of classes when classifying photos. The hyperparameters are clearly explained in Table 2.

Table 2. Model Hyper Parameters

Hyper Parameter	Value
Epochs	100
Batch Size	16
Optimizer	Adam (Learning rate=0.0001)
Loss	Binary Cross Entropy
Activation	Sigmoid

4. Model evaluation and performance analysis

Many metrics have been applied to evaluate our model performance statistically across multiple disease groups using accuracy, precision, recall, sensitivity, specificity, F1 score, and kappa criteria. The model's performance is evaluated using these metrics on an independent test dataset. Additionally, a visual evaluation of the model predictions was performed to analyze possible misclassifications or areas for quality improvement. The performance of the trained model was evaluated by its ability to accurately classify retinal images in health and disease. AUC-ROC is used to comprehensively assess the diagnostic adeptness of our proposed model. Precision measures the accuracy in identifying a particular sample as positive out of all the projected samples. Recall refers to the proportion of correctly identified fundus diseases out of all the actual fundus diseases in the sample. The F1-score, a statistic that combines precision and recall, achieves greater values when both rates are high. The kappa score assesses the degree of agreement between categorized outcomes and their matching ground truth labels. The mathematical expressions for our evaluation metrics can be expressed as follows [34]:

$$Rcall = \frac{TP}{TP+FN} \tag{20}$$

$$precision = \frac{TP}{TP+FP} \tag{21}$$

$$F1_{score} = \frac{2\times Recall \times pescision}{Recall+pescision} \tag{22}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{23}$$

$$\text{Sensitivity} = (\text{TP} + \text{FN}) / (\text{TP}) \tag{24}$$

$$Specificity = TN / (TN + FP) \tag{25}$$

$$Kappa = P_0 - P_e / 1 - P_e \tag{26}$$

$$P_0 = \sum_{i=1}^{r} TP_i / \sum_{i=1}^{r} (TP_i + FN_i) \tag{27}$$

$$P_e = \sum_{i=1}^{r} TP_i + (TP_i + FN_i) / N^2 \tag{28}$$

5. Results and discussion

This work investigates the possibilities of artificial intelligence in solving challenges in the ophthalmology field by building a robust and realistic

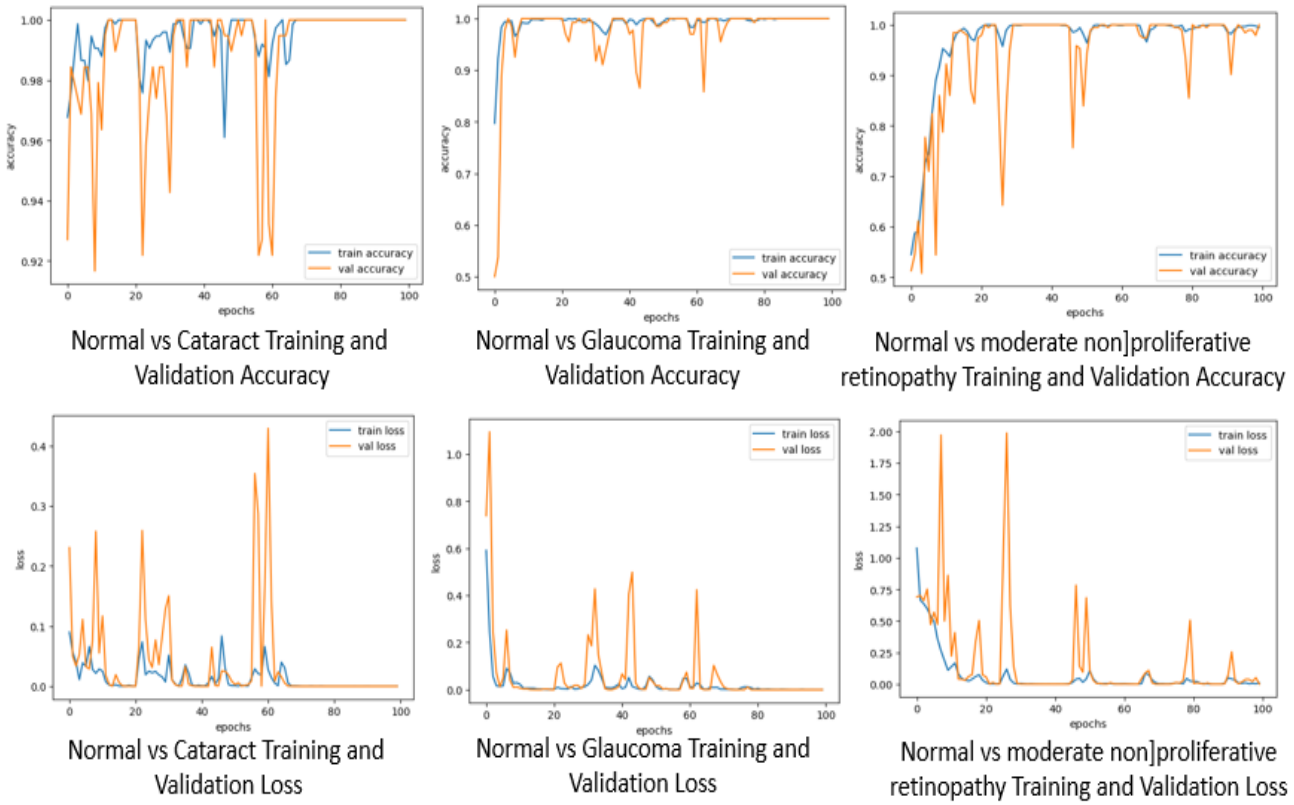

Figure. 6 Model performance with test data included Normal vs Cataract, Normal vs Glaucoma, and Normal vs. Moderate non-proliferative Retinopathy for Accuracy and Loss

Table 3. Model Evaluation Results in Test Data from ODIR included our binary classification

Evaluation Metric on Test Data	Values Normal vs. Cataract	Values Normal vs. Glaucoma	Values Normal vs. Moderate non-proliferative retinopathy	Average performance
Accuracy	1.0	1.0	0.99	0.996
F1 score	1.0	1.0	0.99	0.996
Precision	1.0	1.0	0.99	0.996
Recall	1.0	1.0	0.99	0.996
Sensitivity	1.0	1.0	0.99	0.996
Specificity	1.0	1.0	0.99	0.996
Kappa	0.98	0.96	0.95	0.963

deep-learning model to classify eye diseases automatically. For this purpose, our study uses a lightweight, high-resolution model that combines powerful pre-processing techniques with carefully selected feature extraction methods. These methods include using the PCA algorithm to extract and select features, thereby reducing the dimensionality of the input data, which helps decrease the computation complexity significantly, minimize the model's implementation requirements, such as execution time, and simplify hardware configuration.

The pre-processed features were presented to our proposed classifier, which combines MobileNet as a deep feature extraction technique and SENet, a channel-based feature adaptation method. We recommend applying a deep learning model trained on retinal images using a binary classification technique for classifying eye diseases. After data is pre-processed, it is then grouped depending on the data labeling into three binary classifications: Normal vs. Cataract group, Normal vs. Glaucoma group, and Moderate non-proliferative Retinopathy vs. Normal

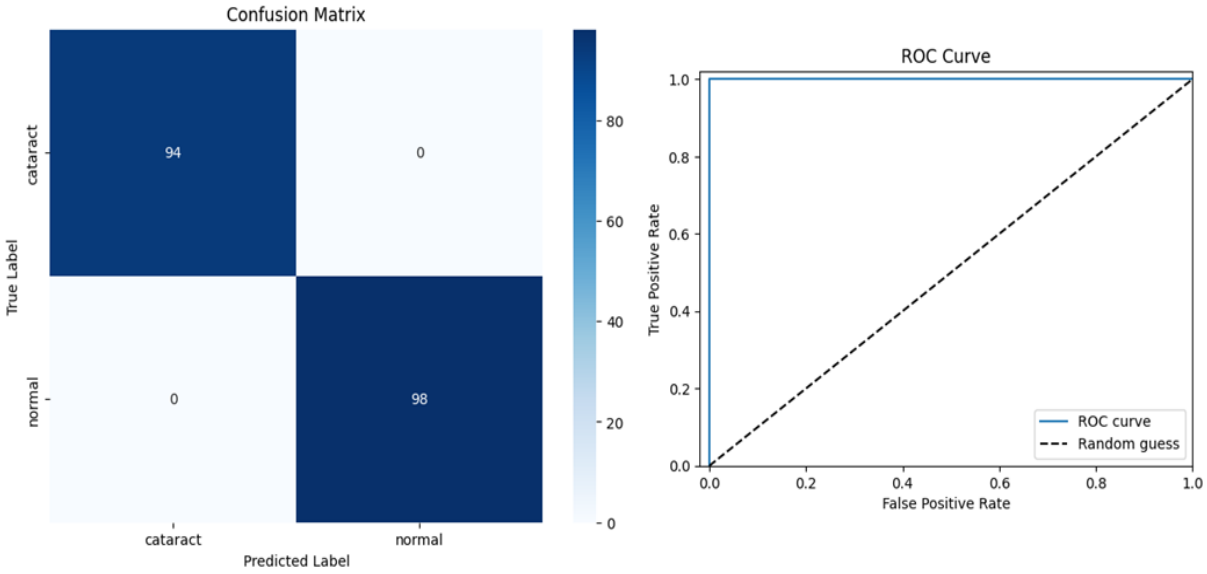

Figure. 7 Confusion Matrix and ROC curve for Cataract vs Normal with ODIR dataset

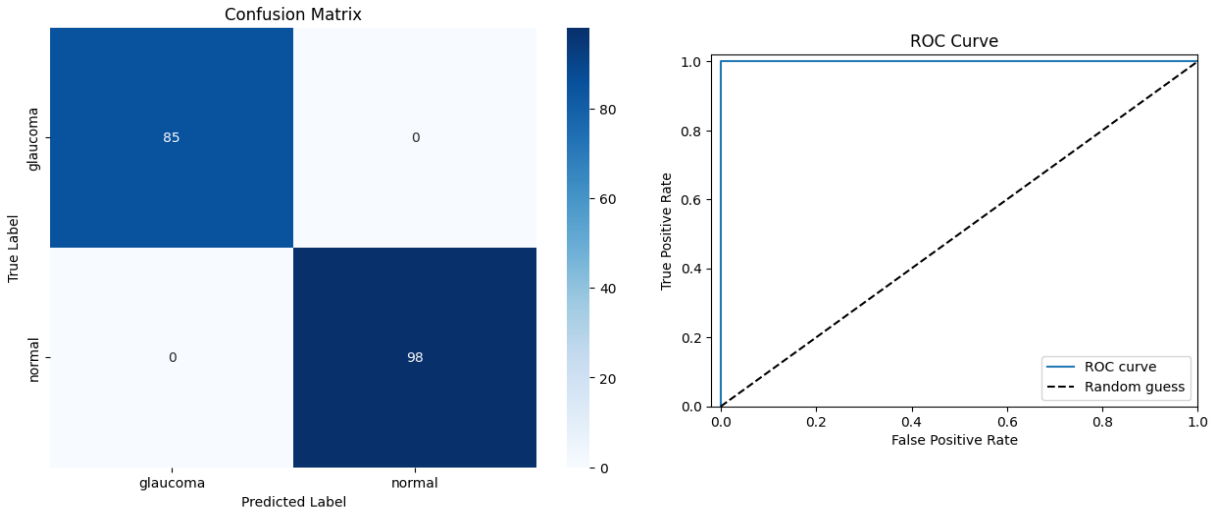

Figure. 8 Confusion Matrix and ROC curve for Glaucoma vs. Normal with ODIR dataset

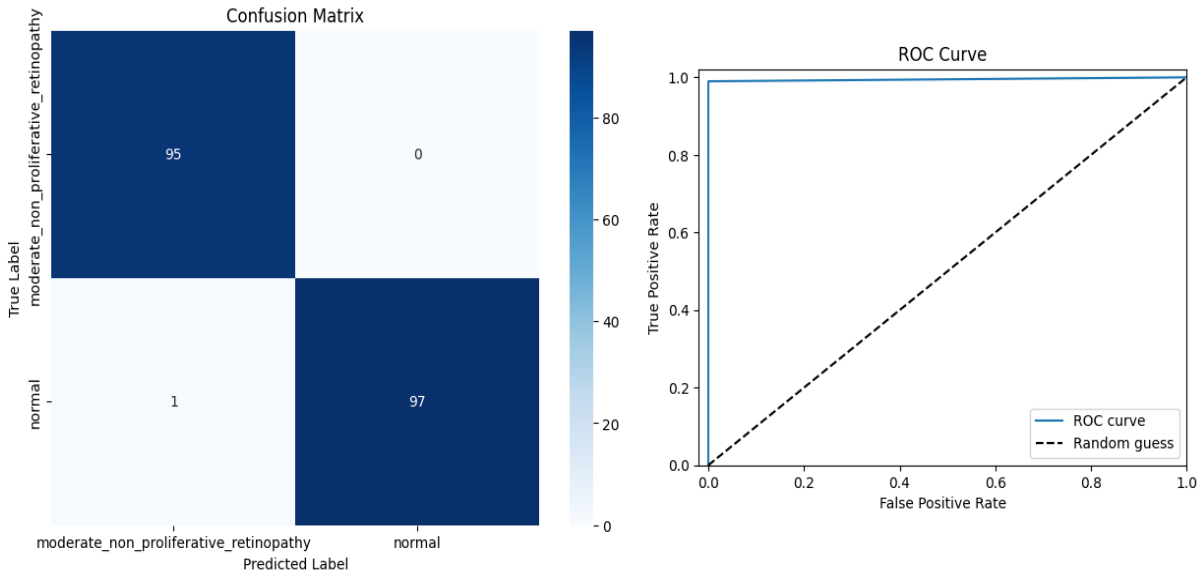

Figure. 9 Confusion Matrix and ROC curve for Moderate non-proliferative retinopathy vs Normal with ODIR dataset

Table 4. Our three binary classifications included model
Evaluation Results with Unseen samples from IDRiD, Oculur recognition, and HRF datasets

Evaluation Metric on Test Data	Values Normal vs. Cataract	Values Normal vs. Glaucoma	Values Normal vs. Moderate non-proliferative retinopathy	Average performance
Accuracy	0.98	1.0	0.99	0.996
F1 score	0.98	1.0	0.989	0.989
Precision	1.0	1.0	1.0	1.0
Recall	0.96	1.0	0.987	0.982
Sensitivity	0.96	1.0	0.981	0.980
Specificity	0.97	1.0	0.994	0.996
Kappa	0.89	0.95	0.94	0.926

group. Our developed models demonstrate outstanding performance, achieving an impressive accuracy of over 99.9% on training and validation datasets. The high accuracy of the algorithm confirmsits ability to accurately classify retinal images into normal and diseased categories.

Moreover, the model demonstrates high confidence in their predictions due to low loss values, around 10e-4. The predicted loss values also confirm the reliability of the models, showing that their predictions closely match the labels observed in practice. The model performance, including the accuracy and the loss rates, are visualized in Fig. 6. The design of the models allows efficient retrieval of relevant information. The robust performance and exceptional accuracy of retinal imaging can be attributed to the features highlighted in Table 3. By including the Squeeze-and-Excitation block as attention approaches, the models become more adept at reducing noise and directing their focus toward relevant information, significantly improving their ability to distinguish and generalize effectively.

The confusion matrixes in Figs. 7 and 8 explained clearly that all cases of cataracts vs. normal glaucoma vs. typical were accurately identified with no false positives or false negatives, providing additional validation for the model's performance. Our model performed well, earning 100% accuracy, precision, recall, and F1 score for cataracts vs normal and glaucoma vs normal prediction. The model's accuracy in detecting these previously mentioned classes is confirmed by the confusion matrix, which shows no instances of misclassification. Similarly, the model's accuracy, precision, recall, and F1 score remained high when tested on the mild non-proliferative retinopathy vs. normal categorization task. The model performed quite well overall, even if there was one misclassification of a normal image, which showed that the model could distinguish between retinal photos with little to no error, as demonstrated in Fig. 9. These figures also explain the ROC curve, demonstrating the model's outstanding discriminatory capacity and efficacy in differentiating between positive and negative situations.

Our proposed model's outstanding performance across multiple assessment metrics indicates how well it can automatically forecast eye diseases. That will help make a big difference in raising the standard of diagnosis and accelerating early clinical action. Even with the encouraging results, significant issues with our study still need to be examined further. The dataset may have biases due to its reliance on historical data. Furthermore, the models' performance may vary under changing imaging conditions or in populations with varied demographic traits, highlighting the importance of robustness testing and validation. We employed image examples that matched our four classes and were gathered from diverse datasets to get around these difficulties, typical in machine learning model generalization. These images are collected from various sources, including IDRiD, Oculur recognition, and HRF datasets. As we have binary models, we consider our binary groups involved Normal vs. each class, Cataract, Glaucoma, and Moderate non-proliferative retinopathy in this evaluation step, where the trained model will test with entirely new and unseen images collected from various datasets.

Figs. 10-12 explain the confusion matrix resulting from the validation stage when our validation samples have been submitted to our transfer learning trained model. The visualized outcomes show stable performance in predicting all classes, whereas Fig. 10 shows only 1 class is wrongly classified as normal.

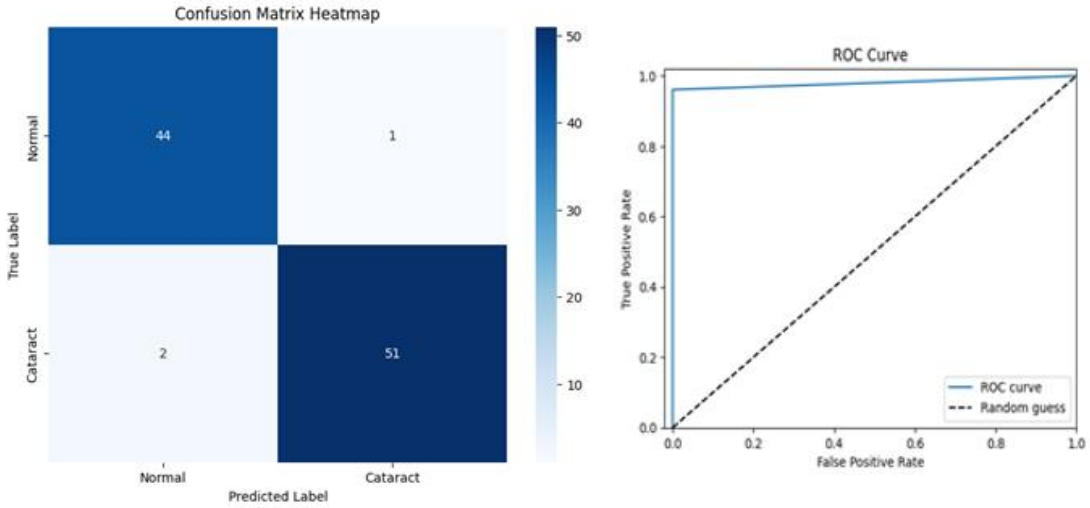

Figure. 10 Confusion Matrix and ROC curve for Normal Vs. Cataract from Unseen samples from IDRiD, Oculur recognition, and HRF datasets

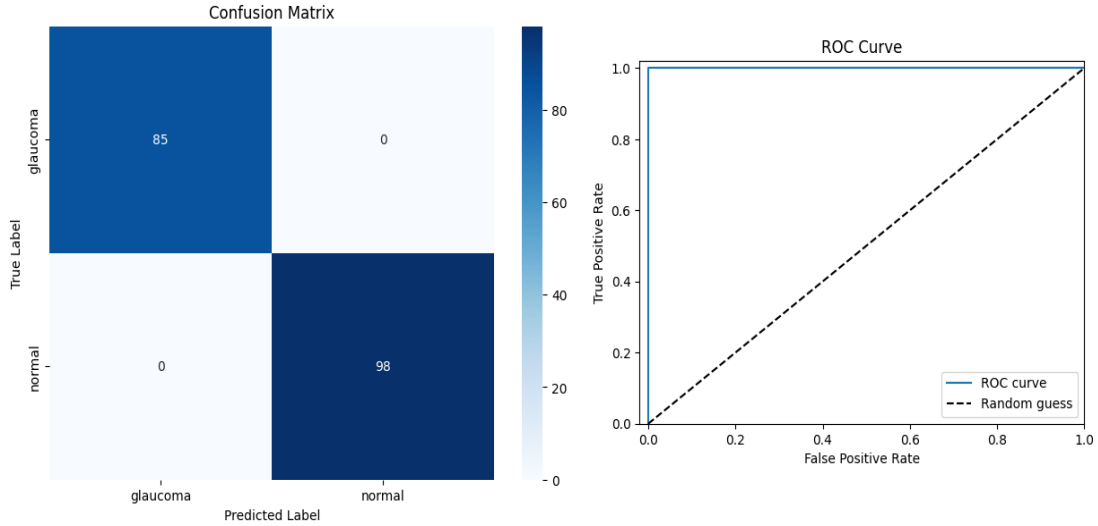

Figure. 11 Confusion Matrix and ROC curve for Normal Vs. Glaucoma from Unseen samples from IDRiD, Oculur recognition, and HRF datasets

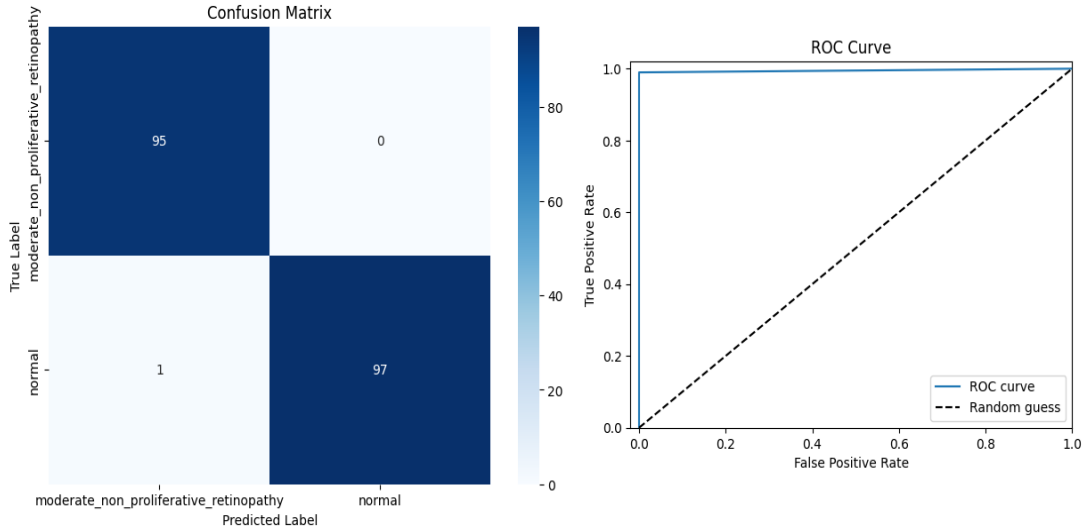

Figure. 12 Confusion Matrix and ROC curve for Normal Vs. Glaucoma from Unseen samples from IDRiD, Oculur recognition, and HRF datasets

Only two images from cataracts are improperly categorized due to the excellent performance of the trained model, which is reflected by its generalization capabilities to give superior performance on the prepared or test data and effectively give good accuracy for unseen test data from different databases. The ROC curve in these Figures proves the model's excellent performance with a True positive and False positive class accuracy score, which is evidence of the model's true generalization abilities. In this analysis, we aimed to test our trained model on unseen databases to evaluate its performance. It's important to note that while a model may be excellent in training if it's overfitting or underfitting, it will not develop true generalization capabilities, which are required as the model when in the real world, would face datasets from different sources. So, while training our model, we ensured that the model had enough fine-tuning so that when any data came, the extracted features would encompass the model to work as it genuinely worked on trained data. So, in this analysis, we tested our model on some new images, and the model's excellent generalization capabilities were proved.

On the other hand, model overfitting and underfitting as critical issues affecting the generalization of machine learning models have been investigated. When a model becomes overfit, its performance on unfamiliar data deteriorates because it has focused on memorizing the training set rather than deriving general principles from it. It is imperative to assess for overfitting, notwithstanding the high accuracy and minimum loss observed in the training and validation datasets. Overfitting can be indicated when the models manifest a serious variance in loss values with higher accuracy on the training data than the validation data. Underfitting occurs when a model's performance quality is reduced with training and validation datasets due to its limitations in recognizing the relevant patterns in the data. Therefore, the experimental finding of our proposed model demonstrates that the absence of noticeable differences in performance between the training and validation datasets indicates that the model is flexible and successfully grasps the essential correlations within the data.

6. Comparative analysis

We comprehensively assessed our proposed model's performance by comparing it to baseline models and state-of-the-art methodologies for identifying ocular disorders to establish its performance as a benchmark. A comparison study evaluated accuracy, processing efficiency, interpretability, and model complexity. Several studies have been conducted to address the difficulties associated with diagnosing eye disorders, such as handling severely imbalanced datasets and managing the high computational demands of models, which consume significant amounts of memory, processing, and time.

Researchers in [26] focused on considering the correlation between left and right eyes for diagnosis. So, A dense correlation network DCNet model was proposed based on a spatial correlation module for feature correlation and a backbone CNN for feature extraction. This work extracts two groups of characteristics from the left and right images. Then, the spatial correlation module records the pixel-wise correlations between the two feature sets. This work attained 97.8 accuracy with a significant limitation of model complexity that required high computation requirements due to its massive number of trainable parameters reaching 25.4 million.

The author in [30] proposed a deep learning model to create an eye disease classification using ShuffleNet V2 with Adam optimizer. This model cleared up fundus image noise and artifacts using denoising autoencoders with the wavelet search method and levy flight as feature extraction methods. This model required 23.5 million trainable parameters to achieve 99.1 classification accuracy. The authors in [35] proposed robust deep learning methods that enhance the detection performance by using a unique combination of a mixture loss function to autonomously identify eye disorders by analyzing retinal fundus color images. The proposed model combined the focal loss and correntropy-induced loss functions in a deep neural network model to enhance the classifier's ability to recognize biomedical data. They suggested this method to overcome the dataset's complexity due to class imbalances. Although the data imbalances have been sorted out in this work, the model saver from high computational complexities with trainable parameters reached 26.7 million with a classification accuracy of 96.5.

Researchers in [36] aimed to categorize an eye illness dataset using hybrid strategies combining feature extraction and fusion algorithms. They suggested enhancing the eye disease classification by applying a classical neural network with deep features extracted by integrated MobileNet and DenseNet121 deep features. This model achieved classification accuracy, reaching 98.7, regardless of its number of trainable parameters of 24.1 million, affecting its implementing requirements.

Compared to all previously discussed works, our suggested model has many contributions that involve a novel transfer learning structure integrated between

Table 5. Comparison of our suggested model with the state-of-the-art model based on the values of various metrics and model complexity

Reference	Model	Accuracy	Sensitivity	Specificity	F1-Score	Kappa	Dataset	learning Rate	Parameters in millions
Ref. [26]	Dense correlation network with spatial correlation features	97.8	96.1	93.1	95.9	93.1	ODIR	0.001	25.4
Ref. [35]	Deep learning with a novel mixture loss function	96.5	94.5	92.4	94.8	94.2	ODIR	0.001	26.7
Ref. [30]	whale optimization algorithm with Shuffle Net V2 model	99.1	98.9	96.3	98.9	96.4	ODIR	0.0001	23.5
Ref. [36]	ANN classifier-based hybrid features	98.7	96.9	91.3	97.0	94.8	ODIR	0.0001	24.1
Our approach	SENet with MobileNet Features Extraction with binary classification	99.96	99.96	99.96	99.96	96.3	ODIR	0.0001	3.6

deep features extracted by MobileNet with Squeeze-and-Excitation Network based on channel-wise feature recalibration concept to classify ocular diseases automatically from retinal pictures. In addition, our model applies significant methods for pre-processing the input data, such as augmentation, normalization, color mode conversion, edge detection, and cropping. Data analysis and dimensionality reduction were performed by applying PCA to our binary groups during the feature extraction step. All these steps contribute significantly to presenting our lightweight classification model with only 3.4 million trainable parameters and a very high performance that outperforms all other mentioned works with accuracy at 99.9%. Around 7 folds have reduced our proposed model's size and computation complexity compared to the state-of-the-art research in Table 5.

7. Conclusion

In summary, this study comprehensively examines the development and evaluation of deep learning models for automatically classifying eye illnesses based on retinal images by employing advanced lightweight model architectures based on the Squeeze-and-Excitation Network (SENet) built on a MobileNet basis. Our model has much fewer parameters compared to the state-of-the-art models. We have achieved remarkable accuracy rates above 99.9% on training and validation datasets across several binary classification tasks. Since the focus is on binary classification tasks to solve the multi-classification tasks and to predict whether an eye image represents a healthy case or one of the eye disease infections, the four classes that represent eye infection cases can be transformed into binary classification tasks. These four classes were grouped into three binary classification tasks: normal vs. cataract, normal vs. glaucoma, and moderate non-proliferative retinopathy vs. normal.

Our work investigates the transfer learning framework capability for automatically categorizing illnesses based on retinal images. Our algorithms' consistent accuracy measurements and low loss values illustrate their reliability and effectiveness in automated sickness identification. Our models consistently distinguish between retinal images that indicate good health and illness, providing essential information for managing ophthalmology patients, planning treatments, and diagnosing diseases early on. Our study highlights the practical usability and potential impact of AI-powered diagnostic tools in the medical domain. They achieve this by implementing automated ailment categorization and streamlining the diagnostic procedure.

Conflicts of Interest

The authors declare no conflict of Interest.

Author Contributions

This work involves many elements related to data preparation, model construction, analysis, software utilization, validation, inquiry, data curation, writing, reviewing and editing, and visualization, and for all these aspects, the authors have equal contributions. The allocation of authorship should be restricted to individuals who have made significant contributions to the reported work.

Notations:

Symbols	Definition of Symbol
$\tilde{x}$	Average of an input vector
x_i	One dimension vector
Φ_i	Subtract the Mean
A	Covariance matrix
C	eigenvalues and eigenvectors
D_G	Output feature map spatial dimensions
D_F	Input feature map spatial dimensions
$D_F \times D_F \times N$	Input feature map size
$D_G \times D_G \times M$	Output feature map size
α	Width Multiplier
ρ	Resolution Multiplier
N	Input channels
M	Output channels
D_F	Feature map size
D_k	Kernel size
K^	Kernel of the depth-wise convolution
DSC	Depth-wise Separable Convolution
H×W	Spatial dimensions
Ftr,	Convolution operator
v_C	Filter parameters
X	Corresponding channel
Z	Squeeze transformation
S	Excitation Operator
σ	Sigmoid activation functions
δ	ReLU activation functions
W_1	Dimensionality-reduction layer parameter
W_2	Dimensionality-increasing layer parameters
R	Reduction ratio
uc	Feature map
sc	channel-wise multiplication
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives

References

[1] Y. Chin, C. Ng, M. Lee, J. Koh, J. Kiew, S. Yang and C. Khoo, "Prevalence of thyroid eye disease in Graves' disease: A meta-analysis and systematic review", *Clinical endocrinology,* Vol. 93, No. 4, pp. 363-374, 2020.

[2] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automated detection of mild and multiclass diabetic eye diseases using deep learning", *Health Information Science and Systems,* Vol. 8, No. 1, pp. 32, 2020.

[3] R. Flaxman, R. Bourne, S. Resnikoff, P. Ackland, T. Braithwaite, V. Cicinelli, A. Das, B. Jonas, J. Keeffe, H. Kempen, and J. Leasher, "Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis", *The Lancet Global Health*, Vol. 5, No. 12, e1221-e1234. 2017.

[4] Y. Tong, W. Lu, Y. Yu and Y. Shen, "Application of machine learning in ophthalmic imaging modalities", *Eye and Vision*, Vol. *7,* No. 1, pp. 1-15, 2020.

[5] W. Lin, J. Chen, M. Chiang, and M. Hribar, "Applications of artificial intelligence to electronic health record data in ophthalmology", *Translational Vision Science & Technology,* Vol. 9, No. 2, pp. 13-13, 2020.

[6] G. An, K. Omodaka, K. Hashimoto, S. Tsuda, Y. Shiga, N. Takada, and T. Nakazawa, "Glaucoma diagnosis with machine learning based on optical coherence tomography and color fundus images", *Journal of Healthcare Engineering*, 2019.

[7] D. TingPeng, A. Varadarajan, P. Keane, P. Burlina, M. Chiang and T. Wong, "Deep learning in ophthalmology: the technical and clinical considerations", *Progress in retinal and eye research,* No. 72, pp. 100759, 2019.

[8] H. Al Abboodi, A. Al-funjan, N. Abd Hamza, A. Abdullah and B. Shami, "Supervised Transfer Learning for Multi Organs 3D Segmentation with Registration Tools for Metal Artifact Reduction in CT Images", *TEM Journal*, Vol.12, No. 3, pp. 1342, 2023.

[9] S. Sengupta, A. Singh, H. Leopold, T. Gulati, and V. Lakshminarayanan, "Ophthalmic diagnosis using deep learning with fundus images-A critical review", *Artificial intelligence in medicine*, Vol. 102, pp. 101758, 2020.

[10] S. Sreng, N. Maneerat, K., "Hamamoto and K. Win, "Deep learning for optic disc segmentation and glaucoma diagnosis on retinal images", *Applied Sciences*, Vol. 10, No. 14, pp. 4916, 2020.

[11] J. Sahlsten, J. Jaskari, J. Kivinen, L. Turunen, E. Jaanio, K. Hietala and K. Kaski, "Deep learning fundus image analysis for diabetic retinopathy and macular edema grading", *Scientific reports*, Vol. 9, No. 1, pp. 10750, 2019.

[12] S. Wagner, D. Fu, L. Faes, X. Liu, J. Huemer, H. Khalid and P. Keane, "Insights into systemic disease through retinal imaging-based comics", *Translational Vision Science & Technology*, Vol. 9, No. 2, pp. 6-6, 2020.

[13] Q. Yan, D. Weeks, H. Xin, A. Swaroop, E. Chew, H. Huang and W. Chen, "Deep-learning-based prediction of late age-related macular degeneration progression", *Nature Machine Intelligence*, Vol. 2, No. 2, pp. 141-150, 2020.

[14] Thompson, A. Jammal, and F. Medeiros, "A review of deep learning for screening, diagnosis, and detection of glaucoma progression", *Translational Vision Science & Technology*, Vol. 9, No. 2, pp. 42-42, 2020.

[15] M. Juneja, S. Singh, N. Agarwal, S. Bali, S. Gupta, N. Thakur, and P. Jindal, "Automated detection of Glaucoma using deep learning convolution network (G-net)", *Multimedia Tools and Applications*, No. 79, pp. 15531-15553, 2020.

[16] V. Bellemo, G. Lim, T. Rim, G. Tan, C. Cheung, S. Sadda and D. Ting, "Artificial intelligence screening for diabetic retinopathy: the real-world emerging application", *Current Diabetes Reports*, No. 19, pp. 1-12, 2019.

[17] L. Qiao, Y. Zhu, and H. Zhou, "Diabetic retinopathy detection using a prognosis of microaneurysm and early diagnosis system for non-proliferative diabetic retinopathy based on deep learning algorithms", *IEEE Access*, No. 8, pp.104292-104302, 2020.

[18] Y. Peng, S. Dharssi, Q. Chen, T. Keenan, E. Agrón, W. Wong and Z. Lu, "DeepSeeNet: a deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus photographs", *Ophthalmology*, Vol.126, No. 4, pp. 565-575, 2019.

[19] R. Poplin, A. Varadarajan, K. Blumer, Y. Liu, M. McConnell, G. Corrado, and D. Webster, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning", *Nature Biomedical Engineering*, Vol. 2, No. 3, pp. 158-164, 2018.

[20] F. Li, H. Chen, Z. Liu, X. Zhang, and Z. Wu, "Fully automated detection of retinal disorders by image-based deep learning", *Graefe's Archive for Clinical and Experimental Ophthalmology*, No. 257, pp. 495-505, 2019.

[21] N. Motozawa, G. An, S. Takagi, S. Kitahata, M. Mandai, Y. Hirami, and Y. Kurimoto, "Optical coherence tomography-based deep-learning models for classifying normal and age-related macular degeneration and exudative and non-exudative age-related macular degeneration changes", *Ophthalmology and therapy*, No. 8, pp. 527-539, 2019.

[22] T. Gadekallu, N. Khare, S. Bhattacharya, S. Singh, P. Maddikunta, and G. Srivastava, "Deep neural networks to predict diabetic retinopathy", *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-14, 2020.

[23] J. M. Ahn, S. Kim, K. Ahn, S. Cho, K. Lee, and U. Kim, "A deep learning model for the detection of both advanced and early glaucoma using fundus photography", *PloS one*, Vol. 13, No. 11, pp. e0207982, 2018.

[24] N. Li, T. Li, C. Hu, K. Wang, and A. Kang, "Benchmark of ocular disease intelligent recognition: One shot for multi-disease detection", *Benchmarking, Measuring, and Optimising Revised Selected Papers*, No. 3, pp. 177-193, 2021.

[25] R. Sundaram, R. Ks, and P. Jayaraman, "Extraction of blood vessels in fundus images of the retina through hybrid segmentation approach", *Mathematics*, Vol. 7, No. 2, pp. 169, 2019.

[26] J. He, C. Li, J. Ye, Y. Qiao, and L. Gu, "Multi-label ocular disease classification with a dense correlation deep neural network", *Biomedical Signal Processing and Control*, No. 63, pp. 102167, 2021.

[27] N. Gour and P. Khanna, "Multiclass multi-label ophthalmological disease detection using transfer learning based convolutional neural network", Biomedical *Signal Processing and Control*, No. 66, pp. 102329, 2021.

[28] M. Junayed, M. Islam, A. Sadeghzadeh, and S. Rahman, "CataractNet: An automated cataract detection system using deep learning for fundus images", *IEEE Access*, No. 9, pp. 128799-128808, 2021.

[29] Y. Xu, M. Hu, H. Liu, H. Yang, H. Wang, S. Lu, and N. Wang, "A hierarchical deep learning approach with transparency and interpretability based on small samples for glaucoma diagnosis", *NPJ digital medicine*, Vol. 4, No. 1, pp. 48, 2021.

[30] A.R. Wahab Sait, "Artificial Intelligence-Driven Eye Disease Classification Model", *Applied Sciences*, Vol. 13, No. 20, p.11437, 2023.

[31] S. Al-Fahdawi, A. S. Al-Waisy, D. Q. Zeebaree, R. Qahwaji, H. Natiq, M. A. Mohammed, J. Nedoma, R. Martine AND M. Deveci, "Fundus-Deepnet: Multi-label deep learning classification system for enhanced detection of multiple ocular diseases through data fusion of fundus image", *Information Fusion*, 102, 102059, 2024.

[32] D. Sachin, "Dimensionality reduction and classification through PCA and LDA", *International Journal of Computer Applications*, Vol. 122, No. 17, 2015.

[33] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks", In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132-7141, 2018.

[34] H. Al Abboodi, A. Al-funjan, A. Aldhahab,W. Salih Abedi, A. Abdullah, A. Siraj, "High-Resolution Model for Segmenting and Predicting Brain Tumor Based on Deep UNet with Multi Attention Mechanism", *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 2, 2024, doi: 10.22266/ijies2024.0430.25.

[35] X. Luo, J. Li, M. Chen, X. Yang, and X. Li, "Ophthalmic disease detection via deep learning with a novel mixture loss function", *IEEE Journal of Biomedical and Health Informatics*, Vol. 25, No. 9 pp. 3332-3339, 2021.

[36] A. Shamsan, E. M. Senan, H. S. A. Shatnawi, "Automatic Classification of Colour Fundus Images for Prediction Eye Disease Types Based on Hybrid Features", *Diagnostics*, Vol. 13, No. 10, pp. 1706, 2023.