



Research article

Cataract and glaucoma detection based on Transfer Learning using MobileNet

Sheikh Muhammad Saqib^a, Muhammad Iqbal^b, Muhammad Zubair Asghar^b,
Tehseen Mazhar^{c,*}, Ahmad Almogren^d, Ateeq Ur Rehman^{e,**}, Habib Hamam^{f,g,h,i}

^a Department of Computing and Information Technology, Gomal University, D.I.Khan 29050, Pakistan

^b Gomal Research Institute of Computing (GRIC), Faculty of Computing, Gomal University, D.I. Khan 29050, Pakistan

^c Department of Computer Science, Virtual University of Pakistan, Lahore, 51000, Pakistan

^d Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh, 11633, Saudi Arabia

^e School of Computing, Gachon University, Seongnam, 13120, Republic of Korea

^f Faculty of Engineering, Uni de Moncton, Moncton, NB, E1A3E9, Canada

^g School of Electrical Engineering, University of Johannesburg, Johannesburg, 2006, South Africa

^h Hodmas University College, Taleh Area, Mogadishu, Somalia

ⁱ Bridges for Academic Excellence, Tunis, Tunisia

ARTICLE INFO

Keywords:

Deep learning
Machine learning
Transfer learning
VeggNet
ResNet
And MobilNet

ABSTRACT

A serious eye condition called cataracts can cause blindness. Early and accurate cataract detection is the most effective method for reducing risk and averting blindness. The optic nerve head is harmed by the neurodegenerative condition known as glaucoma. Machine learning and deep learning systems for glaucoma and cataract detection have recently received much attention in research. The automatic detection of these diseases also depends on deep learning transfer learning platforms like VeggNet, ResNet, and MobilNet. The authors proposed MobileNetV1 and MobileNetV2 based on an optimized architecture building lightweight deep neural networks using depth-wise separable convolutions. The experiments used publicly available data sets with both cataract & normal and glaucoma & normal images, and the results showed that the proposed model had the highest accuracy compared to the other models.

1. Introduction

Numerous machine learning and deep learning techniques have been developed to assist the medical industry in quickly, affordably, and easily identifying diseases [1–3]. Several works were devoted to eye diseases [4–6]. These techniques attempt to emulate human vision in computers, including the ability to recognize things from their photos or videos [7]. Because most patient data from various diseases is useful for automatic learning, machine, and deep learning applications are frequently used in the medical sector. Recent studies have demonstrated that automated algorithms can automatically identify diseases from photos [8,9]. A cataract is a lenticular clouding that blocks the human eye's crystal-clear lens. Usually, the retina receives all of the light from the lens. Poor

* Corresponding author.

** Corresponding author.

E-mail addresses: saqibsheikh4@gu.edu.pk (S.M. Saqib), malikiqbalprince1@gmail.com (M. Iqbal), zubair@gu.edu.pk (M. Zubair Asghar), tehseenmazhar719@gmail.com (T. Mazhar), Ahalmogren@ksu.edu.sa (A. Almogren), 202411144@gachon.ac.kr (A. Ur Rehman), habib.hamam@gmail.com (H. Hamam).

<https://doi.org/10.1016/j.heliyon.2024.e36759>

Received 16 August 2024; Accepted 21 August 2024

Available online 24 August 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

visual acuity results from this light being blocked by the cataract and not reaching the lens. It is the most common eye disease in the world and does not immediately affect vision. But over time, it can affect vision and even lead to vision loss in those over 40 [10]. Early detection can prevent blindness and prevent painful and expensive surgery, depending on the severity of the cataract [11]. The World Health Organization (WHO) estimates that 285 million people worldwide suffer from visual impairment [12]. The remaining have impaired vision and 39 million of them have impaired vision. 33 % of visual impairments and 51 % of blindness were caused by cataracts [13].

Around the world, glaucoma is a well-known contributor to irreversible blindness. Glaucoma is an optic neuropathy that harms the retinal ganglion cells and results in permanent vision loss [14]. The retina's structural changes, particularly those in the optic nerve head (ONH) region, are what lead to this eye disease [15]. The most prevalent type of glaucoma is most likely open-angle eye disease (OAG). It starts by gradually obstructing the drainage system (angle between the iris and the cornea), which causes the eyecup region to enlarge and intraocular pressure to rise [7]. Although there is no cure for glaucoma, early detection can prevent significant vision loss [16]. Due to its lightweight network architecture, MobileNet is one of the most popular deep learning techniques currently in use [17]. The prime focus of the proposed work is to categorize the data as cataract, glaucoma, and normal by using a Transfer Learning model and to produce the best results from the benchmark [18].

The proposed method adds a global pooling layer, three fully connected layers after block layers, and bases the transfer learning process on MobileNetV1 and MobileNetV2. The major objectives of the Proposed Model are given below.

- 1 Constructing Model-1 based on MobileNetV1 and MobileNetV1 for images of cataract & normal eye.
- 2 Constructing Model-2 based on MobileNetV1 and MobileNetV1 for images of glaucoma & normal eye.
- 3 Constructing Model-3 based on MobileNetV1 and MobileNetV1 for images of cataract, glaucoma & normal eye as multi label classification.
- 4 Finally, the paper evaluated the performance of both models using performance parameters.

The objective of this investigation is to assess public databases widely regarded as gold standards in glaucoma analysis. The goal is to provide recommendations aimed at enhancing their effectiveness in aiding neural network learning, thereby facilitating notable progress in the automated screening of glaucomatous papillae. The study is organized into five key areas, encompassing essential aspects of glaucoma, the application of deep learning within the realm of artificial intelligence, the foundational elements of current public databases focused on retinal studies, and the core principles deemed as the gold standard in glaucoma research.

1.1. Problem statement

It is observed that forecasting eye diseases from data presents a significant challenge, mainly due to factors like suboptimal predictor selection, a restricted dataset size, and low image resolution [7,9,19]. Moreover, the effectiveness of Deep Learning models used for predicting eye diseases is hindered by insufficient selection of predictor variables and the absence of hybrid models. To address these issues, our approach involves leveraging transfer learning using MobileNetV1 and MobileNetV2. We aim to tackle the task of predicting eye diseases in the context of a binary-label scenario (1 - cataract, normal; 2 - glaucoma, normal) and a multi-label prediction scenario (cataract, glaucoma, and normal). In this framework, the prediction of eye diseases is performed based on the provided image-disease dataset.

1.2. Research objectives

- RO1.** Setting Architecture of MobileNetV1 and MobileNetV2 for Model-1, Model-2, Model-3.
- RO2.** Predicting Eye Disease as Cataract and Normal Eye through MobileNetV1 and MobileNetV2.
- RO3.** Predicting Eye Disease as Glaucoma and Normal Eye through MobileNetV1 and MobileNetV2.
- RO4.** Predicting Eye Disease as Cataract, Glaucoma and Normal Eye through MobileNetV1 and MobileNetV2.
- RO5.** Assessing the performance of the proposed model in comparison to other benchmark studies.

The primary novelty of this work lies in the innovative application of MobileNetV1 and MobileNetV2 architectures to the specific task of cataract and glaucoma detection, which addresses several critical gaps identified in the existing literature. Our key contributions include.

- **Customized Architecture for Eye Disease Detection:** We have meticulously tailored the MobileNetV1 and MobileNetV2 architectures to cater to the unique challenges of distinguishing between cataract, glaucoma, and normal eyes. This includes the introduction of a global pooling layer and three fully connected layers, enhancing the model's ability to learn and generalize from the data.
- **Multi-label Classification:** Unlike many existing studies that focus solely on binary classification, our approach extends to multi-label classification, effectively predicting multiple eye conditions within a single model. This is a significant advancement in the field, offering a more comprehensive diagnostic tool.

- **Comparative Performance Analysis:** We have conducted an exhaustive evaluation of our proposed models against existing benchmarks, demonstrating superior performance in terms of accuracy and efficiency. This thorough comparative analysis validates the robustness of our approach.
- **Integration of Public Datasets:** We have utilized publicly available datasets for training and validation, ensuring the generalizability and reliability of our models. The integration of these datasets has been methodically documented to facilitate reproducibility and further research.
- **Efficiency for Mobile and Embedded Devices:** Our models are optimized for deployment in resource-constrained environments, such as mobile and embedded devices, without compromising on accuracy. This is achieved through the use of depth-wise separable convolutions, making our solution both lightweight and highly effective.

2. Related work

2.1. Current research

The majority of research on eye diseases has predominantly utilized machine learning and deep learning techniques, with a focus on machine learning literature preceding deep learning studies. In one instance, a computer-assisted cataract classification system, employing fundus images, was introduced in Ref. [20]. Feature extraction involved wavelet transform and sketch-based methods, followed by cataract detection and grading using a multi-class discriminant analysis algorithm. The Correct-Classification-Rates (CCRs) for wavelet transform-based feature extraction were 90.9 % and 77 %, and for sketch-derived features, they were 86.1 % and 70 %. Another study [21] employed the K-Nearest Neighbor (KNN) classifier, utilizing texture features—dissimilarity, contrast, and uniformity—with promising results. In Ref. [22], a competitive cataract detection system was proposed based on statistical texture analysis and Artificial Neural Networks (ANN). Using ANN, the test set was successfully categorized as normal or cataract, achieving an average accuracy of 94.5 %. Detecting glaucoma, authors in Ref. [23] utilized fundus images from the DRISHTI and RIM-ONE V3 datasets, focusing on feature extraction of the ocular cup and intervertebral disc. Meanwhile, an in-vivo automatic Nuclear Cataract detection and classification system was developed in Ref. [24] using machine learning and ultrasound techniques. Cataract images were categorized using an SVM classifier in Ref. [25], achieving a specificity of 93.33 %, while [19] explored a Convolutional Neural Network (CNN) model for automatic glaucoma classification through transfer learning on DRISHTI and RIM-ONE V3 fundus images. Additionally [26], introduced a cataract detection method based on an Android smartphone, employing a single-layer perceptron classifier with an accuracy of 85 %.

In their publication [27], the authors employed five distinct ImageNet-trained architectures, including Xception, ResNet50, VGG19, VGG16, and InceptionV3, for glaucoma detection. Notably, these architectures do not necessitate feature extraction or estimates of optic nerve head (ONH) geometric structures. The achieved scores were 0.8354, 0.8041, 0.8575, and 0.7739, respectively. Deep learning-based approaches have the capability to autonomously learn essential features, seamlessly integrate them into the model creation process, and diminish the reliance on manually designed features, rendering them applicable across various medical imaging modalities [28]. In another study focused on determining the severity of nuclear cataracts, slit-lamp images were utilized in a deep learning-based method. A local filter (CNN) was formulated by combining image fields into a convolutional neural network. Additionally, higher-order features were extracted using a group of recursive neural networks (RNNs) [29]. A deep CNN (DCNN) for cataract detection and grading was developed by utilizing feature maps from the architectures' bundling layers. This approach demonstrated both speed and accuracy, achieving 93.52 % accuracy for cataract detection and 86.69 % accuracy for grading [30]. Furthermore, in their research [31], the authors introduced a trained classifier model based on the ResNet architecture and proposed an automatic cataract detection system utilizing DCNNs.

The proposed approach in Ref. [32] introduces a method for the automatic classification of colposcopy images, leveraging MobileNetV2 networks. The input comprises three distinct image types: VIA, a green lens, and VILI. MobileNetV2's inverted residual and linear bottleneck layer module [17] significantly reduce the memory requirements for processing. In a related study [33], a lightweight model was proposed, employing a transfer learning approach with MobileNetV2. This model was specifically designed for low-power devices and limited resources. They enhanced the basic MobileNetV2 architecture by incorporating an additional convolution and dropout layer. The feature extractor was a five-class fruit dataset with 3213 training images, fine-tuned with the Softmax layer using MobileNetV2, and pre-trained on the ImageNet dataset. The validation set, consisting of fruit images (457 images), achieved an accuracy of 85.12 %. In various domains such as medicine [34,35] agriculture [36–38], text analysis [39–41] and spam analysis [42,43], considerable research has been conducted. However, a majority of papers primarily depend on traditional machine learning techniques, and only a few instances mention using MobileNet approaches for cataract and glaucoma detection. This leads to ongoing challenges, such as the necessity to improve model accuracy while also simplifying by minimizing training parameters, layers, depth, runtime, and the overall model size.

2.2. Research gap

Despite significant advancements in the application of machine learning and deep learning for eye disease detection, several gaps persist in the existing literature. This section identifies these gaps and establishes the foundation for our proposed approach.

- **Suboptimal Architecture Customization:** Existing studies often utilize generic architectures without sufficient customization to address the specific complexities of eye disease detection. There is a notable lack of specialized adaptations in MobileNetV1 and MobileNetV2 architectures for this domain, which our study addresses comprehensively.
- **Limited Focus on Multi-label Classification:** The majority of current methodologies are confined to binary classification tasks. However, in clinical settings, the ability to predict multiple conditions simultaneously is crucial. Our research introduces a robust multi-label classification approach that significantly enhances diagnostic capabilities.
- **Deficit in Comparative Performance Analysis:** While various deep learning models have been proposed, there is a scarcity of comprehensive performance evaluations against established benchmarks. Our study fills this gap by providing an exhaustive comparative analysis, demonstrating the superiority of our models over existing ones.
- **Insufficient Integration of Public Datasets:** Although public datasets are widely available, their integration into deep learning models is often inadequate, leading to suboptimal generalization. Our methodical application of these datasets enhances model reliability and offers a standardized approach for future research.
- **Lack of Comprehensive Objective Assessments:** Many studies do not align their objectives with architectural choices and implementation strategies. Our research explicitly links detailed architectural configurations with multi-label classification objectives, addressing this critical methodological gap.

By addressing these gaps, our study not only advances the state-of-the-art in eye disease detection but also provides a robust framework for future research in this domain.

3. MobileNet architecture (RO1)

Three channels' images of size 224*224 with 3*3 KS (Kernel Sizes) and 32 filters by applying post ZeroPadding2D padding, make up the first block's inputs. The spatial size of the data representation is then decreased (width and height) by applying a Max Pooling (MP) layer with steps of 2 (step size). Since more pixels correspond to more parameters, which requires a huge amount of data, it primarily reduces the image size. Finally, the ReLU activation function is used to activate this block, which means that the negative values of the matrix are considered 0 while the positive values remain unchanged using Equation (1). The processing in the following blocks is the same, but the filters used will be B. 96, 144, 192 and 576. A feature map is created by combining all block outputs and fed into the layers that are fully connected. In comparison to the normal eye, these layers— the flat and dense ones— are intended to detect cataracts or glaucoma. In the last dense layer, the softmax activation function is used to calculate the percentage of recognized images as output. With the probabilities of each value being proportional to the vector's relative scale, the mathematical function Softmax transforms a vector of numbers into one of probability. Model-1, Model-2, and Model-3 contain the same implementation except for the output layer as shown in Table 1. The output Layer for Model-1 and Model-2 contains two neurons because these are binary classifications while Model-3 contains three neurons at the output layer as it is a multiclass problem. The description of the above work is shown in Table 1.

$$f(z) = \max(0, z) \tag{1}$$

Where $z = (\sum_{i=1}^n x_i w_i + b_i)$

$$\text{i.e. Output} = \begin{cases} z, & f(z) \geq 0 \\ 0, & f(z) < 0 \end{cases}$$

Table-1
Architecture for MobileNetV2.

Layer	Output	Stride	Filter	Config
Conv2D	224*224*32	–	32	Relu-KS: 3*3
Max-Pooling	112*112*32	2	–	KS: 2*2
Conv2D	113*113*96	–	96	Relu-KS: 3*3
Max-Pooling	56*56*96	2	–	KS: 2*2
Conv2D	57*57*144	–	144	Relu-KS: 3*3
Max-Pooling	28*28*144	2	–	KS: 2*2
Conv2D	29*29*192	–	192	Relu-KS: 3*3
Max-Pooling	14*14*192	2	–	KS: 2*2
Conv2D	15*15*576	–	576	Relu-KS: 3*3
Max-Pooling	7*7*576	2	–	KS: 2*2
Flatten	62720	–	–	–
Dense	200	–	–	Relu-200
Dense	200	–	–	Relu-200
Output Layer at Model-1				
Dense	2	–	–	Softmax-2
Output Layer at Model-2				
Dense	2	–	–	Softmax-2
Output Layer at Model-2				
Dense	3	–	–	Softmax-3

Fig. 1 shows the working of the proposed model. Where normal eye and cataract images are used to train MobileNetV1 and MobileNetV2 models to recognize a percentage of predicted images. Both models are then trained to predict images of normal eyes and glaucoma eyes.

The proposed method is based on MobileNetV2 [17]. By incorporating linear bottleneck modules and inverted residuals, MobileNetV2 evolved from mobileNetV1 [44]. The depth-wise separable convolution was the foundation of the MobileNet architecture. Standard 2D convolution creates an output channel by directly processing each input channel and convolving in the depth (channel) dimension. The depth-wise convolution separates the input image and filter into separate channels before each input channel converges with the corresponding filter channel. After being produced, the filtered output channels are combined. A separate depth-wise convolution is then used to combine the stacked output channels into one channel, which is then filtered using an 11-fold cone, also known as the point-worst cone. The depth-wise separable convolution achieves the same outcome as the standard convolute because it uses fewer parameters, but it is more efficient [44].

MobileNetV1 consists of 28 layers of convolutions, resulting in an output size of 771280 pixels when counting depth-wise and pointwise convolutions separately. Both MobileNetV1 and MobileNetV2 accept input images with a resolution of 2242243. Therefore, the input images in the dataset are resized and cropped to 224224 pixels. MobileNetV2 follows its first convolution layer with 32 filters by inserting 19 inverted residual bottleneck layers, concluding with a pointwise convolution that generates an output of 77*1280 pixels [17].

To connect data to the lower layers of the network, the remaining block links a convolutional block at its beginning and end using snoop connections. The convolution block of the standard residual block usually has more channels at its start and end compared to its layers in between. In MobileNetV2, the inverted residual block differs significantly from the layers in between, having fewer channels and parameters. The upper section of Fig. 1 illustrates the contrast between the layers of MobileNetV1 and MobileNetV2. Unlike MobileNetV1, MobileNetV2 includes several convolution blocks with skip connections. Fig. 1 depicts the construction of Model-1, Model-2, and Model-3.

3.1. The process of finalizing model parameters

After conducting several experiments to optimize our model’s performance, we finalized the parameters and architecture as follows: The model architecture includes a flatten layer that converts the input data, which is the output of the ReLU activation, into a single-dimensional array of size 62,720. This is followed by a dense layer with 200 neurons, chosen to provide a substantial number of parameters for the network to learn complex patterns, resulting in 12,544,200 parameters. The next dense layer has 100 neurons, helping reduce dimensionality while retaining important features, contributing 20,100 parameters. The final dense layer, with 2 neurons, corresponds to the binary classification task and has 202 parameters. The total number of parameters in the model is 14,822,486, with 12,564,502 being trainable and 2,257,984 non-trainable.

We used the ‘categorical_crossentropy’ loss function, suitable for multi-class classification problems, and the ‘Adam’ optimizer for its efficiency and adaptive learning rate capabilities. To prevent overfitting and improve generalization, we applied data augmentation techniques such as rescaling (1/255), shear range (0.2), and zoom range (0.2). The model was trained using the test set for validation, over 20 epochs, with the number of steps per epoch set to the length of the training set and the number of validation steps set to the length of the test set.

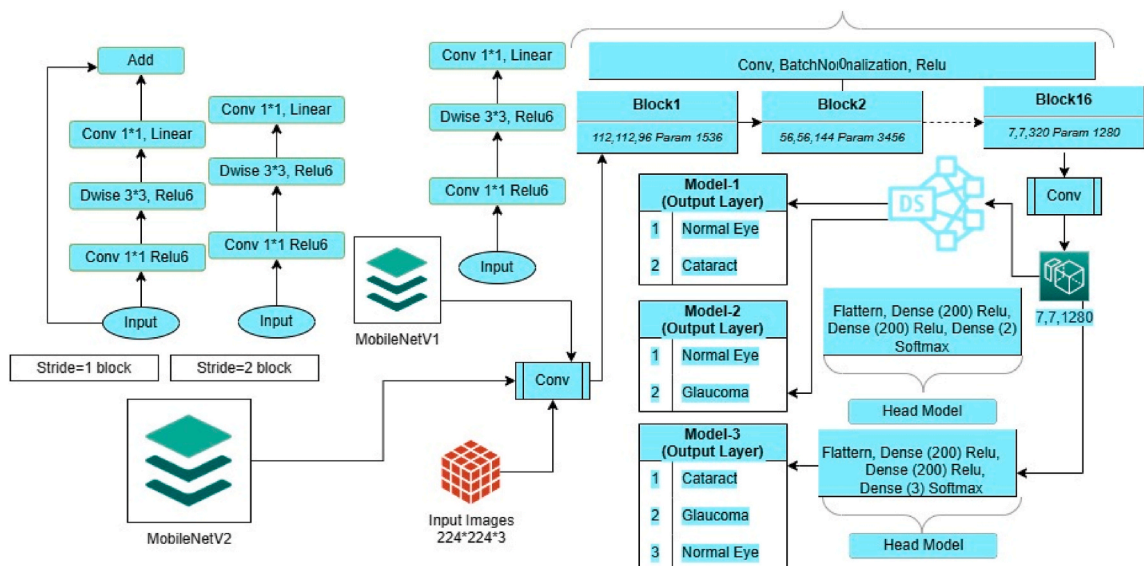


Fig-1. Proposed Work based on MobileNetV1 and MobileNetV2.

The experimentation process began with an initial model design using a basic architecture and default hyperparameters to establish a baseline performance. Various hyperparameters, including the number of layers, number of neurons per layer, activation functions, learning rates, batch sizes, and optimizers, were systematically varied and tested. The model was evaluated based on accuracy, loss, and validation accuracy, aiming to maximize accuracy while minimizing loss and ensuring validation accuracy was consistent with training accuracy to prevent overfitting. Data augmentation techniques were applied to enhance the model's robustness and generalization capability. After numerous iterations, the final architecture and hyperparameters were selected based on their consistent performance across different experiments and datasets. This systematic approach to hyperparameter tuning and model selection proves that the listed parameters and architecture were indeed finalized after extensive experimentation.

3.2. Dataset collection

Public databases [45] provide color images of the retina with a focus on highlighting the optic disc, specifically tailored for glaucoma and cataract research. These databases are curated for research purposes and are conducive to the application of standardized automated methods, including deep learning techniques. Such techniques play a pivotal role in addressing intricate issues within medical imaging, particularly in the automated screening of glaucoma and cataract diseases. The dataset encompasses a total of 500 images, with 300 depicting normal eyes, 100 featuring cataract-afflicted eyes, and 100 showcasing glaucoma-affected eyes. An illustration of this dataset is presented in Fig. 2.

3.2.1. Using the aforementioned data

A folder named 'cataract' contains eye images containing cataract disease in png format, a folder named 'glaucoma' contains eye images containing glaucoma-disease in PNG format and a folder named 'normal' contains eye images containing no disease in PNG format. All folder are stored in the main folder named 'dataset'. For reading image data, the 'os. Walk' command-line interface option is employed, which is a fundamental and versatile tool. The ImageDataGenerator [46] is utilized to separate and rescale the training and testing sets. Specifically, 80 % of the data is allocated for training, while the remaining 20 % is designated for testing, as indicated in Table 2.

3.2.2. Train set

About 80 % of the training data was used for training [47]. This data includes both the outcome identifier (dependent variable) and the input elements (predictor variables).

3.2.3. Testing set

3.3. Reasons to select MobileNetV for efficient model deployment

MobileNetV1 and MobileNetV2 are popular choices for transfer learning due to their efficiency and suitability for resource-constrained environments. These models are designed with a focus on lightweight architecture and computational efficiency, leveraging depth wise separable convolutions to significantly reduce the number of parameters and computations required compared to traditional convolutional networks. This design makes MobileNet models particularly advantageous for deployment on mobile and embedded devices, where storage and processing power are limited. One of the key benefits of MobileNet models is their reduced size, which translates to faster inference times and lower power consumption. MobileNetV2, for instance, introduces innovations such as linear bottlenecks and inverted residuals, which enhance the model's representational power while maintaining efficiency. This results

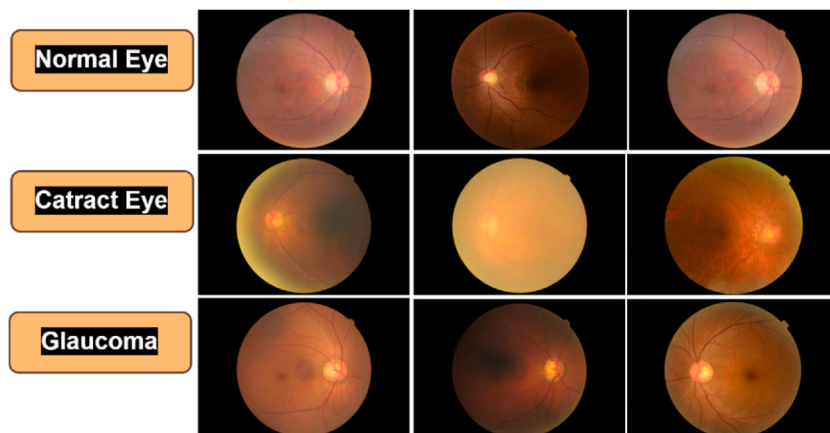


Fig-2. Sample of dataset.

Table 2
Description of training and test Data images.

Class	No of Images	Training Data	Test Data
Normal Eye	300	60 %	40 %
Cataract Eye	100	60 %	40 %
Glaucoma Eye	100	60 %	40 %

in high accuracy that is competitive with larger models, making MobileNetV2 a robust choice for applications requiring real-time performance. Additionally, MobileNet models offer flexibility through adjustable parameters like the width multiplier and resolution multiplier, allowing practitioners to tailor the model to their specific needs. This adaptability is particularly useful for edge computing scenarios, where the balance between model size, latency, and accuracy must be finely tuned. MobileNet models are also compatible with TensorFlow Lite, facilitating their deployment on a variety of devices and platforms, including smartphones, tablets, and IoT devices [17,44].

3.4. Model construction

Our proposed method is written in Python and implemented with Keras. Table 3 shows the hyperparameters we used in our research. We split each dataset into training and test sets in a 60:40 ratio. To prevent the model from overfitting during training, we use 20 epochs and set the total number of images in the test set for validation and the total number of images in the training set for each epoch using the Adam optimizer in Equation (2). Two datasets from Table 1 were used to evaluate the performance of the proposed model. Table 1 shows the architecture of the proposed models.

$$Wt_{t-1} = \frac{\eta}{\sqrt{Sdwt^{Correction} + \epsilon}} \times Vdwt^{Correction} \tag{2}$$

Using Adam, in neural net back propagation, new weight (Wt) can be calculated with old weight by using learning rate η , exponential weighted average V_{dw} and average movement S_{dw} .

$Vdwt^{Correction}$ and $Sdwt^{Correction}$ can be calculated using Eq-3 and Eq-4.

$$Vdwt^{Correction} = \frac{Vdwt}{1 - \beta 1^t} \tag{3}$$

$$Sdwt^{Correction} = \frac{Sdwt}{1 - \beta 2^t} \tag{4}$$

Where,

$$Vdwt = \frac{V_{dw}}{t-1} \times (1 - \beta) \frac{\delta L}{\delta Wt - 1} \tag{5}$$

$$Sdwt = \frac{S_{dw}}{t-1} \times (1 - \beta) \frac{\delta L}{\delta Wt - 1} \tag{6}$$

δL means loss derivative concerning old weight. Eq-7 to Eq-11 can also be used to update bias Bs, as follows:

$$Bs_t = \frac{\eta}{\sqrt{Sdwt^{Correction} + \epsilon}} \times VdBst^{Correction} \tag{7}$$

$$VdBst^{Correction} = \frac{VdBst}{1 - \beta 1^t} \tag{8}$$

Table-3
Parameters for MobileNetV1 and MobileNetV2 for two datasets.

Models	Split	Parameters
MobileNetV1 (Cataract + Normal eyes Dataset)	training Images: 60 % Testing Images: 40 %	13,264,666
MobileNetV2 (Glaucoma + Normal eyes Dataset)	training Images: 60 % Testing Images: 40 %	13,284,566
MobileNetV1 (Cataract + Normal eyes Dataset)	training Images: 60 % Testing Images: 40 %	14,842,786
MobileNetV2 (Glaucoma + Normal eyes Dataset)	training Images: 60 % Testing Images: 40 %	14,822,486

$$SdBst^{Correction} = \frac{SdBst}{1 - \beta 2^t} \quad \text{---} > \quad (9)$$

Where,

$$VdBst = \frac{V_{dw}}{t-1} \times (1 - \beta) \frac{\delta L}{\delta Bst - 1} \quad \text{---} \quad (10)$$

$$Sdwt = \frac{S_{dw}}{t-1} \times (1 - \beta) \frac{\delta L}{\delta Bst - 1} \quad \text{---} \rightarrow \quad (11)$$

Where, β , β_1 , and β_2 are hyperparameters with values ranging from 0 to 1.

3.4.1. Softmax-based prediction scheme

By employing softmax in the hidden layers as the output layer, we determine the likelihood of predicting target labels. These labels correspond to eye conditions, such as 1- cataract or normal, 2- glaucoma or normal, or 3- cataract, glaucoma, or normal. The net input is determined using a formula (Eq. (12)), with the weight vector represented as 'w,' the input vector as 'x,' and the bias as 'b.' Utilizing Eq. (13) enables the computation of the softmax.

$$f(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

Where $z = (\sum_{i=1}^n x_i w_i + b_i)$

$$Output = \begin{cases} 50\%, v1 = 0.5 \\ 40\%, v2 = 0.4 \\ 10\% \quad v3 = 0.1 \end{cases} \quad (13)$$

3.5. Applied example

We conducted multiple computations to predict the occurrence of eye diseases using the latest disease data. Each aspect of the model's operation is thoroughly explained in detail.

3.5.1. Data Preparation

We pick a patient instance from the disease dataset and apply the suggested models for the prediction of eye disease as Model-1: Cataract or Normal, Model-2: Glaucoma or Normal Model-3: Cataract, Glaucoma or Normal. Package 'ImageDataGenerator' of Keras

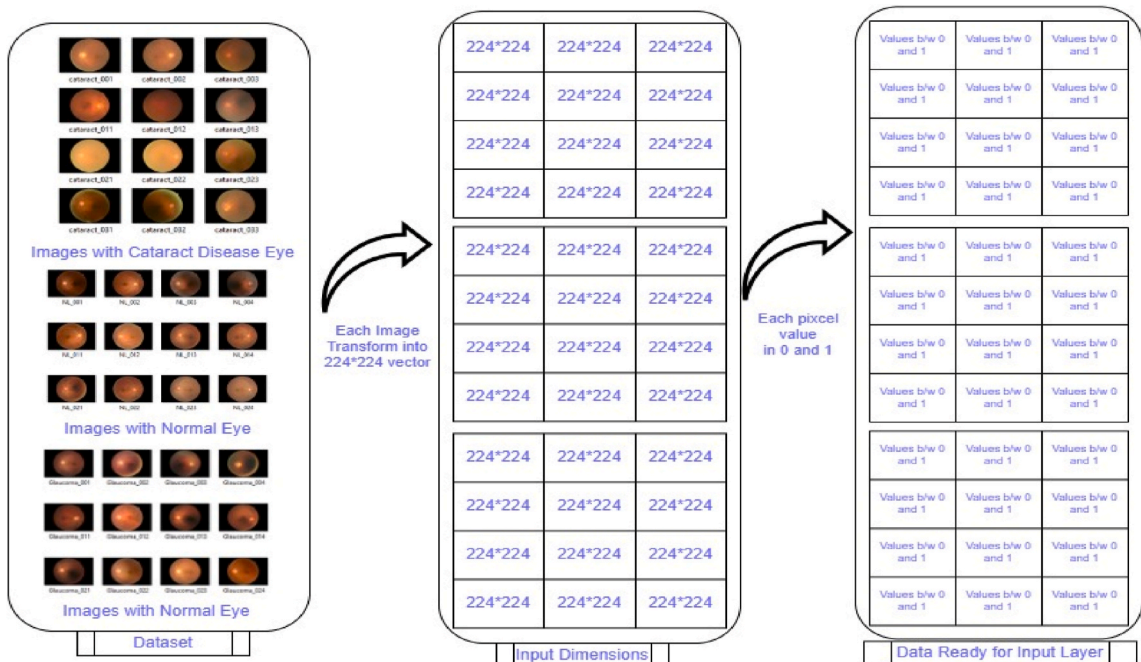


Fig-3. Conversion image dataset into Numeric vector.

will rascal all images into 224 pixels, then will convert them into vectors of normalized values (between 0 and 1) as shown in Fig. 3.

For cataract, glaucoma, and normal eyes, respectively, the images in Fig. 4 are referred to as "catastrophe_001," "NL_00," and "Galumoscope-001." These images will first be transformed into a 224*224 vector, and then each value of these vectors will be divided by 255 to normalize the values between 0 and 1. Models 1 and 3 both receive input in the form of cataract, glaucoma, and normal images, while Model 2 and Model 3 receive it in yellow lines as normal, abnormal, or cataract images.

Model-1 and Model-2 contain two neurons at the last layers while Model-3 contains three neurons at the output layer because Mode-1 and 2 will solve the binary problem and Model-3 will solve the multi-label problem. Model-1, Model-2, and Model-3 have been applied on MobileNetV1 and MobileNetV2 with head sections as shown in Fig. 5.

4. Experimental results and discussion

The first research question has been explained in section 3. This section explains the experiment results used to address the remaining research questions.

4.1. RO2: predicting eye disease as Cataract and Normal Eye through MobileNetV1 and MobileNetV2

Following the application of cataract-eye and glaucoma-eye datasets to both models concerning normal-eye, a sample of predicted values on the said dataset is shown in Table-4, where values 1.0000000e+00, 6.8171773e-14 with a model-MobileNetV1 show that the input image is 100 % detecting as cataract and 0 % as normal eye, and so on.

Table 5 shows two images for each model, one for the cataract eye and one for the normal eye. In Table-5, predicted values 1.000000e+00, 6.817191e-14 from MobileNetV1 and predicted values 1.0000e+00, and 1.3925e-22 from MobileNetV2 indicate that the given image is predicted to be 100 % cataract and 0 % normal eye. While MobileNetV1 predicted values of 1.3467639e-19,

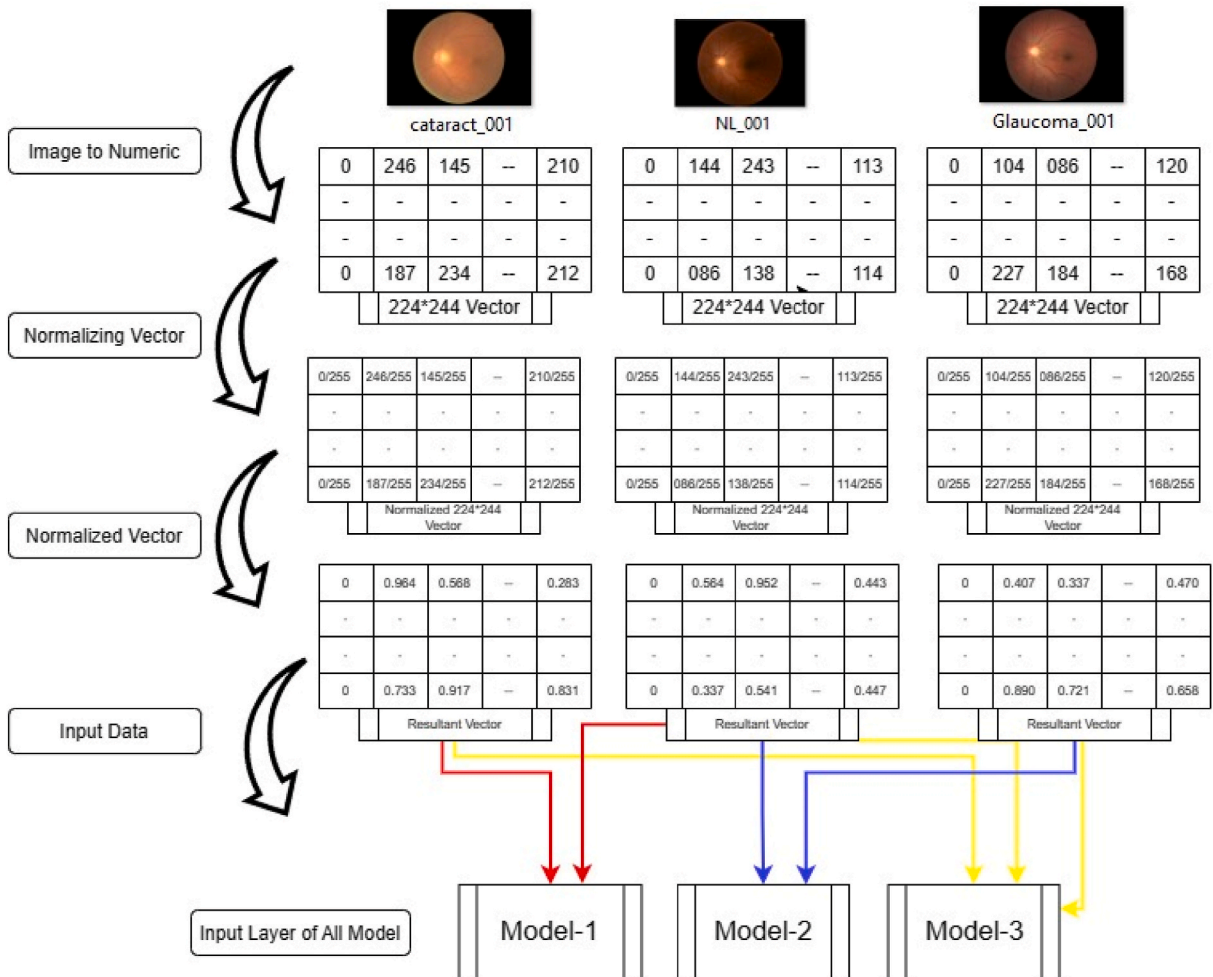


Fig-4. Conversion of Input Image to Model Readable form.

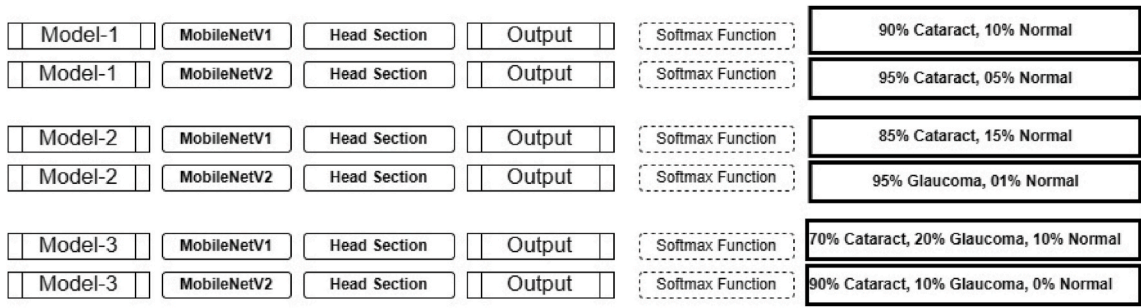


Fig-5. Decision of Model-1, Model-2 and Model-3 based on MobileNetV1 and MobileNetV2.

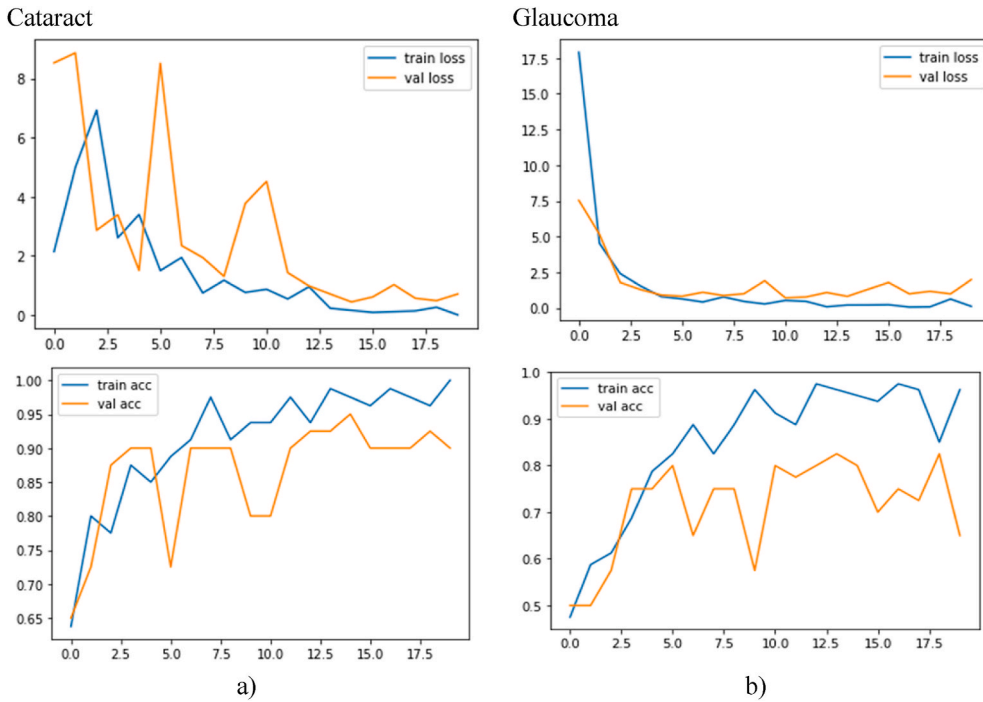


Fig-6. MobileNetV1 loss and accuracy on cataract and glaucoma.

1.0000000e+00 and MobileNetV2 predicted values of 1.424292e-32, 1.0000000e+00 represent that the given image is 100 % considered normal eye. That is an excellent outcome.

4.2. RO3: Predicting eye disease as Glaucoma and Normal Eye through MobileNetV1 and MobileNetV2

Table 6 displays two images for each model, one for the normal eye and one for the glaucoma eye. Table 6 shows that the predicted values 9.999906e-01 9.459773e from MobileNetV1 indicate that the given image is 100 % glaucoma and 0 % normal eye. MobileNetV2 predicted values of 0.9916482 and 0.00835176 show that the given image is 99 % glaucoma and 1 % normal eye. With MobileNet1, predicted values 0.260271, and 0.739729 show that the given image is 26 % glaucoma and 73 % normal eye, while predicted values 0.05109226, and 0.94890773 show that the given image is 5 % glaucoma and 95 % normal eye. That is an excellent result.

4.3. RO4: Predicting eye disease as cataract, Glaucoma and Normal Eye through MobileNetV1 and MobileNetV2

Table-7 shows that with MobileNetV1, the given image is 100 % cataract, 0 % glaucoma, and 0 % normal eye, with predicted values of 1.0000000e+00, 4.6132187e-17, and 7.7960607e-09. The predicted values of 8.2315728e-02, 8.9646542e-01, and 2.1218920e-02 say that the given image has 8 % cataract, 90 % glaucoma, and 2 % normal eye. The predicted values of 8.0872782e-02, 3.6249526e-02, and 8.8287771e-01 show that the given image has 8 % cataract, 4 % glaucoma, and 88 % normal eye. With MobileNetV2, predicted

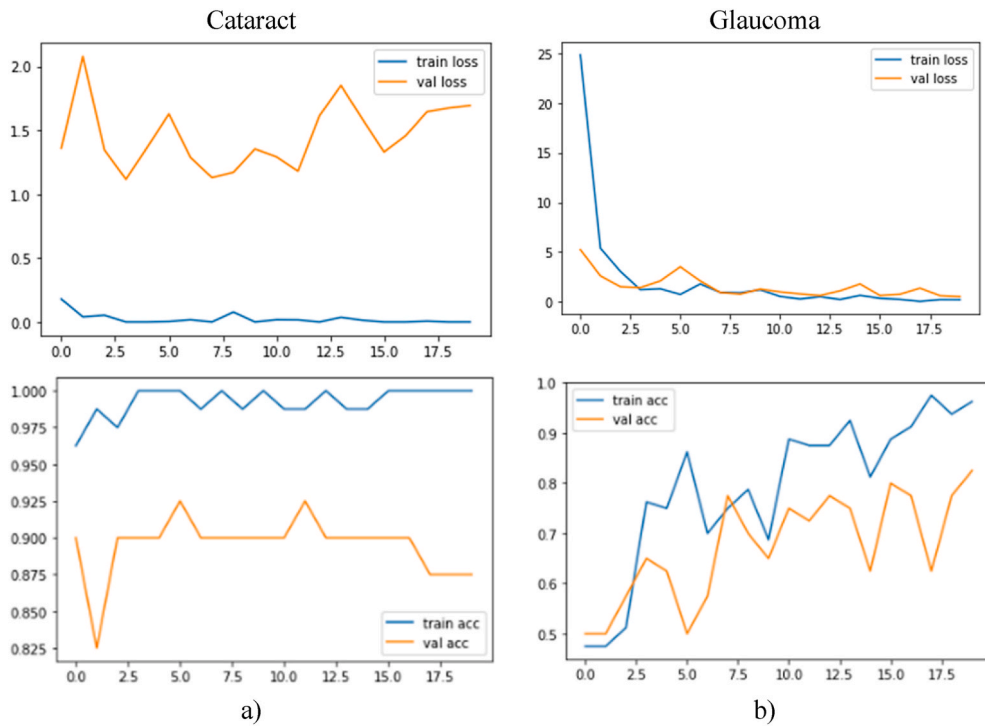


Fig-7. MobileNetV2 loss and accuracy on cataract and glaucoma.

Table 4

Sample of both datasets as predicted values through MobileNetV1 and MobileNetV2.

Models	Cataract: Normal	Glaucoma: Normal
MobileNetV1	[1.000000e+00,6.8171773e-14]	[2.79642493e-01, 7.20357537e-01]
	[9.8119515e-01,1.8804921e-02]	[9.17692721e-01, 8.23073164e-02]
	[1.0000000e+00,5.1756675e-08]	[9.99999285e-01, 7.35101423e-07]
	[1.0000000e+00,2.2646173e-11]	[8.07221532e-01, 1.92778468e-01]
	[1.4882896e-05,9.9998510e-01]	[9.98786747e-01 1.21329841e-03]
MobileNetV2	[6.8655385e-21, 1.0000000e+00]	[9.2175305e-01, 7.8246929e-02]
	[1.0000000e+00, 3.6590593e-09]	[9.9991500e-01, 8.5048421e-05]
	[4.5436084e-15, 1.0000000e+00]	[2.0411807e-04, 9.9979585e-01]
	[0.0000000e+00, 1.0000000e+00]	[9.8578602e-01, 1.4213971e-02]
	[1.0000000e+00, 7.5790901e-10]	[9.8057697e-03, 9.9019426e-01]

Table 5

Predicted values of cataract and normal eye through MobileNetV1 and MobileNetV2.

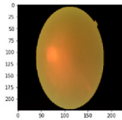
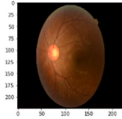
Image Name	Display Image	MobileNetV1	MobileNetV2
Cataract Image		[1.000000e+00, 6.817191e-14]	[1.0000e+00, 1.3925e-22]
Normal Image		[1.3467639e-19, 1.0000000e+00]	[1.424292e-32, 1.000000e+00]

Table 6
Predicted values of glaucoma and normal eye through MobileNetV1 and MobileNetV2.

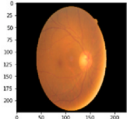
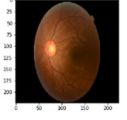
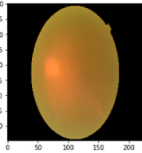
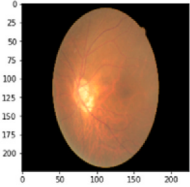
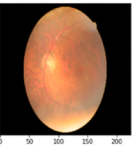
Image Name	Display Image	MobileNetV1	MobileNetV2
Glaucoma Image		[9.999906e-01, 9.459773e-06]	[0.9916482, 0.00835176]
Normal Image		[0.260271, 0.739729]	[0.05109226, 0.94890773]

Table 7
Predicted values of cataract, glaucoma, and normal eye through MobileNetV1 and MobileNetV2.

Model	Cataract	Glaucoma	Normal
Images			
MobileNetv1	[1.000000e+00, 4.6132187e-17, 7.7960607e-09]	[8.2315728e-02, 8.9646542e-01, 2.1218920e-02]	[8.0872782e-02, 3.6249526e-02, 8.8287771e-01]
MobileNetv2	[1.000000e+00, 2.0047671e-19, 1.8522156e-15]	[1.7136662e-03, 9.5648021e-01, 4.1806191e-02]	[1.0107529e-03, 1.9605255e-02, 9.7938401e-01]

values of 1.000000e+00, 2.0047671e-19, and 1.8522156e-15 say that the given image has 100 % cataract, 0 % glaucoma, and 0 % normal eye. The predicted values of 1.7136662e-03, 9.5648021e-01, and 4.1806191e-02 show that the given image has 0 % cataract, 96 % glaucoma, and 4 % normal eye. With predicted values of 1.0107529e-03, 1.9605255e-02, and 9.7938401e-01, the given image has 0 % cataract, 2 % glaucoma, and 98 % normal eye.

4.4. Statistical measures

Confusion matrices are a popular tool for trying to solve classification problems. It can solve both binary classification and multiclass classification problems. The formula below is used to decide the correctness of a model (via a confusion matrix). The confusion matrix includes correctly classified TP values, FP values in the proper class when they should be in another class, FN values in another class when they should be in the right class, and correctly classified TN values in the other class. The performance metrics accuracy (ACC) can be calculated using Eq-2 [48] based on these values. Accuracies of MobileNetV1 and MobileNetV2 on Cataract-Normal and Glaucoma-Normal datasets are shown in Table-8.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad \text{--- (2)}$$

After merging the datasets and categorizing the images as cataract, glaucoma, and normal eye, MobileNetV1 and MobileNetV2 are rearranged for a multiclass problem by changing the output layer with three neurons and the softmax activation function (see Table 9). After 20 epochs, MobileNetV1 achieves 71 % accuracy and MobileNetV2 achieves 72 % accuracy. Because most images in cataract,

Table 8
Accuracy of two models on cataract dataset and glaucoma dataset.

Models	Training Accuracy	Testing Accuracy
MobileNetV1 (Cataract + Normal eyes Dataset)	99 %	90 %
MobileNetV1 (Glaucoma + Normal eyes Dataset)	98 %	72 %
MobileNetV2 (Cataract + Normal eyes Dataset)	99 %	87 %
MobileNetV2 (Glaucoma + Normal eyes Dataset)	96 %	82 %

The accuracy and loss of MobileNetV1 and MobileNetV2 on both datasets, up and down through 20 epochs, are depicted in Figs. 6 and 7, respectively.

Table-9
Accuracies of MobileNetV1 and MobileNetV2 by combining Two Datasets.

Models	Training Accuracy	Testing Accuracy
MobileNetV1 (Cataract + Glaucoma + Normal eyes Dataset)	92 %	71 %
MobileNetV2 (Cataract + Glaucoma + Normal eyes Dataset)	99 %	72 %

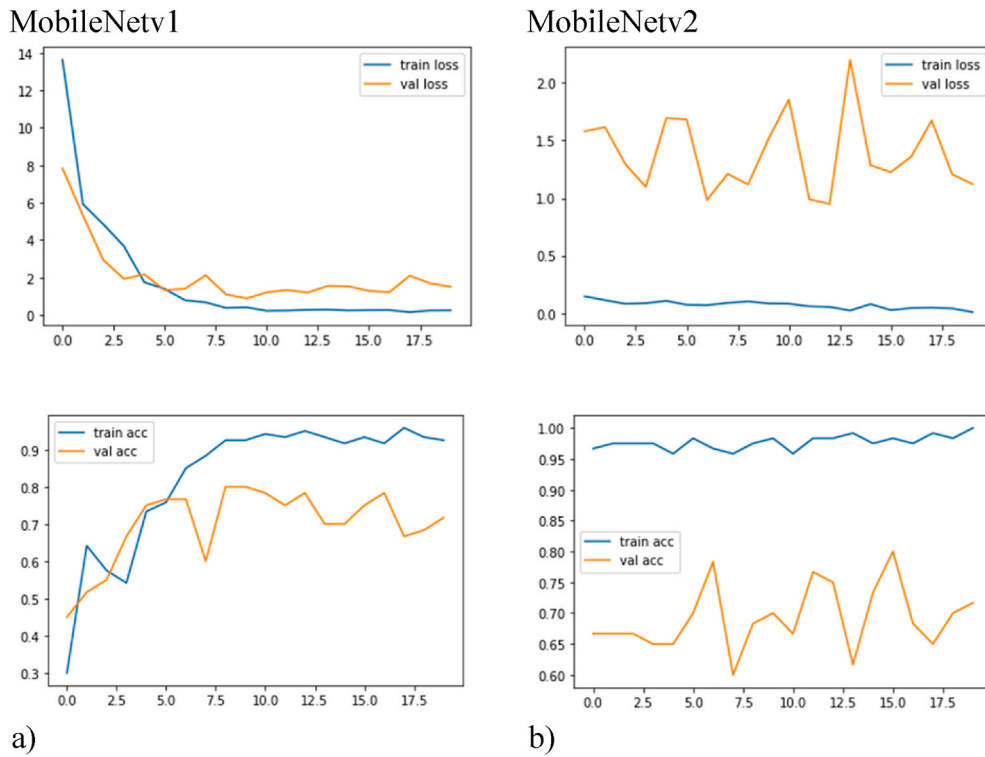


Fig-8. Loss and Accuracy on 3-classes using MobileNetV1 and MobileNetV2.

Table-10
Precision recall and F1-Score for individual and hybrid models.

Models	Classes	Precision	Recall	F1-Score
MobileNetV1 (Cataract + Normal eyes Dataset)	0	95 %	90 %	92 %
	1	90 %	95 %	93 %
	Average	93 %	93 %	92 %
MobileNetV2 (Cataract + Normal eyes Dataset)	0	86 %	95 %	90 %
	1	94 %	85 %	89 %
	Average	90 %	90 %	90 %
MobileNetV1 (Glaucoma + Normal eyes Dataset)	0	76 %	65 %	70 %
	1	70 %	80 %	74 %
	Average	73 %	72 %	72 %
MobileNetV2 (Glaucoma + Normal eyes Dataset)	0	85 %	55 %	67 %
	1	67 %	90 %	77 %
	Average	76 %	72 %	72 %
MobileNetV1 (Cataract + Glaucoma + Normal eyes Dataset)	0	83 %	95 %	88 %
	1	85 %	55 %	67 %
	2	71 %	85 %	77 %
	Average	79 %	78 %	77 %
MobileNetV2 (Cataract + Glaucoma + Normal eyes Dataset)	0	63 %	95 %	76 %
	1	86 %	30 %	44 %
	2	61 %	70 %	65 %
	Weighted Avg	70 %	65 %	62 %

glaucoma, and normal images are so similar that they cannot be distinguished. With a normal eye, cataracts and glaucoma are mostly suitable for binary classification problems. Fig. 8 depicts a graph of accuracy and loss on MobileNetV1 and MobileNetV2 over 20 epochs.

Precision, recall, and F1 score are fundamental metrics used to evaluate the performance of classification models, particularly in the context of imbalanced datasets. Precision, also known as positive predictive value, measures the accuracy of the positive predictions made by the model. It is defined as the ratio of true positive predictions to the total number of positive predictions (both true positives and false positives). In other words, precision tells us how many of the predicted positive instances were actually correct. The formula for calculating precision is given in Eq-3. Recall, also known as sensitivity or true positive rate, measures the model’s ability to correctly identify all relevant instances. It is defined as the ratio of true positive predictions to the total number of actual positive instances (both true positives and false negatives). Recall indicates how well the model captures all the actual positive cases. The formula for calculating recall is given in Eq-4. The F1 score is the harmonic mean of precision and recall, providing a single metric that balances both concerns. It is particularly useful when the dataset has an uneven class distribution, as it combines the strengths of both precision and recall into a single measure. The formula for calculating F1-score is given in Eq-5. Together, these metrics offer a comprehensive view of the model’s effectiveness in making accurate and relevant predictions. Precision, Recall and F1-Score are shown in Table-10.

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{TP}{TP + FN} \tag{4}$$

$$F - Score = 2 * \frac{P * R}{P + R} \tag{5}$$

A confusion matrix provides a visual representation of the performance of a classification model by summarizing the results of predictions versus actual values. It displays the counts of true positives, true negatives, false positives, and false negatives for each class, allowing for a detailed view of where the model is making correct and incorrect predictions. This matrix helps in evaluating the model’s accuracy, precision, recall, and F1 score by breaking down the errors in a way that reveals patterns or biases in the predictions. Visualization of the individual and hybrid models is shown in Fig-9, which makes it easier to understand the distribution of errors and the effectiveness of the models across different classes.

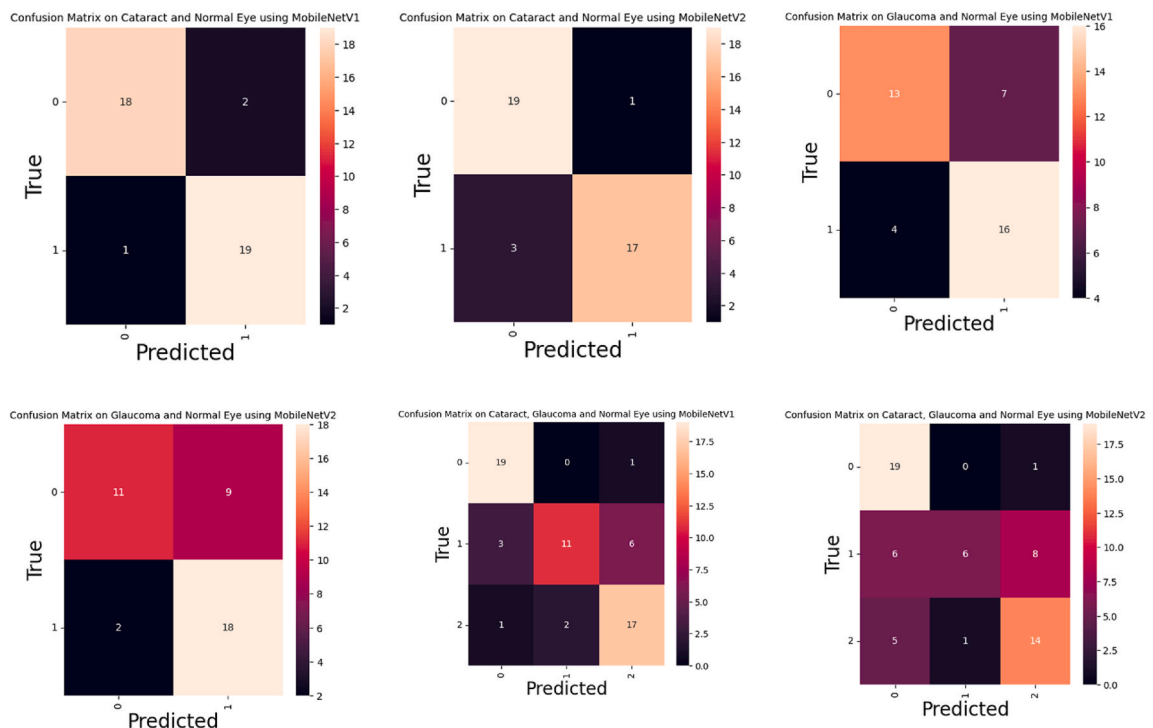


Fig-9. Visual representation of all models.

4.5. Time complexity and actual time for individual and hybrid model

The time complexity of training a machine learning model typically depends on several key factors: the number of epochs (E), the number of steps per epoch (S), the time per step (T), and the overall complexity of the model in terms of the number of parameters (P). For both the hybrid and individual models presented, the time complexity can be generally approximated as $O(E \times S \times T)$, where the duration of training is a function of how many epochs the training runs for, how many steps are in each epoch, and the time it takes to process each step.

The hybrid model's training log provides detailed time per epoch, which allows for a straightforward calculation of the total training time. Summing the times for each of the 20 epochs, the first epoch takes 33 s, while each of the subsequent epochs takes approximately 16–17 s. Specifically, the breakdown of epoch times is as follows: Epoch 1: 33 s; Epochs 2–20: 17, 17, 16, 16, 17, 16, 16, 17, 17, 17, 16, 16, 16, 16, 16, 16, 16, 17 s. Adding these times gives a total training duration for the hybrid model of 356 s, which is approximately 5.93 min. This duration includes all the training epochs and provides a clear indication of the model's training efficiency and speed.

Similarly, the individual model's training log details the time taken per epoch. The individual model also undergoes 20 epochs, with the first epoch taking 22 s and subsequent epochs taking between 10 and 20 s. The detailed breakdown of times for the individual model is: Epoch 1: 22 s; Epochs 2–20: 18, 14, 17, 15, 18, 15, 16, 15, 15, 18, 16, 20, 17, 19, 25, 16, 13, 10, 10 s. Summing these times results in a total training duration of 339 s, or approximately 5.65 min. This shorter duration compared to the hybrid model suggests a slight efficiency advantage, potentially due to a less complex model architecture or faster per-step processing.

In conclusion, the individual model demonstrates a marginally shorter total training time of 5.65 min compared to the hybrid model's 5.93 min. The hybrid model's slightly longer training time could be attributed to its potentially more intricate structure or additional steps required per epoch. Despite this, both models exhibit similar efficiency in training duration, reflecting their computational demands. The major benefits of the individual model include slightly faster training time and potentially simpler architecture, which can lead to easier implementation and troubleshooting. On the other hand, the hybrid model may offer improved performance or accuracy due to its more complex architecture, making it suitable for more demanding applications where these factors are crucial.

4.6. Performance of proposed model on another benchmark dataset

Since MobilneV2 with the Adam optimizer achieved the highest performance metrics, as shown in Table 10, the proposed model's effectiveness is further validated using an additional benchmark dataset <https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification/data>. The results from this evaluation are presented below. Fig. 10 displays the losses and accuracies for both training and testing over epochs. The figure indicates that the model successfully reduced loss and improved accuracy throughout the training process.

The proposed model demonstrated exceptional performance on the test data, achieving an accuracy of 99 % for the cataract + normal eye model, 94 % for the glaucoma + normal eye model, and 89 % for the cataract + glaucoma + normal eye model. These results highlight the model's effectiveness in distinguishing between various eye conditions. Additional performance metrics are presented in Table 11, where it can be observed that most measures exceed 95 %, underscoring the model's robust accuracy and reliability.

To facilitate a deeper understanding of the confusion matrix measures, visual representations are provided in Fig. 11. These visualizations are designed to support scientific investigations by offering a clear and detailed depiction of the model's performance across different classes. The visual aids allow for an easier interpretation of the confusion matrix, highlighting areas of strength and potential improvement in the model's predictions.

4.7. RO5: evaluating the proposed model's performance to that of the other benchmark works

The proposed work stands out in comparison to previous studies on glaucoma and cataract detection through the use of advanced deep learning models, specifically MobileNetV1 and MobileNetV2.

In the work of [19], deep convolutional neural networks were evaluated for glaucoma detection using color fundus images. This study focused on the influence of architecture, data set size, training strategy, and the integration of clinical history data. While thorough, it did not explore lightweight architectures like MobileNet, which are optimized for efficiency and performance on mobile devices.

The work of [9] proposed a novel deep learning method for cataract classification using fundus images, leveraging deep transfer learning with AlexNet. Although it utilized the first five convolutional layers of AlexNet for general feature learning, the study did not incorporate lightweight models like MobileNet, which can significantly reduce the number of parameters and improve efficiency.

The work of [18] utilizes MobileNetV2 for automated glaucoma screening, emphasizing the need for standardized methodologies and addressing challenges like heterogeneous image quality and small sample sizes. Despite its contributions to the structure of public databases, it falls short in providing simultaneous eye images or elements for early diagnosis. In contrast, the proposed work leverages both MobileNetV1 and MobileNetV2 to detect glaucoma and cataracts, optimizing architecture with depth-wise separable convolutions for efficiency and accuracy. Using publicly available datasets and data augmentation techniques, the proposed model achieves superior accuracy compared to previous models, demonstrating its effectiveness in early and accurate detection of these eye diseases. Visualization of individual and hybrid models is shown in Fig-1, reinforcing the proposed approach's robustness.

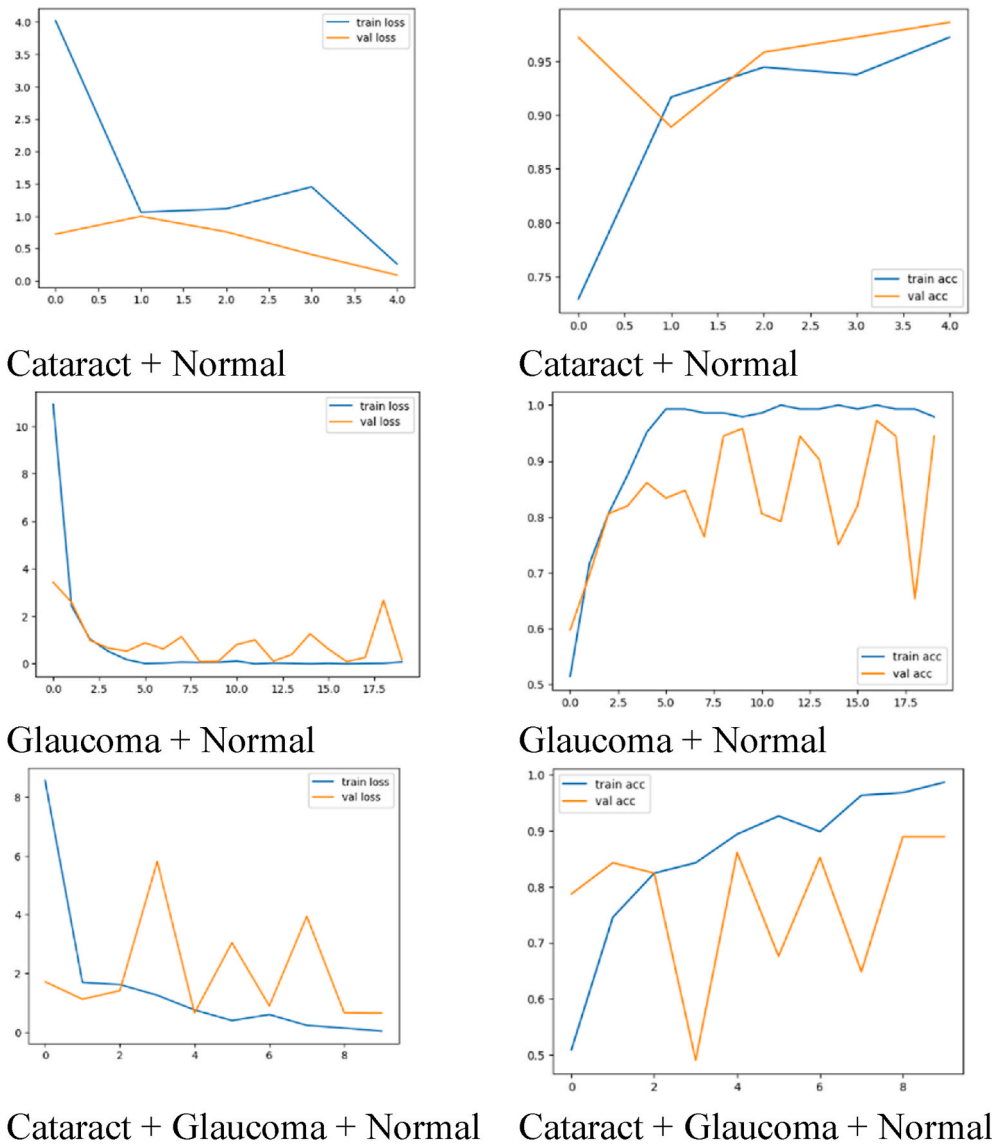


Fig-10. Loss and Accuracy on 3-classes using MobileNetV2 on Dataset-2.

Table-11

Precision recall and F1-Score for individual and hybrid models on Dataset-2.

Models	Classes	Precision	Recall	F1-Score
MobileNetV2 (Cataract + Normal eyes Dataset)	0	100 %	97 %	99 %
	1	97 %	100 %	99 %
	Average	99 %	99 %	99 %
MobileNetV2 (Glaucoma + Normal eyes Dataset)	0	100 %	89 %	94 %
	1	90 %	100 %	95 %
	Average	95 %	94 %	94 %
MobileNetV2 (Cataract + Glaucoma + Normal eyes Dataset)	0	94 %	83 %	88 %
	1	83 %	83 %	83 %
	2	90 %	100 %	95 %
	Weighted Avg	89 %	89 %	89 %

The work of [7] introduced CataractNet, a novel deep neural network for cataract detection, optimized for lower computational cost and running time. However, despite the optimization, this network did not achieve the lightweight efficiency and high accuracy of MobileNetV1 and MobileNetV2 used in the current study.

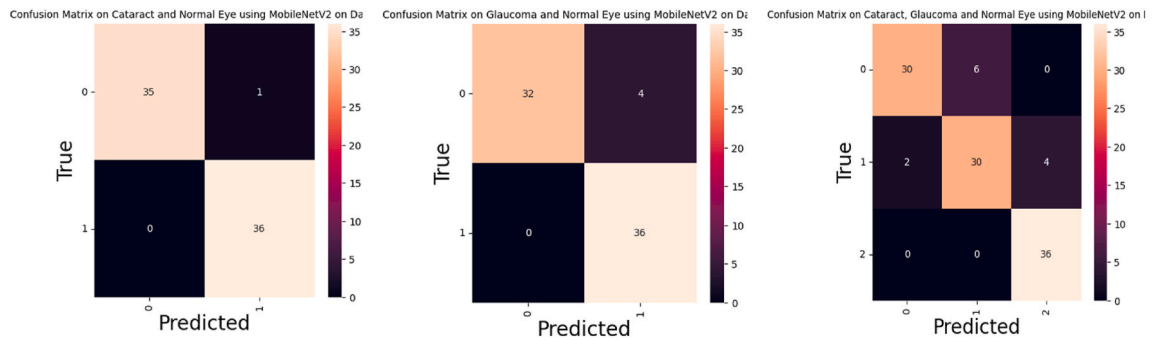


Fig-11. Visual representation of all MobileNetV2 on Dataset-2.

The proposed work addresses these limitations by using MobileNetV1 and MobileNetV2, which are designed to build lightweight deep neural networks with depth-wise separable convolutions. These models are not only efficient but also achieve high accuracy. The use of publicly available datasets for both cataract and glaucoma detection, combined with the optimized architecture, demonstrated superior performance compared to other models.

The work of [45] employs various CNN architectures, such as DenseNet and EfficientNet, to detect eye diseases, achieving notable performance but lacking in optimized efficiency. In contrast, the proposed work uses MobileNetV1 and MobileNetV2, optimized for lightweight deep learning through depth-wise separable convolutions. This approach demonstrated the highest accuracy on publicly available cataract and glaucoma datasets. By incorporating data augmentation techniques, the proposed models not only achieved superior accuracy but also ensured efficiency, making them the best solution for early and accurate eye disease detection.

The work of [46] explores various deep transfer learning models, including Basic CNN, Deep CNN, AlexNet 2, Xception, Inception V3, ResNet 50, and DenseNet121, for predicting multiple eye diseases like diabetic macular edema, choroidal neovascularization, and glaucoma using fundus digital photography and OCT images. While this approach leverages advanced architectures, the proposed work focuses on optimizing MobileNetV1 and MobileNetV2 with depth-wise separable convolutions for early and accurate detection of cataracts and glaucoma. By utilizing publicly available datasets and data augmentation techniques, the proposed models achieved superior accuracy and efficiency compared to the other models, making them the best solution for eye disease detection.

Training involved a categorical cross-entropy loss function and the Adam optimizer, with data augmentation techniques such as rescaling, shearing, and zooming to improve the robustness of the model. The results showed that the proposed model outperformed previous works in terms of accuracy, efficiency, and computational cost, making it the best choice for the detection of these eye diseases. The visualization of individual and hybrid models is shown in Fig-1, further supporting the effectiveness of the proposed approach.

The proposed method is compared to four other models. The classification performance of the four methods is shown in Table-12. It is worth noting that the experimental data for the five methods may not differ. However, the comparison is useful for analyzing the performance of the proposed method. The proposed method outperforms the other three methods, saying that our model is comparable and competitive with the other four models, as shown in Table-12.

5. Conclusion

5.1. Recapitulation

The newly developed transfer learning technique using MobileNetV1 and MobileNetV2 models was created to classify images as cataract and normal, as well as glaucoma and normal. This innovative approach included three dense layers in the head section which

Table 12
Comparison of proposed model with previous works.

Models	Accuracy
AlexNet-Adam [19]	77 % (Cataract + Normal Eye)
ALexNet [47]	92 % (Cataract + Normal Eye)
MobileNetV2 [18]	97 % (Glaucoma + Normal Eye)
CataractNet [7]	98 % (Cataract + Normal Eye)
Deep Learning [45]	89 % (Effected Eye + Normal Eye)
Deep Transfer Learning [46]	98 % (Drusen + Glaucoma + normal + cataracts)
Proposed Work (Adam)-Dataset-1	99 % (Cataract + Normal Eye)
Proposed Work (Adam)-Dataset-2	98 % (Glaucoma + Normal Eye)
	99 % (Cataract + Normal Eye)
	98 % (Glaucoma + Normal Eye)
	89 % (Cataracts Glaucoma + normal)

allowed us to demonstrate its superiority over other state-of-the-art methods like ALexNet and MobileNetV2 using Kaggle datasets. The high accuracies achieved by our models (92 % and 97 %) speak volumes of their effectiveness—especially given the lightweight nature that guarantees efficiency without compromising on adoption potential among clinical settings for eye disease detection and management: cataract or glaucoma.

5.2. Future prospects

In the future, our research will be aimed at broadening the range of eye conditions in the dataset which will help improve the strength and practicality of our model. We also plan to include real-time image processing capabilities that will ensure immediate diagnostic feedback; this feature is especially needed for clinical use. We are also considering applying innovative optimization algorithms during training so that our models can be generalized irrespective of different demographic backgrounds. This upcoming work is designed not only to enhance the detection efficacy but also to make diagnostics more personal: we would take into account peculiar physiological characteristics presented by each patient.

Data availability

On demand.

CRedit authorship contribution statement

Sheikh Muhammad Saqib: Formal analysis, Investigation, Methodology. **Muhammad Iqbal:** Data curation, Investigation, Resources. **Muhammad Zubair Asghar:** Resources, Software, Supervision. **Tehseen Mazhar:** Visualization, Writing – original draft, Writing – review & editing. **Ahmad Almogren:** Data curation, Formal analysis, Resources. **Ateeq Ur Rehman:** Investigation, Resources, Software. **Habib Hamam:** Investigation, Project administration, Resources.

Declaration of competing Interest

The authors have declared no conflict of interest.

Acknowledgment

This work was supported by King Saud University, Riyadh, Saudi Arabia, through Researchers Supporting Project number (RSP2024R184).

References

- [1] A.B. Tufail, et al., Early-stage Alzheimer's disease categorization using PET neuroimaging modality and convolutional neural networks in the 2D and 3D domains, *Sensors* 22 (12) (2022) 4609.
- [2] I. Haq, et al., A deep learning approach for the detection and counting of colon cancer cells (HT-29 cells) bunches and impurities, *PeerJ Computer Science* 9 (2023) e1651.
- [3] A. Echtioui, et al., ? covid19?> Detection Methods of COVID-19, *SLAS TECHNOLOGY: Translating Life Sciences Innovation* 25 (6) (2020) 566–572.
- [4] X.-Q. Zhang, et al., Machine learning for cataract classification/grading on ophthalmic imaging modalities: a survey, *Machine Intelligence Research* 19 (3) (2022) 184–208.
- [5] M.T. Islam, et al., Deep learning-based glaucoma detection with cropped optic cup and disc and blood vessel segmentation, *IEEE Access* 10 (2021) 2828–2841.
- [6] S. Guefrachi, A. Echtioui, H. Hamam, Automated diabetic retinopathy screening using deep learning, *Multimed. Tool. Appl.* (2024) 1–18.
- [7] M.S. Junayed, et al., CataractNet: an automated cataract detection system using deep learning for fundus images, *IEEE Access* 9 (2021) 128799–128808.
- [8] M.A. Syarifah, A. Bustamam, P.P. Tampubolon, Cataract classification based on fundus image using an optimized convolution neural network with lookahead optimizer, in: *AIP Conference Proceedings*, AIP Publishing, 2020.
- [9] S. Hu, et al., ACCV: automatic classification algorithm of cataract video based on deep learning, *Biomed. Eng. Online* 20 (2021) 1–17.
- [10] J. Rana, S.M. Galib, Cataract detection using smartphone, in: *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, IEEE, 2017.
- [11] D. Pascolini, S.P. Mariotti, Global estimates of visual impairment: 2010, *Br. J. Ophthalmol.* 96 (5) (2012) 614–618.
- [12] D. Allen, A. Vasavada, Cataract and surgery for cataract, *Bmj* 333 (7559) (2006) 128–132.
- [13] J.-J. Yang, et al., Exploiting ensemble learning for automatic cataract detection and grading, *Comput. Methods Progr. Biomed.* 124 (2016) 45–57.
- [14] P.J. Foster, et al., The definition and classification of glaucoma in prevalence surveys, *British journal of ophthalmology* 86 (2) (2002) 238–242.
- [15] R.N. Weinreb, T. Aung, F.A. Medeiros, The pathophysiology and treatment of glaucoma: a review, *JAMA* 311 (18) (2014) 1901–1911.
- [16] L. Storgaard, et al., Glaucoma clinical research: trends in treatment strategies and drug development, *Front. Med.* 8 (2021) 733080.
- [17] M. Sandler, et al., Mobilenetv2: inverted residuals and linear bottlenecks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [18] J. Camara, et al., Retinal glaucoma public datasets: what do we have and what is missing? *J. Clin. Med.* 11 (13) (2022) 3850.
- [19] J.J. Gómez-Valverde, et al., Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning, *Biomed. Opt Express* 10 (2) (2019) 892–913.
- [20] L. Guo, et al., A computer-aided healthcare system for cataract classification and grading based on fundus image analysis, *Comput. Ind.* 69 (2015) 72–80.
- [21] Y.N. Fuadah, A.W. Setiawan, T.L. Mengko, Mobile cataract detection using optimal combination of statistical texture analysis, in: *2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*, IEEE, 2015.
- [22] Y.N. Fuadah, A.W. Setiawan, T. Mengko, Performing high accuracy of the system for cataract detection using statistical texture analysis and K-Nearest Neighbor, in: *2015 International Seminar on Intelligent Technology and its Applications (ISITIA)*, IEEE, 2015.
- [23] J. Civit-Masot, et al., Dual machine-learning system to aid glaucoma diagnosis using disc and cup feature extraction, *IEEE Access* 8 (2020) 127519–127529.

- [24] M. Caixinha, et al., In-vivo automatic nuclear cataract detection and classification in an animal model by ultrasounds, *IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng.* 63 (11) (2016) 2326–2335.
- [25] V. Harini, V. Bhanumathi, Automatic cataract classification system, in: 2016 International Conference on Communication and Signal Processing (ICCSP), IEEE, 2016.
- [26] R. Sigit, E. Triyana, M. Rochmad, Cataract detection using single layer perceptron based on smartphone, in: 2019 3rd International Conference on Informatics and Computational Sciences (ICICoS), IEEE, 2019.
- [27] A. Diaz-Pinto, et al., CNNs for automatic glaucoma assessment using fundus images: an extensive validation, *Biomed. Eng. Online* 18 (2019) 1–19.
- [28] E. Uchino, et al., Classification of glomerular pathological findings using deep learning and nephrologist–AI collective intelligence approach, *Int. J. Med. Inf.* 141 (2020) 104231.
- [29] X. Gao, S. Lin, T.Y. Wong, Automatic feature learning to grade nuclear cataracts based on deep learning, *IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng.* 62 (11) (2015) 2693–2701.
- [30] L. Zhang, et al., Automatic cataract detection and grading using deep convolutional neural network, in: 2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC), IEEE, 2017.
- [31] M.R. Hossain, et al., Automatic detection of eye cataract using deep convolution neural networks (DCNNs), in: 2020 IEEE Region 10 Symposium (TENSYP), IEEE, 2020.
- [32] C. Buiu, V.-R. Dănăilă, C.N. Răduță, MobileNetV2 ensemble for cervical precancerous lesions classification, *Processes* 8 (5) (2020) 595.
- [33] Q. Xiang, et al., Fruit image classification based on Mobilenetv2 with transfer learning technique, in: Proceedings of the 3rd International Conference on Computer Science and Application Engineering, 2019.
- [34] X. Chen, et al., Recent advances and clinical applications of deep learning in medical image analysis, *Med. Image Anal.* 79 (2022) 102444.
- [35] Y. Wu, et al., Mutual consistency learning for semi-supervised medical image segmentation, *Med. Image Anal.* 81 (2022) 102530.
- [36] H.-H. Tseng, et al., Rice seedling detection in UAV images using transfer learning and machine learning, *Rem. Sens.* 14 (12) (2022) 2837.
- [37] D. Wang, et al., A review of deep learning in multiscale agricultural sensing, *Rem. Sens.* 14 (3) (2022) 559.
- [38] J.L. Rojas-Aranda, et al., Fruit classification for retail stores using deep learning, in: Pattern Recognition: 12th Mexican Conference, MCPR 2020, Morelia, Mexico, June 24–27, 2020, Proceedings, vol. 12, Springer, 2020.
- [39] N.C. Dang, M.N. Moreno-García, F. De la Prieta, Sentiment analysis based on deep learning: a comparative study, *Electronics* 9 (3) (2020) 483.
- [40] H. Luo, N. Liu, C. Feng, Question and answer classification with deep contextualized transformer, in: Advances in Information and Communication: Proceedings of the 2021 Future of Information and Communication Conference (FICC), vol. 2, Springer, 2021.
- [41] S. Fekri-Ershad, Bark texture classification using improved local ternary patterns and multilayer neural network, *Expert Syst. Appl.* 158 (2020) 113509.
- [42] Q. Yaseen, Spam email detection using deep learning techniques, *Procedia Computer Science* 184 (2021) 853–858.
- [43] G. Shahariar, et al., Spam review detection using deep learning, in: 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE, 2019.
- [44] A.G. Howard, et al., Mobilenets: efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861 (2017).
- [45] G. Arslan, Ç.B. Erdaş, Detection of cataract, diabetic retinopathy and glaucoma eye diseases with deep learning approach, *Intelligent Methods In Engineering Sciences* 2 (2) (2023) 42–47.
- [46] Y. Kumar, S. Gupta, Deep transfer learning approaches to predict glaucoma, cataract, choroidal neovascularization, diabetic macular edema, drusen and healthy eyes: an experimental review, *Arch. Comput. Methods Eng.* 30 (1) (2023) 521–541.
- [47] H. Zhang, et al., Automatic cataract grading methods based on deep learning, *Comput. Methods Progr. Biomed.* 182 (2019) 104978.