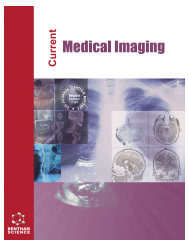
BENTHAM SCIENCE

CrossMark

Current Medical Imaging

Content list available at: https://benthamscience.com/journals/cmir



RESEARCH ARTICLE

An Artificial Intelligence Driven Approach for Classification of Ophthalmic Images using Convolutional Neural Network: An Experimental Study

Shagundeep Singh[1], Raphael Banoub[4], Harshal A. Sanghvi[1,2,5,*], Ankur Agarwal[1], K.V. Chalam[3,5], Shailesh Gupta[4,5] and Abhijit S. Pandya[1,5]

[1]*Department of CEECS, Florida Atlantic University, FL, USA*
[2]*Charles E. Schmidt College of Medicine, Florida Atlantic University, FL, USA*
[3]*Department of Ophthalmology, Loma Linda University, CA, USA*
[4]*Department of Ophthalmology, Broward Health, FL, USA*
[5]*Department of Technology and Clinical Trials, Advanced Research, FL, USA*

Abstract:

Background:

Early disease detection is emphasized within ophthalmology now more than ever, and as a result, clinicians and innovators turn to deep learning to expedite accurate diagnosis and mitigate treatment delay. Efforts concentrate on the creation of deep learning systems that analyze clinical image data to detect disease-specific features with maximum sensitivity. Moreover, these systems hold promise of early accurate diagnosis and treatment of patients with common progressive diseases. DenseNet, ResNet, and VGG-16 are among a few of the deep learning Convolutional Neural Network (CNN) algorithms that have been introduced and are being investigated for potential application within ophthalmology.

Methods:

In this study, the authors sought to create and evaluate a novel ensembled deep learning CNN model that analyzes a dataset of shuffled retinal color fundus images (RCFIs) from eyes with various ocular disease features (cataract, glaucoma, diabetic retinopathy). Our aim was to determine (1) the relative performance of our finalized model in classifying RCFIs according to disease and (2) the diagnostic potential of the finalized model to serve as a screening test for specific diseases (cataract, glaucoma, diabetic retinopathy) upon presentation of RCFIs with diverse disease manifestations.

Results:

We found adding convolutional layers to an existing VGG-16 model, which was named as a proposed model in this article that, resulted in significantly increased performance with 98% accuracy ($p<0.05$), including good diagnostic potential for binary disease detection in cataract, glaucoma, diabetic retinopathy.

Conclusion:

The proposed model was found to be suitable and accurate for a decision support system in Ophthalmology Clinical Framework.

Keywords: Artificial intelligence, Deep learning, Ocular disease, Detection, Diagnosis, Algorithms.

Article History	Received: December 01, 2023	Revised: February 27, 2024	Accepted: March 11, 2024

1. INTRODUCTION

Artificial intelligence (AI) is a rapidly growing field that aims to develop computer systems and algorithms capable of thinking, learning, and making decisions reminiscent of human intelligence [1 - 3]. Anisotropic angle distribution learning for head poses estimation and attention understanding in human-computer interaction. Numerous studies have been conducted on the utilization of deep learning systems and computer vision [1 - 3]. AI is increasingly transforming healthcare and allowing medical practitioners to make more informed and complex medical decisions, detect and diagnose diseases, create appropriate treatment plans, and minimize avoidable errors that could harm patients [2]. Within medicine, deep learning has recently seen an upsurge in its application, particularly in fields such as oncology, neurology, and cardiology, where available

* Address correspondence to this author at the Department of CEECS, Florida Atlantic University, FL, USA; E-mail: harshaPsanghvi7@gmail.com

electronic MRI, CT, and PET radiographic image data can be leveraged for mass screening and to mitigate morbidity related to delayed diagnosis of devastating diseases [3, 4]. One area of medicine where AI has been particularly impactful is ophthalmology [5, 6]. This focus is due in part to the reliance on cross-sectional tomography and RCFIs, as these provide magnified high-resolution images and documentation of ophthalmoscopic examination findings within electronic medical record systems [5 - 8]. AI systems in ophthalmology cover a wide range of areas, from the diagnosis of external eyelid lesions to the interpretation of congenital and acquired retinal diseases [7]. Of particular interest is the neurosensory retina, which is located within the posterior pole of the eye, detects light from the outside world, and transmits visual information to the brain *via* the optic nerve. The retina is histologically complex, containing photoreceptors, glial cells, pigment cells, neurons, and arterial and venous vascular systems [8]. Due to its location and prerequisite requirement for normal function permissive of intact visual acuity, the retina is highly susceptible to an exceedingly great number of congenital dystrophic, metabolic, vasculopathic, anemic, infectious, neuroinflammatory, neuro-atrophic, oncologic, and other retinal diseases that may affect one or more histologic structures or functions within the neurosensory retina resulting in partial or complete loss of vision.

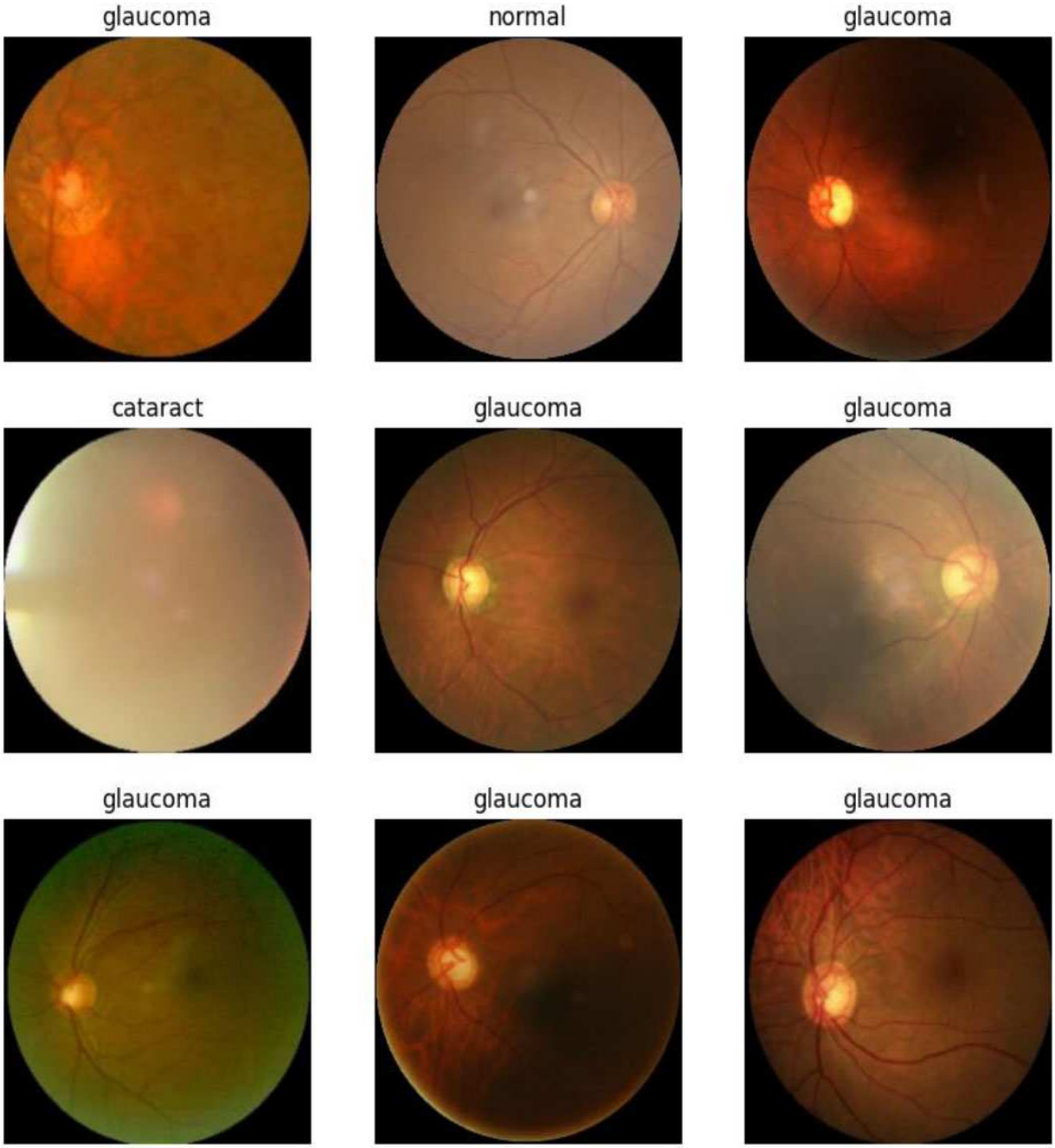


Fig. (1). Fundus images from the dataset [31].

Fig. (**1**) indicates the fundus images from the dataset and Fig. (**2**) indicates the heatmap of the sample fundus image from the dataset. This study focused on Type 2 Diabetes Mellitus (T2DM), a chronic metabolic disease that leads to chronically elevated blood sugar levels and can cause damage to various organs and systems in the body, including the blood vessels, eyes, nerves, heart, kidneys, and brain. This damage can lead to severe complications, such as hypertension, retinopathy, peripheral neuropathy, myocardial infarction, heart failure, chronic kidney disease, end-stage renal failure, stroke, and even death [9, 10]. The healthcare costs associated with T2DM are estimated to be approximately USD 327 billion [11, 12], with projections showing this figure will continue to rise as the prevalence of T2DM increases. It is estimated that within the next five years, the number of diabetic patients developing diabetic retinopathy will increase by 33% [6]. Patients with T2DM are also at an increased risk of developing cataracts, glaucoma, and diabetic retinopathy with or without macular edema [6]. Previous studies have shown that Deep Learning (DL) systems can accurately and precisely detect referable diabetic retinopathy from RCFIs [3, 6, 9]. However, few studies to date have created a model capable of classifying multiple ocular diseases with high accuracy. In this study, we sought to create and evaluate a deep learning CNN model that analyzes a dataset of shuffled retinalcolor fundus images (RCFIs) from eyes with various ocular disease features (cataract, glaucoma, diabetic retinopathy) or control. In the proposed framework, we determine the relative performance of our finalized deep learning model in classifying randomly shuffled RCFIs from eyes with various ocular diseases (cataract, glaucoma, diabetic retinopathy) or control, when compared with existing models. Endpoints were accuracy (Acc), precision (Prc), recall (Rcl), and F1-score. Then, we determined the relative diagnostic potential of a finalized deep learning model to serve as a binary screening test for the preliminary diagnosis of each disease: cataract, glaucoma, or diabetic retinopathy. After training, the dataset on randomly shuffled RCFIs containing each of the included disease states as well as normal images. The endpoints were sensitivity (Sn) and specificity (Sp). This paper is divided into five sections.

- Section II. Related Works
- Section III. Methodology and Framework
- Section IV. Results
- Section V. Discussion and Future Work

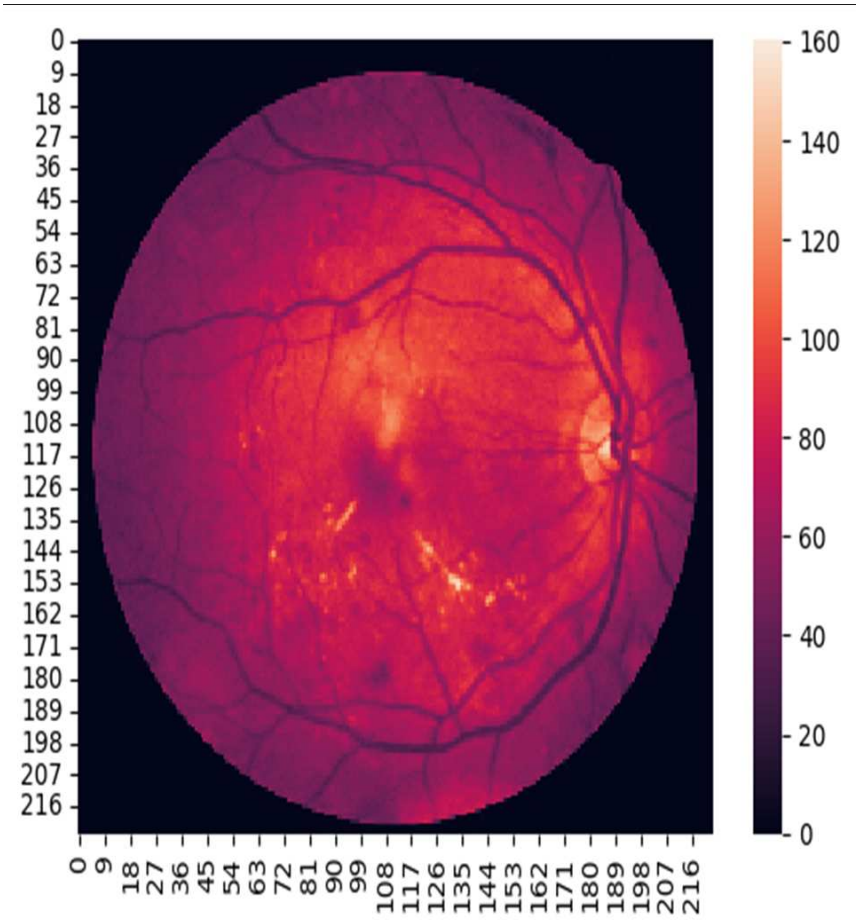


Fig. (2). Heat map of sample fundus image from the dataset [31].

2. RELATED WORKS

Numerous studies have been conducted on the utilization of deep learning systems and computer vision in the detection and diagnosis of important diseases within healthcare [12, 13]. Kaya *et al*. demonstrated that a CNN (AlexNet) could detect and distinguish hemorrhagic and ischemic stroke based on non-contrast CT image films, proposing its implementation for preliminary treatment stroke management plans to improve patient outcomes [14]. Singh *et al*. used CNN on CT scan images to detect pulmonary infiltration and differentiate pneumonia from viral and bacterial etiologies, specifically detecting pneumonia related to the SARS-CoV-2 virus (COVID-19) [15]. Seth *et al*. reviewed various literature for the detection of breast cancer using MRI scans with a CNN [16]. In ophthalmology, deep learning has been shown in numerous studies to be an effective method for developing automated systems. These systems detect and classify ocular diseases such as diabetic retinopathy and glaucoma from RCFIs with high accuracy (87.9% - 96.8%) using ML support vector machines, k-nearest neighbors, and various architectures of CNN such as VGG-16, ResNet, DenseNet, Inception, and AlexNet [4 - 6, 17 - 21].

In early studies on deep learning convolutional neural networks, Abràmoff *et al*. and Kose *et al*. conducted similar experiments where they trained a CNN on 120,000 and 10,000 pre-labeled RCFIs to detect diabetic retinopathy with a sensitivity of 99% and an accuracy of 94% [22, 23]. Leong *et al*. analyzed AI applications in Ophthalmology, identifying 1,762 publications on its potential benefits. They also found 70 review articles, with 24 each on DR and glaucoma detection, and the rest on OCT and retinal CFI features [18]. Lee *et al*. trained a DL CNN on 1,289 OCT macula images to detect IRF and diagnose macular edema causes, achieving a Dice coefficient of 0.911 compared to clinical professional assessments [19]. In a separate study, Gulshan *et al*. evaluated the performance of a pre-existing CNN (ResNet) on 128,000 RCFIs for the detection of both diabetic retinopathy and diabetic macular edema and found ResNet was able to classify diabetic retinopathy with macular edema, diabetic retinopathy without macular edema, and control with an accuracy of 96% [24]. Sarki *et al*. used two pretrained CNN models on ImageNet for mild multiclass diabetic eye diseases and multiclass diabetic eye diseases and employed performance enhancement techniques on VGG16 to obtain an accuracy of 88.3% [25]. Lee *et al*. trained a CNN on 1,289 OCT macula images (1,919,680 macula sections) to detect intraretinal fluid (IRF) and diagnose macular edema etiologies, achieving a maximal cross-validation Dice coefficient of 0.911 when compared to the diagnostic assessments of clinical professionals [26]. Nazir *et al*. used Fast Region-based CNN to detect diabetic retinopathy in RCFIs. CNN was trained on a dataset of over 2,500 images and was able to accurately classify images as normal or abnormal with a sensitivity of 92% and a specificity of 91% [27]. In addition, Luo *et al*. created and trained a CNN with the base EfficientNet-B3 to classify 1,973 RCFIs to differentiate normal retina from diabetic retinopathy and determined the relative performance of their model *versus* a baseline model. They found their CNN to have approximately 10% higher accuracy compared to the previous model (89.6% *vs* 80.0%). “In the pursuit of advanced diagnostic techniques for eye diseases, Quellec *et al*. introduced a transfer learning method for DR detection in retinal images. Their model, trained on an extensive set of fundus color images, attained an impressive accuracy of 94.9% in identifying DR [28].

Deep learning has been widely applied in ophthalmology, one such application is in the detection of early and or occult glaucoma with CNN architectures [21]. Milea *et al*. used pre-existing AI to detect and classify the presence of optic nerve edema due to increased intracranial pressure (papilledema) and other optic disc abnormalities from 1505 ocular fundus photographs in external testing and achieved an AUC of 96%, sensitivity of 96.4%, and specificity of 84.7% for papilledema. Quellec *et al*. proposed a transfer learning approach for detecting diabetic retinopathy in RCFIs, where the model was trained on a large dataset of RCFIs and achieved an accuracy of 94.9%. Bhaskaranand *et al*. proposed a deep learning model for the detection of diabetic retinopathy from RCFIs, trained on a large dataset of over 40,000 images, and achieved a high sensitivity of 90% and specificity of 63.2% for detecting diabetic retinopathy [29]. In 2020, Gargeya and Leng created a retinal diabetic retinopathy detection deep learning model that achieved 94% sensitivity after training on 120,000 pictures [30]. In the rapidly evolving field of ophthalmic technology, Grzybowsky *et al*. delved into the implications of AI. They explored the pros and cons of AI in Ophthalmology, including concerns about data quality and privacy. The authors stressed the importance of clear regulations to ensure patient safety and ethical AI application [20].

3. METHODS

3.1. Dataset

Training of the proposed model and existing CNN models first began by obtaining a dataset of 4,217 RCFIs (1038 cataract, 1098 diabetic retinopathy, 1007 glaucoma, and 1074 control images) from the open-source database “eye_diseases_classification” https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification [31].

This dataset contains quality images. Since all the images are the same color but different in shade, the images were converted to Grayscale. This process reduces training time, facilitates internal computation, and increases accuracy.

3.2. Methods

3.2.1. Baseline Methods

Each 512 x 512-pixel RGB image from the RCFI dataset was rescaled to 224 x 224 pixels. These rescaled images were batched with a batch size of 64, resulting in 66 total batches. The batches were then distributed for training (70%), validation (20%), and testing (10%) to create an ImageNet dataset containing the RCFI images. Pixels from each RCFI image were fed into the input layer for the training of all the deep learning convolutional neural network (CNN) models.

During training, the model was fed labelled images consisting of input images and corresponding class labels so

that it could learn the optimal weights for all trainable layers. The model adjusted its weights through backpropagation and gradient descent, aiming to minimize the loss function and improve accuracy on the training data.

Once training was complete, the model was used to make predictions on new, unseen input images. The model took each input image, processed it through the layers of the network, and produced class probabilities as the output. The predicted class was selected as the one with the highest probability. Thus, the model produced a predicted class for each new input image fed into it.

3.2.2. Transfer Learning Methods

Transfer learning models were used as a means of comparison during the determination of our proposed model performance. Transfer learning models VGG16, ResNet, and DenseNet were selected, based on factors of increased relative performance compared to other models and to leverage each model's large volume of pre-existing model gain through extensive pre-training on expansive datasets. Our intention during the selection of the transfer learning model was to organize a competitive comparison between existing models' and our model's performance. To conduct a proper comparison of performance between transfer learning models that were created for routine use and our model, we first needed to align the transfer learning models with our specifically defined task. Fine-tuning of each transfer learning model, VGG16, ResNet, and DenseNet, was hence achieved through addition of one Flatten Layer, one Dense Output Layer, and several Custom Layers (as outlined below). The addition of the custom layers was permitted prior to performance evaluation of each of the transfer learning models to elucidate higher-order details and nuances in the performance of our created model.

VGG16 contained 16 Depth Layers that were generated using ImageNet transfer learning weights. To fine-tune our model for analysis of RCFIs, seven custom layers were added to VGG16: 1 Flatten, 2 Dense, 2 Dropout, 1 Batch Normalization, 1 Dense Output. The total number of parameters after VGG16 model preparation was 138,357,544.

ResNet contains 50 Depth Layers that were generated using ImageNet transfer learning weights. To fine-tune our model for analysis of RCFIs, seven custom layers were added to ResNet: 1 Flatten, 1 Dropout, 1 Batch Normalization, and 1 Dense Output. The total number of parameters after DenseNet model preparation was 25,642,714. DenseNet contained 121 Depth Layers that were generated using ImageNet transfer learning weights. To fine-tune our model for analysis of RCFIs, seven custom layers were added to DenseNet: 1 Flatten, 1 Dropout, 1 Batch Normalization, and 1 Dense Output. The total number of parameters after DenseNet model preparation was 8,068,506.

3.3. The Proposed Convolutional Neural Network

The proposed convolutional neural network model was built upon the foundation of the pre-existing VGG16 transfer learning model. Additional custom layers were added in the following manner:

First, the output from the VGG16 base model was passed through a series of custom layers stacked on top. A Flatten layer flattened the 3D tensor output from VGG16 into a 1D tensor. This flattened tensor was fed into a fully connected layer with 512 units and Rectified Linear Unit (ReLU) activation function. Next, a Dropout layer with a 0.3 dropout rate was applied, where 30% of the units were randomly set to 0 during training to prevent overfitting. Batch normalization was then used to normalize the activations from the previous layer across each batch to stabilize learning. After that, another fully connected layer with 128 units and ReLU activation was added. This was followed by another Dropout layer with a 0.3 dropout rate. For multi-class classification, a final fully connected layer with 4 units and Softmax activation was added, representing the number of classes. For binary classification, the final layer had 2 units and Softmax activation. In both cases, the Softmax activation function produced output probabilities for each class. The output from the last dense layer was the predicted probabilities for each class. The model outputs a probability distribution over the classes for each input image. Finally, the convolutional neural network model was ensembled with the LeNet5 model to complete the proposed architecture.

4. RESULTS

4.1. Evaluation Metrics

Performance was determined using standard computational metrics accuracy (Acc), precision (Prc), recall (Rcl) and F1-score (F1) for each CNN model based on the correct classification of shuffled RCFIs into 'cataract', 'glaucoma', 'diabetic retinopathy', or 'control' bins (Fig. **3**) [30, 31].

$$Acc = (TPc + TNc)/ (TPc +FPc+TNc+FNc)$$

$$Prc =TPc / (TPc+FPc)$$

$$Rcl = TPc / (TPc+FNc)$$

$$F1 = 2*(Prc*Rcl)/(Prc+Rcl)$$

[Classification] TPc: True Positive, TNc: True Negative, FPc: False Positive, FNc: False Negative.

Classification true positive (TP_c) and true negative (TN_c) scores were contingent on the model placing RCFIs into the correct of three disease classes (cataract, glaucoma, diabetic retinopathy) correctly identifying control. Hence, TP_c events were only counted when an RCFI from an eye with an ocular disease was presented and classified under the correct ocular disease. Given three disease classes, one TP_c rate was determined for each ocular disease. Similarly, TN_c events were counted when an RCFI from an eye with control features was presented correct non-detection of all three ocular diseases was achieved, thereby conferring correct identification of 'control.' Given that TN_c events only occurred after ruling out all ocular diseases, only one TN_c rate was calculated in each model's performance analysis.

Fig. (**3**) again is the matrix which shows the baseline approach for the proposed model. The counting of classification false positive (FP_c) and false negative (FN_c) involved consideration of whether an error was absolute

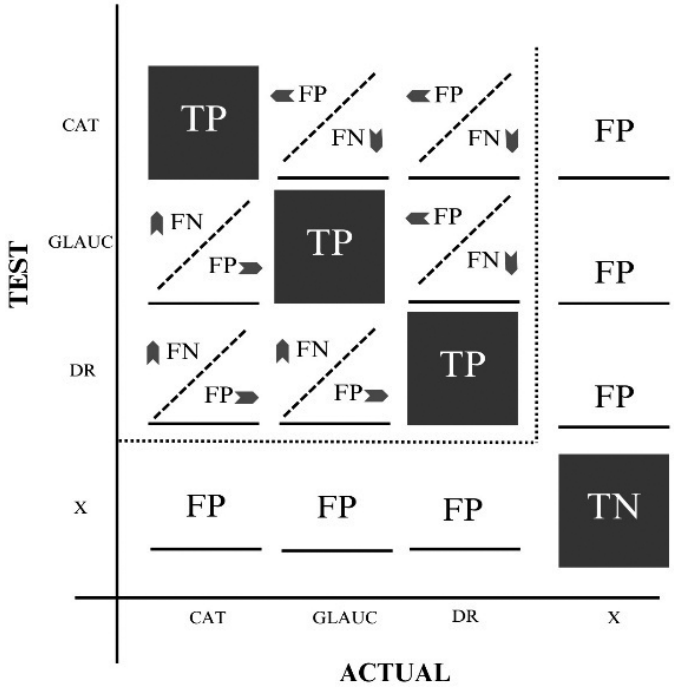


Fig. (3). Matrix showing the base approach for the proposed model.

(against normal control) or relative (against another disease). Absolute FN_c events occurred when an RCFI from an eye with any ocular disease was incorrectly classified as control due to the failed detection of one and correct non-detection of two ocular diseases. Analogously, relative FPc events occurred when an RCFI from a control eye was incorrectly classified into the bin for one ocular disease due to an invalid detection of that disease but correct non-detection of the other two ocular disease state classes. Thus, one absolute FNc and one absolute FPc were calculated for each ocular disease. For each event involving an RCFI from an eye with one ocular disease that was incorrectly classified into the bin for another ocular disease, both a relative FN_c and a relative FP_c event resulted due to the failed and invalid detections, respectively. These six relative errors represented every combination pair for three ocular disease classes, each with two possible false positive and two possible false negative relative errors. An analysis of variance (ANOVA) was performed for CNN models' accuracy, precision, recall, F1-score to determine whether significant differences in model performance exist.

4.2. Sparse Categorical Cross Entropy Loss Function

Given the dataset multiclass, sparse categorical cross entropy [30] loss function was deployed and defined where t_i is the truth label, and $_i$ was the probability, which was the output of SoftMax function for i^{th} class.

$$L_{CE} = -\sum_{i=1}^{n} t_i \log(p_i), for\ n\ classes$$

The cross-entropy loss function was used to predict the output probability with the actual label, which penalizes the model based on the deviation of models' categorization of disease state compared with RCFI dataset's diagnosis label for that image. In the sparse categorical cross-entropy loss function, one-hot encoding of labels was not required, as it was in traditional categorical cross entropy [refer-CCE]. The labels were marked as integers (1,2,3). Data augmentation was avoided in each algorithm for this study, as the data included high-resolution fundus color images. The detection model was trained on Google Collaboratory's default RAM runtime and was implemented using Python.

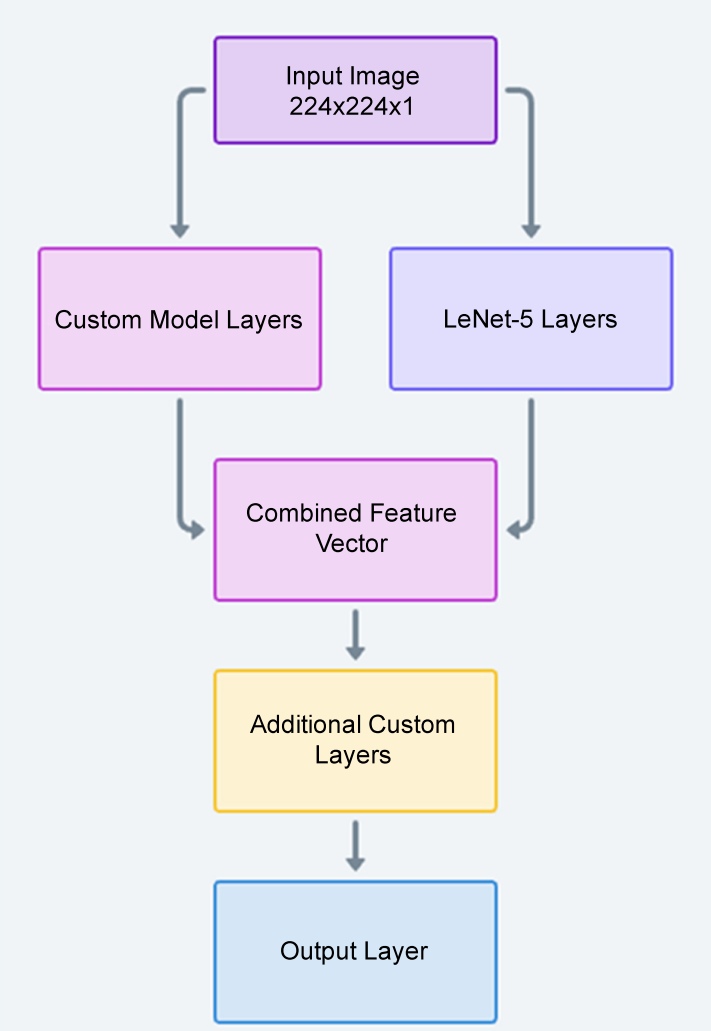
To enhance the process of disease diagnosis in our model, we used two loss functions. The Sparse Categorical Cross Entropy loss function was utilized in our deep learning model's initial stage. This function was extremely useful for classification jobs where the classes are mutually exclusive. In our example, the model was attempting to categorize the input into one of four classes, one of which denotes normal eyes. The difference between the true distribution and the anticipated probability distribution was quantified using sparse categorical cross-entropy, with an emphasis on the right class. Given that this model worked well in situations with numerous class labels and was more computationally efficient when dealing with sparse labels, it was an advantageous option in this situation.

4.3. Binary Cross Entropy Loss Function

Given the dataset multiclass, binary cross entropy [30] loss function was deployed and defined where t_i was the truth label and p_i was the probability, which was the output of SoftMax

function for i_{th} class. Binary cross-entropy loss function maps were used to predict the output probability with the true label and penalize in logarithmic function based on how far the output from the actual label was. In sparse categorical cross-entropy loss function, one-hot encoding of the labels is like in traditional categorical cross entropy [refer-CCE]. Labels that can be marked as integers (1,2,3) were not required. Data augmentation was avoided in each algorithm for this study, as the data included high-resolution fundus color images. The detection model was trained on Google Collaboratory's default RAM runtime and was implemented using Python. To determine whether a significant difference in model diagnostic potential exists, an analysis of variance (ANOVA) was performed for CNN models' sensitivity and specificity.

The model then used a series of binary classifications to identify a specific disease state and distinguish it from healthy eyes in the second stage of the procedure. The Binary Cross-Entropy loss function was used for these jobs. Binary classification tasks use the binary cross-entropy loss function, sometimes referred to as log loss. The cross-entropy loss between true labels and anticipated probability was calculated using this loss function. In essence, it assessed how well a classification model performs given that its output is a probability value between 0 and 1. The loss is -log(p) when the true label is 1.plus, -log(1-p) when the true label is 0, where p is the predicted probability. Therefore, the loss grew as the projected probability departed from the actual label. Given that the difference was calculated between the model's predictions and the actual data in situations when there are two class labels, this loss function was a good option for binary classification issues. The Binary Cross-Entropy function measures the difference between the projected likelihood and the actual label, offering a reliable technique to enhance the ability of our model to identify the presence or absence of disease. The model achieved high accuracy in this key binary decision-making phase of the diagnostic by using this function.



Flowchart-1 Architecture of the proposed model

All CNN models converged to the 100th epoch, except for the VGG-16 model, which only converged to the 47th epoch before failure. Hence, the VGG-16 model was excluded from all statistical comparisons. Overall, the proposed model served as a better diagnostic screening test for preliminary diagnosis of all ocular diseases using RCFIs when compared to existing models after dedicated disease training [Sn/Sp] (cataract: 0.9953/0.9769, glaucoma: 0.9897/0.9374, diabetic retinopathy: 1.000/0.9098) (P-VALUE). The sensitivity and specificity for existing models were: DenseNet (cataract: 0.96/0.96, glaucoma: 0.96/0.95, diabetic retinopathy: 0.91/0.91), ResNet (cataract: 0.9/0.85, glaucoma: 0.49/0.72, diabetic retinopathy: 0.31/0.51). In general, there were neither significant differences in true positive or negative rates nor false positive or negative rates between models.

4.4. Confusion Matrix

The following section describes the results obtained in terms of Confusion Matrix, Accuracy *vs* Epoch Curves, Loss *vs* Epoch Curves, and calculations of various parameters. Figures describe the confusion matrix, which states the classification as Cataract *vs* Non-Cataract, Glaucoma *vs* Non-Glaucoma, Diabetic Retinopathy *vs* Non-Diabetic Retinopathy for the proposed model, DenseNet Model and ResNet Model. Figs. (**4 - 15**) denote Accuracy *vs* Epoch, Loss *vs* Epoch Curves. Tables **1 - 6** show the calculation of various parameters for the Proposed Model, DenseNet Model, and ResNet Model.

4.5. Proposed Model

4.5.1. Architecture

Flowchart 1 illustrates the outlined architecture, and how an embedded neural network operates by taking an input image and processing it through two separate streams: one comprising the custom model layers and the other consisting of LeNet-5 layers. The outputs of these streams are then combined into a single feature vector, which is further processed by additional custom layers, culminating in the final output layer.

Fig. (**7**) represents the Accuracy *vs* Epoch and Loss *vs* Epoch Curves for the proposed model.

Table 1. Calculation of parameters for the proposed model.

Accuracy	**0.980**		
Training	0.993		
Validation	0.982		
-	CAT	GL	DR
Precision	0.940	0.890	0.980
Recall	0.960	0.860	0.910
F1 Score	0.950	0.870	0.940

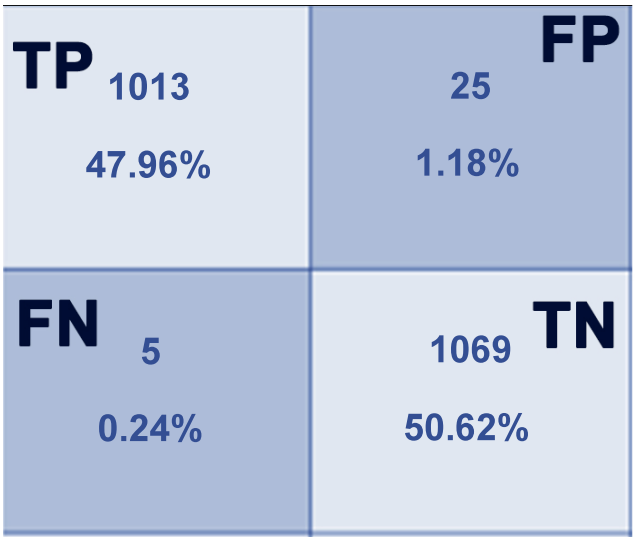


Fig. (4). Confusion matrix (cataract *vs* on cataract).
Fig. (**4**) displays the statistics of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values.
TP: 1013 cases were correctly classified as having cataracts.
FP: 25 cases were incorrectly classified as having cataracts when they did not.
TN: 1069 cases were correctly classified as not having cataracts.
FN: 5 cases were incorrectly classified as not having cataracts when they did.

Table 2. Calculation of various parameters for densenet model.

Accuracy	**0.790**		
Training	0.836		
Validation	0.792		
-	CAT	GL	DR
Precision	0.850	0.740	0.90
Recall	0.890	0.710	0.740
F1 Score	0.870	0.720	0.810

Table 3. Calculation of various parameters for resnet model.

Accuracy	**0.630**		
Training	0.663		
Validation	*0.635*		
-	CAT	GL	DR
Precision	0.830	0.760	0.840
Recall	0.780	0.710	0.690
F1 Score	0.800	0.730	0.750

Table 4. Calculation of parameters for proposed model.

Ocular Condition	**TP** **TP + FN**	**Sn**	**TN** **TN + FP**	**Sp**
Cataract	1013 1013+ 5	0.995	1069 1069+25	0.977
Glaucoma	1063 1063+11	0.989	944 944+63	0.937
Diabetic Retinopathy	1074 1074 + 0	1.000	999 999 + 99	0.909

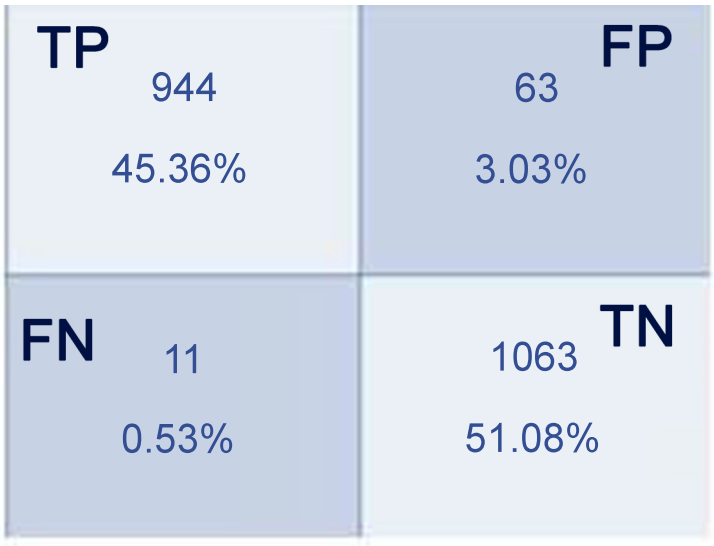


Fig. (5). Confusion matrix (glaucoma *vs* non-glaucoma).
Illustrates the statistics of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values.
TP: 944 cases were correctly classified as having glaucoma.
FP: 63 cases were incorrectly classified as having glaucoma when they did not.
TN: 1063 cases were correctly classified as not having glaucoma.
FN: 11 cases were incorrectly classified as not having glaucoma when they did.

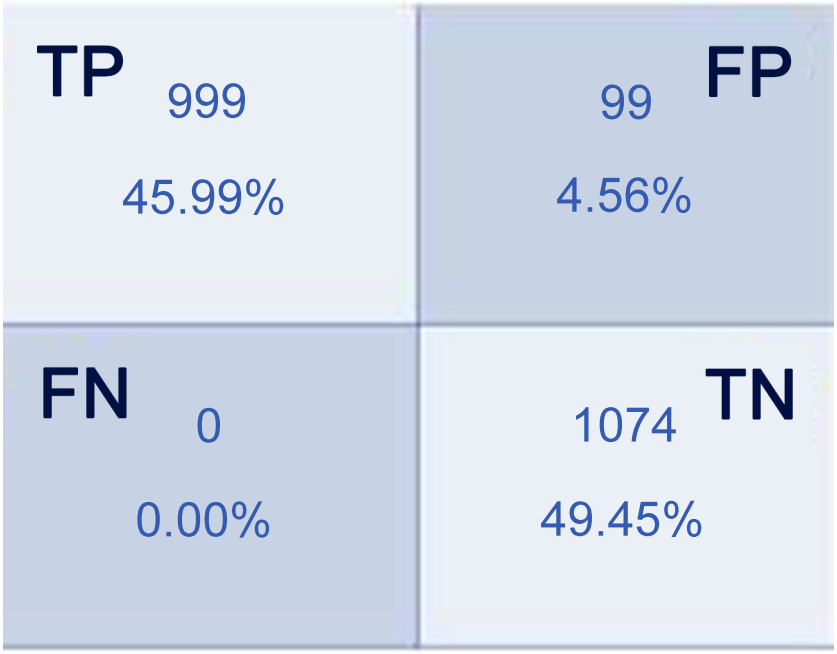


Fig. (6). Confusion matrix (diabetic retinopathy *vs* non-diabetic retinopathy).
Illustrates displays the statistics of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values.
TP: 999 cases were correctly classified as having diabetic retinopathy.
FP: 99 cases were incorrectly classified as having diabetic retinopathy when they did not.
TN: 1074 cases were correctly classified as not having diabetic retinopathy.
FN: 0 cases were incorrectly classified as not having diabetic retinopathy when they did.

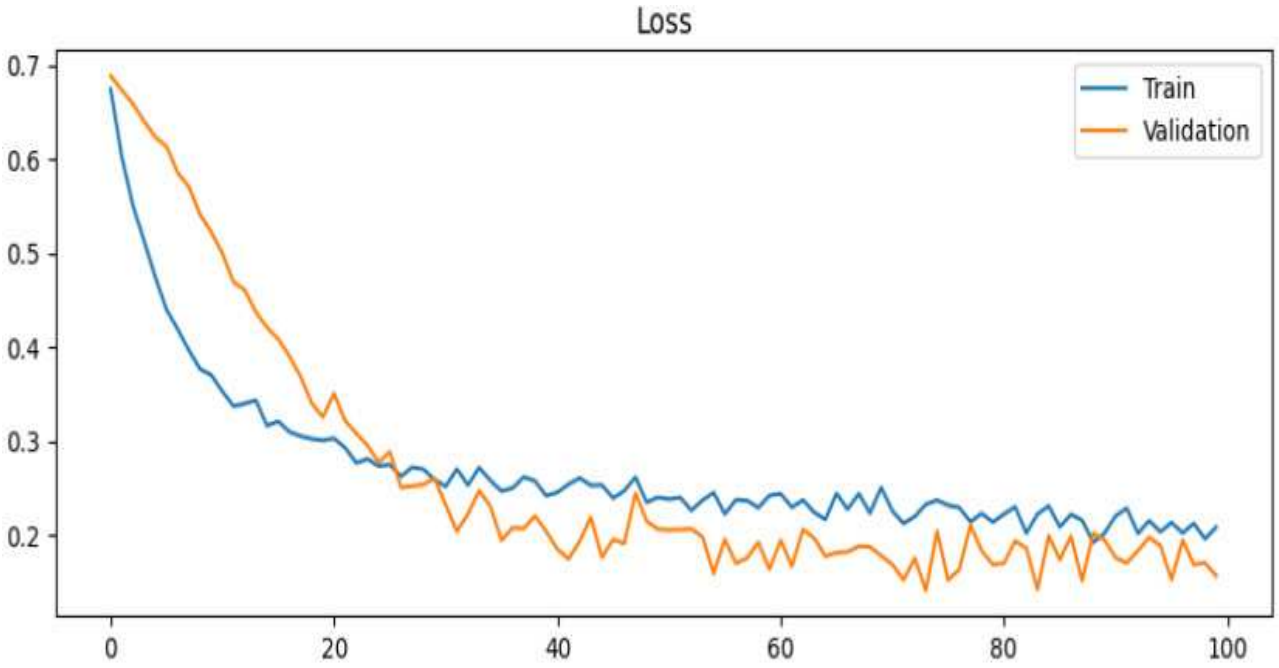


Fig. 7 contd.....

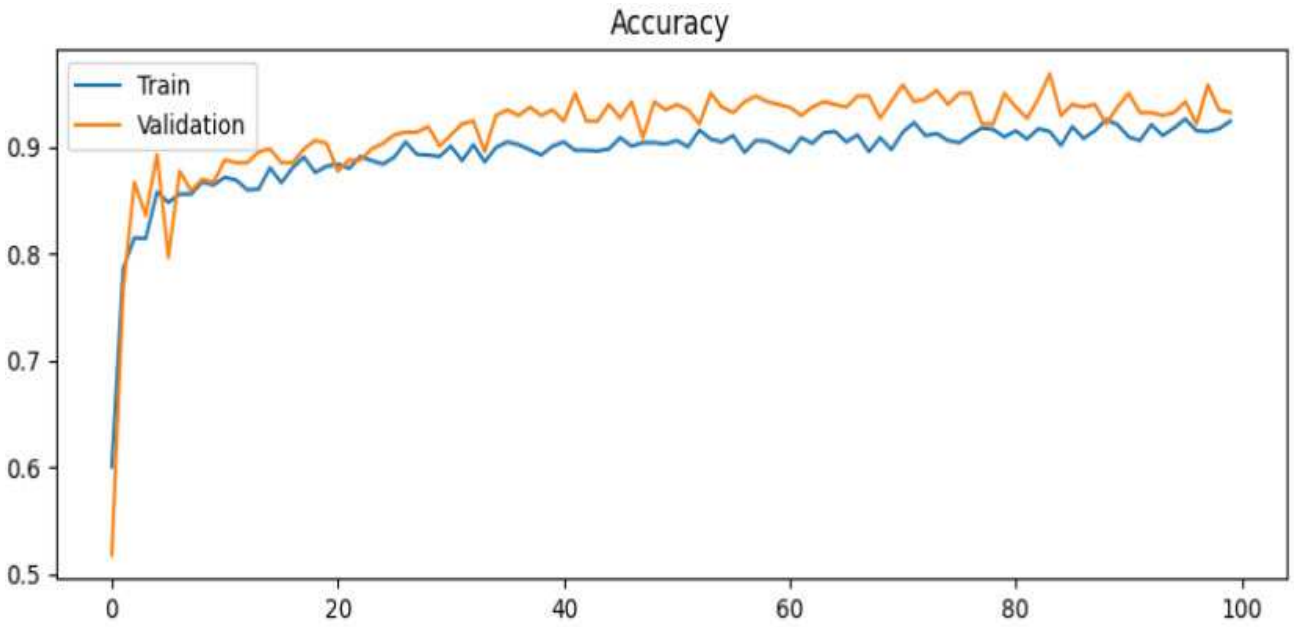


Fig. (7). Accuracy *vs* Epoch and Loss *vs* Epoch Curves for the Proposed model.

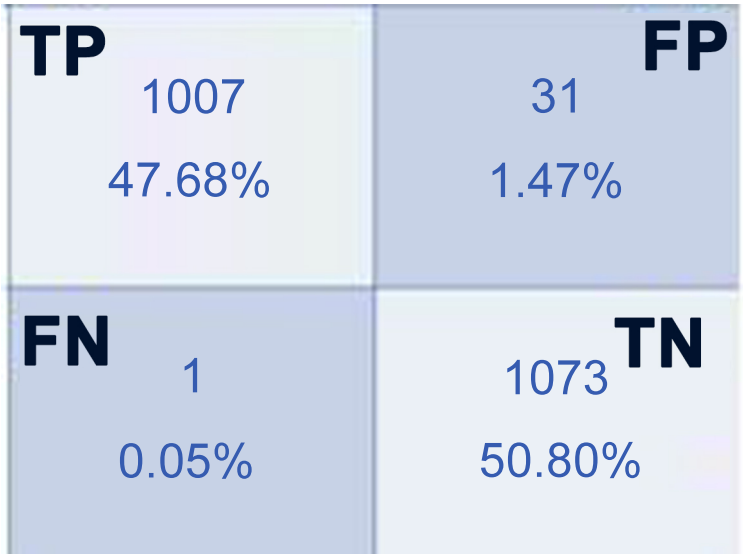


Fig. (8). Confusion matrix (cataract *vs* on cataract).
Shows the statistics of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values.
TP: 1007 cases were correctly classified as having cataracts.
FP: 31 cases were incorrectly classified as having cataracts when they did not.
TN: 1073 cases were correctly classified as not having cataracts.
FN: 1 case was incorrectly classified as not having cataracts when they did.

4.6. Loss Plot

The training loss starts high and rapidly decreases during the initial epochs and slowly decreases over the rest of the epochs. The validation loss also starts high but decreases at a similar rate to the training loss. However, after the initial decrease, it begins to oscillate and does not show a consistent downward trend. This suggests that the model might be overfitting to the training data.

4.7. Accuracy Plot

The training accuracy starts low but rapidly increases in the initial epochs. Afterward, it continues to grow but at a slower rate. The validation accuracy also starts lower and then increases, but it starts to plateau and oscillates after the initial rise. This plot again hints at potential overfitting, as the model's performance on the training data continues to improve while its performance on the validation data remains stable.

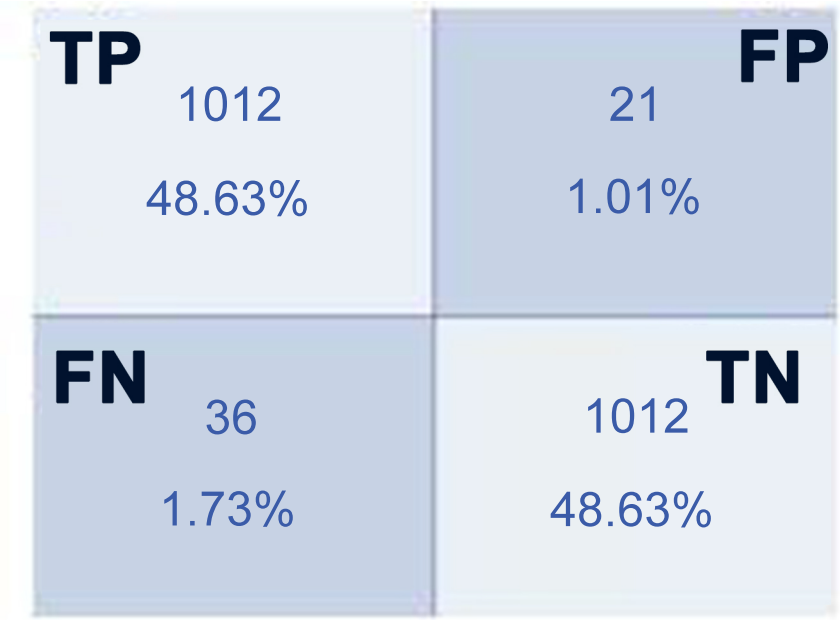


Fig. (9). Confusion matrix (glaucoma *vs* non-glaucoma).
Shows the statistics of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values.
TP: 1012 cases were correctly classified as having glaucoma.
FP: 21 cases were incorrectly classified as having glaucoma when they did not.
TN: 1012 cases were correctly classified as not having glaucoma.
FN: 36 cases were incorrectly classified as not having glaucoma when they did.

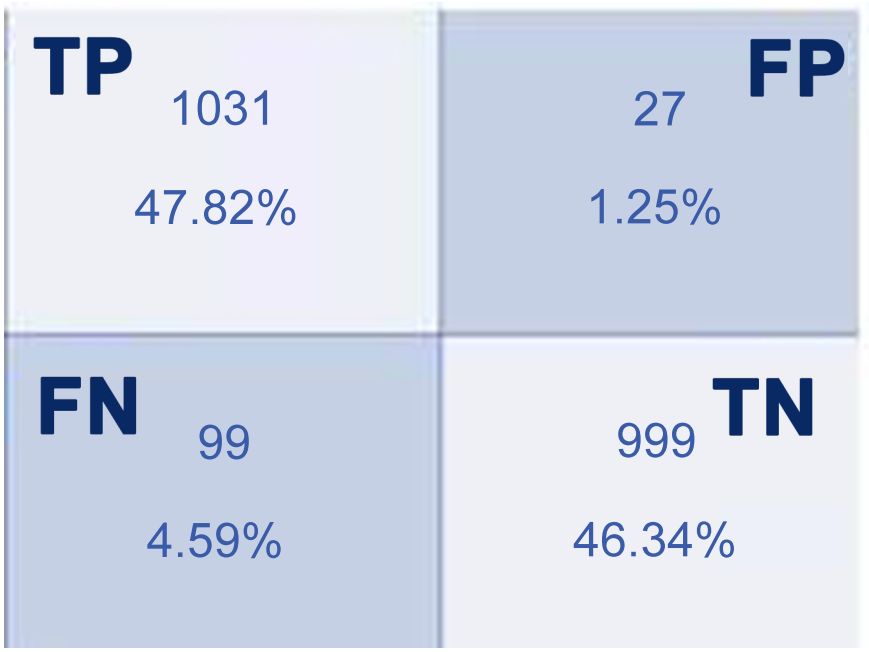


Fig. (10). Confusion matrix (diabetic retinopathy *vs* non-diabetic retinopathy).
Shows the statistics of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) Values.
TP: 1031 cases were correctly classified as having diabetic retinopathy.
FP: 27 cases were incorrectly classified as having diabetic retinopathy when they did not.
TN: 999 cases were correctly classified as not having diabetic retinopathy.
FN: 99 cases were incorrectly classified as not having diabetic retinopathy when they did.

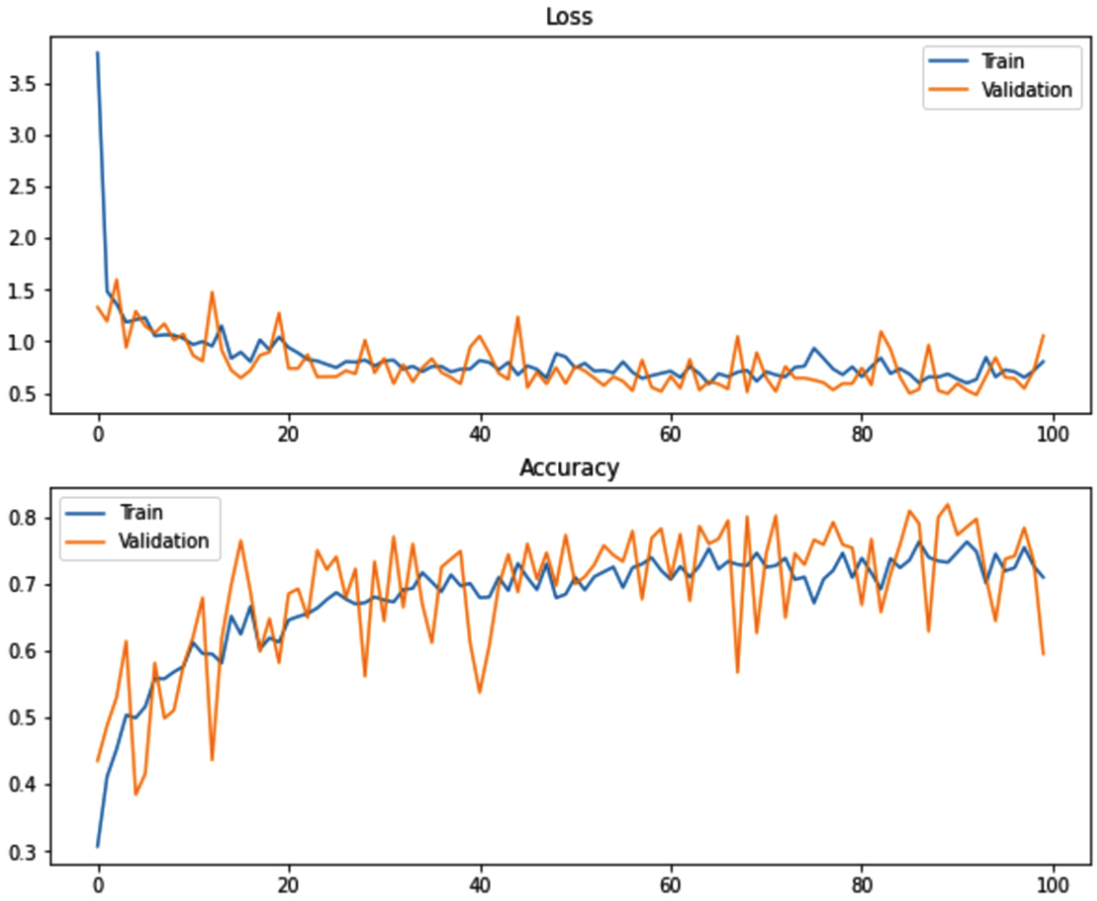


Fig. (11). Accuracy *vs* epoch and loss *vs* epoch curves for densenet.

The graph displays the training and validation loss and accuracy for an AI model over 100 epochs. The loss graph portrays a typical convergence pattern, with both the training and validation loss decreasing sharply in the initial epochs, followed by a gradual descent. The training loss decreases more consistently than the validation loss, which fluctuates but remains within a similar range after about 20 epochs, suggesting good generalization without significant overfitting. The accuracy graph exhibits an increase in both training and validation accuracy over time, with training accuracy being slightly higher throughout, indicating the model was learning effectively and maintaining a consistent performance on unseen data.

4.8. DenseNet Model

Fig. (**11**) represents accuracy *vs* epoch and loss *vs* epoch curves for densenet model.

4.8.1. Loss

The training loss starts extremely high but quickly decreases, while the validation loss initially decreases but then fluctuates around a certain value without significant further improvement. This process could indicate overfitting as the training loss continues to decrease, but the validation loss does not show the same trend.

4.8.2. Accuracy

The training accuracy displayed an upward trend, but the validation accuracy seemed to plateau and fluctuate around a certain value, again hinting toward overfitting.

The graph presents a more erratic pattern of training and validation loss and accuracy. The loss for both training and validation decreases initially but shows significant fluctuation throughout the training process, which could indicate a model struggling with convergence or being affected by a noisy dataset. The accuracy graph indicates a high degree of variability in both training and validation accuracy, with both lines crossing frequently. This suggests the model might benefit from further hyperparameter tuning, more data, or a review of the data quality to improve stability and performance.

4.9. ResNet Model

Fig. (**15**) represents accuracy *vs* epoch and loss *vs* epoch curves for resnet model.

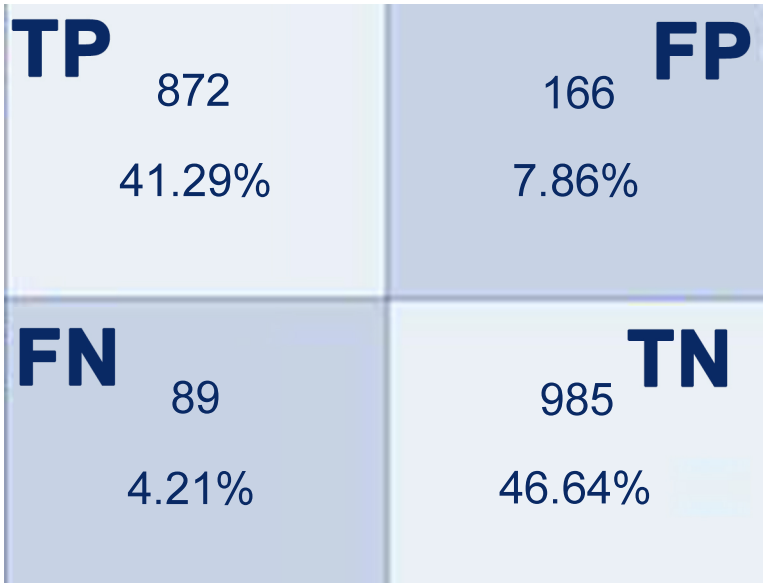


Fig. (12). Confusion matrix (cataract *vs* on cataract).
The statistics of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values.
TP: 872 cases were correctly classified as having cataracts.
FP: 166 cases were incorrectly classified as having cataracts when they did not.
TN: 985 cases were correctly classified as not having cataracts.
FN: 89 cases were incorrectly classified as not having cataracts when they did.

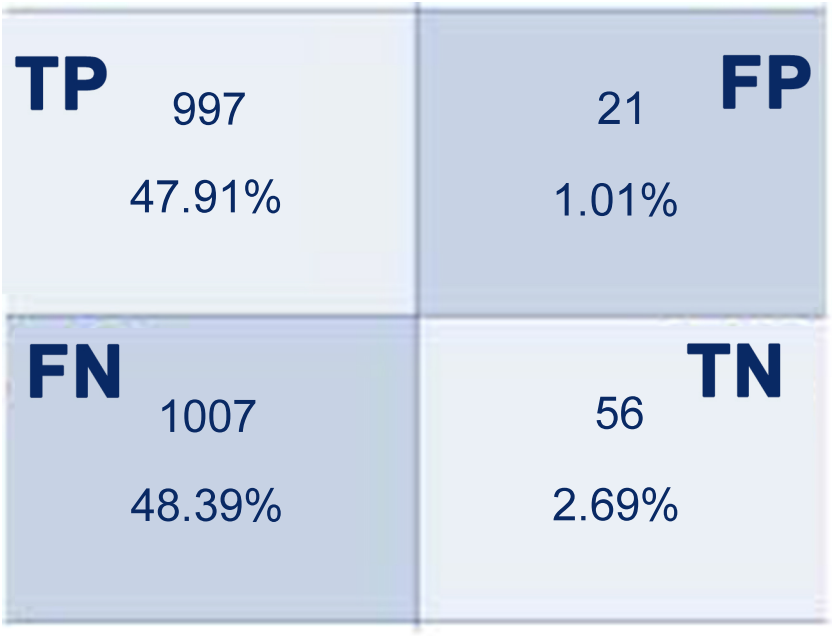


Fig. (13). Confusion matrix (glaucoma *vs* non-glaucoma).
The statistics of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values.
TP: 997 cases were correctly classified as having glaucoma.
FP: 21 cases were incorrectly classified as having glaucoma when they did not.
TN: 1007 cases were correctly classified as not having glaucoma.
FN: 56 cases were incorrectly classified as not having glaucoma when they did.

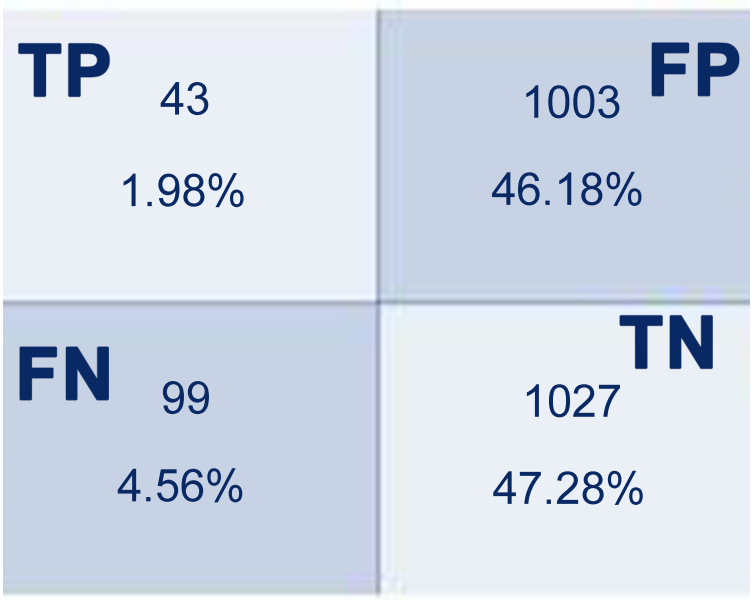


Fig. (14). Confusion matrix (diabetic retinopathy *vs* non-diabetic retinopathy).
The statistics of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values.
TP: 43 cases were correctly classified as having diabetic retinopathy.
FP: 1003 cases were incorrectly classified as having diabetic retinopathy when they did not.
TN: 1027 cases were correctly classified as not having diabetic retinopathy.
FN: 99 cases were incorrectly classified as not having diabetic retinopathy when they did.

4.9.1. Loss

Both training and validation loss are depicted with minor decreases over the epochs. However, the scale of the loss changes was quite narrow, suggesting that the model may have plateaued and is not learning much more from the data. The validation loss is volatile, with fluctuations that do not converge to a significantly better value.

4.9.2. Accuracy

Training accuracy fluctuates quite a bit, which might be an indicator of instability in the training process. The validation accuracy also fluctuates and does not show a clear trend of improvement over epochs. This lack of improvement and the erratic pattern can be a sign of either a model struggling to capture the patterns in the data or a noisy dataset.

The third graph illustrates a model that exhibits a gentle downward trend in both training and validation loss, with the validation loss showing more fluctuations. This could be a sign of the model beginning to overfit, as the validation loss does not decrease as steadily as the training loss. The accuracy graph reflects considerable volatility, particularly with validation accuracy, which does not show a clear upward trend. This could imply that the model's capacity may not be sufficient to capture the complexity of the data, or the model may need better regularization techniques to enhance its generalization abilities.

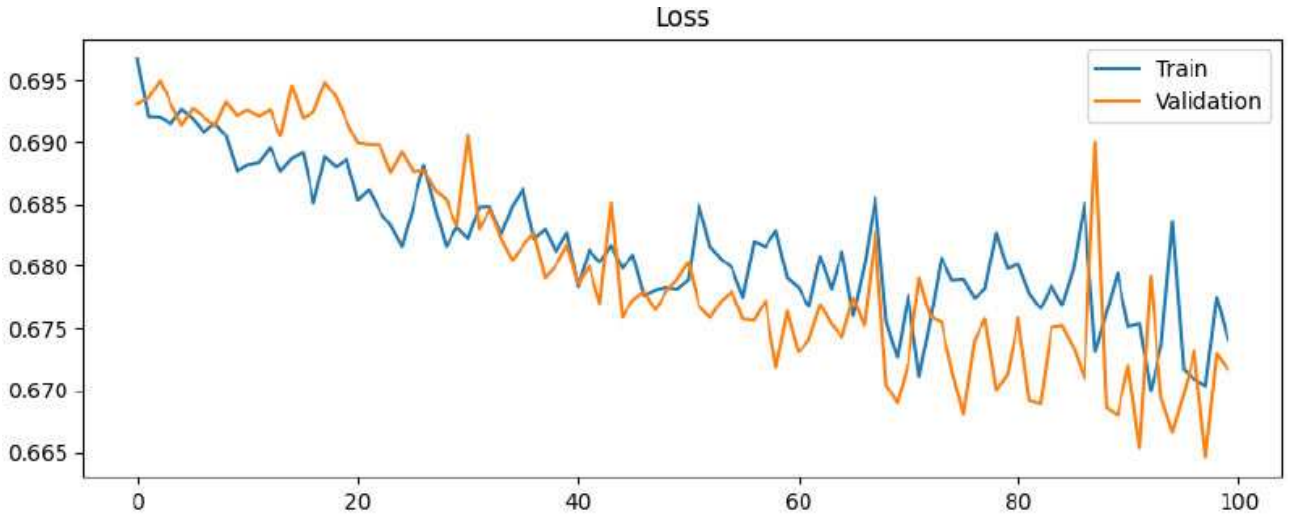


Fig. 15 contd.....

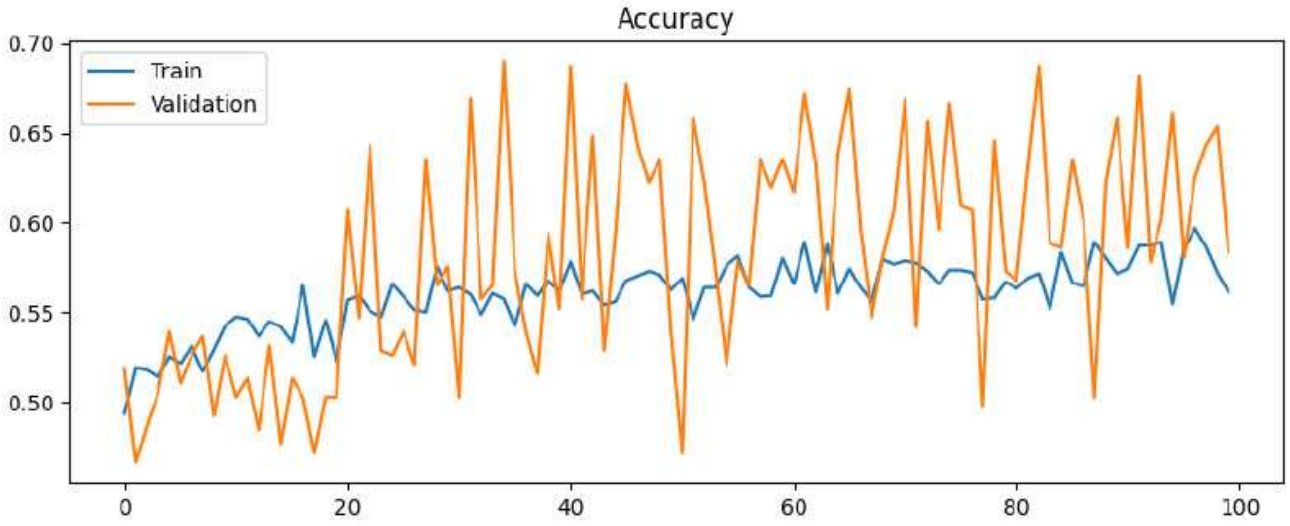


Fig. (15). Accuracy *vs* epoch and loss *vs* epoch curves for resnet model.

4.10. Overall Findings

• The proposed model demonstrated excellent training and validation performance, with both metrics closely mirroring each other. This suggested good generalization. The model is well-optimized for the given task. The model had high accuracy for this classification, given the high TP and TN values and low FP and FN values. The percentages further confirmed this point. The proposed model demonstrates satisfactory performance for this classification with relatively low FP and FN values. The proposed model appeared particularly strong in identifying cases of diabetic retinopathy, given the zero FN.

• The DenseNet Model appeared to have reached a plateau in both loss and accuracy, suggesting convergence. The model may benefit from regularization or hyperparameter tuning to improve validation performance.

• The ResNet Model displayed signs of instability, especially in the validation accuracy. This might indicate the model was sensitive to the specific validation data or that the model was not generalized enough. Techniques such as dropout, regularization, or data augmentation might help.

4.11. Preliminary Clinical Diagnostic Test

The diagnostic potential was determined with standard clinical diagnostic metrics sensitivity and specificity for each CNN model for cataract, glaucoma, and diabetic retinopathy using disease-specific ImageNet training cycles. For example, during cataract training cycles, CNN models were presented with shuffled RCFIs with either cataracts, glaucoma, diabetic retinopathy, or control features. CNN diagnostic potential for cataracts was determined for each CNN based on the rate of RCFI assignment with the correct binary cataract score (1 = cataract, 0 = Not cataract). Analogously, each CNN's diagnostic potential for glaucoma and diabetic retinopathy of CNNs were firmly based on rates of RCFI assignments with the correct binary glaucoma score (1 = glaucoma, 0 = Not glaucoma) or the correct diabetic retinopathy (1= diabetic retinopathy, 0 = diabetic retinopathy) during each respective and distinct disease-specific training cycle.

$$Sn = (TP)/(TP + FN)$$

$$Sp = (TN)/(TN + FP)$$

Sn: Sensitivity. Sp: Specificity.

Diagnostic potentials were evaluated using standard binary clinical testing, given that comorbid conditions frequently exist within the medical field, despite the presentation of shuffled RCFIs containing 4 possible ocular disease states, true positive (TP) and true negative (TN) scores were founded upon binary determinations for each individual disease. A TP event was totaled when the disease under training was correctly detected. An FP event was counted when the disease was incorrectly detected despite its absence. Regardless of whether other disease features from ocular conditions were present, a TN event was only calculated if the model correctly determined the absence of the disease of interest. Finally, FN events were reckoned when the disease of interest was presented and recognized.

In addition, the overall performance of the proposed model was increased across all metrics when compared to existing CNN models. The proposed model achieved an increased overall accuracy 98.87% (training accuracy: 0.993, validation accuracy: 88.34%) as shown in Fig. (**6**) and Table **1** when compared with ResNet as shown in Fig. (**14**) and Table **3** [overall: training, validation] (79%: 0.836, 0.792%) and DenseNet as noted in Fig. (**10**) and Table **2** (0.63%: 0.663%, 0.635%) (P-VALUE). The proposed model detected cataracts, glaucoma, and diabetic retinopathy with increased precision (0.94, 0.89, 0.98) compared with DenseNet (0.85, 0.74, 0.90) and ResNet (0.83, 0.76, 0.84) (P-VALUE) as shown in Fig. (**15**). In addition, the proposed model demonstrated increased recall for cataract, glaucoma, and diabetic retinopathy (0.96, 0.86, 0.91) compared with DenseNet (0.89, 0.71, 0.74) and ResNet (0.78, 0.71, 0.69) ($p<0.05$), as well as a better F1 score proposed model (0.95, 0.87, 0.94) compared with DenseNet (0.87, 0.72, 0.81) and ResNet (0.8, 0.73, 0.75) (P-VALUE).

Table 5. Calculation of parameters for densenet model.

Ocular Condition	TP TP + FN	Sn	TN TN + FP	Sp
Cataract	1013 1013+ 5	0.995	1069 1069+25	0.977
Glaucoma	1063 1063+11	0.989	944 944+63	0.937
Diabetic Retinopathy	1074 1074 + 0	1.000	999 999 + 99	0.909

Table 6. Calculation of parameters for ResNet model.

Ocular Condition	TP TP + FN	Sn	TN TN + FP	Sp
Cataract	1013 1013+ 5	0.995	1069 1069+25	0.977
Glaucoma	1063 1063+11	0.989	944 944+63	0.937
Diabetic Retinopathy	1074 1074 + 0	1.000	999 999 + 99	0.909

Abbreviations: = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

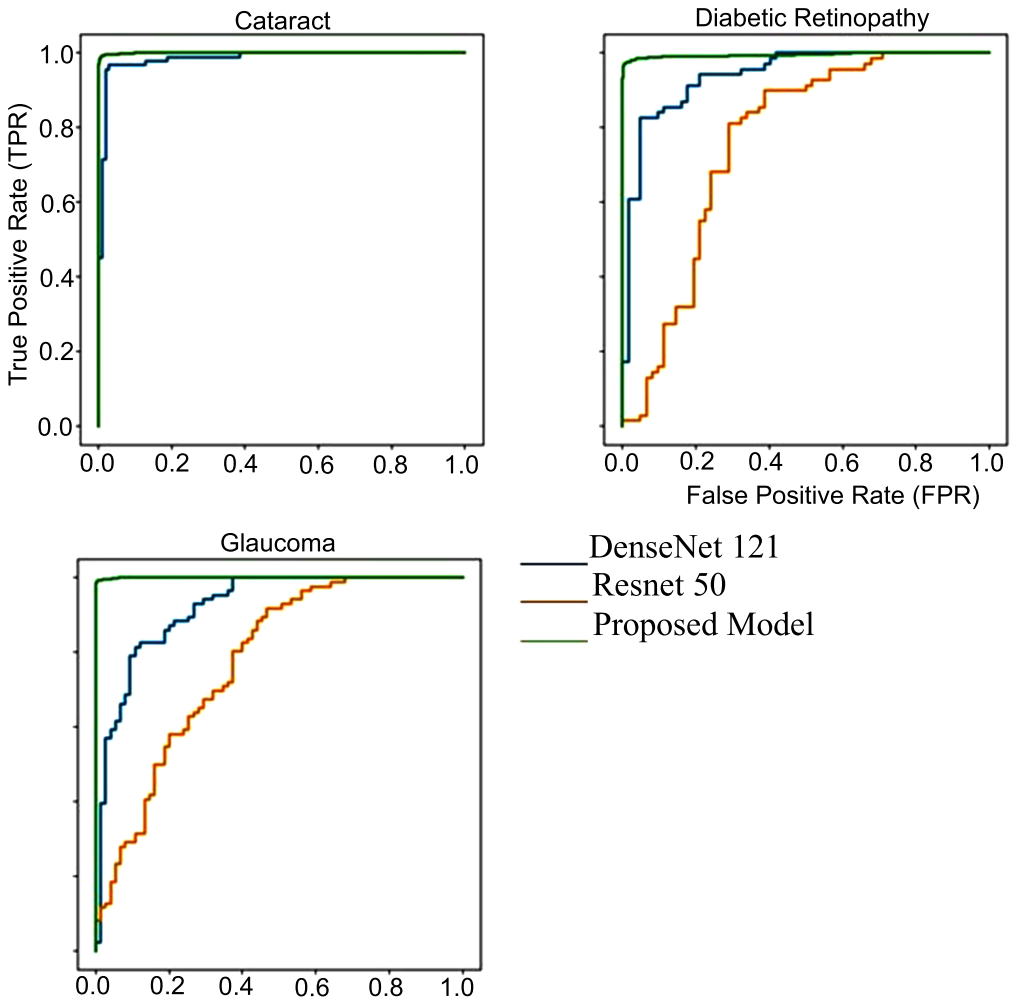


Fig. (16). Performance comparison of the models by analysis through ROC curve.

4.12. Performance of Algorithm

Fig. (**16**) shows Receiver operating characteristic (ROC) curves. ROC curves are commonly used to evaluate the performance of binary classification models by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) at various thresholds.

4.12.1. Cataract

The three models (DenseNet121, ResNet50, Proposed Model) had identical performance for detecting cataracts. Their ROC curves reach a TPR of almost 1.0 with an exceptionally low FPR, indicating excellent performance.

4.12.2. Diabetic Retinopathy

The proposed model appeared to be the most performant, reaching higher TPR values at lower FPRs compared to the other two models. This suggests the proposed model was better at detecting diabetic retinopathy with fewer false positives.

DenseNet121 and ResNet50 indicate similar performance up to an FPR of around 0.2, after which DenseNet121 has a slightly higher TPR.

4.12.3. Glaucoma

The proposed model again seemed superior, particularly in the mid-range of FPR. It achieves a higher TPR at a lower FPR compared to the other two models.

ResNet50 and DenseNet121 were closer in performance, with ResNet50 having a slight advantage in the mid-range of the FPR.

4.12.4. Overall Observations

The proposed model consistently performed better or comparably to the other two models across all three conditions. It often achieved higher TPRs at lower FPRs, suggesting better diagnostic capability with fewer false positives.

ResNet50 and DenseNet121 displayed competitive performance, with one sometimes edging out the other depending on the specific condition being diagnosed.

5. DISCUSSION

AI holds great promise as a future tool in medicine. Already, ML systems have shown potential as a means for medical professionals to augment the detection of rare, life-threatening diseases, such as malignant cancers or ischemic strokes, through analysis of radiographic imaging. In the future, machine learning may take on more clear-cut roles within medicine, such as large-volume biomedical image data analysis to standardize disease screening, assist disease classification, and distinguish diseases that can be hard to differentiate. Clinically, great interest exists in the creation of specialized deep learning systems that enable earlier diagnosis and decrease treatment delay for progressive, irreversible diseases. In this study, we demonstrated that deep learning convolutional neural networks can be fitted with additional convolutional layers to achieve increased performance compared to non-modified predecessors. In particular, the authors sought to quantify the relative performance gained while completing an especially demanding task; in this case: classifying a shuffled dataset of CRFIs obtained from eyes with multiple ocular diseases or control.

However, in designing this study, we also wanted to create training conditions that simulated circumstances that become relevant upon utilization of the model, specifically within the clinical environment. In this effort, we appreciated that the comparison of diagnostic tests was standardized and reported in 2x2 tables with binary diagnostic test results *versus* either a clinical result or gold-standard diagnostic test result. Yet, regardless of how diagnostic tests were compared, patients often presented with more than one related (comorbid) medical condition at a time or coexisting diseases that nevertheless impacted routine disease screenings. Thus, disease diagnoses were not mutually exclusive, and, moreover, the presence of a comorbid disease that was unrelated may nevertheless disrupt and create inaccuracy within diagnostic screening for the disease of interest and hence, indirectly hinder quality care. In both examples, the medical principle of comorbid disease highlights how patient medical complexity may complicate appropriate computational analysis by deep learning systems. For this reason, our study design aimed to maintain the use of ImageNet's dataset containing randomly shuffled RCFIs from eyes with manifestations from multiple diseases, even performing dedicated training for one disease at a time. Despite the increased potential margin for matrix confusion, the same modified deep learning model used in this study classified three ocular diseases simultaneously, and achieved high sensitivity and specificity when serving as a binary screening test for preliminary diagnosis of individual diseases while analyzing a complex RCFI dataset.

Fig. (**17**) denotes the framework of the proposed innovative process. The technician uploads the fundus images to the software. The software is accessed by the physician with their own user credentials and the clinician can see the images. The physician uploads the image to the software. The first step in the software is to clean the patient information from the images and then the images are labelled with a unique patient's identification number. The image is then checked along with patients' identification number according to HIPAA. Furthermore, the image is then sent to the backend cloud model for computation. The classification is generated in the form of a PDF report, which classifies images as Cataract or Non-Cataract, Diabetic Retinopathy or Non-Diabetic Retinopathy, Glaucoma or Non-Glaucoma. The report is sent to the physician and the comments of the physician are saved into the software as patient notes.

In clinical settings, various handheld devices exist that will assist physicians to capture the photographs of the front and back of the eye. These recorded images will then be stored in a HIPAA compliant cloud-based system. The proposed framework can be utilized as a testing and validation dataset for the proposed framework and the biofeedback loop stores the physician's diagnosis into the proposed framework. The model will learn about the images, and the system will be more

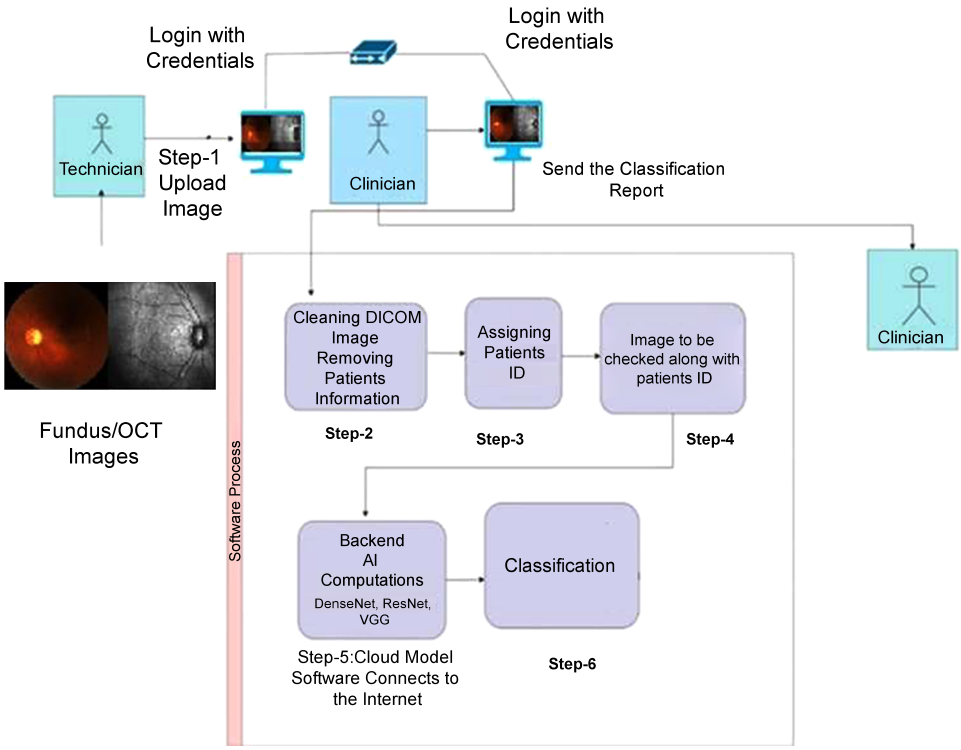


Fig. (17). Framework for classification of ophthalmic images.

robust and will be available to treat patients at remote locations. The limitation of the system is computational architectures, which may not be available when applied in remote locations. The handheld devices may be able to address these limitations and shall be economically available in order to reduce medical costs and thus give treatment to the patients.

The physicians have validated the system. The system was also presented at many conferences, which motivated the authors creating a more robust framework, which was a cloud-based system where the images were stored on a HIPAA Compliant Cloud based system and the analytics was carried out. The visual representation of the framework is shown in Fig. (**17**).

CONCLUSION

AI systems cover a wide range of areas, including the diagnosis of external eyelid lesions and the interpretation of congenital and acquired retinal diseases. This study provided an experimental setup on Type 2 Diabetes Mellitus (T2DM), a chronic metabolic disease that leads to elevated blood sugar levels and damage to organs and systems. The healthcare costs associated with T2DM are estimated to be around USD 327 billion, with projections showing it will continue to rise as the prevalence of T2DM increases. Artificial intelligence has the potential to revolutionize disease detection and referral processes in ophthalmology. By leveraging AI as a screening tool, objective parameters can be utilized to identify patients who require further evaluation for diabetic retinopathy and glaucoma, leading to timely interventions and improved patient outcomes. This study represents a significant contribution by demonstrating the creation of a deep learning convolutional neural network model capable of analyzing retinal color fundus images and differentiating and class. This study leads to the development of decision support systems for clinicians, which will assist them in the current settings as well as the development of telehealth applications.

LIMITATIONS OF THE STUDY

The dataset that the authors have utilized was an unstructured dataset, which was formatted and structured to be utilized in a specific way. The data set needs more labelling and classification.

FUTURE APPROACH

The model would be trained on multiple datasets provided by previously curated and labelled images and the framework would increase the possibility of diagnosing further retinal disease. The framework will also include the software for the proposed approach and direction to system development.

AUTHORS' CONTRIBUTIONS

It is hereby acknowledged that all authors have accepted responsibility for the manuscript's content and consented to its submission. They have meticulously reviewed all results and unanimously approved the final version of the manuscript.

LIST OF ABBREVIATIONS

RCFIs	=	Retinal Color Fundus Images
CNN	=	Convolutional Neural Network
MRI	=	Magnetic Resonance Inmaging
CT	=	Computed Tomography

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data and supportive information are available within the article.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Russell SJ, Norvig P. Artificial intelligence is a modern approach. In: Wikimedia Foundation, Inc. London: Prentice Hall 2010; p. 1136.

[2] Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018; 2(10): 719-31. [http://dx.doi.org/10.1038/s41551-018-0305-z] [PMID: 31015651]

[3] Jiang F, Jiang Y, Zhi H, *et al.* Artificial intelligence in healthcare: Past, present and future. Stroke Vasc Neurol 2017; 2(4): 230-43. [http://dx.doi.org/10.1136/svn-2017-000101] [PMID: 29507784]

[4] Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med 2022; 28(1): 31-8. [http://dx.doi.org/10.1038/s41591-021-01614-0] [PMID: 35058619]

[5] Bogunovic H, Montuoro A, Baratsits M, *et al.* Machine learning of the progression of intermediate age-related macular degeneration based on oct imaging. Invest Ophthalmol Vis Sci 2017; 58(6): BIO141-50. [http://dx.doi.org/10.1167/iovs.17-21789] [PMID: 28658477]

[6] Ting DSW, Pasquale LR, Peng L. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol 2019; 103(2): 167-75. [PMID: 30361278]

[7] Ting DSW, Cheung CYL, Lim G, *et al.* Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017; 318(22): 2211-23. [http://dx.doi.org/10.1001/jama.2017.18152] [PMID: 29234807]

[8] Davson Hugh. Human Eye Anatomy Encyclopedia Britannica 2024. Available from: https://www.britannica.com/science/human-eye (Accessed on: 23 March 2024)

[9] Luo X, Li J, Chen M, Yang X, Li X. Ophthalmic disease detection *via* deep learning with a novel mixture loss function. IEEE J Biomed Health Inform 2021; 25(9): 3332-9. [http://dx.doi.org/10.1109/JBHI.2021.3083605] [PMID: 34033552]

[10] Ellahham S. Artificial intelligence: The future for diabetes care. Am J Med 2020; 133(8): 895-900. [http://dx.doi.org/10.1016/j.amjmed.2020.03.033] [PMID: 32325045]

[11] WHO. World health organization. 2024. Available from: https://www.who.int/

[12] CDC. Health and economic benefits of diabetes interventions. In: National Center Chronic Disease Prev Health Prom. 2022. Available from: https://www.cdc.gov/chronicdisease/programs-impact/pop/diabetes.htm

[13] Gao J, Yang Y, Lin P, Park DS. Computer vision in healthcare applications. J Healthc Eng 2018; 2018: 1-4. [http://dx.doi.org/10.1155/2018/5157020] [PMID: 29686826]

[14] Kaya B, Önal M. A CNN transfer learning□based approach for segmentation and classification of brain stroke from noncontrast CT images. Int J Imaging Syst Technol 2023; 33(4): 1335-52. [http://dx.doi.org/10.1002/ima.22864]

[15] Singh D, Kumar V, Vaishali , Kaur M. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks. Eur J Clin Microbiol Infect Dis 2020; 39(7): 1379-89. [http://dx.doi.org/10.1007/s10096-020-03901-z] [PMID: 32337662]

[16] Sheth D, Giger ML. Artificial intelligence in the interpretation of breast cancer on MRI. J Magn Reson Imaging 2020; 51(5): 1310-24. [http://dx.doi.org/10.1002/jmri.26878] [PMID: 31343790]

[17] Honavar S. Artificial intelligence in ophthalmology - Machines think! Indian J Ophthalmol 2022; 70(4): 1075-9. [http://dx.doi.org/10.4103/ijo.IJO_644_22] [PMID: 35325987]

[18] Leong YY, Vasseneix C, Finkelstein MT, Milea D, Najjar RP. Artificial intelligence meets neuro-ophthalmology. Asia Pac J Ophthalmol 2022; 11(2): 111-25. [http://dx.doi.org/10.1097/APO.0000000000000512] [PMID: 35533331]

[19] Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. Biomed Opt Express 2017; 8(7): 3440-8. [http://dx.doi.org/10.1364/BOE.8.003440] [PMID: 28717579]

[20] Grzybowski A. Artificial intelligence in ophthalmology: Promises, hazards and challenges.Artificial Intelligence in Ophthalmology. Cham: Springer International Publishing 2021; pp. 1-16. [http://dx.doi.org/10.1007/978-3-030-78601-4_1]

[21] Abbas Q. Glaucoma-deep: Detection of glaucoma eye disease on retinal fundus images using deep learning. Int J Adv Comput Sci Appl 2017; 8(6): 41-5. [http://dx.doi.org/10.14569/IJACSA.2017.080606]

[22] Abràmoff MD, Folk JC, Han DP, *et al.* Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmol 2013; 131(3): 351-7. [http://dx.doi.org/10.1001/jamaophthalmol.2013.1743] [PMID: 23494039]

[23] Köse C, Şevik U, İkibaş C, Erdöl H. Simple methods for segmentation and measurement of diabetic retinopathy lesions in retinal fundus images. Comput Methods Programs Biomed 2012; 107(2): 274-93. [http://dx.doi.org/10.1016/j.cmpb.2011.06.007] [PMID: 21757250]

[24] Gulshan V, Peng L, Coram M, *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016; 316(22): 2402-10. [http://dx.doi.org/10.1001/jama.2016.17216] [PMID: 27898976]

[25] Sarki R, Ahmed K, Wang H, Zhang Y. Automated detection of mild and multi-class diabetic eye diseases using deep learning. Health Inf Sci Syst 2020; 8(1): 32. [http://dx.doi.org/10.1007/s13755-020-00125-5] [PMID: 33088488]

[26] Milea D, Najjar RP, Jiang Z, *et al.* Artificial intelligence to detect papilledema from ocular fundus photographs. N Engl J Med 2020; 382(18): 1687-95. [http://dx.doi.org/10.1056/NEJMoa1917130] [PMID: 32286748]

[27] Nazir T, Irtaza A, Javed A, Malik H, Hussain D, Naqvi RA. Retinal image analysis for diabetes-based eye disease detection using deep learning. Appl Sci 2020; 10(18): 6185. [http://dx.doi.org/10.3390/app10186185]

[28] Quellec G, Charrière K, Boudi Y, Cochener B, Lamard M. Deep image mining for diabetic retinopathy screening. Med Image Anal 2017; 39: 178-93. [http://dx.doi.org/10.1016/j.media.2017.04.012] [PMID: 28511066]

[29] Bhaskaranand M, Ramachandra C, Bhat S, *et al.* Automated diabetic retinopathy screening and monitoring using retinal fundus image analysis. J Diabetes Sci Technol 2016; 10(2): 254-61. [http://dx.doi.org/10.1177/1932296816628546] [PMID: 26888972]

[30] Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. Ophthalmology 2017; 124(7): 962-9. [http://dx.doi.org/10.1016/j.ophtha.2017.02.008] [PMID: 28359545]

[31] Doddi GV. 2022. Available from: https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification