



Transfer Learning for Automated Classification of Eye Disease in Fundus Images from Pretrained Model

Praveen Kumar Mannepalli^{1*}, Vishwa Deepak Singh Baghela², Alka Agrawal³, Prashant Johri², Shubham Satyam Dubey², Kapil Parmar²

¹ Department of Computer Science and Engineering, Chandigarh University, Mohali 140413, India

² School of Computing Science and Engineering, Galgotias University, Uttar Pradesh 203201, India

³ Electronics and Communication Department, GLA University, Mathura 280406, India

Corresponding Author Email: praveen.e17513@cumail.in

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410520>

ABSTRACT

Received: 14 September 2023

Revised: 14 March 2024

Accepted: 20 August 2024

Available online: 31 October 2024

Keywords:

transfer learning, eye diseases, prediction, multiclass classification, deep learning, DenseNet-121, fundus images

In today's environment, diseases are on the rise and manifest in various ways, harming numerous bodily parts. In recent years, sophisticated analysis of retinal pictures has enabled the development of computerised methods for diagnosing various disorders. These tools save us both time and money. The analysis of retinal fundus pictures is essential for the early diagnosis of eye disorders. For years, fundus images have also been utilised to detect retinal disorders. If the two picture modalities were merged, the resulting image would be more useful since fundus images detect abnormalities such as drusen, geographic atrophy, and macular haemorrhages. At the same time, OCT scans reveal these abnormalities' specific shapes and locations. Images of the fundus are an essential diagnostic tool for numerous retinal diseases. This research investigates the effectiveness of transfer learning using the pre-trained model for predicting eye diseases from fundus images. In this study, the fundus images dataset is used, which is collected from the GitHub repository. This dataset has multiple classes, i.e., Maculopathy, Healthy, Glaucoma, Retinitis Pigmentosa and Myopia. To perform the classification, a transfer learning-based pre-trained DenseNet-121 model is proposed. Python has been employed for simulations. The results of the experiment show that, when applied to the test data, the suggested model had the highest precision (97%), sensitivity (92%), and particularity (98%) and outperformed the existing model in terms of all performance metrics.

1. INTRODUCTION

The eyes are an essential part of human life. Every person relies on the eyes to see and sense the world around them. Sight is one of the most important senses because it provides us with 80% of the information they process. By taking good care of our eyes, they can reduce the chance of going blind and losing our eyesight. They should also be on alert for any problems that may develop in our eyes, such as cataracts and glaucoma. Due to population ageing, it is anticipated that the number of patients with eye diseases will substantially rise shortly. Most people experience eye issues at some point in time. Some of the eye issues are minor and simple to cure at home which will go away on their own, other major eye issues need assistance from the expert doctors. When these eye diseases are accurately diagnosed at an early stage, only then the progression of these eye diseases be stopped. These eye diseases have a wide range of visually discernible symptoms. To accurately diagnose eye illnesses, it is required to analyse a wide range of symptoms [1].

In this situation, maintaining eyesight and improving quality of life are primarily achieved by early detection and appropriate care of eye diseases. Deep AI integration in ophthalmology may be useful in achieving this goal because it

can quicken the diagnostic procedure and lower the need for human beings [1]. A subdiscipline of computer science known as artificial intelligence (AI) focuses on utilising computers to create algorithms that mimic human intellect. In 1956, the idea of artificial intelligence was first proposed. The field has advanced significantly since then, and is now referred to as "the newest and greatest phase of industrialisation in human existence." At times, the terms DL, artificial intelligence (AI), and ML have been used interchangeably; nonetheless, it is crucial to make a distinction between the three terms (Figure 1).

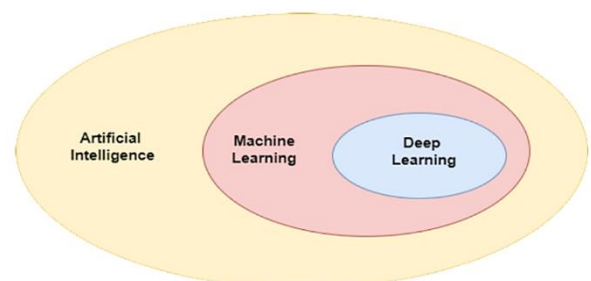


Figure 1. Overview of artificial intelligence (AI) and its subfields [2]

Automatic identification of ocular disorders using deep learning approaches has recently been implemented [3-5]. There was a great deal of diagnostic sensitivity and specificity shown by the detection data. Fundus picture categorisation using Deep Learning (DL) has arrived at cutting-edge levels, but the method is still limited by the amount of training data and processing time required.

One way to learn anything new is through transfer learning, which involves taking what you know about one activity and applying it to another. It aids in getting decent results with little data and less training time required to get better results. By applying the characteristics learnt on one job to another (fundus pictures), the transfer learning technique helps ease the limits discussed earlier. Fundus picture categorisation can benefit from the characteristics learned by neural networks trained on generic images through the application of transfer learning. Applying transfer learning as a feature extraction and classifier has recently led to the introduction of one or two kinds of eye diseases [6-8].

To determine and differentiate between different aspects of fundus pictures for the purpose of diagnosing each eye condition, a deep neural network (DNN) with several layers and filters is utilised. Misdiagnosis or the failure to identify the illness can occur when the fundus image is damaged by pixels that are either too dark or too light, or by pixels that are ambiguous or unevenly illuminated. As a result, picture quality is an essential first step in fundus image categorisation and a crucial attribute of fundus pictures. In order to remove low-quality photos before diagnosing eye diseases, past research did not employ integrated quality evaluations. The only things done so far are quality evaluations and fundus classifications. So, this is the first research on integrated quality evaluation for multiclass fundus image categorisation.

The current research environment lacks a clear paradigm for effectively incorporating AI into ophthalmic clinical practice. This gap prevents progress in the early diagnosis and treatment of eye illnesses. However, the present research environment lacks a defined framework for incorporating AI into ophthalmic clinical practice, resulting in a gap in eye disease identification and management. Using artificial intelligence (AI), this study seeks to close this gap, notably the DenseNet-121 model, to improve the accuracy and efficiency of eye illness prediction. Our research is motivated by the crucial necessity to detect eye disorders early on in order to prevent visual damage and blindness. Manual screening for eye disorders is time-consuming and error-prone. Automated approaches are required for accurate and rapid detection of fundus pictures, which aids clinicians in early diagnosis.

“Research Question or Hypothesis: Can the DenseNet-121 model, along with transfer learning techniques, considerably improve the accuracy and efficiency of predicting eye illnesses using fundus images?”

Our research intends to contribute to this determined by:

- **Method Improvement for Accuracy and Efficiency:** Presenting a method that improves the accuracy and efficiency of eye disease prediction by utilising transfer learning and the DenseNet-121 model.
- **Preprocessing for Enhanced Visibility:** Using a series of preprocessing procedures, such as image resizing and contrast-limited adaptive histogram equalisation (CLAHE), to significantly improve blood vessel visibility in fundus images, capturing critical characteristics for disease detection.

- Fine-tuning a pre-trained DenseNet-121 model involves freezing existing layers and adding specialised classification layers appropriate to the fundus picture classification challenge.
- **Model Evaluation and promise:** demonstrated model's accuracy and efficacy on the test dataset, showing its promise as an efficient detection tool for the early diagnosis of eye illnesses.

The following is the structure of this research for the sections that follow: In the first Section 1 provides the background, problem statement, and contributions without clear demarcation. I recommend the authors revise the introduction to clearly outline the background, identify the gap in the current research, state the research question or hypothesis, and then clearly state the contributions of their work. Next Section 2 delivers the literature review that is relevant to our research area by highlighting the advancements and limitations that their study aims to address. Then Section 3 provides the research methodology for this eye disease detection using a transfer learning model; After this, Section 4 presents the results and discussion with performance measures, as well as a comparative analysis with base and proposed models. In last Section 5, cover the summary of this work with future work.

2. LITERATURE REVIEW

Alternatively, employing both uninfected and diseased DR images, handcrafted image recognition techniques have produced respectable results and accuracy regarding medical DR image illness identification. Deep learning (DL) is a branch of machine learning that is also a branch of artificial intelligence (AI). It deals with procedures that are modelled after the finely built restrictions and model parameters of brain neurons, and its restricted detection effectiveness is caused by the high maintenance costs of these models. Due to its ability to directly learn features from training data, deep learning-based transfer learning models have drawn attention from a variety of research domains in recent years.

This section examines existing diagnostic methods for ocular diseases (ODs). They talk about the current restrictions and emphasise the key directions and fixes offered in the suggested system to address the present issues. For instance, Aranha et al. [2] provided a network of CNN that was transferred-learned on 38,727 high-resolution fundus pictures. The suggested method was able to diagnose diabetic retinopathy and classify cataracts, excavation, and blood vessels from low-quality pictures with accuracies of 87.4%, 90.8%, 87.5%, and 79.1%, respectively.

In 2023, one of the most sought-after DL methods, DR, is categorised into five stages using DenseNet. The hyperparameters are adjusted, and the batch size, learning rate, and epochs' maxima are discovered. These hyperparameters vary depending on the kind and volume of data and are typically application-specific [9].

In 2023, the majority of the dataset used in this study is made up of high-resolution fundus pictures from people with diabetic retinopathy (DR), glaucoma, cataracts, and healthy eyes. According to the experimental results, the proposed system outperforms existing methods for classifying eye illnesses and achieves 97% accuracy using a modified efficientNetB3 model [10].

In 2022, the features of the eye are extracted for efficient

categorisation using suggested deep learning architectures such as CNN, hybrid CNN with ResNet, and hybrid CNN with DenseNet 2.1. The accuracy of the models was 96.22%, 93.18%, and 75.61%, in that order. The hybrid CNN with DenseNet architecture is highlighted as the ideal deep learning classification model for automated DR detection in the paper's comparative analysis of the CNN, hybrid CNN with ResNet, and hybrid CNN with DenseNet architectures [11].

In 2022, Shamia et al. [12] introduced a cloud-based, deep convolutional neural network expert system for illness diagnosis. Detection rates for DR were 91%, while those for cataract and glaucoma were 90% and 86%, respectively. The technique allowed the development of an understandable and straightforward GUI for usage in digital environments.

In 2022, the study by Meenakshi and Thailambal [3] focused on proposed ensemble-based strategies for diabetic retinopathy categorisation and prediction. Images are first processed using a boosting-based ensemble learning algorithm for diabetic retinopathy prediction. The photos of diabetic retinopathy are then sorted into their distinct phases using a CNN. Both DR image prediction and picture grading accuracy of 96% were achieved by the proposed ensemble-based CNN.

In 2021, the suggested approach [4] is created and designed to make it simple for individuals to diagnose diseases of the retina, glaucoma, and cataract. For detection, the techniques RF, Logistic Regression, SVM, and Gradient Boosting are utilised. In terms of algorithms, their effectiveness makes them stand out from the competition, and in our situation, gradient boosting delivers the greatest outcomes for the cataract-affected eye with 90% accuracy. Consequently, the accuracy of the supervised algorithms logistic regression and RF is 89.1% and 86.1%, respectively.

In 2021, Raza et al. [5] provided a model based on a deep learning classification scheme for four distinct digital retinal image (DRI) formats. They use a Kaggle dataset totaling 602

DRI and weighing 1.67GB to test the Inception v4 model's invariant. The outcomes are quite satisfactory as they attain a 96% accuracy rate.

In 2021, Chelaramani et al. [6] evaluated the prediction of eye diseases using different methods for developing deep neural classifiers with small labeled data. They also put forth a unique MTL-based teacher ensemble approach for distilling information. Their method obtains an 83% overall accuracy, 75% top-5 accuracy, and 48 BLEU throughout tasks T1, T2, and T3 on a dataset of 7212 labelled and 35854 unlabeled photographs from 3502 cases. Even with only 15% training data, our technique outperforms baselines for each of the three tasks by 8.1, 3.2, and 11.2 points, respectively.

In recent years, the field of fundus image categorisation has made significant advances using deep learning (DL) algorithms. Notably, (CNNs) convolutional neural networks and other deep learning architectures have shown significant progress in automating eye illness detection, with accuracy rates above 90%. These accomplishments demonstrate the promise of deep learning models in illness diagnosis. However, present approaches face issues such as high maintenance costs, poor detection performance, and the necessity for application-specific hyperparameter adjustment. The present research intends to expand on these advances by presenting a new system that uses the DenseNet paradigm. This technique focuses on improving accuracy and efficiency in fundus image categorisation, notably in distinguishing between different stages of diabetic retinopathy. By tackling these constraints and scrutinising existing deep-learning approaches, the study hopes to help improve diagnostic tools. The ultimate objective is to create a more robust and effective method for early detection and categorisation of ocular disorders. The following Table 1 provides the comparative analysis of the previous study based on its performance and techniques.

Table 1. Comparison different research on their accuracy and techniques

Ref	Title	Methodology	Accuracy	Future Work
[7]	DL analysis of eye fundus images to support medical analysis	convolutional neural network, deep learning	88.72%	Future work will include evaluating our analysis on larger datasets, as they are currently too small.
[8]	Recognition of ocular diseases with ensemble of dl models	VGG16, DenseNet201 and Resnet50	72.8%	The work can be extended further by more data from local hospitals all over India and building a large dataset for the models.
[13]	Classification of retinal detachment using DL through retinal fundus images	neural networks like Densenet-201, InceptionV3, Googlenet, VGG16, ResNet50, and VGG19	96.16	As a result of this effort, physicians have a better environment to work in, which in turn benefits both the approach to treatment and the clinical decision.
[14]	An expert system to foresee eye disorder with deep CNN	SVM, Decision tree, KNN, Logistic Regression and ANN	92.78%	The future work will be implemented on the next technique and achieve better performance.
[15]	Retina diseases diagnosis using DL	(VGG-16, Inception V3, MobileNet, ResNet-50, and Xception)	96.21	To analyse information for hidden trends in order to improve future choices.

3. RESEARCH METHODOLOGY

In this section, the problem statement is discussed. Then, to overcome this problem, a transfer learning model is proposed that is discussed in this section with some research steps.

3.1 Problem statement

The difficulty in using biomedical information systems to diagnose eye disorders and diseases with relatively reversible symptoms of Glaucoma. This disease frequently results in severe damage to the optic nerves that can be located and

identified because of its developing complexity. However, there is currently no automated expert system that uses deep learning to identify glaucoma. These conditions, which are frequently referred to as the worst eye disease, can be extremely harmful and upsetting for patients. As a result, as was already said, the ideal course of action is to put in place an expert system that might anticipate the inevitable outcome that, given enough time and resources, an eye will inevitably be able to identify glaucoma. When this happens, it is crucial that the discovery be quickly identified and eradicated or quarantined before any real harm is done.

3.2 Proposed methodology

The DenseNet-121 model is applied to carry out image preprocessing and classification on fundus images as part of the research method. The dataset obtained from a GitHub repository contains 250 fundus images belonging to five classes, separated into train (202 images) and test (48 images) subfolders. To improve the image quality and blood vessel visibility, image preprocessing is used. The preprocessing steps include scaling the pictures to 224×224 pixels, which is the standard dimension, dividing the images into RGB channels, analysing the characteristics of each channel, and applying CLAHE to the GREEN channel to improve contrast. By combining the three channels, the enhanced colour fundus images are then reconstructed. The DenseNet-121 model with pre-trained dimensions is utilised for classification. The layers of the model are frozen to preserve the pre-trained knowledge, and new classification layers are added to modify the model to the specific fundus image classification task. The hyperparameters are configured using an optimiser named Adam, a loss function named categorical cross-entropy, a batch size named 4 and epochs named 30. The model's ability to accurately and efficiently categorise fundus images into their corresponding classes is assessed using the test dataset. The overall working strategy are visualised in Figure 2 in the form of research steps which is shown below.

(<https://github.com/gcowen/fundusimageclassification>). There are five classes are given in this dataset given in Table 2 with their quantity of images. The dataset contains two subfolders, i.e., train and test, in which 202 images are training data and 48 images are testing data, and the total images are 250. Figures 3 and 4 show the sample image and also the sample image of color conversion.

Table 2. Classification of the images in the database

Categories	Number of Images
Glaucoma	75
Maculopathy	72
Myopia	49
Retinitis pigmentosa	22
Healthy	32

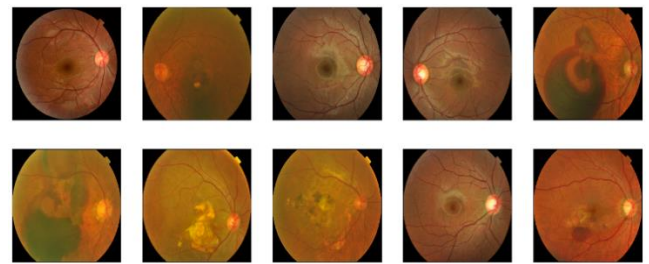


Figure 3. Sample images of the dataset

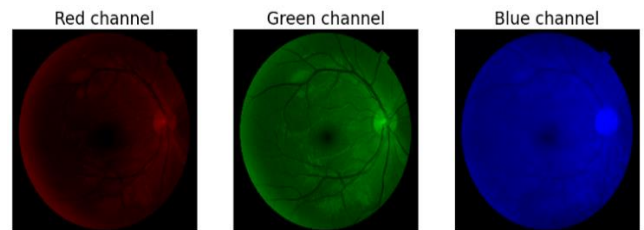


Figure 4. Sample image of colour conversion

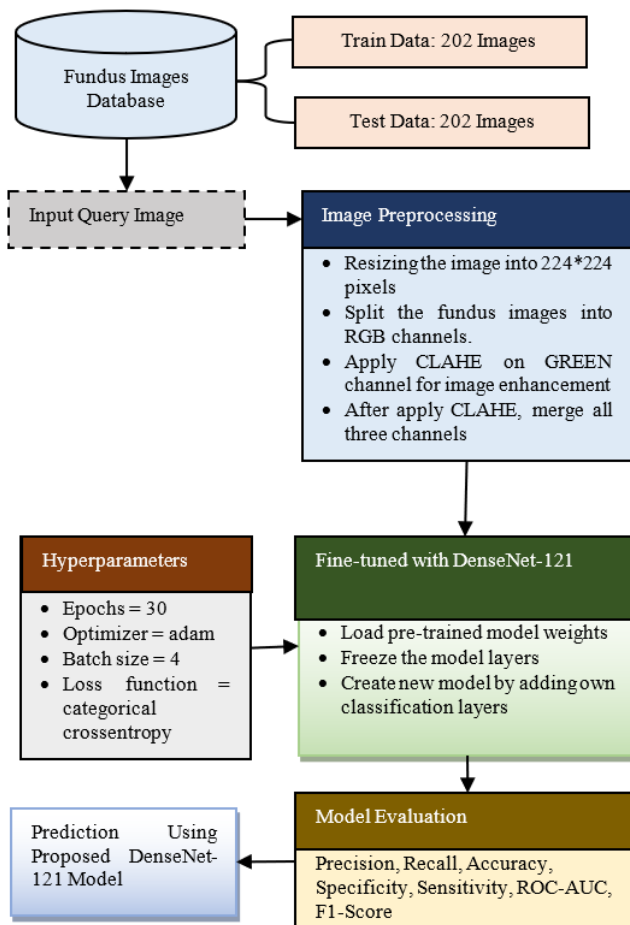


Figure 2. A conceptual framework of the proposed methodology

3.2.1 Data collection

In this research, fundus images dataset is used which is collected from the GitHub repository

3.2.2 Image preprocessing

The purpose of image preprocessing in computer vision and image analysis the objective is to increase the image's quality by reducing or rejecting noise and to get the data ready for use in subsequent analysis or machine learning techniques. Depending on the goal and the content of the images, the preprocessing procedures used to prepare them for further analysis might differ [16]. The following are some of the steps involved in preparing an image:

- Resizing the original fundus image to a typical dimension is 224 by 224 pixels is the first step in the image preparation. Since many DL models need inputs of fixed sizes, this is generally done to maintain consistency in image dimensions.
- The fundus image consists of three different channels, or colours: red, green, and blue. They may analyse and process the image in three distinct ways, one for each of these channels, to solve problems unique to each of them.
- Specific characteristics are observed upon analysis of the individual channels. Blood vessel visibility may be compromised due to noise or saturation in the RED channel. It's possible that the BLUE channel has poor contrast and not enough data. However, it seems that the GREEN channel

provides the finest contrast for blood vessels.

- The GREEN channel undergoes CLAHE method to increase the visibility of blood vessels and improve contrast. In order to improve local contrast and details in an image, CLAHE is employed to more evenly distribute the intensity values of pixels. CLAHE is a modification of AHE designed to reduce the effect of excessive contrast. CLAHE works on "tiles," or small, independent parts of the image, rather than the whole thing. Next, bilinear interpolation is used to merge neighbouring tiles, thus eliminating the false barriers [17]. This method can be used to improve image contrast. CLAHE may also be used on colour images; Usually, the luminance channel is used for this, And the results are significantly superior for an HSV image after modifying the brightness channel than for a BGR image after modifying all channels.
- The improved colour fundus image is reconstructed by merging back all three channels after CLAHE has been applied to the GREEN channel. The final outcome now has better contrast in the GREEN channel, allowing for clearer visualisation of blood vessels. Figure 5 shows the Before and after applying CLAHE images.

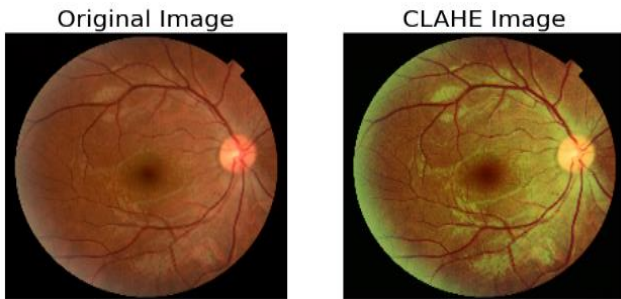


Figure 5. Before and after applying CLAHE

The image preprocessing based on the aforementioned details includes resizing the image, analysing individual colour channels, applying CLAHE to the GREEN channel for enhancement, and merging the channels to obtain an improved and visually enhanced colour fundus image in which blood vessels are more distinguishable and contrast is increased. This preprocessing phase is essential for optimising input data for subsequent analysis and enhancing the efficacy of following image processing or deep learning algorithms.

3.2.3 Proposed densenet-121 model

The finetune process of the proposed DenseNet-121 Model is implemented in this work. To recognise the structures of the eye, the DenseNet-121 model is trained using a dataset of fundus images. DenseNet-121 is a robust neural network with extensive experience in image recognition and other computer vision tasks. The DenseNet block diagram, shown in Figure 6, makes use of a dense block with five layers. The name "DenseNet-121" comes from the total number of neural network layers, which is 121. DenseNet-121 typically consists of a combination of numerous layers [18].

- **Input Layer:** This layer provides input to the convolutional layers. It implements specific adjustments like feature scaling, data

augmentation, and mean subtraction efficiently.

- **Convolution Layer:** The operational unit segments that are most determined are convolutional filters, pooling, and activation.
- **Normalisation Layer:** This layer should be applied to all feature maps of the corresponding layer in order to obtain a more comprehensive interpretation of the input given.
- **Dropout Regularisation Layer:** As a consequence, the network's over-fitting is diminished, and more reliable features can be learned.
- **Fully Connected Layer:** One vector was produced as a result of this method of layer input processing.
- **The Dense Layer:** That all of the neurons in the layer below are linked to the neurons in this one. The majority of neural network architectures use this layer.
- **SoftMax Layer:** The output layer immediately precedes this layer. This stratum assigns a decimal probability of victory to each class. The probability ranges in between 0 & 1 is a decimal number.

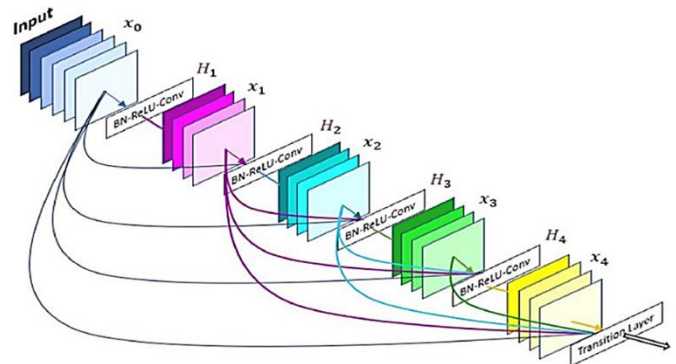


Figure 6. Block diagram of DenseNet

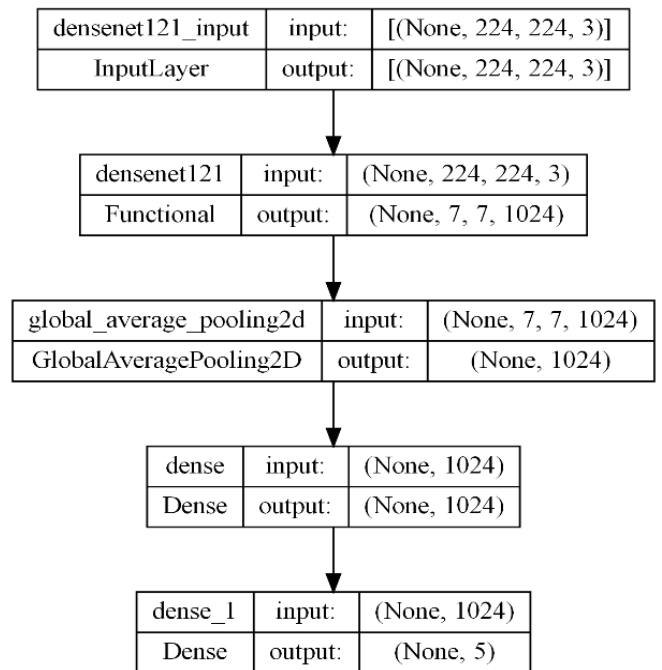


Figure 7. The layered architecture of the DenseNet-121 model

Model: "sequential"

Layer (type)	Output Shape	Param #
densenet121 (Functional)	(None, 7, 7, 1024)	7037504
global_average_pooling2d (GlobalAveragePooling2D)	(None, 1024)	0
dense (Dense)	(None, 1024)	1049600
dense_1 (Dense)	(None, 5)	5125

=====
 Total params: 8,092,229
 Trainable params: 1,054,725
 Non-trainable params: 7,037,504

Figure 8. Model summary

DenseNet is built on a convolution and pooling layer. After a dense block, a layer of transition comes next, a transition layer serves another dense block, and a classification stage precedes a dense block [19].

As simple as it may appear, DenseNets effectively bind each layer to its neighbouring layers. This is the most crucial aspect to determine. In DenseNet, the input of a layer is formed by concatenating the feature mappings of antecedent layers. The following Figures 6, 7 and 8 show the block diagram and architecture of DenseNetmodel.

The model is summarised above as a Sequential model architecture in Figure 8, which has several layers. Layer one is a priorly trained deep convolutional neural network for image classification known as DenseNet-121. Feature maps having a shape of (None, 7, 7, 1024) are generated by this layer from the input images. Because of its pre-trained weights, the DenseNet-121 model includes around 7,037,504 parameters that cannot be changed. A Global Average Pooling 2D layer follows the DenseNet-121 layer and concatenates the 7×7 feature mappings into a single vector of length 1024. In the subsequent Dense layers, each with 1024 and 5 neurons, the retrieved features are further transformed, and a prediction of the final output is made. There is a total of 1,049,600 trainable parameters due to the Dense layers. The total set of parameters in the model is 8,092,229, with 1,054,725 being trainable and 7,037,504 non-trainable. Image classification tasks with five output classes are a good fit for this architecture, and the pre-trained DenseNet-121 basis offers a useful feature extraction backbone, allowing for successful training with less labelled data.

This model is implemented utilising various hyperparameter configurations, including a number of epochs, optimiser, batch size, and loss function.

- **Batch Size (4)**

The term "batch size" describes how many input samples are assessed collectively throughout each training cycle. With a batch size of 4, the model will take in four samples at a time and adjust its parameters depending on the gradients from those samples. Batch size has a significant effect on memory usage and training efficiency, which are the main metrics in deep learning. Hardware constraints, data sizes, and model complications are the key factors to the determination of the batch size.

- **Epochs (30)**

The training cycle referred to as "epoch" is spelled out here. After that, model is going to train itself by repeating 30 times for the whole data set. We always run experiments to come up with optimal number epochs to avoid overfitting.

- **Adam Optimizer**

Optimisation methods, e.g., "Adaptive Moment Estimation" or Adam, are commonly used in deep learning models being trained. By merging AdaGrad and RMSprop into Adam, two other common optimisation methods, Adam presents a speedier and more efficient convergence during training. This is because it adjusts the learning rate for each parameter uniquely dependent on previous gradients so that it can be flexible in regard to changing parameter values. Adam works effectively with complex and large-scale models because of its flexibility; it can quickly reach the optimal solution and avoid problems like gradients that expand or disappear.

- **Categorical Crossentropy Loss Function**

Each input sample often belongs to one of many classes, and the categorical cross-entropy is a loss function frequently employed for such tasks. It evaluates the dissimilarity between the expected probability distribution (the model's output) and the actual probability distribution (the one-hot encoded target labels) in the context of deep learning. The projected probabilities (i.e., the correct class should have a probability near 1, and other classes should have probabilities close to 0) should closely match the true probabilities in order to minimise the categorical cross-entropy during training. Accurate predictions and proper classification of input data are achieved by the model by minimising the categorical cross-entropy.

Rationale for selecting DenseNet-121: Several variables impacted the use of DenseNet-121 in this investigation. First, the extensive connection structure promotes feature reuse, allowing the model to successfully capture complicated patterns and relationships within fundus images. DenseNet-121's parameter efficiency, achieved by shared feature maps, decreases the danger of overfitting and facilitates training with less labelled data. The model's pre-trained weights on large-scale datasets improve its capacity to generalise features from a variety of images, which is especially useful in medical imaging where obtaining vast labelled data is problematic. Additionally, DenseNet-121 has shown cutting-edge performance in a variety of computer vision tasks, highlighting its applicability for our fundus image categorisation goals.

3.3 Proposed algorithm

Here is a proposed algorithm for fundus image classification using the transfer learning model fine-tuning steps with image preprocessing.

Input: Fundus Images Dataset

Output: Eye Disease Prediction

Procedure:

Download and install the software tool, i.e., Python and Jupyter Notebook editor.

Import some required Python libraries such as NumPy, Pandas, Matplotlib, Keras, Sklearn, and TensorFlow.

Step 1. Collect and load the fundus images dataset from the GitHub repository consisting of 250 images belonging to five classes, i.e., healthy, glaucoma, myopia, maculopathy, and retinitis pigmentosa (train: 202 images, and test: 48 images).

Step 2. Perform the image preprocessing with several required phases:

- Resize all images to a typical dimension of 224×224 pixels to ensure uniformity for input to the model.
- Split the fundus images into three channels Red, Green, Blue (RGB).

- Analyze the characteristics of every channel to determine to which channel the contrast enhancement technique should be applied.

Step 3. To enhance image quality using CLAHE:

- Employ CLAHE to the GREEN channel to enhance the fundus images' contrast.
- Reconstruct the improved colour fundus images by combining the GREEN channel with the RED and BLUE channels from the original images.

Step 4. Model architecture and training

- Use pre-trained weights (ImageNet weights) to load the DenseNet-121 model in order to take use of the features that have been learnt.
- Frozen DenseNet-121 model layers for maintenance of pre-trained knowledge.
- Add more layers of classification to the DenseNet-121 model to tailor it to the particular fundus image classification task.

Step 5. Configure the hyperparameters as follows:

- Adam Optimizer
- Categorical Cross-Entropy Loss Function
- Batch Size=4
- Number of Epochs =30

Step 6. Model Training

- With the preprocessed training dataset, instruct the DenseNet-121 model modification.
- The weights of the pre-trained DenseNet-121 layers remain frozen throughout training, while the newly added classification layers are modified to meet the individual task.

Step 7. Evaluate the trained model on the preprocessed test dataset to assess classification accuracy and efficacy.

Step 8. Calculate the f1-score, specificity, model's accuracy, sensitivity, recall, and precision, as well as any other relevant performance metrics.

Step 9. Examine the model assessment outcomes to ascertain the model's performance in classifying fundus images.

Step 10. Interpret the performance measures and assess them against existing model.

Step 11. Based on the outcomes of evaluation and analysis, conclude the research.

4. EXPERIMENTAL FINDINGS AND ANALYSIS

The Jupyter Notebook editor and the Python simulation tool are used for this study's implementation. Python is great for fast study and real-world application. It was Guido van Rossum who developed Python, a highly productive high-level programming language. The open-source online program Jupyter Notebook allows users to create and share documents, including visualisations, live code, equations, or narrative prose. This section covers a variety of topics, including performance markers, simulation outcomes, and dataset descriptions. The illustration provided a clear visual representation of the experimental outcomes in different graphs, and tables.

4.1 Dataset statistics

Kaggle's public dataset

(<https://www.kaggle.com/linchundan/fundusimage1000/>, accessed on 19 October 2021). provided the medical pictures. They selected a subset of the dataset that included retinal specialist-verified images of four distinct diseases as well as healthy eyes. Only 250 fundus images representing 5 different disease types (72 maculopathy, 22 retinitis pigmentosa, 75 glaucoma, 49 myopia and 32 normal images showing no disease) make up the whole collection shown in Table 3. 48 samples from each set were used for testing, and a total of 202 instances were used for training. About 2000 by 3000 pixels is the original image resolution. As to accommodate the pre-trained networks, they are resized to 224 by 224 pixels. The original data's colour information is maintained. The following Figures 9, 10, and 11 are the analysis and visualisation of datasets.

Table 3. Image classification in the database

Classes	No. of Images
Glaucoma	75
Maculopathy	72
Myopia	49
Retinitis pigmentosa	22
Healthy	32

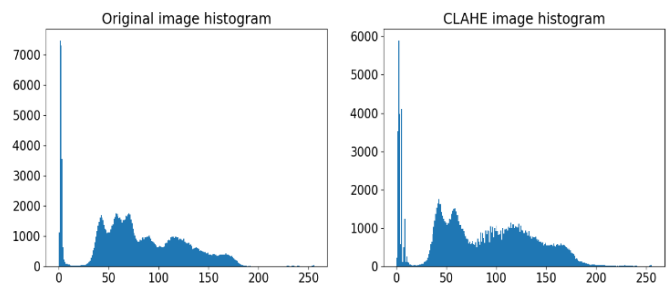


Figure 9. Histogram graph

Figure 9 shows histograms comparing the original image and the CLAHE-enhanced image. The original image has low contrast with intensity values concentrated on the lower end, while the CLAHE histogram is more spread out, indicating improved contrast and better visibility of details.

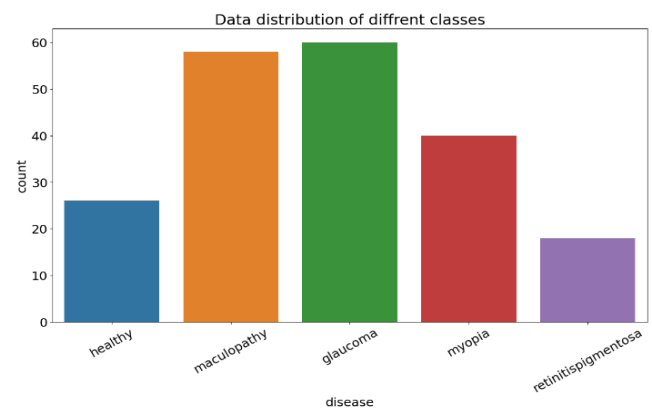


Figure 10. Data distribution bar graph of different classes

Figure 10 displays a bar graph of the data distribution for the various classes. This graph depicts the proportion of a dataset's samples or instances that fall into various classes. In this instance, there are five distinct categories: healthy, glaucoma, myopia, maculopathy, and retinitis pigmentosa.

The number of occurrences in a given class is expressed by the height of the bar representing that class. These classes are depicted on x-axis, and no. of counts are depicted on y-axis. According to this figure, the glaucoma class has maximum images than the other classes.

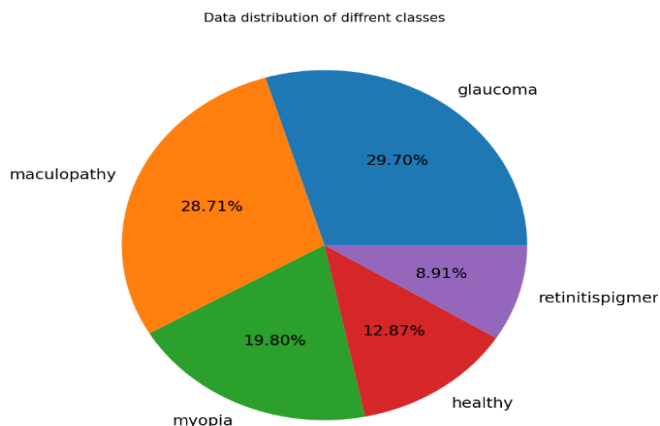


Figure 11. Data distribution pie graph of different classes

A data distribution pie graph is shown in Figure 11. A pie graph, sometimes called a pie chart, illustrates the relative size of several classes within a given dataset. To be more specific, the dataset is broken down into the following five classes: healthy (12.87%), glaucoma (29.70%), myopia (19.80%), maculopathy (28.71%), as well as retinitis pigmentosa (8.91%). The size of the slice of the pie that represents a given class indicates the class's relative importance to the whole dataset. The pie chart clearly illustrates the distribution of instances throughout the classes and gives a quick visual overview. Nearly 30 per cent of the data is associated with glaucoma, followed by 28.71 per cent from maculopathy, 19.80 per cent from myopia, 12.87 per cent from healthy, and 8.91 per cent from retinitis pigmentosa. This graphical representation facilitates rapid comprehension of the data's class distribution and identification of any potential class imbalances, which can be crucial for model development and analysis-related decisions.

4.2 Performance evaluation metrics

This section defines the many measures that are used to assess the performance of a classification model and suggests which metrics work best for that purpose [20].

4.2.1 Confusion matrix

When a classification model can be tested using true values, the classifier's prediction results are represented graphically in the confusion matrix. The binary and multiclassification types of classification are covered here.

Figure 12 shows the confusion matrix of a binary classifier. Actual values are represented by true (1) and false (0), but predicted values are represented by negative (0) and positive (1). Estimates of the potential categorisation models are derived using the words TP, TN, FP, and FN in the confusion matrix.

A confusion matrix for the categorisation of four distinct classes is shown in Figure 12. Four-class categorisation is the division of examples (instances) into four groups. There are four classes: "Class A, B, C, and D."

Class designation		Actual class	
		True (1)	False (0)
Predicted class	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure 12. Confusion matrix of binary classification

- **TP (True Positive):** A data point in the confusion matrix is TP if and only if the result matches what was anticipated.
- **FP (False Positive):** In the confusion matrix, a false positive will be displayed if a positive outcome is anticipated but a negative consequence really transpires.
- **FN (False Negative):** In the confusion matrix, a false negative arises when an anticipated negative consequence really transpires as a positive outcome.
- **TN (True Negative):** This situation is referred to as TN when there is a negative predicted consequence and a negative actual outcome.

4.2.2 Accuracy

One of the simplest classification metrics to compute is accuracy, which is the proportion of correct predictions to all guesses.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

4.2.3 Precision

The precision ratio is calculated by dividing the total number of expected positive samples by the percentage of real positive samples.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

4.2.4 Recall (Sensitivity)

Recall, on the other hand, refers to the proportion of true positive samples predicted against total number of available positive samples. It's also known as sensitivity or hit rate.

$$Recall / Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

4.2.5 F1-Score

An algorithm for binary classification model's positive class predictions is called an F1 Score or F-score. Recall and accuracy are used in calculations. It's a composite score for recall and accuracy. An ideal F1 Score would be the harmonic average of accuracy and recall.

$$F1-Score = 2 * \frac{precision * recall}{precision + recall} \quad (4)$$

4.2.6 Specificity

The ratio of True negatives and TN+FP is the measure of specificity. They desire a high value for specificity. Its value is between 0 and 1.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

4.2.7 ROC (Receiver operating characteristic) area

The compromise between TP Rate and FP Rate is seen on the ROC curve in Figure 13. For each cutoff, they compute and display the TP Rate and FP Rate. Lowering the FP Rate and raising the TP Rate causes the respective thresholds to rise. An individual ROC curve's area under the curve (AUC) can be used to measure how good it is. How well the model ranks predictions is measured by the AUC and ROC scores. Shows how likely it is that a positive outcome will be rated higher than a negative one based on a random selection. It is shown between the X-axis for FPR and the Y-axis for TPR. The model performs much worse than a random guessing model if the value is less than 0.5.

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN} \quad (7)$$

The key metrics for assessing the ocular illness classification model are recall, f1-score, accuracy, precision, Specificity, Sensitivity, as well as ROC-AUC, which are important for predicting negative cases accurately, capturing true positives, and minimising false positives, accordingly. These metrics are consistent with the medical context, emphasising proper recognition of positive as well as negative instances for successful eye diagnosis of illness.

4.3 Proposed model results and discussion

The training model consists of a pair of phases, which they called the fine-tuning step and transfer learning phase. The model converges swiftly across a number of epochs (Figures 13 and 14), demonstrating the benefit of transfer learning in accelerating the learning curve.

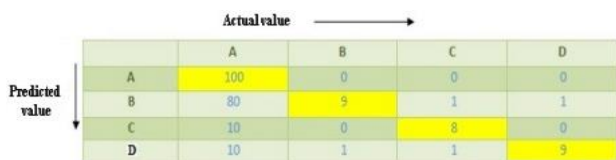


Figure 13. Four-class classification problem confusion matrix

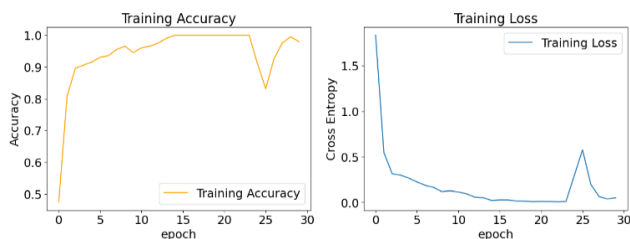


Figure 14. Training accuracy graph of DenseNet-121 model

Figure 14 displays the training accuracy line graph for the suggested DenseNet-121 model. The y-axis in this picture displays the accuracy, and the epochs values are shown in % on the x-axis. The training accuracy is low at the initial epoch,

then the accuracy is increased at 2nd epochs up to 90% approx. Then, the training accuracy is increased consistently up to 30 epochs, and it reaches around 97% accuracy.

After the trained model was obtained, it was used to analyse the left-over test data. The model accurately distinguished between the five groups tested, as shown in Figures 15 and 16.

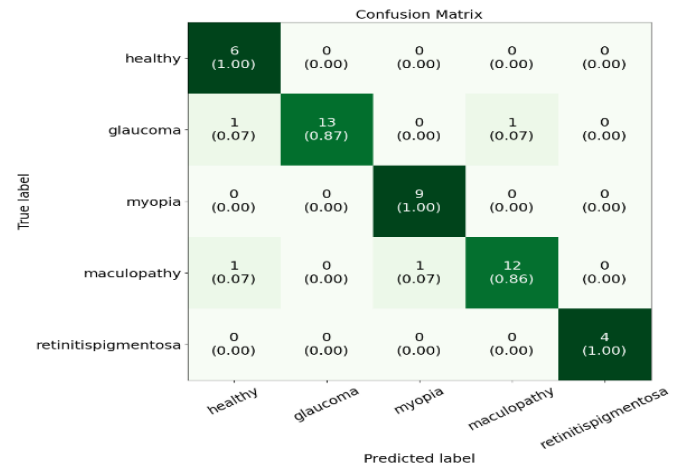


Figure 15. Confusion matrix of proposed DenseNet-121 model

Figure 15 shows the confusion matrix for the mentioned DenseNet-121 model. This confusion matrix shows the multi-classification for multiple data classes, i.e., healthy, glaucoma, myopia, maculopathy, and retinitis pigmentosa. It provides the count of FP, TP, and FN predictions made for each class. The number of correct predictions provided for each class appears along the diagonal of this matrix.

	precision	recall	f1-score	support
healthy	0.75	1.00	0.86	6
glaucoma	1.00	0.87	0.93	15
myopia	0.90	1.00	0.95	9
maculopathy	0.92	0.86	0.89	14
retinitispigmentosa	1.00	1.00	1.00	4
accuracy			0.92	48
macro avg	0.91	0.94	0.92	48
weighted avg	0.93	0.92	0.92	48

Figure 16. Classification report of proposed DenseNet-121 model

The suggested DenseNet-121 model's classification report for the fundus pictures dataset is shown in Figure 16 above. This dataset has multiclass such as healthy, glaucoma, myopia, maculopathy, and retinitis pigmentosa. There are different performance parameters results are given in this report. According to this report, the highest results are 100% precision for glaucoma, and retinitis pigmentosa, 100% recall, for healthy, myopia, and retinitis pigmentosa, 100% f1-score for retinitis pigmentosa, 94% of macro average, 93% of weighted average and the overall accuracy is 92%, respectively.

The outcomes in terms of accuracy, sensitivity, and specificity were outstanding. Figure 17, depicts the ensuing confusion matrix. The findings show that when it comes to transfer learning, DenseNet-121 can classify many eye diseases using fundus images with astounding accuracy using

very small labelled data.

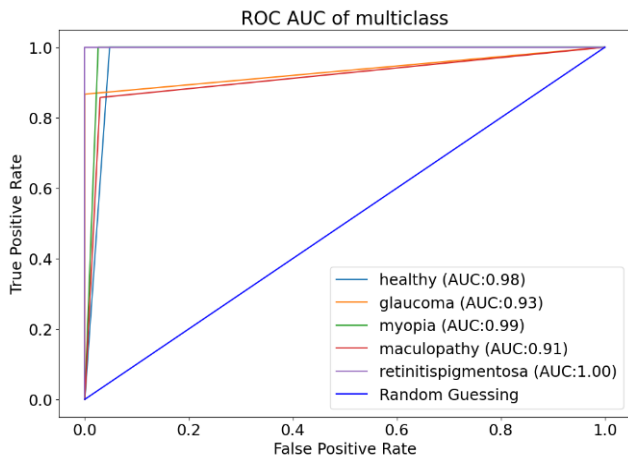


Figure 17. ROC-AUC graph of multiclass

The above Figure 17 shows a ROC-AUC graph of multiclass for proposed DenseNet-121 model. The FPR, which shows the percentage of false-negative classifications, is shown on this graph's x-axis, and the y-axis displays the TPR which reflects correctly classified positive instances of the proportion. Typically, the ROC curve begins at (0, 0), which represents the lowest threshold at which all instances are classified as negative, and ends at (1, 1), which represents the highest threshold at which all instances are classified as positive. The curve illustrates the efficacy of the model across various thresholds. In this ROC curve, the AUC values for every class are 93% for glaucoma, 98% for healthy, 99% for myopia, 100% for retinitis pigmentosa and 91% for maculopathy, respectively.

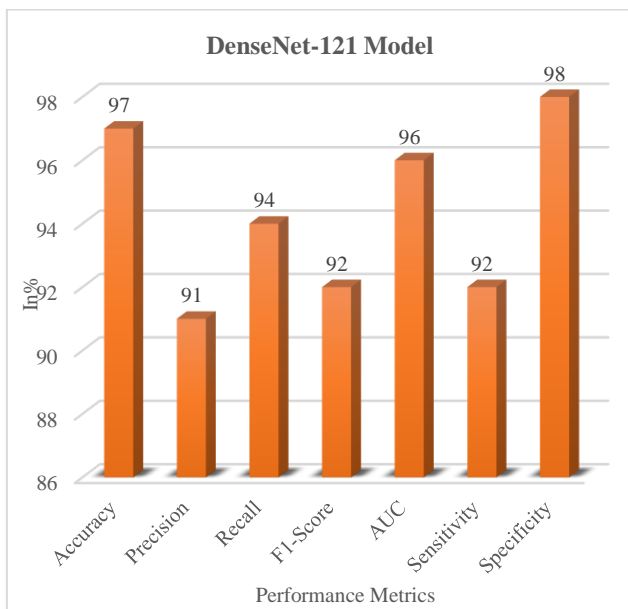


Figure 18. Performance measuring graph of DenseNet-121 model

Figure 18 shows a bar graph that represents the evaluation of performance results of proposed DenseNet-121 model. In the given graph, x-axis denotes the different performance metrics, and y-axis denotes the quantity in percentage. All assessment criteria were met by the DenseNet-121 model, which showed off flawless scores of 97% for accuracy, 91%

for precision, 94% for recall, 92% for F1-Score, 96% for AUC, 92% for sensitivity, and 98% for specificity. In conclusion, the DenseNet-121 model performed very well in terms of all performance metrics. The simulation results of proposed model are shown in Table 4.

Table 4. Simulation results of the proposed model

Model	Accuracy	Sensitivity	Specificity
DenseNet-121 Model	97	92	98

Table 4 shows the simulation results of proposed densenet-121 model in terms of precision, sensitivity, and specificity. The DenseNet-121 achieved highest 97% testing accuracy, 92% sensitivity, and 98% specificity. Similarly, Table 5 depicts the other performance results, i.e., precision, recall, f1-score, and AUC for the proposed model.

Table 5. Performance results of DenseNet-121 model

Model	Precision	Recall	F1-Score	AUC
DenseNet-121 Model	91	94	92	96

The DenseNet-121 model obtained extraordinary performance across all evaluation metrics, with highest 91% precision, 94% recall, 92% F1-Score, and 96% AUC. In summary, the DenseNet-121 model performed flawlessly, correctly recognizing positive instances and differentiating between positive and negative samples, making it very dependable for the specified task.

4.4 Comparative analysis between base and proposed models

Table 6 provide the comparison between base and proposed models in terms of performance measures for fundus image classification.

Table 6. Comparative result base and propose model

Model	Base Model	Propose Model
Accuracy	95	97
Precision	91	91
Recall	87	94
F1-Score	88	92
AUC	92	96
Sensitivity	87	92
Specificity	96	98

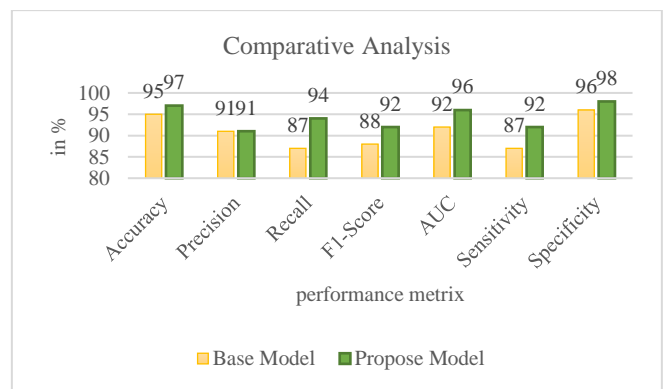


Figure 19. Comparison graph between base and proposed model

Figure 19 and Table 6 demonstrate the performance

comparison results between base and proposed models. The effectiveness of both the "Base" and "Proposed" models for multiclass classification has been evaluated in this analysis. The "Base" model's accuracy was 95%, demonstrating that 95% of the predictions were accurate. The precision was satisfactory (91%), meaning that of the cases assigned to classes, that many were accurately predicted, and recall is 87%. An F1-score of 88% indicates a good compromise between precision and recall. The ROC curve, or receiver operating characteristic, showed strong overall discriminating performance involving a region under the curve (AUC) of 92%. The model's sensitivity reached 87%, demonstrating its propensity to properly categorize positive examples, while its specificity, reflecting its propensity to correctly categorize negative examples, reached 96%. In contrast, the "Proposed" model performed well in every respect, coming in at 92% sensitivity, 94% recall, 96% AUC, 98% specificity, and 91% precision. When evaluated on the provided multiclass dataset, these perfect results demonstrate that the "Proposed" model accomplished the classification assignment admirably.

4.5 Discussion

Following transfer learning and fine-tuning, the proposed DenseNet-121 model demonstrated excellent performance in the classification of fundus images. Performance indicators for multiclass fundus image classification show a significant improvement with the proposed DenseNet-121 model over the baseline model. While the classic model's 95% accuracy was respectable, its sensitivity (87%), recall (87%), and precision (91%), were all below average. Based on both memory and accuracy, the 88% F1-score offered a solid balance. Even though the specificity rate was 96%, the ROC curve overall demonstrated extremely high discrimination with an AUC of 92%. With a 97% accuracy rate, 91% precision rate, 94% recall rate, and 92% F1-Score, the proposed DenseNet-121 model performed much better across the board. Strong discriminating skills were shown by the AUC, which reached an astounding 96%. In addition, the proposed model demonstrated outstanding proficiency in both positive and negative sample categorization with sensitivity of 92% and specificity of 98%. In general, the proposed model demonstrated notable improvements, offering a more precise and trustworthy answer to the challenging task of multiclass classification within the framework of detecting eye diseases using fundus images.

5. CONCLUSION, LIMITATIONS AND FUTURE WORK

This research has a fundus image or eye disease classification task used the pretrained DenseNet-121 model we did as well. The eye model used to predict whether there was any disease in an eye. This methodology entailed many processes, such as scaling and using CLAHE on GREEN channel, which were expected to give fundus imaging higher capability for detecting blood vessels. The DenseNet-121 model was improved in its suitability for classifying fundus images through the process of pausing its layers and incorporating additional classification layers. By accurately and successfully classifying fundus images into their corresponding categories, the model demonstrated its viability and potential as an effective screening tool for early detection

of eye diseases based on its performance on the test dataset. Moreover, the execution surpasses all expectations. The proposed DenseNet-121 model achieved the highest 97% accuracy, 91% precision, 94% recall, 92% f1-score, 96% AUC, 92% sensitivity, and 98% specificity.

Notwithstanding the beneficial findings, more methodological improvement is required. Data augmentation tactics, or just adding more data, might improve the model's performance and resilience right from the start. To reduce the likelihood of biased predictions and guarantee adequate representation for all diseases, it is necessary to deal with the possibility of class imbalance.

Limitations: As a fundus image classification method, it has a number of drawbacks. With such a little dataset (only 250 photos), the model may not be able to generalise well or include all kinds of features. The model's capacity to generalise across populations is called into question by the existence of biases in the dataset. There is a lack of investigation into alternate preprocessing approaches and the algorithm's sole dependence on CLAHE for contrast enhancement, which might limit its capacity to handle a variety of image features. There is a need for an alternate evaluation approach that can consider potential biases and impartiality. The hyperparameters are not assessed exhaustively, so their performance-influencing effect remains unknown. What is more, the comparison based research skips the criterion of models interpretability and computational effectiveness of them. Through increase the size of the dataset, involving the various populations, deciding on different preprocessing techniques, and performing a complete hyperparameter tuning investigations, could be the main criteria that would improve the approach's robustness and might be recognized practically.

Future work: Subsequent investigations might explore a range of subjects to enhance the proposed fundus image categorization system. First, expanding the dataset with information from more populations will improve the generalizability of the model. To improve the model's learning and feature extraction, it might be helpful to investigate hyperparameter tweaking in detail and look at preprocessing methods other than CLAHE. Several organisations may aim for more equal estimates if they remove dataset biases and include fairness standards into the assessment process. Apart from this, modelling both efficiency and the interpretability of the algorithm and carrying research on the new developments in the interpretability would generately help build our understanding. In addition, the efficiency of the application can be validated using independently collected datasets that will test its durability and reliability in different medical fields. In general, the future of algorithms development should engage in replacing may weaknesses with strengths, assessing ethical aspects and widening the application in real-world conditions.

REFERENCES

- [1] Sattigeri, S.K., Gowda, D.N., Ullas, K.A., S, A.M., Professor, A. (2022). Eye disease identification using deep learning. *International Research Journal of Engineering and Technology*, 9(7): 974-978.
- [2] Aranha, G.D., Fernandes, R.A., Morales, P.H. (2023). Deep transfer learning strategy to diagnose eye-related conditions and diseases: An approach based on low-

- quality fundus images. *IEEE Access*, 11: 37403-37411. <https://doi.org/10.1109/ACCESS.2023.3263493>
- [3] Meenakshi, G., Thailambal, G. (2022). Categorisation and prognostication of diabetic retinopathy using ensemble learning and CNN. In 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, pp. 1145-1152. <https://doi.org/10.1109/ICOEI53556.2022.9777156>
- [4] Ramanathan, G., Chakrabarti, D., Patil, A., Rishipathak, S., Kharche, S. (2021). Eye disease detection using machine learning. In 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, pp. 1-5. <https://doi.org/10.1109/GCAT52182.2021.9587740>
- [5] Raza, A., Khan, M.U., Saeed, Z., Samer, S., Mobeen, A., Samer, A. (2021). Classification of eye diseases and detection of cataract using digital fundus imaging (DFI) and inception-V4 deep learning model. In 2021 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, pp. 137-142. <https://doi.org/10.1109/FIT53504.2021.00034>
- [6] Chelaramani, S., Gupta, M., Agarwal, V., Gupta, P., Habash, R. (2021). Multi-task knowledge distillation for eye disease prediction. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, pp. 3983-3993. <https://doi.org/10.1109/WACV48630.2021.00403>
- [7] Parra, R., Ojeda, V., Vázquez Noguera, J.L., García Torres, M., Mello Román, J.C., Villalba, C., Facon, J., Divina, F., Cardozo, O., Castillo, V.E., Castro Matto, I. (2021). Automatic diagnosis of ocular toxoplasmosis from fundus images with residual neural networks. In *Public Health and Informatics*, IOS Press, pp. 173-177. <https://doi.org/10.3233/SHTI210143>
- [8] Sanya, S. (2021). Detection of ocular diseases using ensemble of deep learning models. *International Journal of Engineering Research & Technology*, 10(9): 637-640. <https://doi.org/10.17577/IJERTV10IS090221>
- [9] Gulati, S., Guleria, K., Goyal, N. (2023). Classification of diabetic retinopathy using pre-trained deep learning model-DenseNet 121. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, pp. 1-6. <https://doi.org/10.1109/ICCCNT56998.2023.10308181>
- [10] Sharma, R., Gangrade, J., Gangrade, S., Mishra, A., Kumar, G., Gunjan, V.K. (2023). Modified EfficientNetB3 deep learning model to classify colour fundus images of eye diseases. In 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), Hamburg, Germany, pp. 632-638. <https://doi.org/10.1109/ICCCMLA58983.2023.1034676>
- 9
- [11] Raja Sarobin M., V., Panjanathan, R. (2022). Diabetic retinopathy classification using CNN and hybrid deep convolutional neural networks. *Symmetry*, 14(9): 1932. <https://doi.org/10.3390/sym14091932>
- [12] Shamia, D., Prince, S., Bini, D. (2022). An online platform for early eye disease detection using deep convolutional neural networks. In 2022 6th International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, pp. 388-392. <https://doi.org/10.1109/ICDCS54290.2022.9780765>
- [13] Yadav, S., Roy, N.K., Sharma, N., Murugan, R. (2022). Classification of retinal detachment using deep learning through retinal fundus images. In 2022 IEEE India Council International Subsections Conference (INDISCON), Bhubaneswar, India, pp. 1-6. <https://doi.org/10.1109/INDISCON54605.2022.9862901>
- [14] Ahmed, M.R., Ahmed, S.R., Duru, A.D., Uçan, O.N., Bayat, O. (2021). An expert system to predict eye disorder using deep convolutional neural network. *Academic Platform-Journal of Engineering and Science*, 9(1): 47-52. <https://doi.org/10.21541/apjes.741194>
- [15] Elsharif, A.A.E.F., Abu-Naser, S.S. (2022). Retina diseases diagnosis using deep learning. *International Journal of Academic Engineering Research (IJAER)*, 6(2): 11-37.
- [16] Gonzalez, S.L., Woods, R.C., Eddins, R.E. (2018). *Digital Image Processing Using MATLAB*. Gatesmark Publishing.
- [17] Fadzil, M.H.A., Nugroho, H.A., Nugroho, H., Iznita, I.L. (2009). Contrast enhancement of retinal vasculature in digital fundus image. In Proceedings - 2009 International Conference on Digital Image Processing, ICDIP 2009. <https://doi.org/10.1109/ICDIP.2009.32>
- [18] Sivasangari, A., Ajitha, P., Bevisjenila, Vimali, J.S., Jose, J., Gowri, S. (2022). Breast cancer detection using machine learning. In: Shakya, S., Bestak, R., Palanisamy, R., Kamel, K.A. (eds) *Mobile Computing and Sustainable Informatics. Lecture Notes on Data Engineering and Communications Technologies*, Springer, Singapore, 68(07): 693-702. https://doi.org/10.1007/978-981-16-1866-6_50
- [19] Saini, A., Guleria, K., Sharma, S. (2023). An efficient deep learning model for eye disease classification. In 2023 International Research Conference on Smart Computing and Systems Engineering (SCSE), Kelaniya, Sri Lanka, pp. 1-6. <https://doi.org/10.1109/SCSE59836.2023.10215000>
- [20] Vujović, Ž. (2021). Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications*, 12(6): 599-606. <https://doi.org/10.14569/IJACSA.2021.0120670>