Prediction of cardiovascular risk factors from retinal fundus images using deep learning algorithms

A Project Report

Submitted by

Washim Akram (Scholar No. 2014085)

Gitanjali Deka (Scholar No. 2014108)

Kuppireddy Varshith Reddy (Scholar No. 2014130)

Under the supervision of

Dr. R. Murugan

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY SILCHAR,

SILCHAR, ASSAM, INDIA-788010

MAY, 2024

DECLARATION

We hereby declare that the project entitled "Prediction of cardiovascular risk factors from retinal fundus images using deep learning algorithms" submitted for the B. Tech. (ECE) degree is our original work and the project has not formed the basis for the award of any other degree, diploma, fellowship, or any other similar titles.

Signature of Students

Place:

Date:

CERTIFICATE

This is to certify that the project titled "Prediction of cardiovascular risk factors from retinal fundus images using deep learning algorithms" is the bona fide work carried out by WASHIM AKRAM, Registration no. 2014085, GITANJALI DEKA, Registration no. 2014108 and KUPIREDDY VARSHITH REDDY, Registration no. 2014130, students of BTech (ECE) of National Institute of Technology Silchar (An Institute of National Importance under MHRD, Govt. of India), Silchar, Assam, India during the academic year 2023-24, in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology (Electronics and communication Engineering) and that the project has not formed the basis for the award previously of any other degree, diploma, fellowship or any other similar title.

Name and Signature of the Supervisor

Place:

Date:

NATIONAL INSTITUTE OF TECHNOLOGY SILCHAR, ASSAM, INDIA

ABSTRACT

According to estimates from the World Health Organisation, around 17 million fatalities worldwide occur as a result of cardiovascular disorders each year. It is a class of diseases that includes aortic aneurysms, carditis, valvular heart disease, congenital heart disease, hypertensive heart disease, stroke, heart failure, rheumatic heart disease, cardiomyopathy, abnormal heart rhythms, peripheral artery disease, thromboembolic disease, and venous thrombosis. Risk factors include habits like smoking, behaviours, or conditions that can increase a person's risk of developing cardiovascular disease like age, blood pressure levels, haemoglobin levels etc. Hence early and automatic prediction tools are critical for cardiovascular risk predictions.

Convolutional Neural Networks are the class of deep learning neural network architecture, predominantly used for computer vision and image classification tasks, which is why it finds its purpose in medical image processing and detection as well. CNN is a great feature extractor, thus using it to categorize medical images will help us avoid difficult and expensive feature engineering and might be a strong alternative to provide computerized and non-invasive pre diagnosis model for patient risk retrieval.

The proposed CNN based architecture is trained on a dataset of annotated retina fundus images, consisting of both normal and abnormal samples i.e., with confirmed cardiovascular diseases and the network learns to extract relevant features and patterns after feeding the training and testing input image datasets. Our experimental results intend to demonstrate the effectiveness of the deep learning model, and draw comparative analysis among the prediction results and evaluation metrics of all three pre-trained CNN architecture used, namely Inception v3, Resnet-50 and VGG-19. The proposed system holds great promise for assisting clinicians in the non-invasive detection and diagnosis of heart disease, potentially improving patient outcomes, and reducing healthcare costs. Moreover, the CNN-based approach can be extended to other types of medical image analysis tasks, contributing to the advancement of computer-aided diagnosis systems in the field of cardiology.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all those who have contributed to the completion of this project report. First and foremost, we would like to thank our supervisor Dr. R. Murugan, for their invaluable guidance, support, and encouragement throughout the project. Their expertise and feedback have been instrumental in shaping our research and improving the quality of our work. We would also like to thank all the panellists for their contributions to the project. Their input, insights, and assistance have been greatly appreciated and have helped us to complete the project successfully. We are also grateful to the National Institute of Technology, Silchar for providing us with the necessary resources and facilities to carry out our research. Finally, we would like to express our heartfelt thanks to our family and friends for their constant support and encouragement, and for being a source of inspiration and motivation. Thank you all for your contributions and support.

Washim Akram

Gitanjali Deka

Kuppireddy Varshith Reddy

LIST OF FIGURES

Table of Contents

Chapter 1 Introduction

1.1 Problem Definition

Cardiovascular diseases (CVDs) remain the leading cause of death globally, contributing to increased healthcare costs in numerous countries. The World Health Organisation (WHO) estimates that cardiovascular diseases (CVDs) account for 32% of all deaths globally, with low- and middle-income countries accounting for two thirds of these deaths. Moreover, cardiovascular disease (CVD) accounted for 38% of all early deaths in people under 70 due to non-communicable diseases (NCDs). Indians are more likely to have a higher relative risk burden, an earlier age of initiation, a higher case fatality rate, and a higher incidence of premature deaths due to the CVD pandemic. According to a 2014 World Economic Forum and Harvard School of Public Health research, CVD will cost India's economy almost $2.17 trillion between the years 2012 and 2030.

In the Prospective Urban Rural Epidemiology (PURE) study, participants from lower-income countries (primarily Indians, 83%) had significantly higher rates ($p < 0.001$) of major cardiovascular events and mortality (7.39 and 9.84 per 1000 person-years, respectively) than participants from high-income countries (3.64 and 2.19 per 1000 person-years, respectively) during the assessment of cardiovascular risk in 156,424 individuals using the INTERHEART risk score.

Blood pressure, cholesterol, hyperglycemia, and smoking behaviours are among the common factors linked to CVDs. These factors can be managed with medication and by taking certain preventative measures. However, variables like age, race, and family history of CVDs do not alter when taking medicine; therefore, complexity and nonlinearity of risk factors must be taken into account for reliable CVD prediction.

Few of the traditional CVD risk calculators are-

• The Systematic Coronary Risk Evaluation (SCORE) calculator, which forecasts the risk of fatal cardiovascular events over a ten-year period using factors such as gender, age, total cholesterol, systolic blood pressure, and smoking status.

• The Framingham risk score is a risk assessment tool that, like SCORE, is similar in that it estimates a patient's 10-year risk of myocardial infarction and cardiac death by considering risk factors like gender, age, total cholesterol, high-density lipoprotein (HDL) cholesterol, systolic blood pressure, and smoking habits.

According to WHO statistics, quitting smoking, maintaining a healthy diet, and engaging in regular physical activity can prevent 80% of early heart attacks and strokes. One of the most important aspects of preventing CVD is routinely monitoring and managing risk factors for heart disease and stroke, such as high blood pressure, high cholesterol, high blood sugar, or diabetes.

Early detection of CVD can help prevent hospitalisation and deterioration of CVD as well the irreversible damage to the heart that can occur when heart failure is left unmanaged. It can also help alert people and help them adopt better lifestyles and behaviours for CVD prevention before worsening conditions.

Some common types of cardiovascular diseases are-

- **Coronary heart disease**- The major blood channels that supply our heart with blood are the coronary arteries. Atheroma, or the accumulation of fat in the coronary arteries, can cause blockages or interruptions in the blood supply to the heart muscle, leading to coronary heart disorders. Atheroma-induced narrowing of the coronary arteries can limit blood flow to the heart muscle, resulting in angina (chest pain), and a heart attack can occur when a coronary artery is totally stopped.

- **Stroke-** It is a medical disorder that arises from disruptions in the blood circulation to the brain. Because blood gives our brains the oxygen and nutrition they need to function properly, any interruption or restriction in blood flow will cause our brain cells to begin to die, potentially leading to brain damage or even death. Thus, a stroke is a medical emergency that needs to be treated right away.

- **Peripheral arterial disease-** Also referred to as peripheral arterial disease, this condition arises when there is a blockage in the arteries that supply your limbs, mainly your legs. The most common symptom of peripheral arterial disease is pain in your legs when walking. This is usually in one or both of your thighs, hips, or calves.

- **Aortic disease-** The aorta, the body's major blood conduit, transports blood from the heart to the body's other organs. Aortic aneurysms, in which the aortic wall weakens and protrudes outward, are among the most prevalent types of aortic disorders. This can cause pain in our chest, back or abdomen (tummy).

In conclusion, prediction and early detection of cardiovascular diseases is critical in improving outcomes and increasing the chances of successful treatment. The earlier cardiovascular diseases can be predicted or diagnosed, the more likely it is that the patient can be more careful about the risk factors that might be a result of their lifestyles.

Following an accurate assessment of the risks associated with cardiovascular diseases, specific interventions can be implemented to lower that risk. These interventions may include the use of statins and dietary modifications, and they can also help patients make clinical decisions by supporting primary CVD prevention efforts.

1.1 Project overview/specifications

Deep learning, a branch of artificial intelligence and subclass of machine learning, has developed into a workable methodology for analysing complicated sources of data, such as medical images, due to the advancement of computing systems in recent years. Emerging studies on identification and prediction from ocular biomarkers of systemic diseases and implementation of deep learning tools in oculomics, shows promising nature of deep learning as pre-diagnosis medical tool. Also, along with retinal morphology analysis, deep learning can enhance our understanding of complex disorders, such as musculoskeletal diseases, traumatic brain injury, cardiovascular disease, renal impairment, Alzheimer’s disease, or anaemia detection.

Lesion identification, automatic diagnosis, medical picture segmentation, and medical image synthesis are some of the machine learning applications for images in the medical industry. Deep learning algorithms combined with medical imaging techniques can give rise to cost effective and more accessible pre-diagnostic tools.

To get a two-dimensional (2D) depiction of the three-dimensional (3D) retinal tissues projected onto the imaging plane, fundus imaging technologies use an elaborate microscope with an implanted image sensor that records the reflected light from the internal surface of the eye. Prior research has demonstrated a robust association between risk factors and anatomical features found in retinal fundus imaging, including blood vessels and the optical disc.

Changes in retinal fundus image due to cardiovascular diseases that form the basis of their detection are:

1.**Retinal Vascular Changes**: The venular widening and retinal arteriolar narrowing have been associated with a wide range of CVD risk factors in patients.

2.**Microvascular Abnormalities:** Structural or functional irregularities within the small blood vessels (micro vessels) of the retina.

3.**Arteriovenous Nicking**: If the retinal veins appear compressed at arteriovenous crossings, it can be indicative of hypertension, which is linked to CVD.

4.**Atherosclerosis:** Atherosclerosis, the buildup of plaque in arteries lead to narrowed and irregularly shaped vessels of retina.

5.**Embolisms:** Small blood clots can be visible in retinal arteries which suggest an increased risk of clot-related conditions like stroke or heart attack.

Researches too show the retinal vasculature and blood vessels in eye can help learn about the cardiovascular functions and potential risk factors for heart health. Systemic cardiovascular diseases like arterial hypertension, coronary heart disease, as well as obesity are all associated with structural vascular changes in the retina. Other reports evidenced changes in retinal structure as an early indicator of the presence and severity of coronary artery disease. In our study, we take into account only retinal vasculature changes and effects on Optic Cup to Optic Disc Ratio (CDR). Various diseases and their progressive stages can manifest different effects on both arteries and veins.

For instance, arterial narrowing is a recognized occurrence linked with hypertension, while venous dilation is associated with stroke and cardiovascular ailments. Hence, the classification of blood vessels as arteries and veins is very relevant to our work since the arterio-venous ratio (AV ratio) is strongly related to cardiovascular problems as it is typically used to identify hypertension, one of the most prevalent cardiovascular diseases. In ophthalmology and optometry, the cup-to-disc ratio (often abbreviated CDR) is a metric mostly used to evaluate the course of glaucoma. The "blind spot" in the eye, where blood vessels enter the retina and the optic nerve leaves, is anatomically located in the optic disc. The optic disc may have some typical cupping or it may be completely flat. The cup-to-disc ratio compares the diameter of the cup portion of the optic disc with the total diameter of the optic disc. The normal cup-to-disc ratio is less than 0.5 (0.3-0.5). Higher cup-to-disc ratio may imply glaucoma or other pathologies. If this CDR ratio exceeds or reduces from the range of 0.3-0.5, it determines the abnormal condition that indicates the CVD presence.

In our study, we focus on and propose the amalgamation of both these advance techniques- retinal fundus imaging and deep learning algorithm to predict the presence of cardiovascular risk factors. To detect cardiovascular diseases using CNNs, the algorithm is trained on datasets of labelled retinal fundus images, including both normal and diseased images (retinal images of patients suffering cardiovascular diseases). The CNN then learns to identify features and patterns in the images that are useful for distinguishing between different classes of images and predict the presence of CVD with utmost accuracy.

We selected these techniques considering their non-invasive nature and cost effectiveness of the proposed model. In conclusion, we intend on devising a model at par with the performance of SCORE and the Framingham risk score and account for the data anomalies present in the previous studies by focusing on confirmed diagnosis of other medical history. Finally, we want to design a model that can be generalized for a larger population of the world.

1.3 Hardware Specification

Processor: Intel Core i5 processor.

Memory: Up to 8GB of RAM memory.

Storage: Solid state drive (SSD) 256GB.

Operating System: Windows 11.

1.4 Software Specification

1. Language: Python

2. **Deep Learning Framework:** TensorFlow

3. Libraries: Keras, NumPy, OpenCV, pandas, sklearn, matplotlib

4. **Pre-trained Models:** VGG-19, ResNet-50, Inception V3

5. **Image Preprocessing:** Resize, Normalize, Data Augmentation

6. **Training Dataset:** Retinal fundus images dataset with labelled normal and cardiovascular diseases. (80%)

7. Validation Dataset: Separate retinal fundus images dataset with labelled normal and cardiovascular diseases. (10%)

8. Test dataset: Retinal fundus images dataset with unlabelled images to test the performance of the model based on evaluation metrics. (10%)

9. **Metrics for Evaluation:** Accuracy, Precision, Recall, F1-Score, Confusion Matrix.

Chapter 2
Literature Review

2.1 Existing System

[1] Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning.

Methodology-

The algorithm used in this article is the Inception-v3 neural-network architecture. It was used to predict various labels such as smoking status, gender, age, SBP, BMI, DBP, and HbA1c. The images were pre-processed and the neural network was trained using a distributed-stochastic-gradient- descent implementation.

Limitations-

- Future research might investigate whether the results of this study apply to images with smaller or bigger fields of view, as this study only examined images with a 45° field of view.
- Larger dataset or a population with more cardiovascular events required as overall dataset is small for deep learning.
- The databases lacked several crucial inputs for the cardiovascular risk calculators that were already in use, such as lipid panels and precise diabetes diagnoses. HbA1c and blood pressure were risk variables that were only present in one of the datasets.
- Some variables might be biassed because they were self-reported, such smoking status.

[2] Validation of a deep learning based retinal biomarker (Reti-CVD) in the prediction of cardiovascular disease: data from UK Biobank.

Methodology-

Binary classification deep-learning algorithm is used. It is based on a probability score derived from retinal photographs and CT scans. The algorithm assigns a probability score ranging from zero to one, with a high probability score meaning a high probability of the presence of coronary artery calcium (CAC).

Limitations-

- Misclassification Bias
- Variability in Retinal Photograph Quality
- Limited Generalizability
- Lack of External Validation
- Limited Comparison to Other Risk Assessment Tools.

[3] Development and validation of a deep learning algorithm using fundus photographs to predict 10-year risk of ischemic cardiovascular diseases among Chinese population.

Methodology-

Dataset of 411,518 individuals from Tongren Hospital in Beijing, Shibei Hospital in Shanghai was used, that contained fundus images of both eyes and medical information, including total cholesterol, BMI, sex, age, SBP and suffering from diabetes or not. Inception-Resnet-v2 was used with input size of 299×299. For the input, Inception-Resnet-v2 returned an output with a dimension of 8×8×1536 which was forwarded to 8×8 size average pooling layer and then flatten into a 1×1536-dimension vector. Next, a fully connected layer with a 32-unit dimension was applied. The Adam method was utilised to optimise network parameters, and data augmentation techniques such as random rotation (±30°), random horizontal flipping, and random cropping were applied. Area under the receiver operating characteristic curve (AUC) and coefficient of determination (R2) were used to assess the algorithm's performance.

Limitations-

• The 10-year ICVD risk was not derived from actual ICVD events in longitudinal investigations, but rather from cross-sectional data, which reduced the algorithm's dependability.

• The development dataset lacked information regarding the smoker's present status.

[4] Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs.

Methodology-

The algorithms were trained and validated using 216,152 retinal photos from five datasets sourced from Singapore, the UK, South Korea, and Singapore. To predict the likelihood of the existence of CAC (RetiCAC), a deep learning system was trained on a single dataset from a South Korean health screening centre. RetiCAC scores were categorised into tertiles, and using external test sets from South Korea, Singapore, and the UK Biobank, Cox proportional hazards models were used to assess how well RetiCAC may predict cardiovascular events. Aggregated analysis and guided backpropagation were used to create the saliency maps.

Limitations-

• The requirement for additional validation among various racial and national groups.

• Since the training set was sourced from health screening facilities, there may be selection bias in the data, which could lead to underrepresentation of the overall population.

• Due to the nature of administrative data, the survival models that include hospital inpatient and mortality data for the incident cardiovascular disease may have introduced some misclassification in the SEED and UK Biobank.

• To assess the performance changes for CAC prediction and directly compare the risk prediction of cardiovascular disease between CT-measured CAC score and RetiCAC, more research with a larger sample size of CAC data are required.

[5] Diagnosing Cardio Vascular Disease (CVD) using Generative Adversarial Network (GAN) in Retinal Fundus Images.

Methodology-

Network Model: To enable retraining of the current ImageNet for usage in addressing customised picture classification tasks, a modified Inception V3 architecture DL model evaluates the created synthetic retinal images. DL models developed using the Keras API and TensorFlow frameworks.

Stochastic gradient descent is used to train the suggested model on a batch size of 32 pictures. Adam optimizer is selected to implement the model in order to maximise its learning rate to 0.001.

A retinal vessel is used as the input picture by the U-Net Generator Model. To down sample an input image over a few layers up to the bottleneck level, it is fed to both the encoder-decoder generator and the U-Net model. The output contains the finished image at the appropriate size. The suggested U-Net made the task easier by bringing the links to be of same size between network layers.

Limitations-

- Lack of Clinical Validation
- Limited Sample Size
- Lack of Comparative Analysis
- Lack of External Validation
- Lack of External Expertise

[6] Cardiovascular Risks Detection Using Fundus Image

Methodology-

100 people were surveyed and the examination included taking of numerous fundus photographs of the persons left and right eyes, measuring the blood pressure, SCORE risk stratification and pulse wave. Additionally, small questionnaire was employed. Pre-processing methods like Manual Cropping, Vessel Enhancement and Outlier Removal employed. Pre-trained networks involved and model architectures considered are Mobile Net, Mobile Net V2, NASNet Mobile and InceptionV3.

Limitations-

Segmenting retinal structures became more difficult due to the disparity in image quality between retinographs and the EFS prototype, which is now under development at Fraunhofer. This meant that in order to minimise image flaws like reflections and blurring, manual cropping was required. The development of an automated pipeline for fundus picture capture and real-time cardiovascular risk prediction was hampered by the manual cropping step. Overall, the fundus photos from a younger population had superior vascular contrast and, most crucially, definition. It's probable that picture contrast had an impact on the results since younger individuals typically have healthier cardiovascular risk factors, which lowers the risk of cardiovascular disease. To validate this, more research and a larger dataset are needed. The labels were obtained using the SCORE calculator, which is a glaring drawback. Because of this, the findings are only as good as this calculator, which has an AUC of 0.72.

2.2 Proposed System

From the acquired dataset of 588 retinal images of cardiovascular disease patient and 518 normal retinal images, it is split into training, testing and validation dataset in the ratio 80%:10%:10% respectively. ResNet50, InceptionV3 and VGG19 are convolutional neural networks (CNNs) with different architectures and varying numbers of convolutional and fully connected layers. While VGG19 uses 16 convolutional and three fully connected layers with a small filter size of 3x3 to extract and classify features, InceptionV3 uses 48 convolutional and three fully connected layers and employs inception modules to combine different-sized filters, resulting in a more efficient network with fewer parameters. On the other hand, ResNet50 uses 50 convolutional and three fully connected layers and solves the vanishing gradient problem by using residual blocks.

Owing to the limited datasets we acquired till now, we drew comparative analysis of all three pre-trained models namely- VGG19, Resnet50 and Inception v3 to compare results of evaluation metrics and figure out the best classifier suitable for our proposed idea. Furthermore, each of the three models is individually trained on a training dataset, and then their predictions are combined on a validation dataset to create the ensemble model.

Inception-v3 resulted in a whopping 100% accuracy, proving to be the best classifier compared among all 3 tested architectures. And the Ensemble model gave an accuracy of 91%.

2.3 Feasibility Study

Three pre-trained deep learning models make up the ensemble model in our project: VGG19, ResNet50, and InceptionV3. Every one of these models has advantages and disadvantages in recognizing patterns and features in images. The ensemble model can increase picture classification accuracy by utilizing the advantages of each individual model by merging their predictions. These predictions can be combined using a variety of methods, such as weighting or average depending on performance on a training dataset.

Nonetheless, it might be difficult to train and operate ensemble models because to their increased complexity and computing demands, but the improved accuracy and robustness of the ensemble model can make the additional effort worthwhile.

Chapter

3

SYSTEM ANALYSIS & DESIGN

3.1 Flowchart

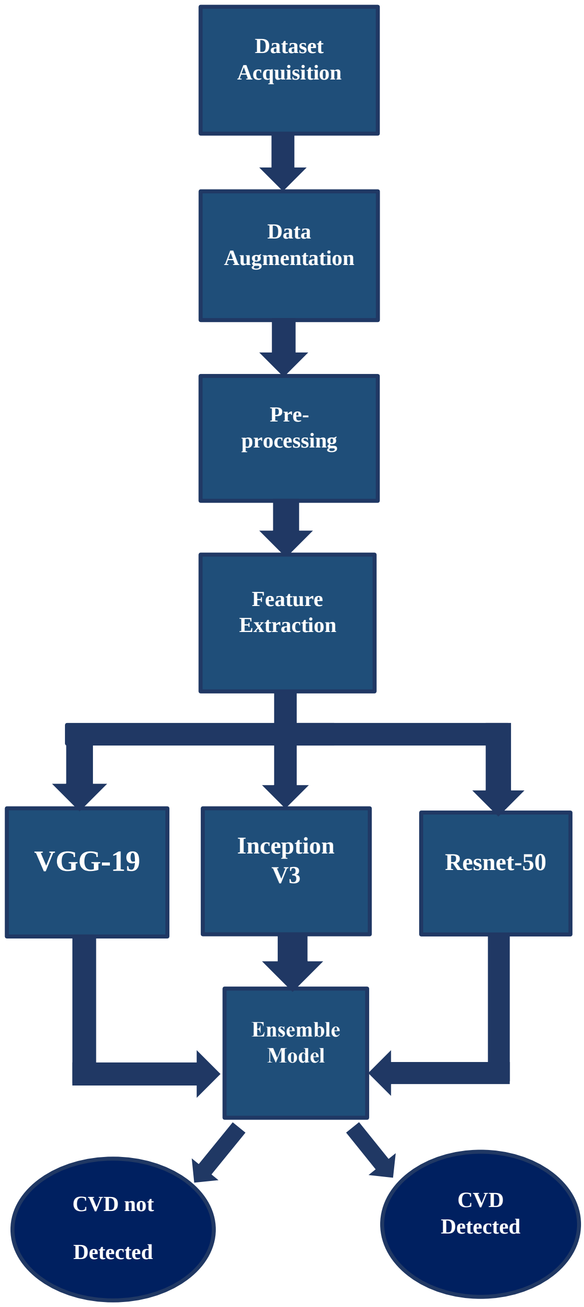


Fig 1: Workflow of the Project

3.2 Design and Test Steps / Criteria

In this step, we acquired color retinal images of people with both cardiovascular diseases present and not present, from publicly available datasets on the internet. The rear surface of the eye, or fundus, is its interior which is comprised of retina, macula, optic disc, fovea, and blood vessels. Convolutional Neural Network (CNN) training and testing will be conducted using these images.

Sources of Dataset Acquisition: MESSIDOR, CHASEDB1, STARE dataset.

Total Images Collected- 770 images

Total Images after Data Augmentation- 3400 images

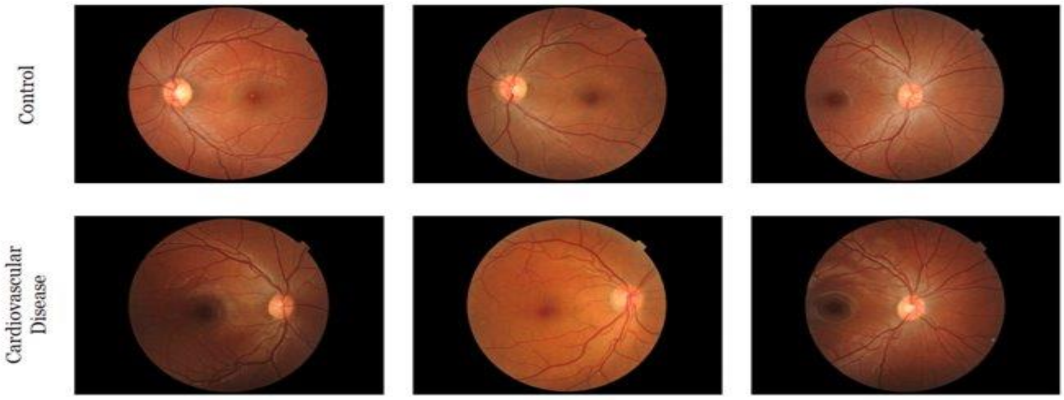


Fig 2: Control (Normal) and CVD affected retinal fundus image

Pre-Processing Techniques-

To make better predictions and build robust model, the fundus images must undergo some form of pre-processing technique. All the images must adhere to a certain standard before they are fed to the CNN for training. For image standardization, first we resize the images to a fixed resolution and then normalize the pixel range values. We resized images to 256 x 256 pixels to match the CNN input layer requirements.

3.3 ALGORITHMS- CNN architectures implemented in this project:

- VGG-19
- ResNet-50
- Inception V3 classification.
- Ensemble

3.3.1 VGG-19

The Visual Geometry Group at Oxford University created the convolutional neural network (CNN) known as VGG19 in 2014. It has 19 layers total, comprising 16 convolutional layers, 3 fully connected layers, and a softmax output layer. It is an expansion of the VGG16 network. VGG19 takes an input image of size 224 x 224 pixels and passes it through a series of convolutional layers, where each layer extracts a set of features from the image.

These features are then passed through a series of fully connected layers, which produce a final output prediction. The architecture of VGG19 is shown below:

• Input layer: The VGG19's input layer accepts a picture with a size of 224 x 224 x 3, where 3 stands for the RGB channels of the image.

• Convolutional layers: VGG19 has 16 convolutional layers with a stride of 1 pixel and a 3 × 3 filter size. As we go further into the network, the number of filters in each layer rises from 64 to 512.

• Max-pooling layers: VGG19 has a max-pooling layer with a 2 x 2 filter size and a stride of 2 pixels after every two convolutional layers. Max-pooling lowers overfitting and the spatial dimension of the feature maps.

• After the final convolutional layer, VGG19 employs three fully linked layers, each containing 4096 neurons. A softmax activation function is used to the output of the final fully connected layer to provide the final prediction.

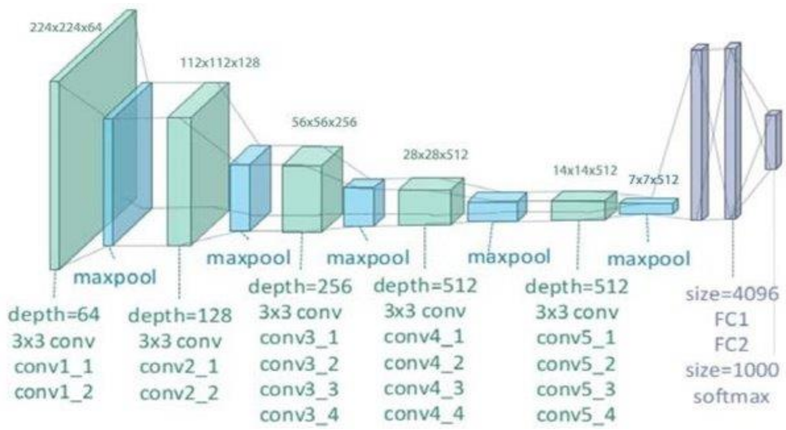


Fig 3: VGG-19 Architecture

One of the key benefits of VGG19 is its high accuracy on a wide range of image classification tasks, even when trained on limited data. This is due in part to the use of small convolutional filters (3x3) that enable the model to effectively learn and represent features at different scales. VGG19 can be fine-tuned on a new dataset by retraining the final layers of the network on the new data, while keeping the pre-trained layers fixed. This process of transfer learning can significantly reduce the amount of data required to train an effective image classifier.

Overall, VGG19 is a powerful and highly effective pretrained model that can be leveraged for a wide range of image classification tasks, including academic misconduct detection in online exams.

3.3.2 RESNET-50

Convolutional neural network (CNN) developers at Microsoft Research created ResNet50 in 2015. It has 50 layers total, including 49 convolutional layers and 1 fully connected layer, and it belongs to the ResNet family of networks. ResNet50's operation is as follows: ResNet50 takes a 224 × 224-pixel input picture and runs it through a succession of convolutional layers, each of which extracts a specific collection of features from the image. The network may then learn increasingly complicated characteristics from the picture by passing these features through a succession of residual blocks. The ResNet50 architecture is displayed below:

• Input layer: The input layer of ResNet50 accepts a picture with dimensions of 224 × 224 x 3, where x3 denotes the RGB channels of the image.

• Convolutional layers: Each of the 49 convolutional layers in ResNet50 has a 3 x 3 filter size and a 1-pixel stride. As we move further into the network, the number of filters in each layer rises from 64 to 2048.

• Residual blocks: ResNet50 employs a residual block after every two convolutional layers. Two convolutional layers, a batch normalisation layer, and a ReLU activation function are found in a residual block. After adding the output of the second convolutional layer to the block's input, a further ReLU activation function is applied. As a result, deeper networks may be trained, and the problem of disappearing gradients is helped by the network learning residual connections.

• Global average pooling layer: ResNet50 employs a global average pooling layer to create a 2048-element feature vector following the last residual block.

• Fully connected layer: A fully connected layer with 1000 neurons processes the output of the global average pooling layer to create the final prediction.

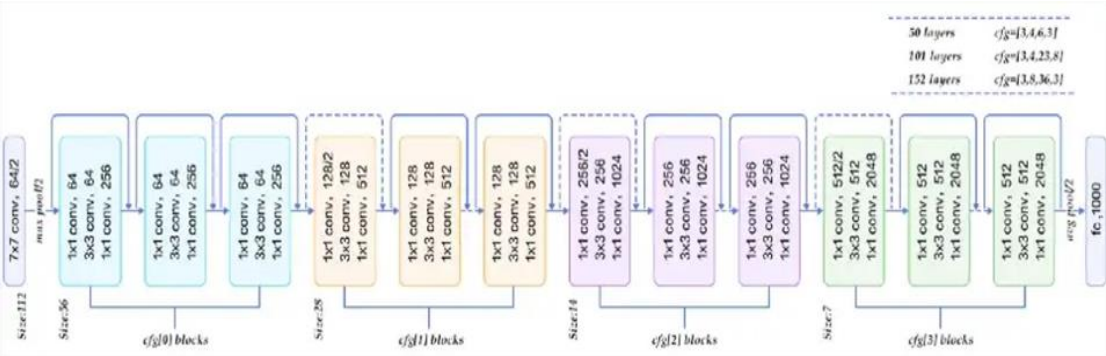


Fig 4: ResNet 50 Architecture

• **Skip Connection:**

Adding the initial input to the convolutional block's output. ResNet trains on F(X), unlike all the other algorithms, which train on the output 'Y'. To put it another way, ResNet seeks to make F(X)=0 such that Y=X. A direct link called a skip connection bypasses several model levels. This skip connection causes the output to differ. In the absence of the skip connection, input 'X is multiplied by the layer weights, and then added to a bias term. The activation function F () is then applied, and the result is F(w*x + b) (=F(X)). However, the output of the skip connection approach is F(X)+x. There are two kinds of blocks in ResNet-50,

1. Identity Block
2. Convolutional Block

If and only if the Input Size equals the Output Size, the value of "x" is added to the output layer. If not, we include a "convolutional block" in the shortcut path to make the input and output sizes equal. There are two ways of making the input size equal to the output size–

• Padding the input volume

• Performing 1*1 convolutions

Size of output layer is calculated using: [{n+2p-f)/s} +1] ^2

Where, n = input image size

p = padding

s = stride

f = number of filters.

For, 1*1 convolutional layer, size of output layer is equal to (n/2) *(n/2) given the input size is n.

3.3.3 INCEPTION V3

Convolutional neural network (CNN) developers at Google created InceptionV3 in 2015. It belongs to the Inception family of networks and has 48 layers total, including a softmax output layer, three fully linked layers, and 42 convolutional layers. The workings of InceptionV3 are as follows: InceptionV3 takes a 299 × 299-pixel input picture and runs it through a succession of convolutional layers, each of which extracts a specific collection of characteristics from the image. The network may then learn more complicated characteristics from the picture by passing these features through a number of Inception modules. Below is a diagram of InceptionV3's architecture:

• Input layer: InceptionV3's layer of input accepts a picture with dimensions of 299 X 299 X 3, where 3 stands for the RGB channels of the image.

• Convolutional layers: InceptionV3 has 42 convolutional layers, each with a different filter size and stride. There are between 32 and 1024 filters in each layer.

• Inception modules: To extract information from the input image, InceptionV3 employs a number of Inception modules. Multiple parallel convolutional layers with various filter sizes and strides can be found in an Inception module. A batch normalisation layer and a ReLU activation function are applied after concatenating the outputs of these convolutional layers.

• Max-pooling layers: InceptionV3 employs a max-pooling layer with a 3 x 3 filter size and a stride of 2 pixels after some of the Inception modules. Max-pooling lowers overfitting and the spatial dimension of the feature maps.

• After the last Inception module, InceptionV3 employs three completely linked layers, each containing 2048 neurons. A softmax activation function is used to the output of the final fully connected layer to provide the final prediction.

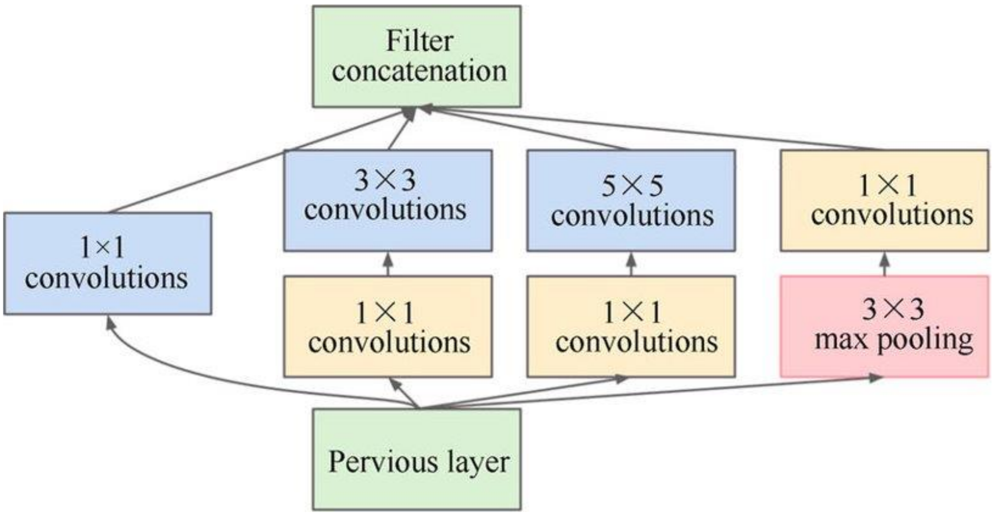


Fig 5: Inception V3 Architecture

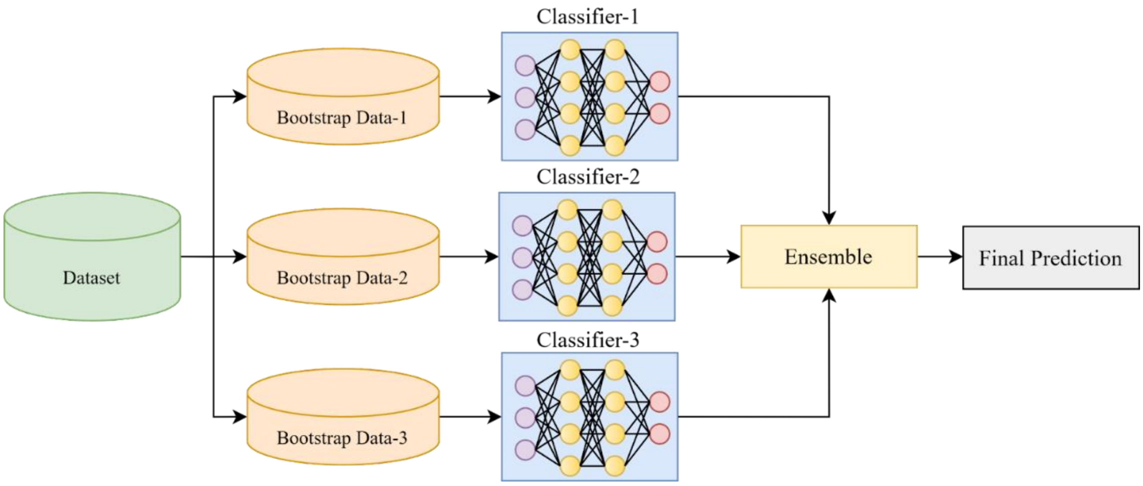
On several benchmarks, including the ImageNet dataset, which consists of more than 1.2 million pictures divided among 1000 classes, InceptionV3 has achieved state-of-the-art performance. Its usage of Inception modules, which make it easier to learn features at different sizes and resolutions and let it to learn more complicated features from the input picture and improve accuracy on image classification tasks, is attributed with its success.

3.3.4 ENSEMBLE MODEL-

An ensemble model involves training multiple models to collaborate and generate a unified prediction, with the premise that their combined forecast can surpass the accuracy of any single model. In the realm of computer vision, this technique can be implemented by integrating predictions from diverse deep learning models such as VGG19, ResNet50, and InceptionV3. By amalgamating these models, overall prediction accuracy can be enhanced due to the varied strengths and weaknesses of each. To construct an ensemble model with VGG19, ResNet50, and InceptionV3, each model undergoes individual training on a designated training dataset. Subsequently, their predictions are amalgamated using a validation dataset, resulting in the creation of the ensemble model.

Various approaches can be employed to merge the predictions, with a common method being simple averaging, where the final prediction is the average of the predictions generated by the three models.

Once the ensemble model is trained, it can be utilized to make predictions on a test dataset by applying each of the three models to the dataset and aggregating their predictions in a similar manner as done during validation.



` **Fig 6: Ensemble Model Architecture**

Advantages of the Ensemble model:

• Increased Accuracy: The ensemble model can outperform any single model in accuracy by integrating the predictions of other models.

• Robustness: Compared to a single model, an ensemble model is more resistant to biases or mistakes.

• Diversity: Because the three models were combined, the ensemble model was able to learn a more varied collection of characteristics. Each of the three models has strengths and shortcomings of its own.

• Generalization: The ensemble model, which has learned from a variety of sources, is better equipped to generalise to new data.

Chapter

4

Results of ResNet50, VGG-19, Inception-v3 and Ensemble Model

4.1 Terminologies

Plot loss (or error) vs. Epoch or Accuracy vs. Epoch graphs: To evaluate the convergence of neural network models, it is typical to analyse learning curve graphs. These graphs typically depict the progression of loss (or error) versus epoch, or accuracy versus epoch. It is customary to anticipate a decline in loss and an increase in accuracy with an increase in epoch during training. Eventually, we anticipate that both loss and accuracy will reach a plateau.

Confusion matrix: Measuring the performance of a classification problem with multiple output classes can be done easily with a confusion matrix. This type of matrix is essentially a table with two dimensions - "Actual" and "Predicted". Each dimension has four possible outcomes - "True Positives (TP)", "True Negatives (TN)", "False Positives (FP)", and "False Negatives (FN)", which can be seen in the example below.

Explanation of the terms associated with confusion matrix are provided below –

• **True Positives (TP)** − TP is the situation when both the actual class and the predicted class of a data point are 1.

• **True Negatives (TN)** − TN is the situation when both the actual class and the predicted class of a data point are 0.

• **False Positives (FP)** − FP is the situation when the actual class of a data point is 0 and the predicted class is 1.

• **False Negatives (FN)** − FN is the situation when the actual class of a data point is 1 and the predicted class is 0.

Precision

In the context of document retrieval, precision refers to the accuracy of our ML model in returning relevant documents. The calculation of precision can be derived using the confusion matrix and the formula: **Precision = TP/(TP+FP)**

Recall

The concept of recall refers to the count of positive results that are generated by our machine learning model. This can be determined using a confusion matrix and the application of the accompanying formula: **Recall = TP/(TP+FN)**

F1 Score

F1 Score is computed as the harmonic mean of these two values, and ranges from 0 (worst) to 1 (best). The formula to calculate F1 score is a weighted average of precision and recall, which can be expressed mathematically:

F1 = 2x (precision x recall) / (precision + recall)

F1 score is having equal relative contribution of precision and recall, hence it measures the balance between precision and recall.

ROC Curve

The Receiver Operating Characteristic (ROC) curve is a tool used to assess the effectiveness of binary classification algorithms, like machine learning models, by varying their discrimination threshold. The ROC curve shows the True Positive Rate (TPR) versus the False Positive Rate (FPR) at different threshold settings, where TPR represents the proportion of correct positive predictions out of all actual positive cases, and FPR represents the proportion of incorrect positive predictions out of all actual negative cases. Through the ROC curve, one can determine how well the algorithm can differentiate between positive and negative cases. An ideal classifier would show a TPR of 1 and an FPR of 0, positioned at the top-left corner of the ROC curve. Meanwhile, a random classifier would have an ROC curve that is diagonal, and an area under the curve (AUC) of 0.5. The AUC is an extensively used metric to compare binary classifiers' performance. Its value ranges between 0 and 1, where a higher AUC value indicates better classifier performance. A random classifier's AUC is 0.5, while a perfect classifier's AUC is 1.

To sum it up, the ROC curve is a beneficial tool that can provide insights into how well a binary classification algorithm is able to distinguish between positive and negative cases.

4.2 Overview and Performance metrics of various model implemented

Table: 1- Performance table of various implemented models

Parameters	ResNet50(%)	VGG-19(%)	Inception-v3(%)	Ensemble Model (%)
Accuracy	89.88	82.35	87.29	92.71
Precision (weighted average)	92.30	90.90	100	92.30
Recall	92.30	83.33	92.85	92.30
F1-score (weighted average)	92.30	86.95	96.29	92.30

Performance metrics of VGG-19 architecture-

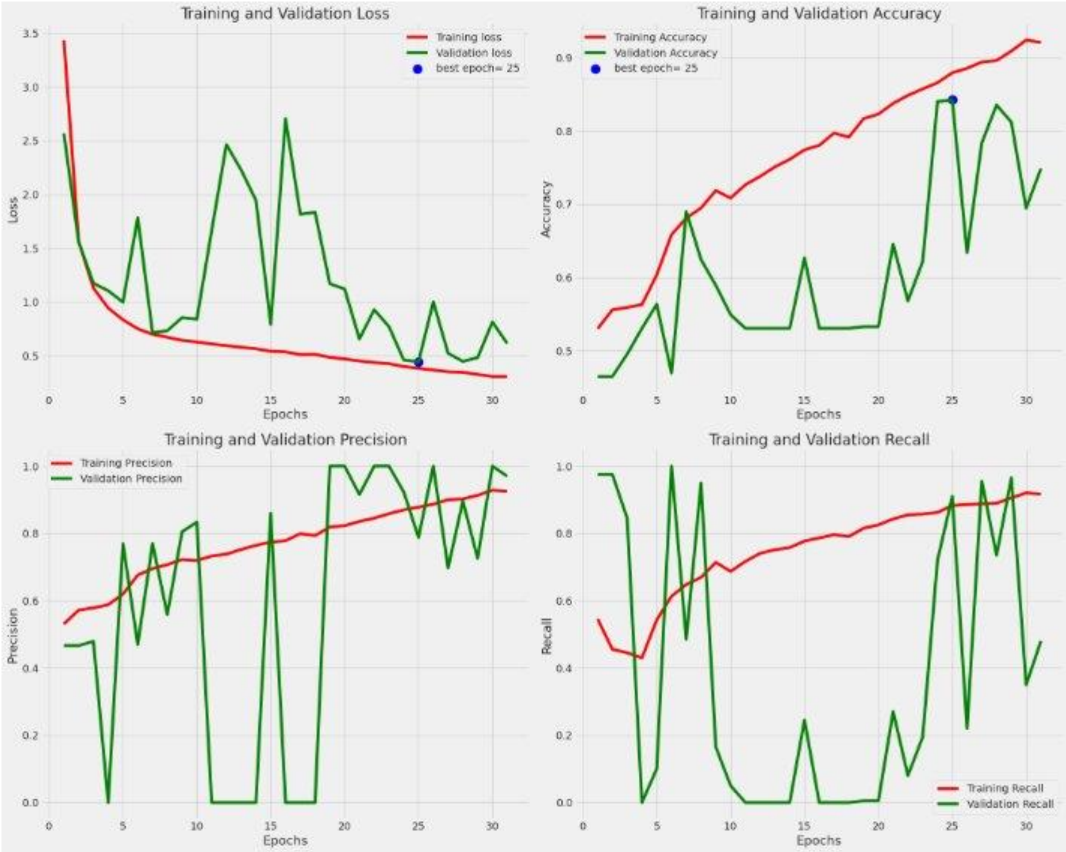


Fig-7: - Accuracy, Loss, Recall and Precision graphs of VGG19 Model

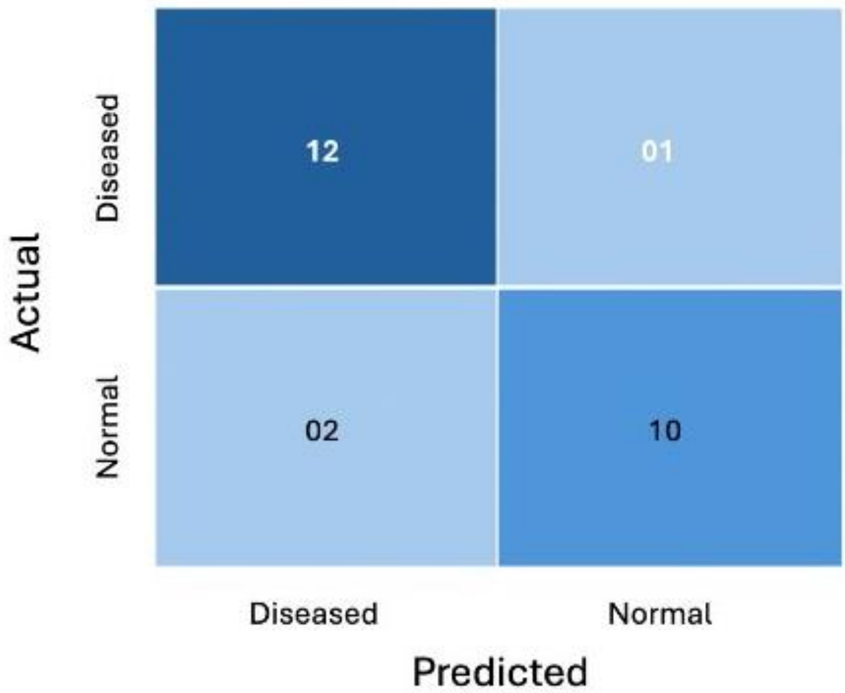


Fig 8: - Confusion Matrix of VGG-19 Model

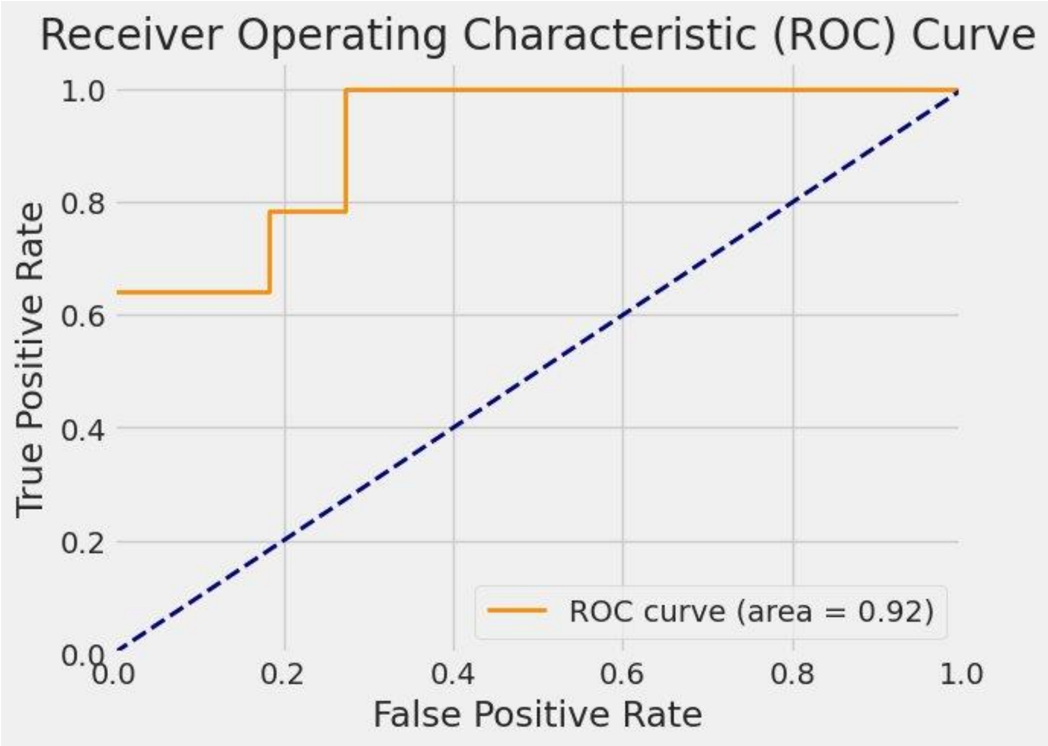


Fig 9:- ROC curve of VGG-19 model with parameters (Precision,Recall,F1-Score)

Performance metrics for ResNet50 architecture:

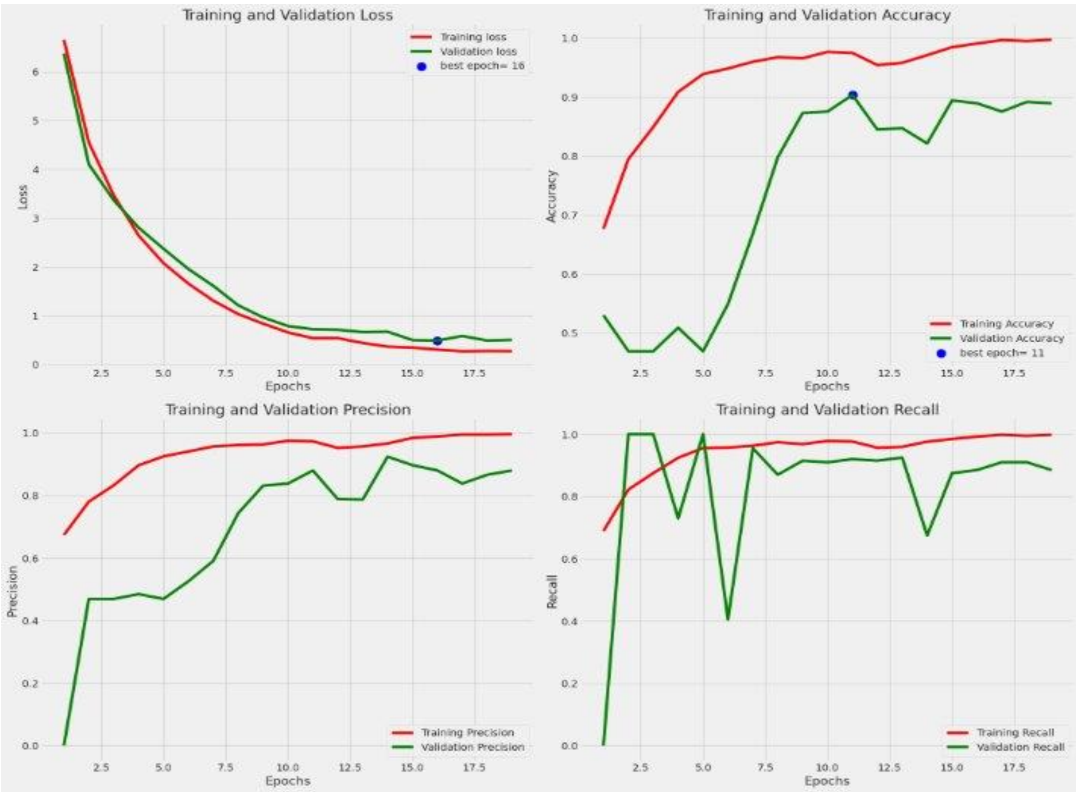


Fig-10: - Accuracy, Loss, Recall and Precision graphs of ResNet50 Model

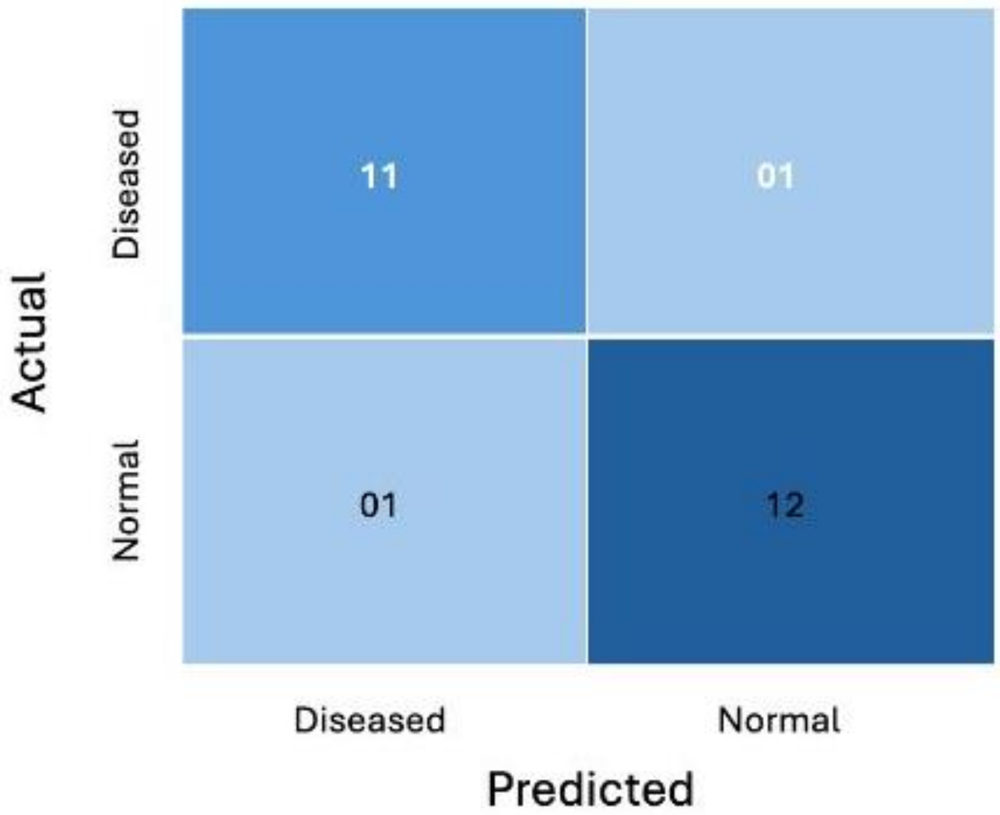


Fig 11:- Confusion Matrix of Resnet-50 model

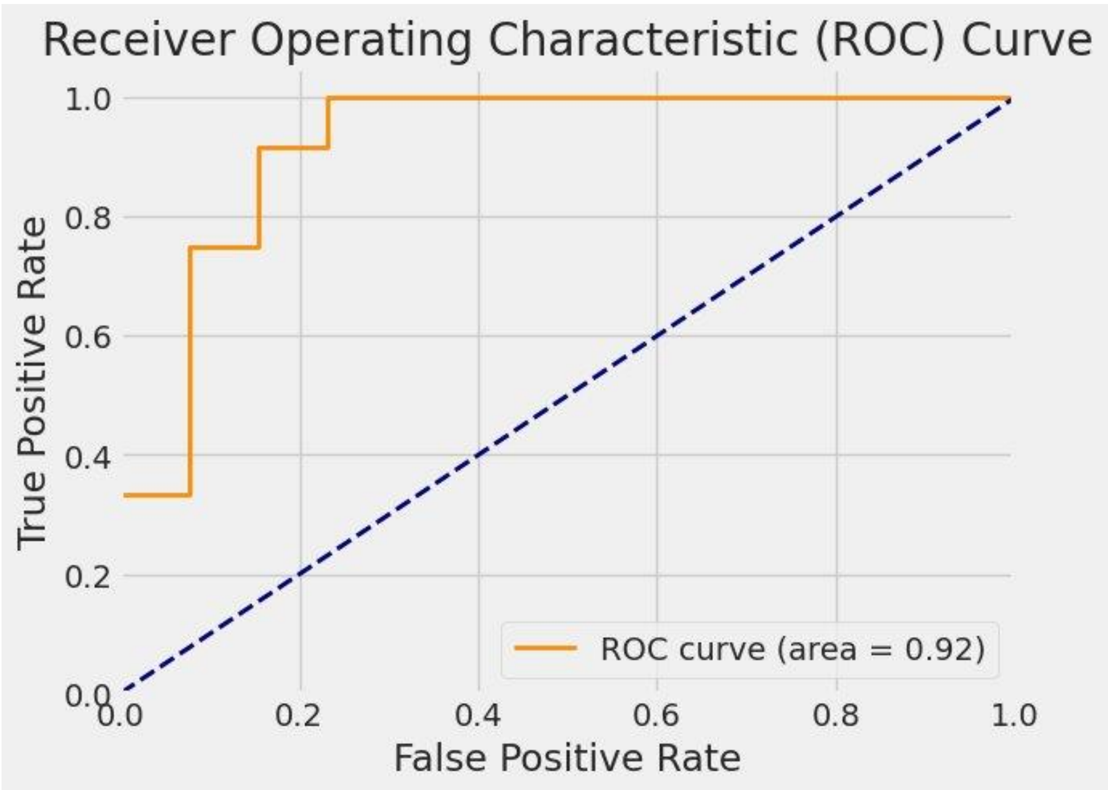


Fig 12: - ROC curve of Resnet-50 model with parameters (Precision,Recall,F1-Score) 50 model with parameters (Precision,Recall,F1-Score)

Performance metrics for Inception V3 architecture:

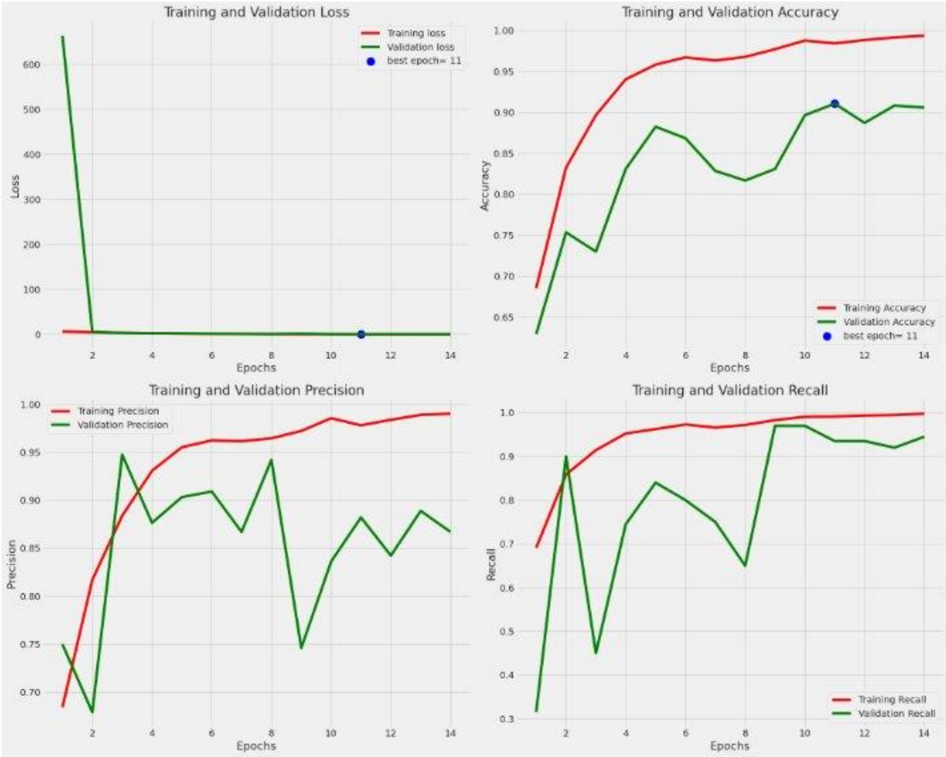


Fig-13: - Accuracy, Loss, Recall and Precision graphs of Inception -V3 Model

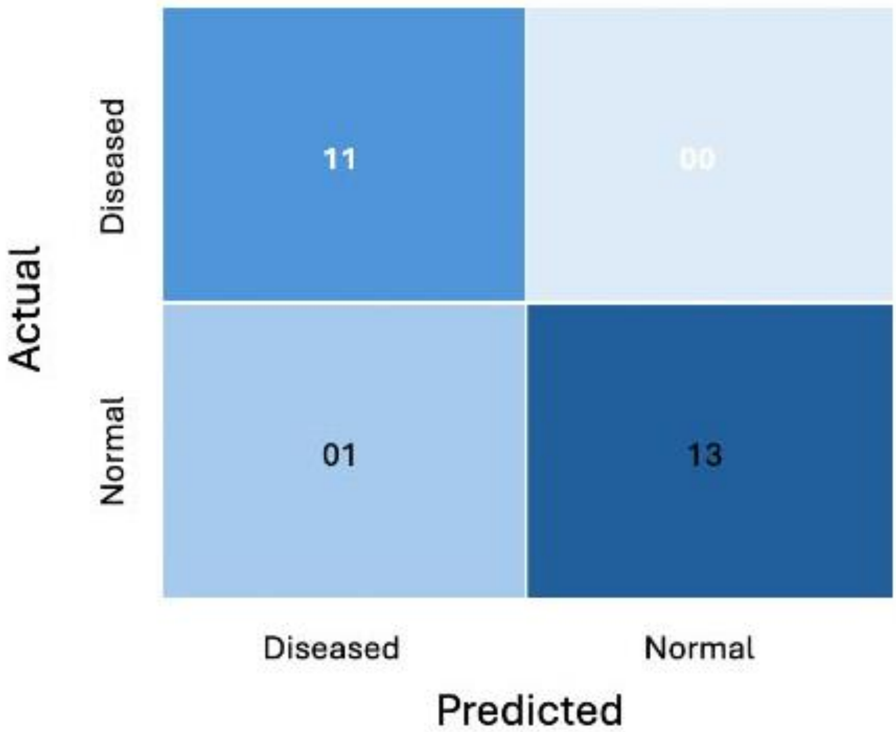


Fig 14:- Confusion Matrix of Inception-V3 model

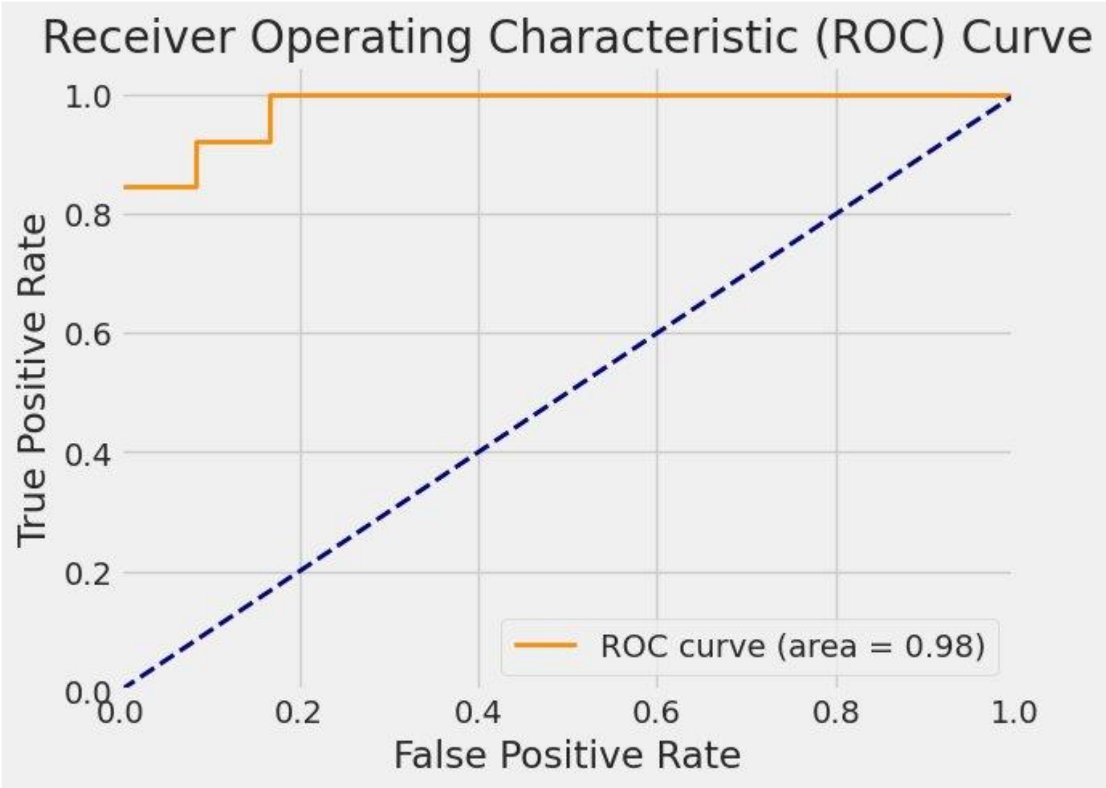


Fig 15:- ROC curve of Inception V3 model with parameters (Precision,Recall,F1-Score)

Performance metrics for Ensemble architecture:

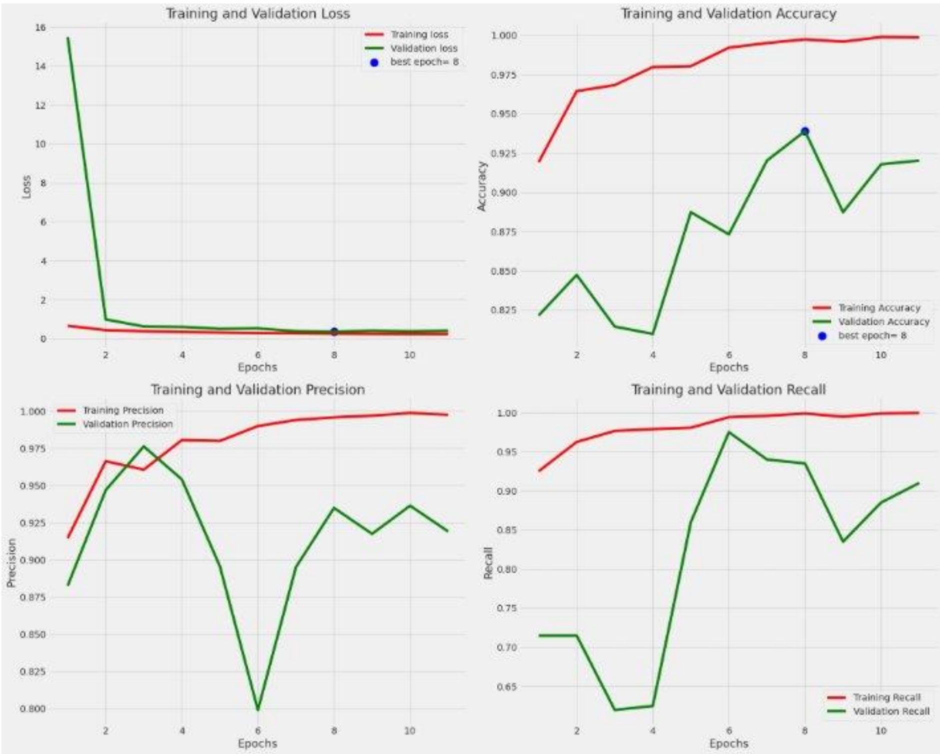


Fig 16: - Accuracy, Loss, Recall and Precision graphs of Ensemble model

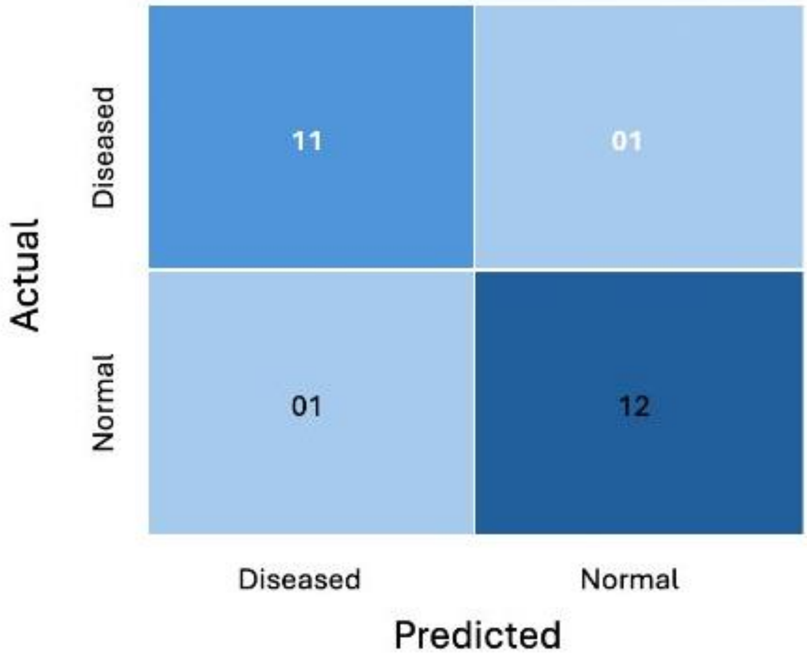


Fig 17:- Confusion Matrix of Ensemble model

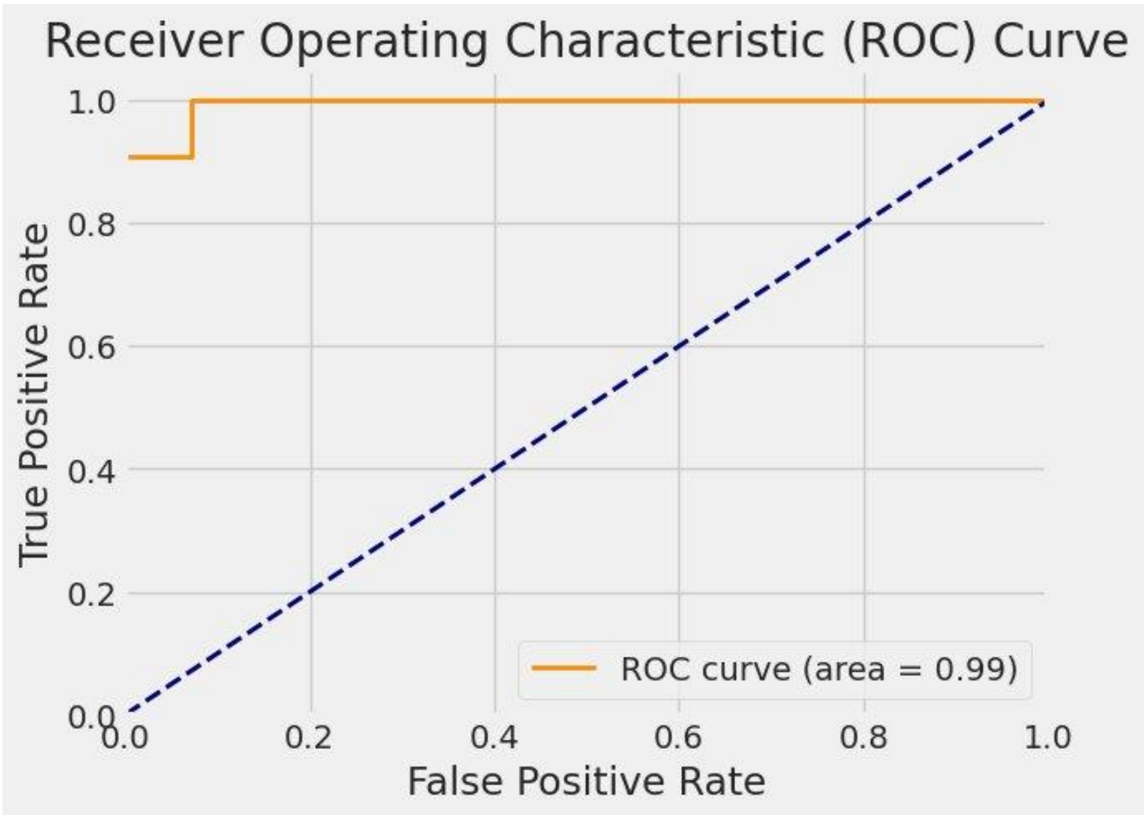


Fig 18:- ROC curve of Ensemble model with parameters (Precision,Recall,F1-Score)

Chapter 5

Conclusion and Future Scope of Work

In conclusion, the use of deep learning algorithms for the detection of cardiovascular risk prediction is a promising area of research. We have explored the effectiveness of three popular convolutional neural network architectures: ResNet-50, VGG-19, and Inception V3, and have shown that all three models can achieve high accuracy in detecting cardiovascular diseases from retinal fundus images. ResNet-50 model has achieved the highest accuracy of 89.88% among all these models on testing dataset. Furthermore, we have increased the accuracy of our predictions to 92.71% by integrating these models using an ensemble approach (average of the pre-trained weights of the preceding three models). The ensemble model demonstrates the importance of leveraging multiple models for complex tasks such as medical diagnosis, where combining the strengths of individual models can help compensate for their respective weaknesses.

Our results show that using deep learning algorithms for cardiovascular diseases and its risk detection has the potential to significantly improve the speed, accuracy, and efficiency of diagnosis, reducing the risk of missed or delayed diagnosis and improving patient outcomes. Additionally, our approach is non-invasive, affordable, and easily scalable, making it a promising tool for use in clinics and hospitals.

However, we are faced with the ultimate challenge of discovering an affordable dataset of retinal fundus images of cardiovascular diseases patients that we are yet to overcome. Regrettably, our efforts to obtain pertinent datasets through outreach to local hospitals and online researchers within our field have been unfruitful thus far. Consequently, we are compelled to utilize publicly available datasets for retinal vessel segmentation (STARE), as we have no alternative resources at our disposal.

Furthermore, the existing publicly accessible online datasets do not include the categorization of retinal images into those depicting of cardiovascular disease patients or patients with high risk of cardiovascular diseases but also other multi-diseases like diabetic retinopathy among others.

This deficiency arises from the limited availability of records tracking cardiovascular patients or acquiring their retinal fundus images, which is not a widespread practice.

Both individual models and the ensemble model has high scope of improvement to achieve the highest obtainable accuracy, thereby ensuring reliable diagnostic outcomes. Upon creation of the final model, further attempt to optimize the performance of our models for different populations and clinical settings is plausible, as well as the output of the developed diagnostic model will be required to be evaluated and validated with the aid of experts for its proper implementation in the health sector.

REFERENCES

[1] R. Poplin *et al.*, "Prediction of cardiovascular risk factors from retinal fundus photographs via Deep Learning," *Nature Biomedical Engineering*, vol. 2, no. 3, pp. 158–164, 2018. doi:10.1038/s41551-018-0195-0

[2] R. M. Tseng *et al.*, "Validation of a deep-learning-based retinal biomarker (Reti-CVD) in the prediction of cardiovascular disease: Data from UK Biobank," *BMC Medicine*, vol. 21, no. 1, 2023. doi:10.1186/s12916-022-02684-8

[3] Y. Ma *et al.*, *Development and validation of a deep learning algorithm using fundus photographs to predict 10-year risk of ischemic cardiovascular diseases among Chinese population*, 2021. doi:10.1101/2021.04.15.21255176

[4] T. H. Rim *et al.*, "Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs," *The Lancet Digital Health*, vol. 3, no. 5, 2021. doi:10.1016/s2589-7500(21)00043-1

[5] T. K. Revathi, B. Sathiyabhama, and S. Sankar, "Diagnosing Cardio Vascular Disease (CVD) using Generative Adversarial Network (GAN) in Retinal Fundus Images." in *Annals of the Romanian Society for Cell Biology*, vol. 25, Association of Cell Biology Romania, 2021, pp. 2563–2572

[6] David Joel Nogueira Azevedo. "Cardiovascular Risks Detection Using Fundus Image." 19 July 2019.