International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

PROTECTING VISION IN THE WORKPLACE: A DEEP LEARNING APPROACH TO DETECT DAIBETIC RETINOPATHY

***Mr. A. Ravi Kishore*[1], *M. Madhuri*[2], *N. Shravani*[3]**

[1]Assistant Professor, *Information Technology GMR Institute of Technology* Rajam, India, ravikishore.a@gmrit.edu.in
[2]Student *Information Technology GMR Institute of Technology* Rajam, India, 20341A1267@gmrit.edu.in
[3]Student *Information Technology GMR Institute of Technology* Rajam, India, 20341A1275@gmrit.edu.in

ABSTRACT:

Diabetes is a universal condition that has created an impact on majority of people in recent years. According to estimates, 77 million Indians have diabetes. Diabetes causes a condition called diabetic retinopathy, which leads to loss of vision. The occurrence of diabetic retinopathy among diabetic patients is estimated to be 27%. In this work, we focus on three sectors of working professionals: the IT sector (75% of working people are at high risk), the pharmaceutical sector (9.5% of adults are at high risk) and the private transportation sector (5.2% of working men in metropolitan cities are at high riskThe use of five models—InceptionV3, VGG19, ResNet50, DenseNet121, and DenseNet169—for the diagnosis of diabetic retinopathy is examined in this experiment. The data collection "Asia Pacific Tele-Ophthalmology Society (APTOS)"2019 blindness detection is used in this paper. There are 3662 samples in the collection, and each one has an image resolution of 3216 × 2136. Five categories will be created from the output: None, mild, moderate, severe, and proliferative disease states. These models achieved accuracy of 97.43, 96.64, 95.71, 97.12, 95.73 respectively. The main aim of our work is to aware the people to detect the spreading of disease early which helps them not to get into a critical situation. The best accuracy observed is 97.43 with Inception V3.

Keywords: Diabetes, Diabetic Retinopathy, Deep Learning, Image Classification, APTOS blindness detection.

INTRODUCTION :

Introduction of Diabetic Retinopathy:

Diabetes is the most common disease occurring nowadays in most of the individuals irrespective of their age. Long-term diabetes has induced DR, a degenerative illness that is a direct effect of mellitus. Most diabetics go blind due to late identification of diabetic retinopathy. If DR is identified and treated right away, patients' sight can usually be spared. However, an ophthalmologist must manually review the photos in order to diagnose DR, which is costly and takes a lot of time and effort. The retina's blood vessels inflate up as a result of DR, causing the retina to separate and the blood vessels to get damaged. These minuscule blood vessels begin to hemorrhage, which results in floaters, blurriness, black spots in the vision, and a challenge seeing colors. Nowadays people are trusting in computer more than the human beings for accurate results. The five stages of DR can be identified as:

1. NPDR: Non-proliferative diabetic retinopathy, it is the normal condition of an eye without DR.
2. Mild NPDR: Microaneurysms, which are microscopic balloon-like swellings in the retina's blood vessels, constitute the first stage of DR and are indicative of this condition.
3. Moderate NPDR: It is characterized by damage to parts of the retina's blood vessels, which causes blood and fluid to leak into the tissue of the retina.
4. Severe NPDR: If diabetes is not managed properly, more blood vessels will get damaged and obstructed, and there will be an increase in the amount of blood and fluid that leaks into the retina.
5. Proliferative DR: This is the final DR stage, where the disease has advanced significantly and is very problematic to one's vision.

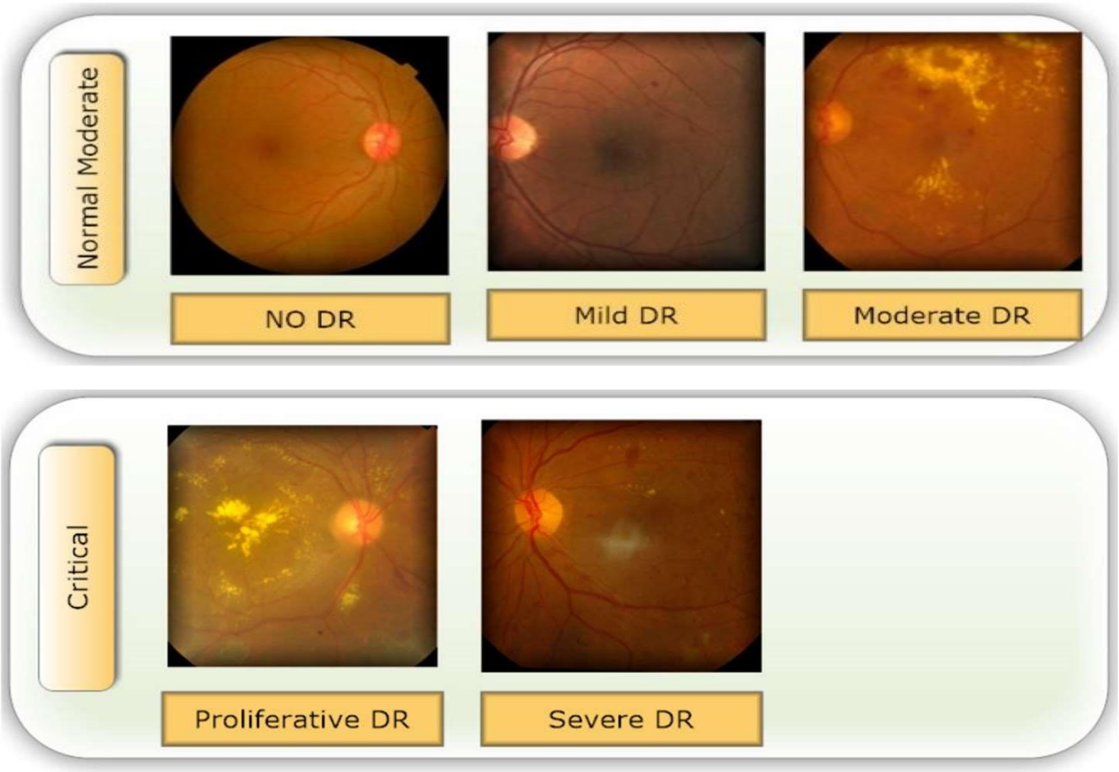

Fig 1.1: Types of DR based on severity

Significance of Machine learning and Deep learning:

In the field of healthcare, the significance of machine learning and deep learning techniques, such as those used in your project for diabetic retinopathy detection, cannot be overstated. These advanced technologies have revolutionized the way medical conditions are diagnosed and treated, offering more accurate and efficient solutions.

By leveraging ML and DL models trained on large datasets, your project aims to automate the detection of diabetic retinopathy, a critical step in preventing vision loss among individuals with diabetes. The use of transfer learning with pre-trained models like VGG19, InceptionV3, ResNet50, DenseNet121, and DenseNet169 enhances the performance of the detection system by leveraging knowledge learned from vast image datasets like ImageNet.

The ability of these models to analyze complex patterns and features in medical images allows for early and accurate diagnosis of diabetic retinopathy, enabling timely intervention and treatment. This offers automated screening options, which not only improves patient outcomes but also lessens the workload for healthcare workers.

Overall, the integration of ML and DL technologies in healthcare projects like yours showcases their profound impact on improving diagnostic accuracy, patient care, and ultimately, saving lives.

Major challenges

The literature identifies two primary challenges: first, while numerous methods exist for detecting diabetic retinopathy, there is a shortage of qualified professionals to provide effective screening and treatment of vision-threatening DR; second, there is limited opportunity for large-scale dataset training.

Solutions to these challenges

To address these challenges our proposed work seeks to provide an efficient solution that identifies the production of diabetic retinopathy in early stages and the work mainly targets on most prone zone of working individuals which yields a better society and also our proposed work has been trained on large dataset that is APTOS 2019 blindness detection dataset.

Overview of the proposed work

In this Work, we provide an overview of our proposed scheme, which encompasses a holistic approach to detection of Diabetic Retinopathy. Our model is designed to made an comparative analysis on 5 models Inception V3, VGG19, ResNet50, DenseNet121 and DenseNet169. The output of the proposed includes classification of diabetic retinopathy into 5 classes namely 0No_DR, 1-Mild_DR, 2-Moderate_DR, 3-Severe_DR, 4-Proliferative_DR. The proposed work includes training of the proposed models on APTOS(Asia Pacific Tele-ophthalmology Society) 2019 Blindness detection dataset. This dataset consists of 3662 samples in which 1805 images constitute No_DR class , 999 images constitute Moderate_DR, 370 images constitute Mild_DR, 295 images constitute Proliferative_DR and 193 images constitute Severe_DR. So the model aims to predict the classes earlier and accurate.

METHODOLOGY :

About the Model:

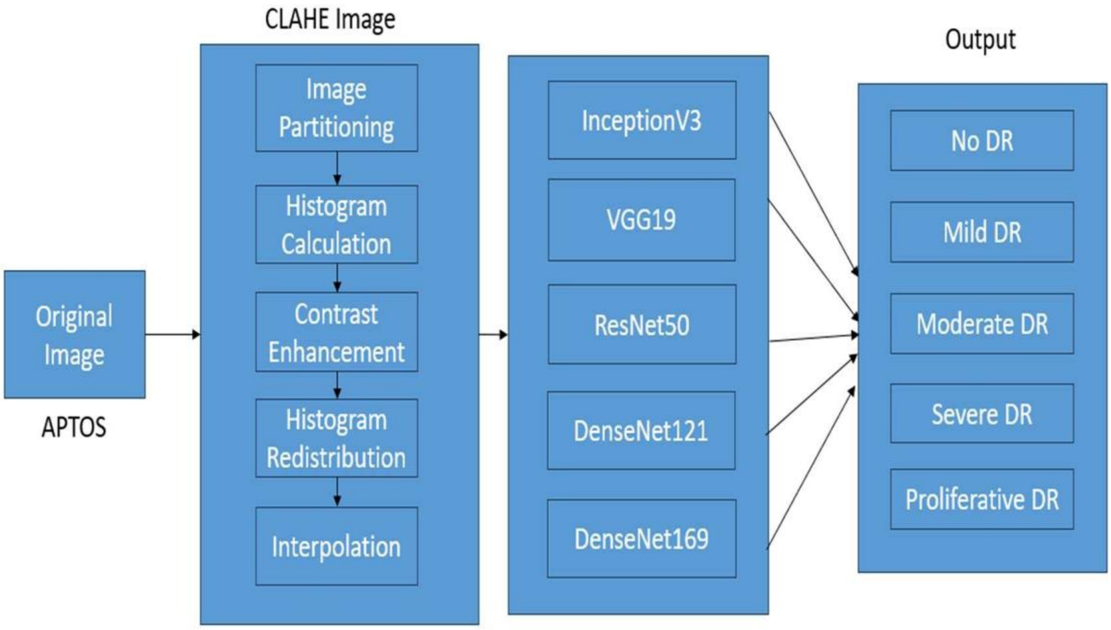

Fig 4.1: Flow Diagram of the work process

1. We begin by using the APTOS (Asia Pacific Tele-Ophthalmology Society) dataset as our input. This dataset contains retinal pictures with varying degrees of diabetic retinopathy severity.
2. To achieve good accuracy in this work, all photographs will be resized to the same size of 224*224 pixels.
3. To improve image contrast and details, this study makes use of CLAHE (Contrast Limited Adaptive Histogram Equalisation). This helps the models to capture better features in the images.
4. CLAHE is a technique for image equalization. Contrast over-amplification is avoided via CLAHE, a variant of Adaptive Histogram Equalization (AHE).
5. CLAHE has three main tasks to perform Tile generation, Histogram Equalization, Bilinear interpolation.
6. A neural network architecture called DenseNet (Densely Connected Convolutional Networks) was created to address the vanishing gradient problem and enhance feature reuse and compactness.
7. In this work we make a comparative analysis on 5 models namely InceptionV3, VGG19, ResNet50, DenseNet121, DenseNet169.
8. These models are trained on APTOS 2019 blindness detection dataset and a comparative analysis for the best classification has been made.
9. Finally the output can be classified into 5 types as No DR, Mild DR, Moderate DR, Severe DR, Proliferative DR

Inception V3:

With over 78.1% accuracy on the ImageNet dataset, Inception v3 is a very accurate image recognition model. It is the result of combining the thoughts of several different researchers. The architecture is composed of fully connected layers, convolutions, average pooling, max pooling, concatenations, dropouts, and symmetric and asymmetric building pieces. Activation inputs are frequently subjected to batch normalisation. The Softmax function is used to calculate losses. Google's Inception v3 is a powerful convolutional neural network architecture designed for image categorization applications. Key features include Inception modules with parallel convolutions of different sizes, 1x1 convolutions for dimensionality reduction, and factorization for efficiency. The model also integrates auxiliary classifiers for stable training, utilizes batch normalization for faster convergence, and employs global average pooling in place of fully connected layers. Inception v3 demonstrated exceptional performance on the ImageNet dataset, proving its proficiency in recognizing a wide range of objects and scenes.

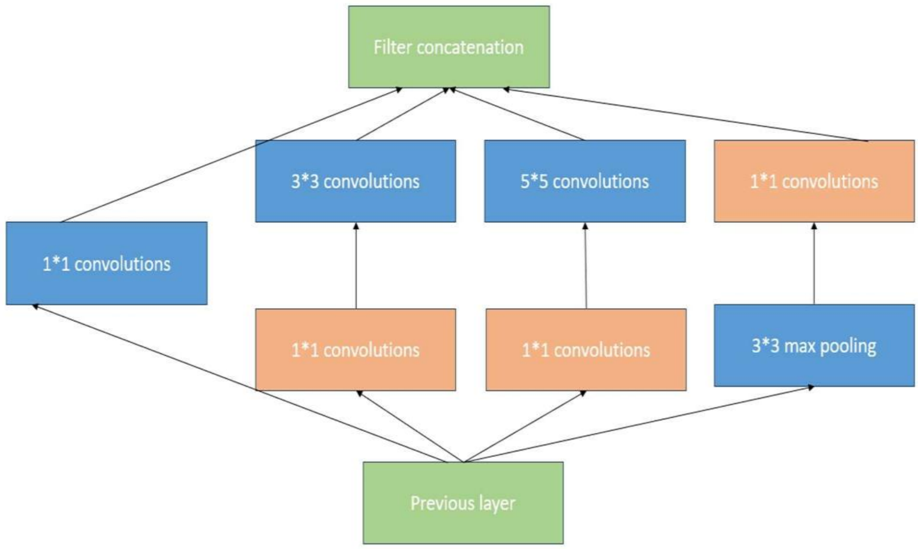

Fig 4.2: Inception V3 architecture

Overview the Inception V3 architecture:

Basic Building Block – Inception Module:

Inception Modules are the cornerstone of Inception V3. They consist of several parallel convolutional layers with various filter dimensions. The model can learn characteristics at different spatial resolutions thanks to these parallel branches, which allow it to record data at both fine and coarse scales. The outputs of these branches are then concatenated or “stacked” together.

Factorization of Convolutions:

Inception V3 employs factorized convolutions, which involves decomposing a larger kernel into smaller kernels (e.g., a 5x5 convolution is decomposed into two consecutive 3x3 convolutions). This reduces the number of parameters and computational cost while preserving expressive power.

1x1 Convolutions:

1x1 convolutions are used extensively in Inception V3. They are applied to reduce the number of channels in order to save computational resources and allow the network to learn more compact and meaningful representations.

Auxiliary Classifiers:

Inception V3 introduces auxiliary classifiers at intermediate layers of the network. These classifiers are used to inject additional gradients during training, which can help with combating the vanishing gradient problem. During inference, these auxiliary classifiers are typically ignored.

Global Average Pooling:

Inception V3 employs global average pooling at the network's endpoints rather than completely connected layers. With the channel information preserved, this technique shrinks the feature maps' spatial dimensions to 1x1. By doing so, the number of parameters is greatly decreased and overfitting is less likely.

Batch Normalization and ReLU Activations:

After every convolutional layer, batch normalisation is done to stabilise and accelerate training. The network as a whole uses Rectified Linear Unit (ReLU) activations to add non-linearity.

Regularization Techniques:

Inception V3 uses techniques like L2 weight regularization and dropout to prevent overfitting.

Optimization and Loss Function:

The cross-entropy loss function, which calculates the discrepancy between expected and actual class probabilities, is commonly used to train the model.

VGG-19:

A deep convolutional neural network called VGG19 is well-known for being easy to use and efficient when classifying images. It is relatively deep because it has three fully linked layers and sixteen convolutional layers. The network primarily uses 3x3 convolutions and employs max-pooling layers for dimensionality reduction. ReLU activation functions introduce non-linearity, and dropout layers are used to reduce overfitting. VGG19 expects input images to be 224x224 pixels. It was trained on the ImageNet dataset and performed well in the 2014 ImageNet Large Scale Visual Recognition Challenge. While considered a classic, newer architectures have surpassed it in terms of both accuracy and computational efficiency.

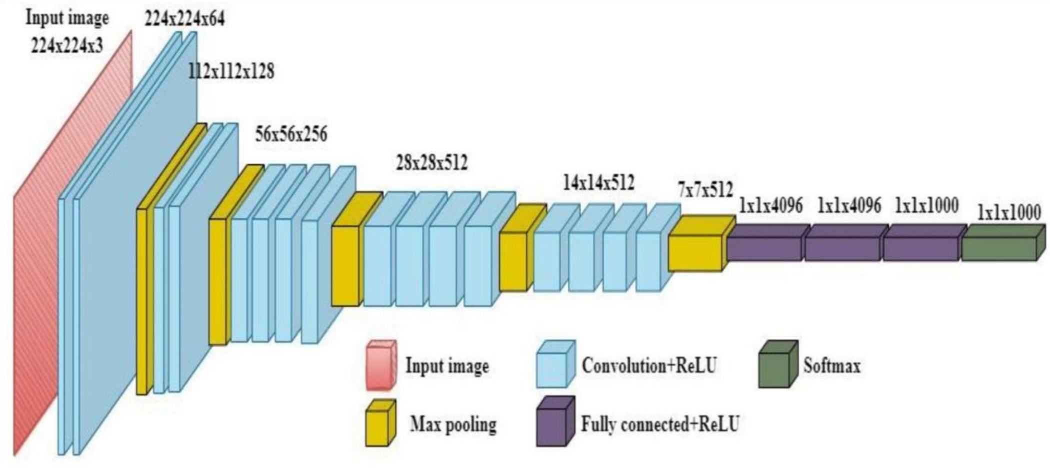

Fig 4.3: VGG19 architecture

Overview the Inception VGG-19 architecture:

Input Layer:

The network receives a fixed-size input image, typically 224 by 224 pixels.

Convolutional Layers:

The 16 convolutional layers that make up VGG19 are each followed by an activation function known as a rectified linear unit (ReLU). The network can learn a variety of features at various scales since the convolutional layers employ tiny 3x3 filters with a stride of 1.

Max Pooling Layers:

A max pooling layer with a 2x2 window and a stride of 2 comes after every two successive convolutional layers. Max pooling preserves crucial information while shrinking the feature maps' spatial size.

Fully Connected Layers:

Three fully connected layers, each with a ReLU activation function, make up the final layer of VGG19. The final fully connected layer has 1,000 neurons, which is equal to the number of classes in the ImageNet dataset (the dataset that was used to train VGG19 initially). The first two fully connected layers contain 4,096 neurons apiece.

Softmax Layer:

A softmax activation function in the last layer generates class probabilities.. It transforms the outputs of the previous layer into probabilities that sum to 1.

Dropout (Optional):

Some variants of VGG models, including VGG19, may include dropout layers after the fully connected layers to reduce overfitting during training

Output Layer:

As many neurons as there are classes in the classification task make up the output layer. It provides the final predicted probabilities for each class.

Loss Function and Optimization:

The cross-entropy loss function, which gauges the difference between expected and actual class probabilities, is commonly used to train the model.

ResNet50:

ResNet-50 is a deep neural network known for effective image recognition. It uses “residual connections” to alleviate training difficulties in very deep networks. The architecture consists of residual blocks, employs a bottleneck design, and utilizes global average pooling instead of fully connected layers. Unlike some models, it doesn’t use dropout or batch normalization within blocks. It is typically pre-trained on ImageNet for a broad understanding of visual concepts. This innovative architecture has greatly influenced subsequent deep learning models.

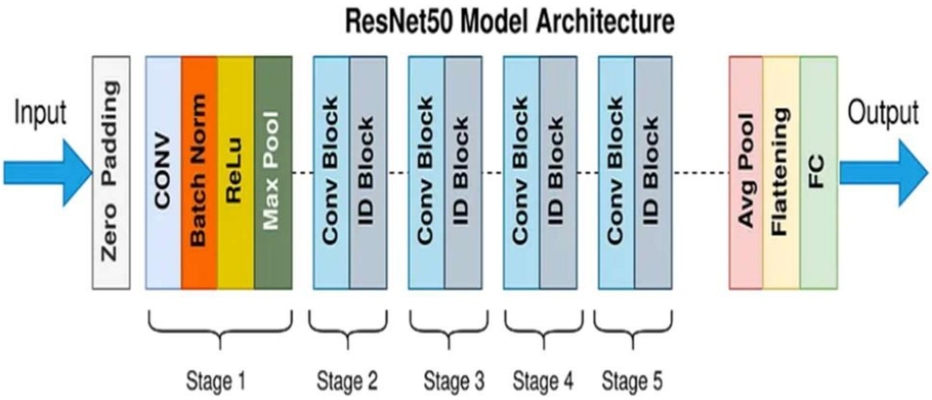

Fig 4.4: ResNet50 Architecture

Overview the Inception VGG-19 architecture:

Input Layer:

The network receives a fixed-size input image, typically 224 by 224 pixels.

Initial Convolutional Layers:

Starting with a typical 7x7 convolutional layer with 64 filters, batch normalisation and a ReLU activation function are applied to the network. To lower the spatial dimensions, a 3x3 max-pooling layer with a stride of 2 is used subsequently.

ResNet-50 is distinguished by the widespread utilisation of residual blocks. Multiple convolutional layers, usually with 3x3 filters, make up each block. The input is first processed through a number of convolutional layers inside a block. Then, a "shortcut connection" is formed by adding the output of these layers to the initial input. By doing this, the network is able to learn the residual, or the variation between the input and the intended result, which can help with training very deep networks.

Bottleneck Architecture:

ResNet-50 uses a “bottleneck” architecture in its residual blocks to reduce the computational cost. This involves using 1x1, 3x3, and 1x1 convolutions successively, which helps in reducing the number of parameters and computation.

Stacking Blocks:

ResNet-50 is constructed by stacking multiple residual blocks. The exact number of blocks depends on the specific ResNet variant (in this case, 50 indicates the number of layers, including the convolutional and fully connected layers).

Global Average Pooling (GAP):

ResNet-50 uses global average pooling at the network's endpoints in place of fully connected layers. With the channel information preserved, this technique shrinks the feature maps' spatial dimensions to 1x1.

Fully Connected Layer:

With as many neurons as there are classes in the classification problem, the final layer is a fully linked layer. It provides the final predicted probabilities for each class.

Softmax Layer:

The output layer uses a softmax activation function to transform the final layer's outputs into class probabilities.

Loss Function and Optimization:

The cross-entropy loss function, which calculates the discrepancy between expected and actual class probabilities, is commonly used to train the model.

DenseNet:

A neural network architecture called DenseNet (Densely Connected Convolutional Networks) was created to solve the vanishing gradient problem while enhancing feature reuse and compactness.

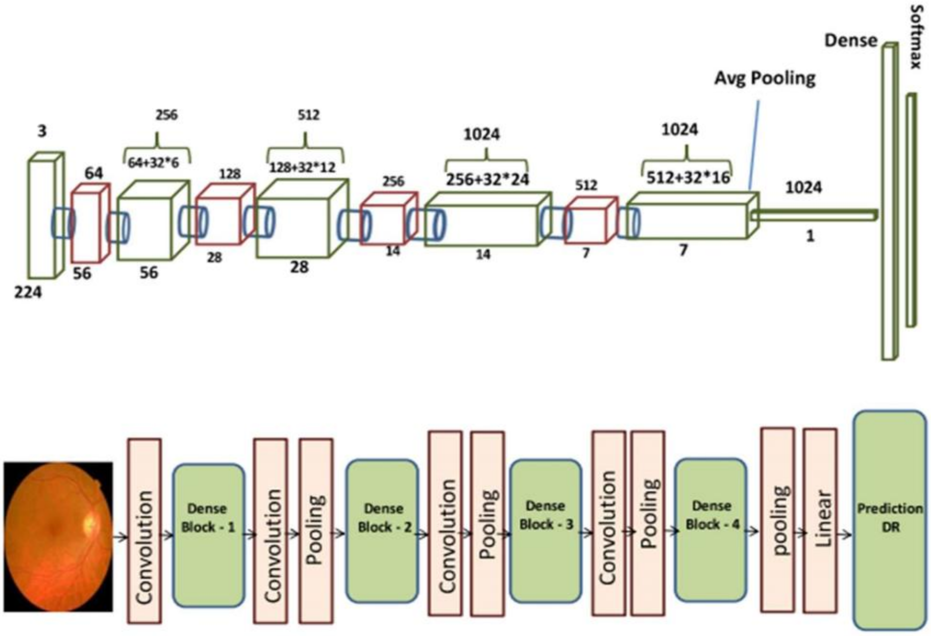

Fig 4.5: DenseNet Architecture

Overview the Inception VGG-19 architecture:

Input Layer:

The network receives a fixed-size input image, typically 224 by 224 pixels.

Initial Convolutional Layers:

Starting with a typical 7x7 convolutional layer with 64 filters, batch normalisation and a ReLU activation function are applied to the network. To lower the spatial dimensions, a 3x3 max-pooling layer with a stride of 2 is used subsequently. DenseNet is distinguished by its dense block structure. Typically, a dense block consists of six to twelve layers, each of which receives feature maps from all layers that came before it. Each layer in a dense block usually comprises of a sequence of batch normalisation, ReLU activations, and 3x3 convolutions.

Bottleneck Layer:

Each layer within a dense block uses a "bottleneck" architecture. It involves using 1x1 convolutions to reduce the number of input channels, followed by 3x3 convolutions, and then expanding the channels again with another 1x1 convolution. This bottleneck design helps in reducing computational cost and para.

Transition Layers:

After each dense block, a transition layer is introduced to reduce the spatial dimensions and compress feature maps before passing them to the next dense block. The transition layer consists of a 1x1 convolution for dimensionality reduction followed by 2x2 average pooling.

Global Average Pooling (GAP):

DenseNet uses global average pooling at the network's end, just like ResNet topologies do, to shrink the feature maps' spatial dimensions to 1x1 while maintaining channel information.

Fully Connected Layer:

With as many neurons as there are classes in the classification task, the last layer is a fully linked layer. It provides the final predicted probabilities for each class.

Softmax Layer:

The output layer uses a softmax activation function to transform the final layer's outputs into class probabilities.

Loss Function and Optimization:

The cross-entropy loss function, which calculates the discrepancy between expected and actual class probabilities, is commonly used to train the model. Stochastic gradient descent (SGD) and its more sophisticated variations, such as Adam, are frequently used in optimisation.

DENSENET 121:

A deep convolution neural network architecture called DenseNet-121 was created specifically for image classification applications. Its distinct "dense connectivity" pattern—in which every layer gets input from every layer before it—is what distinguishes it. Strong feature reuse and gradient flow are encouraged during training as a result. The model consists of dense blocks with multiple layers, and transition layers that reduce spatial dimensions. Unlike some models, DenseNet-121 typically doesn't use dropout layers, relying on its dense connectivity for regularization. It employs global average pooling instead of fully connected layers, significantly reducing parameters.

DENSENET 169:

It is deeper and more complex. DenseNet-169 also includes transition blocks between dense blocks. These transition blocks serve two main Purposes: Reducing spatial dimensions and Controlling the number of feature maps being passed.

RESULTS & DISCUSSIONS :

The InceptionV3, VGG19, ResNet50, DenseNet121, DenseNet169 are examined in the test dataset and displayed 97.43, 96.64, 95.71, 97.12, 95.73 respectively. The retinal pictures in a mild form of DR exhibit red patches known as hemorrhages. A number of epochs in the model can analyze variables including the presence of yellow specks known as hard exudates and microaneurysms that appear in proliferatve DR. The best accuracy is 97.43 and is achieved by Inception V3.

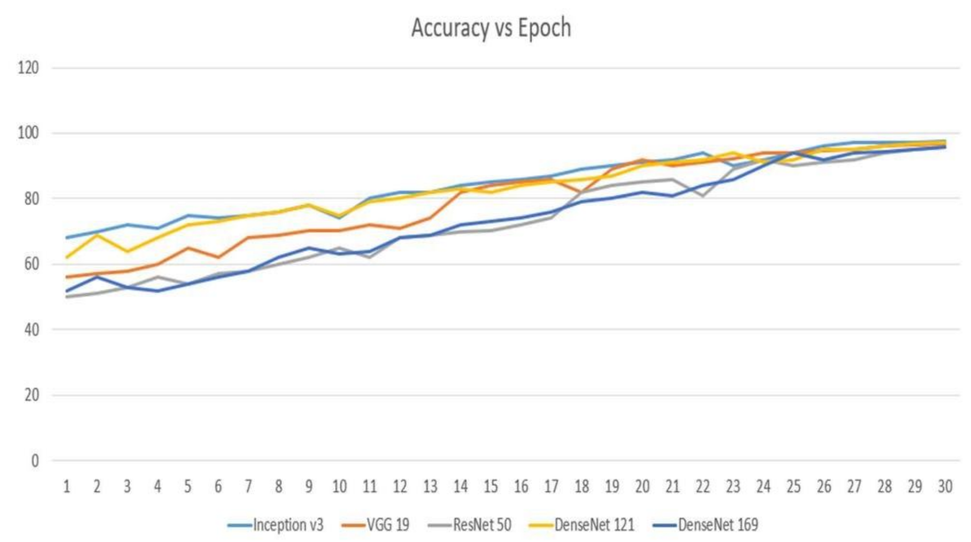

Fig 6.1: Accuracy vs Epoch plot of proposed models

WEB INTERFACE DEPLOYMENT:

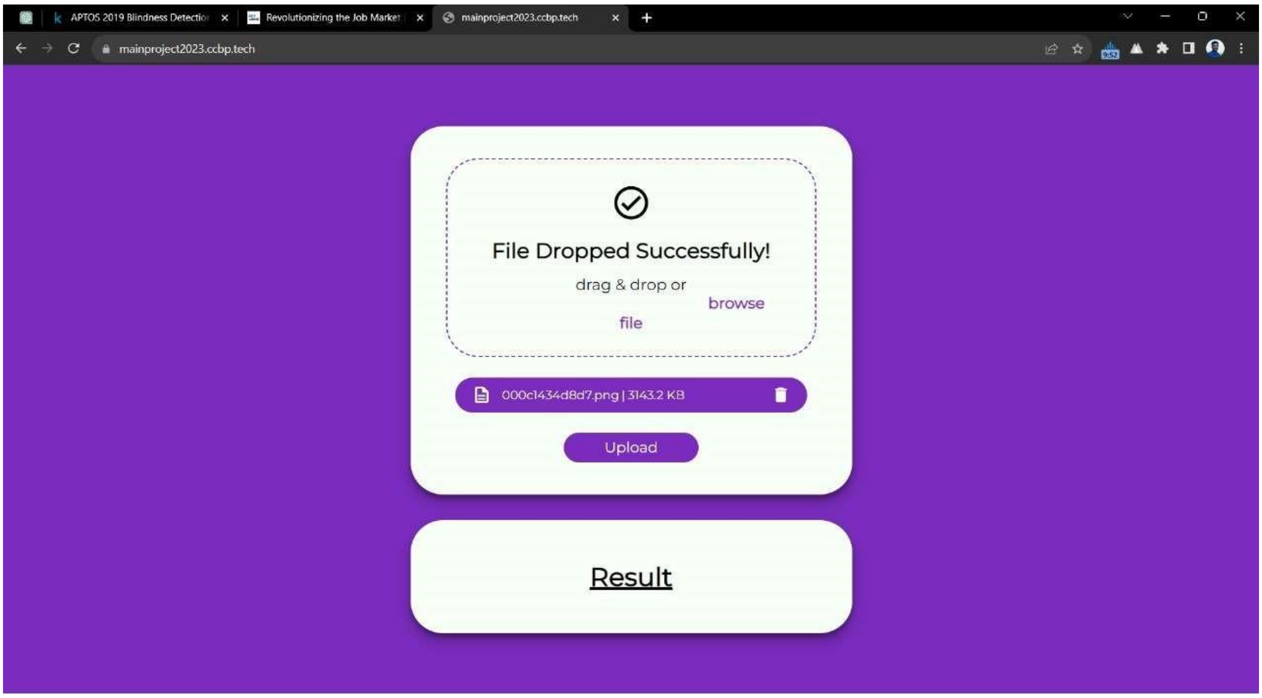

Fig 6.2: Uploading the image

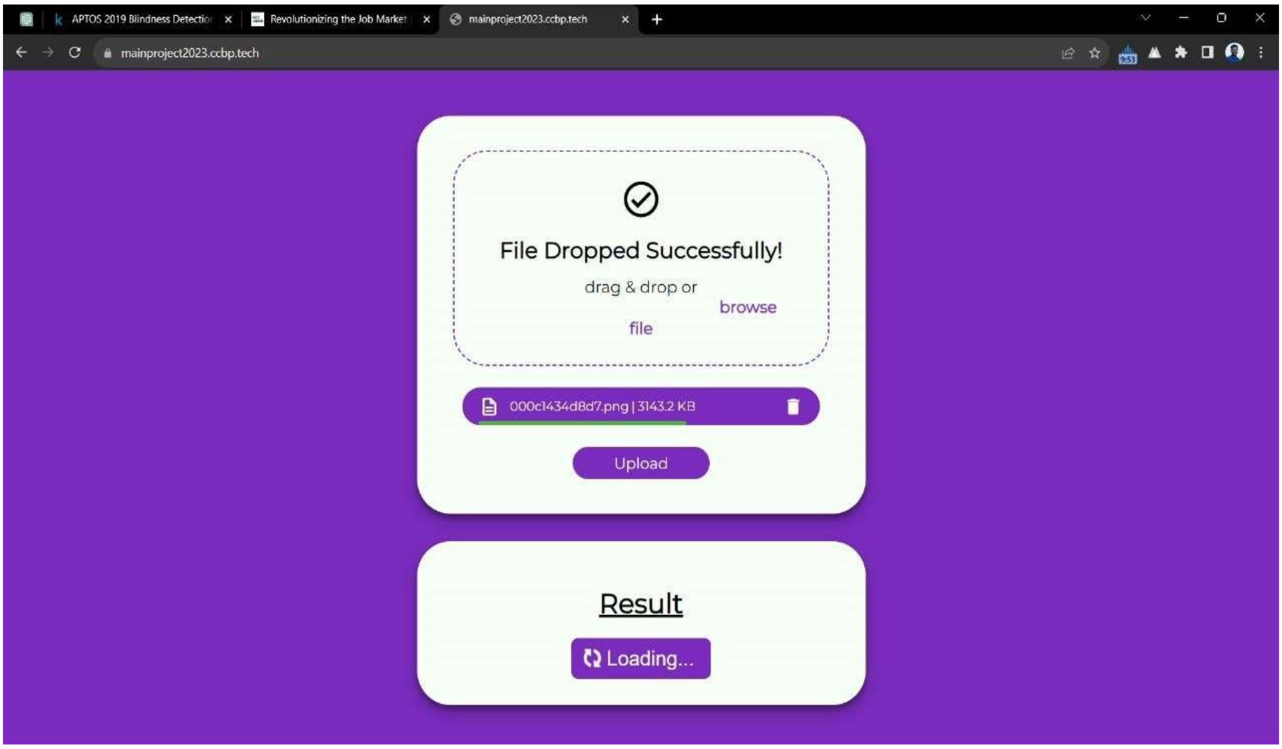

Fig 6.3: Loading the image on processing

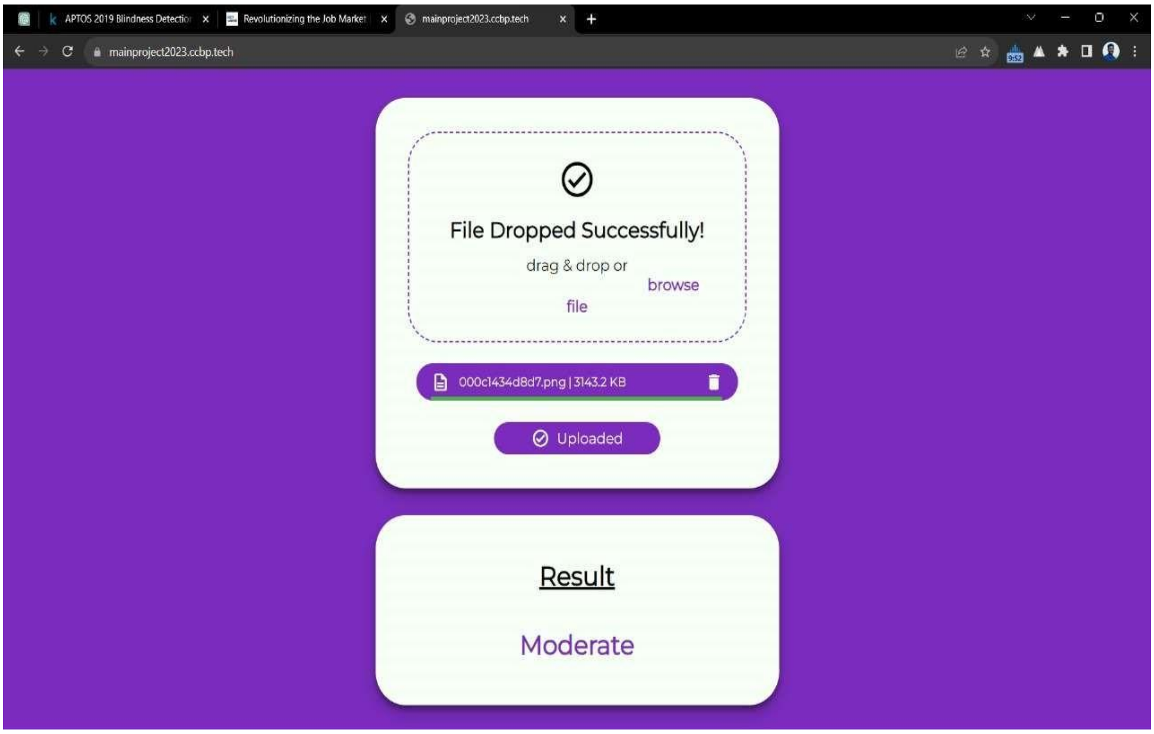

Fig 6.4: Result

Conclusion :

In conclusion, the project focusing on diabetic retinopathy detection using deep learning models has shown promising results in automating the identification of this vision-threatening condition. By leveraging advanced technologies like machine learning and deep learning, we have successfully developed models that can classify diabetic retinopathy into different stages with high accuracy.

The utilization of transfer learning with pre-trained models such as VGG19, InceptionV3, ResNet50, DenseNet121, and DenseNet169 has significantly improved the efficiency and performance of the detection system. We have established the potential of these models to support healthcare professionals in early diagnosis and intervention, ultimately leading to improved patient outcomes, through the project's training and evaluation process.

The successful implementation of these deep learning models in diabetic retinopathy detection underscores the importance of integrating cutting-edge technologies into healthcare practices. By automating the screening process and providing accurate results, these models can potentially revolutionize the way diabetic retinopathy is diagnosed and managed, leading to better patient care and outcomes