ESTIMATING THE SEVERITY OF KNEE OSTEOARTHRITIS USING DEEP CONVOLUTION NETWORK FROM X-RAY IMAGE

A PROJECT REPORT

Submitted By

Mohammed Ashiq H 195001064

Rakesh J V 195001087

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

Department of Computer Science and Engineering

Sri Sivasubramaniya Nadar College of Engineering
(An Autonomous Institution, Affiliated to Anna University)
Kalavakkam - 603110

April 2023

Sri Sivasubramaniya Nadar College of Engineering

(An Autonomous Institution, Affiliated to Anna University)

BONAFIDE CERTIFICATE

Certified that this project report titled **“ESTIMATING THE SEVERITY OF KNEE OSTEOARTHRITIS USING DEEP CONVOLUTION NETWORK FROM X-RAY IMAGE”** is the *bonafide* work of “**Mohammed Ashiq H(195001064)** and **Rakesh J V (195001087)**” who carried out the project work under my supervision.

Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Dr. T.T. Mirnalinee
Head of the Department
Professor,
Department of CSE,
SSN College of Engineering,
Kalavakkam - 603 110

Dr. B. Bharathi
Supervisor
Associate Professor,
Department of CSE,
SSN College of Engineering,
Kalavakkam - 603 110

Place:
Date:

Submitted for the examination held on...........

Internal Examiner **External Examiner**

ACKNOWLEDGEMENTS

I thank GOD, the almighty for giving me strength and knowledge to do this project.

I would like to thank and deep sense of gratitude to my guide **Dr. B. BHARATHI**, Associate Professor, Department of Computer Science and Engineering, for his valuable advice and suggestions as well as his continued guidance, patience and support that helped me to shape and refine my work.

My sincere thanks to **Dr. T.T. MIRNALINEE**, Professor and Head of the Department of Computer Science and Engineering, for her words of advice and encouragement and I would like to thank our project Coordinator **Dr.B. BHARATHI**, Associate Professor, Department of Computer Science and Engineering for her valuable suggestions throughout this project.

I express my deep respect to the founder **Dr. SHIV NADAR**, Chairman, SSN Institutions. I also express my appreciation to our **Dr. V. E. ANNAMALAI**, Principal, for all the help he has rendered during this course of study.

I would like to extend my sincere thanks to all the teaching and non-teaching staffs of our department who have contributed directly and indirectly during the course of my project work. Finally, I would like to thank my parents and friends for their patience, cooperation and moral support throughout my life.

Mohammed Ashiq H **Rakesh J V**

ABSTRACT

Osteoarthritis (OA) of the knee is a common degenerative joint condition that affects millions of people worldwide. We present a new approach employing deep convolutional neural networks (CNNs) to estimate OA severity from X-ray imaging by ensembling them and masking is performed on those images to segment the part of injury to find the distance between knee bones. The Network will be trained by features related with various degrees of knee OA using a huge dataset of X-ray images labelled with severity scores. We also propose a technique of segmentation for calculating the joint distance, which is a critical sign of joint health. The CNN models we used are; EfficientNet, InceptionV3, DenseNet and MobileNet and we obtained an accuracy of 94%, 94%, 93% and 93% respectively. Then after ensembling we obtained an accuracy of 99%. We used mask RCNN for segmentation and distance was calculated. The obtained IOU score for the mask r cnn model we trained was 0.73

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Osteoarthritis is becoming more common in middle-aged men and women around the world. Majority of those affected are ignorant of their condition. They ignore the discomfort because they believe it will pass. They don't feel severe discomfort until a certain time, which is already too late. If it is not diagnosed sooner, one may not lead a happy or healthy life. The severity of the condition must be determined manually by the doctors by going through the X Ray images in order to advise the patients. Therefore, we develop a deep learning model that can accurately predict the severity of knee osteoarthritis among the patients from their x-ray image as healthy, moderate and severe. Additionally, in our work, we are going to compute the distance between the knee bones from the X ray images to determine how prone the patient is to falling into the next grade.

The severity of this disease will be determined by the distance between knee bones and can be classified into five categories namely Grade 0, Grade 1, Grade 2, Grade 3, and Grade 4. Generally, grade 0,1 and 2 are considered healthy, grade 3 is moderate and grade 4 is severe. The distance between the knee joints, Joint Space Width(JSW) is one of the crucial factor to determine the severity. For grade 1, the JSW will be usually greater 4 mm, for grade to the JSW is around 2 to 4 mms. Then the JSW for grade 3 will be less than 2mm and for grade 4 the disease already severe and the knee joints will be touching together.

The first part of our work is classification. The X ray images will be classified according to their grade of injury. Convolutional Neural Networks (CNNs) have been employed for the respective task. CNNs have been widely used in various computer vision applications due to their superior performance in image recognition tasks especially on medical images. There are several new CNN architectures which perform exceptionally good on medical images. EfficientNet, DenseNet, MobileNet, and InceptionV3 are such models which we use here in our work for classification.

A collection of CNN models called EfficientNet[2] performs at the cutting edge of image classification tasks. It employs a compound scaling technique that balances the trade-off between accuracy and processing economy by optimizing the depth, width, and resolution of the network. EfficientNet achieves good accuracy while using fewer parameters, running faster than many other models.

DenseNet[3] is a CNN model with feed-forward connections between every layer, resulting in high parameter sharing and a less number of parameters required to train the network. DenseNet has attained cutting-edge performance on a number of benchmarks, including object detection and image classification.

MobileNet[4] is a group of compact CNN models created to be computationally effective in mobile and other low-power applications. In order to reduce the number of parameters while retaining a high level of accuracy, MobileNet uses depthwise separable convolutions.

InceptionV3[5] is a CNN model that combines convolutional layers with various kernel sizes to collect data at various scales. To enhance the performance of the model, InceptionV3 also uses additional methods such batch normalization, factorized convolutions, and label smoothing. InceptionV3 is frequently utilized in both research and business since it has attained cutting-edge performance on numerous picture categorization tasks.

In our work, we are preparing an ensemble model with all these model to classify the X Ray images as three categories: Healthy, Moderate and Severe. The data is pre-processed before using for classification. Median noise removal and Contrast Limited Adaptive Histogram Equalization(CLAHE) are applied on images to remove noise and enhance it so that features can be learnt easily and effectively.

The second part of our work is segmentation and calculating the distance. Masks for all the images is created manually and fed into Mask R CNN for segmentation process. After taking the segmented part, the central perpendicular distance of the segmented part will be calculated as the JSW distance. Using this distance, we can determine how close the patient is to falling into next stage.

Mask R-CNN enhances the well-known Faster R-CNN object recognition model by adding a second branch for pixel-level segmentation, enabling it to create excellent object masks. A unique ROIAlign layer that permits precise feature extraction and input pixel alignment is one of its significant contributions, as is a multi-task loss function that enables training of both segmentation and detection tasks at the same time.

1.2 Motivation

Arthritis is the medical term for swelling or joint inflammation. Osteoarthritis is the most common type of arthritis affecting millions of individuals worldwide.

Each bone in a healthy joint has its own layer of articular cartilage, a form of connective tissue that enables the two bones to slide against each other without any friction. Osteoarthritis is characterised by the gradual loss of articular cartilage, which results in a large increase in friction between the two bones. This friction causes inflammation, which in turn triggers pain in the nerve endings in this joint region.

This most frequently affects knee joints as the cartilage between two knee bones start to rupture. The severity of this disease will be determined by the distance between knee bones and can be classified into five categories namely Grade 0, Grade 1, Grade 2, Grade 3, and Grade 4. Generally grade 0,1 and 2 are considered healthy, grade 3 is moderate and grade 4 is severe.

Osteoarthritis affects 10% of males and 18% of women who are 60 years or older. Even younger people are affected by this due to the shift in lifestyle. The majority of those affected are ignorant of their condition. They don't experience extreme pain until a certain point, by which it is too late. Hence, We create a model which able to determine the severity of Knee osteoarthritis and also measure the distance between two bones.

The Network will be trained by features related with various degrees of knee OA using a huge dataset of X-ray images labelled with severity scores.We also propose a technique of segmentation for calculating the joint distance, which is a critical sign of joint health. We will use a dataset of X-ray images from individuals with different levels of knee OA severity to evaluate our model. Our results will demonstrate that the method we have suggested is capable of segmenting the knee joint and properly evaluating OA severity, offering a valuable tool for clinical diagnosis and therapy planning.

1.3 Problem Statement

Osteoarthritis is becoming more common in middle-aged men and women around the world. Majority of those affected are ignorant of their condition. They ignore the discomfort because they believe it will pass. They don't feel severe discomfort until a certain time, which is already too late. If it is not diagnosed sooner, one may not lead a happy or healthy life. The severity of the condition must be determined manually by the doctors by going through the X Ray images in order to advise the patients. Therefore, we develop a deep learning model that can accurately predict the severity of knee osteoarthritis among the patients from their x-ray image as healthy, moderate and severe. Additionally in our work, we are going to compute the distance between the knee bones from the X ray images to determine how prone the patient is to falling into the next grade.

1.4 ASSEMBLING OF THE REPORT

The rest of this report is divided into the following sections: • Chapter 2 outlines the literature search done regarding the projects and to identify knowledge gaps as well as applicable theories and concepts • Chapter 3 contains the system's architecture along with the design considerations that each component of the architecture was given. • Chapter 4 has the details regarding the implementation, experimental strategy, and conclusions are discussed. • Chapter 5 concludes the project and gives the limitation, further development for the future.

CHAPTER 2

Literature Survey

2.1 General Need

At the moment, diagnosing knee osteoarthritis severity usually involves combining clinical examination, imaging tests like X-rays, and patient-reported symptoms. These approaches, however, are arbitrary and open to inaccuracy. Deep convolutional neural networks have the potential to provide a more accurate and objective evaluation of the severity of knee osteoarthritis.

The creation of a computer-aided diagnosis system that can automatically detect and assess the severity of knee osteoarthritis from X-ray pictures could be one of the practical applications of the project. By increasing the accuracy of the diagnosis and perhaps saving time and resources, this could result in better patient treatment outcomes.

The initiative might also have an impact on the creation of individualised treatment strategies. Clinicians could more effectively treat patients and lower the risk of problems by utilising deep convolutional neural networks to precisely estimate the severity of knee osteoarthritis in each individual patient.

Overall, the idea has the potential to offer clinical staff and patients alike significant practical advantages in the identification and management of knee osteoarthritis.

2.2 PRIOR EXPLORED IDEAS AND RELATED WORKS

Yaorong Xiao et.al.[10] wrote a study titled ”Classification of Osteoarthritis Severity Using Machine Learning Algorithms” .

The study’s objective was to create a machine learning model using clinical and radiographic data, that could reliably predict the severity of osteoarthritis. To do this, the authors compared the k-NN algorithm, naive bayes classifier, and logistic regression, three different machine learning algorithms.

The dataset for the study was madeup of 108 cases of osteoarthritis patients who had each completed a clinical examination and radiographic evaluation. Age, gender, body mass index, joint space narrowing, and osteophyte production were only a few of the clinical and radiographic data that the authors employed as input variables for the machine learning algorithms.

The k-NN method outperformed the other two, with an accuracy of about 65% in predicting the severity of osteoarthritis, according to the authors’ testing and training of the three algorithms. The scientists came to the conclusion that while machine learning algorithms would be helpful in forecasting the severity of osteoarthritis, more study was required to increase the models’ accuracy.

Jaynal Abedin et al. [6] did a research regarding knee osteoarthritis severity, the goal of the study was to use machine learning models based on patient data and plain X-ray images to predict the severity of knee osteoarthritis. Three models used for the research are convolutional neural network (CNN), an ElasticNet model, and a random forest (RF) model were evaluated in terms of performance by the authors.

The study's dataset included X-ray pictures of 143 participants with knee osteoarthritis and clinical information. The clinical information included demographic details, pain ratings, and radiographic results. The Kellgren-Lawrence grading system, with grades ranging from 0 to 4, was employed by the authors to evaluate the severity of knee osteoarthritis. Grades range from 0 to 4, with 4 denoting severe osteoarthritis.

The three machine learning models were developed, tested, and their performance was assessed using root mean squared error (RMSE) as the performance metric. The outcomes demonstrated that the CNN model, with an RMSE of 0.77, outperformed the RF model, which had an RMSE of 0.94, and the ElasticNet model, which had an RMSE of 0.97.

The CNN model outperformed the other models, according to the authors, who also proposed using machine learning models built on X-ray images and patient data to create precise tools for the early detection and treatment of knee osteoarthritis.

Overall, the study provides light on the potential of machine learning models for estimating the severity of knee osteoarthritis using X-ray images and patient data, and it offers insights into how well various machine learning models perform in comparison.

Rohit Kumar Jain et al. [7] study's goal was to use a convolutional neural network (CNN) based on the HR Net architecture to categorise the severity of knee osteoarthritis. The Kellgren-Lawrence grading method, which ranges from 0 to 4, was employed by the authors to grade the knee joints of 140 patients with various stages of knee osteoarthritis.

The dataset was split into training and testing sets by the authors, who used 70% of the data for training and 30% for testing. They made advantage of the cutting-edge CNN architecture known as HR Net, which combines high-resolution and multi-resolution capabilities.

The accuracy metric was used to assess the model's performance on the testing set. The findings demonstrated that the HR Net-based CNN model classified the severity of knee osteoarthritis with an accuracy of 71.74%.

The authors also acknowledged the limitations of their study, including the limited size of the dataset and the lack of diversity in the patient group, and they contrasted their findings with those of earlier studies. They came to the conclusion that their findings showed the potential of deep learning approaches for automating the classification of the severity of knee osteoarthritis.

Overall, the research demonstrates the potential of CNN-based methods for X-ray image-based knee osteoarthritis severity classification and offers insights into the use of HR Net architecture to this task. However, additional research with bigger and more varied datasets is required to validate and enhance the precision of these models.

Next further study was done by Sozan Mohammed Ahmed et al. [8] the goal was to use a combination of deep learning and machine learning models to grade the severity of knee osteoarthritis. The Kellgren-Lawrence grading method, which ranges from 0 to 4, was employed by the authors to grade the knee joints of 400 patients with various stages of knee osteoarthritis.

The dataset was split into training and testing sets by the authors, who used 70% of the data for training and 30% for testing. They combined machine learning models like Support Vector Machines (SVMs) and Random Forests with deep learning models like Convolutional Neural Networks (CNNs) and ResNet.

The accuracy metric was used to assess the model's performance on the testing set. According to the findings, the suggested model was 90.8% accurate in classifying the severity of knee osteoarthritis.

The authors came to the conclusion that their findings showed promise for recognising the severity grading of knee osteoarthritis using a combination of deep learning and machine learning models.However, additional research with bigger and more varied datasets is required to validate and enhance the precision of these models.

2.3 SURVEYING ABOUT THE TECHNOLOGIES INVOLVED

[2] A deep neural network architecture called EfficientNet was created by Google in 2019 and has produced state-of-the-art results on a number of computer vision applications, including segmentation, object detection, and picture categorization.

The architecture is meant to achieve great accuracy while being extremely resource-efficient in terms of computation. Due to its efficiency, EfficientNet is especially well suited for applications that require a little amount of computational power, including those that run on mobile devices or in cloud computing settings.

Medical image analysis is only one of the many computer vision jobs that EfficientNet has been used for.In particular, a number of studies have investigated the application of EfficientNet for X-ray image analysis, including the detection and classification of various diseases and abnormalities in the lungs and other organs.

A reserach on Ensemble of efficientNet for diagnosis of Tuberculosis [12] was conducted on the year 2021. The research got an accuracy of 94.35% and it was improved through Ensemble learning to an accuracy of 97.44%.

Another work on Automated medical diagnosis of COVID-19 [11] through EfficientNet in the year 2020. The accuracy of this work was 99.62% for binary classification and 96.70% for multiclass classification.

A deep neural network architecture called MobileNet was created for use in embedded and mobile vision applications. The concept was first described in their study [4] ”MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications” by Menglong Zhu et.al.

The authors of the paper put forth a brand-new convolutional neural network (CNN) architecture that is intended to be highly accurate while still being computationally and memory-efficient. Using depthwise separable convolutions, which factorise a typical convolutional layer into two distinct layers—a depthwise convolution and a pointwise convolution—is the main innovation of MobileNet.

The pointwise convolution combines the output of the depthwise convolution using a 1x1 convolution, whereas the depthwise convolution applies a single filter to each input channel independently. In comparison to a typical convolutional layer, this factorization requires less parameters and computational resources while preserving spatial and channel-wise correlations.

Authors also suggest a set of hyperparameters to manage the trade-off between model size and precision, making it simple to adapt MobileNet for various embedded and mobile vision applications.

In 2022, A work on COVID-19 detection from chest x-ray using MobileNet and residual separable convolution block was carried out by V. Santhosh Kumar Tangudu et al. This work utilizes pre-trained MobileNet for binary image classification on two publicly available datasets COVID5K, and COVIDRD. It has given an accuracy of 99% accuracy on both datasets.

Gao Huang et.al.introduced the deep neural network architecture [3] known as DenseNet in 2017. The fundamental goal of DenseNet is to use dense connections across layers to circumvent the vanishing gradient issue that arises in very deep neural networks. The output of the layer before it is used as the input in conventional neural networks. In DenseNet, every layer establishes a direct connection with every other layer by accepting the output of all preceding layers as its input.

These close connections enhance the network's information flow, cut down on the amount of parameters needed, and encourage feature reuse, among other advantages. DenseNet may accomplish state-of-the-art performance on a variety of computer vision applications while using fewer parameters and less processing when compared to other deep neural network architectures since it reuses characteristics learnt by prior layers.

The authors of the research also offer a brand-new architecture called DenseNet-BC, which enhances the DenseNet architecture with bottleneck layers and compression to further minimise the number of parameters and computation needed.

Since then, DenseNet has gained popularity as an architecture for computer vision tasks like semantic segmentation, object detection, and image classification. It has been employed in many different applications, including as natural language processing, autonomous driving, and medical imaging.

A work on Chest X-Ray Image to Classify Lung diseases in Different Resolution Size using DenseNet-121 Architectures by v et al on 2021. This study is done with four different datasets: tuberculosis dataset, pneumonia dataset, cardiomegaly dataset, and COVID-19 dataset. It is experimentally shown that DenseNet121 model achieves the highest accuracy in classification about accuracy of 89%, 90.4%, 89.8% and 98.6%, respectively.

Deep neural network architecture known as InceptionV3 was created by Google . It was first mentioned in the research article [5] by Christian Szegedy et.al. in 2016.

With the use of a combination of convolutional layers of various sizes, the major goal of InceptionV3 is to address the trade-off between accuracy and processing efficiency in deep neural networks. The authors update the Inception architecture with a number of additional features.

Since then, InceptionV3 has gained popularity as an architecture for computer vision tasks like semantic segmentation, object identification, and image classification. It has been employed in many different applications, including as natural language processing, medical imaging, and autonomous driving.

[9] ”Deep Learning for the Classification of Lung Nodules Using a Convolutional Neural Network” by Ardila et al. (2019) is InceptionV3 to medical picture classification.

In this study, the authors suggest a deep learning framework for classifying lung nodules in computed tomography (CT) scans using InceptionV3. A deep learning architecture classified lung nodules as malignant or benign with a high accuracy of 94.4% and a sensitivity of 96.6%. The authors also contrasted their strategy with other cutting-edge approaches, demonstrating that their framework performed better than them in terms of accuracy and sensitivity.

CHAPTER 3

Explanation about the Used Models

3.1 CNN

Convolutional Neural Networks (CNNs) are a type of deep learning neural network architecture specifically designed for image and video recognition tasks.They have been actively utilized for a variety of computer vision applications, including segmentation, object identification, and image classification.

The convolutional layer, which applies the mathematical operation of convolution to the input data, is the fundamental component of a CNN. The convolution operation involves sliding a small filter (also called a kernel or weight matrix) across the input data, element-wise multiplying each part of the input by the filter, and summing the results to produce a single output value. For each position of the filter across the input data, this operation is repeated numerous times, producing a feature map, or set of output values. A CNN can extract progressively complicated features from the input data by stacking many convolutional layers on top of one another.

The pooling layer, which decreases the spatial dimensions of the feature maps created by the convolutional layers, is another key component of a CNN. The most popular pooling operation is max-pooling, which replaces the entire region of the feature map with the maximum value taken from one or more regions of the feature map. This makes the network more resilient to small variations in the input data and lowers the computational cost of the network.

Finally, CNN's output layer is created to be specific to the problem it is solving, such as a softmax activation function for image classification, or a sigmoid activation function for binary classification.

Overall, CNNs are one of the most often used neural network architectures in this field and have shown to be quite efficient for a number of computer vision tasks. They are an effective tool for tackling a variety of issues due to their capacity to learn hierarchical features from input data and their robustness to input variations.

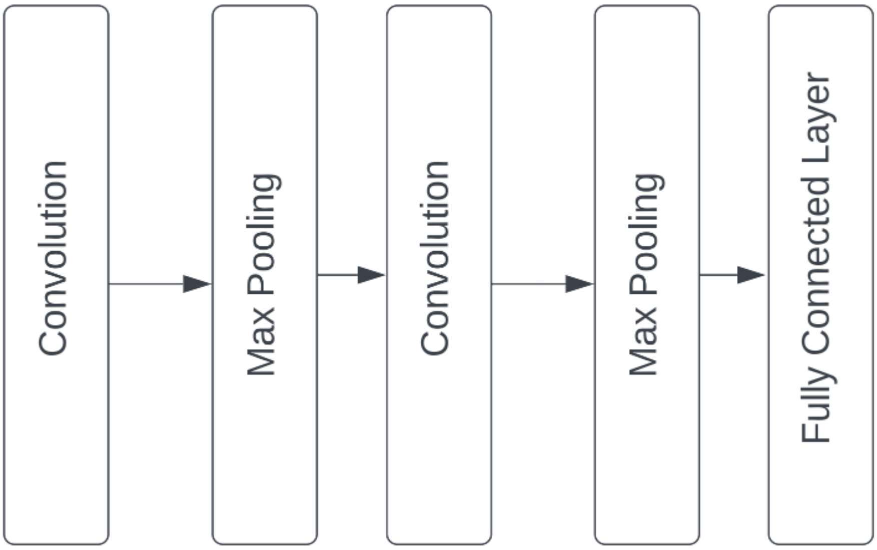


FIGURE 3.1: CNN ARCHITECTURE

3.2 EfficientNet

EfficientNet is an architecture of CNN used for Image Classification developed by google. With fewer parameters, it is intended to attain a high level of precision. In order to scale up the network size and resolution while preserving accuracy and efficiency, EfficientNet architecture was developed. This makes them useful for a variety of computer vision applications, including semantic segmentation, object identification, and picture classification.

More layers tend to improve the performance of neural networks. Depth scaling is a technique that can be used to accomplish this. Depth Scaling is technique of increasing the number of layer in the network so that the accuracy is increased and the network is more powerful and more number of features can be identified. However, more layers create the issue of vanishing gradients. Vanishing gradient is nothing, but if there are more layers added, at some point the partial derivative vanishes and the loss function value is zero. Typically, step connections in ResNet are used to solve the vanishing gradient problem.

The vanishing gradient problem has now been handled by using step connections, however because there are many layers, this requires a lot of computing and other tasks. As a result, in EfficientNet, we also scale the width and resolution along with depth. Resolution scaling refers to the increase in pixels, while depth scaling refers to the increase in channels, breadth scaling to the increase in feature maps.By increasing the number of feature maps more characteristics from the image can be obtained and by resolution scaling more features can be learned and algorithm works better.Therefore, we scale the depth, width, and resolution evenly to avoid the vanishing gradient problem.

Now the scaling factor is required i.e, how much scaling is required.Now this is done by a technique called compound scaling which is uniquely implemented in efficientNet. Compound Scaling is done as follows:

$$f = d.w^s.r^s$$

f refers to how much we can scale, d refers to depth scaling factor, w refers to width scaling factor, r refers to resolution scaling factor, s value is obtained after grid search and value is 1.

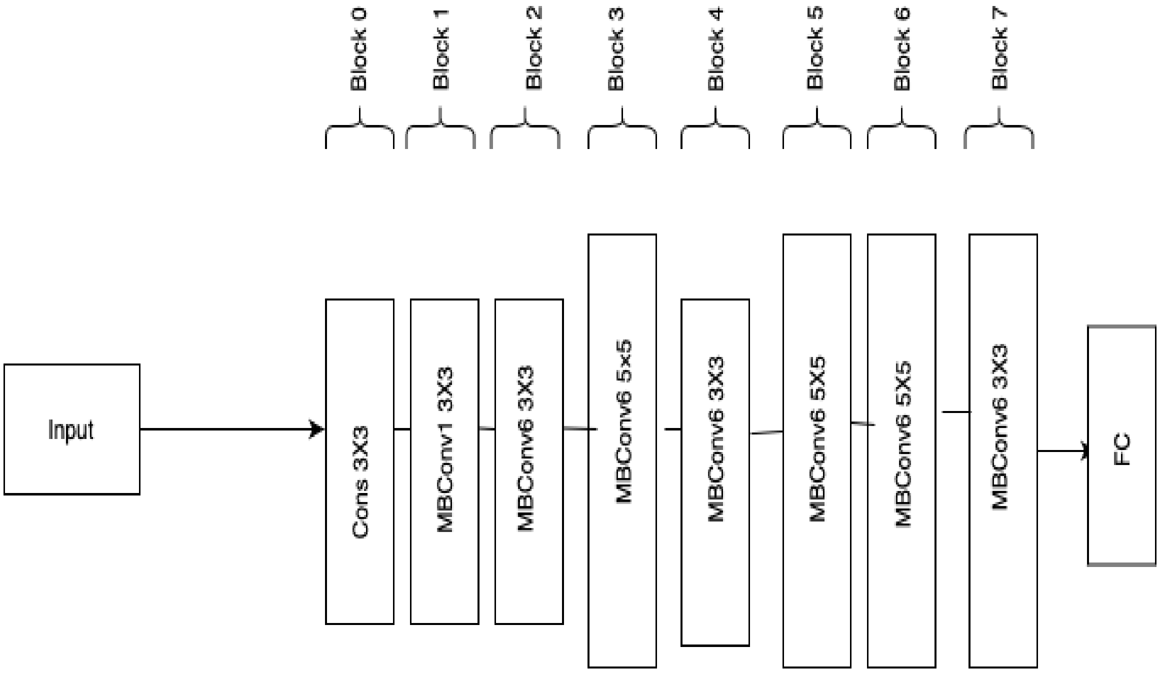


FIGURE 3.2: Architecture of EfficientNet

3.2.0.1 Architecture of EfficientNet

EfficientNet consists of 8 blocks.

- Block 0: Has Conv3X3 Convolutional layer
  Pooling layer
  Convolutional layer
  Pooling layer

- Block 1: Depthwise separable convolution
  Pointwise convolution
  Activation function

- Block 2: Depthwise separable convolution
  Pointwise convolution
  Activation function

- Block 3: Depthwise separable convolution with stride 2
  Pointwise convolution
  Activation function

- Block 4: Depthwise separable convolution
  Pointwise convolution
  Activation function

- Block 5: Depthwise separable convolution with stride 2
  Pointwise convolution
  Activation function

- Block 6: Depthwise separable convolution
  Pointwise convolution
  Activation function

- Block 7: Depthwise separable convolution with stride 2
  Pointwise convolution
  Activation function

Each block in the network is designed to extract features from the output of the previous block, and the combination of all the blocks produces a rich feature representation that is used by the head of the network to make a prediction. The use of depthwise separable convolutions and pointwise convolutions in each block helps to keep the network computationally efficient while still capturing important information from the input image.

MBconv, or Mobile Inverted Residual Bottleneck Convolution, is a building block used in many recent deep learning models, including the EfficientNet architecture. MBconv is a variation of the Inverted Residual Bottleneck Convolution (IRB) block, which is widely used in mobile networks due to its computational efficiency and effectiveness in feature extraction.

Conv layer, in EfficientNet is a standard layer used to extract features from the input data. By applying filters that are convolved over the input, convolutional layers are used to extract regional information from the input image. The activations created by the convolution operation are then used as input for successive layers in the network once the filters learn to recognise particular features in the input, such as edges, textures, or forms.

3.3 INCEPTION V3

Inception-v3 is aarchitecture of convolutional neural network (CNN) that was developed by Google researchers by combining convolutional layers with various filter sizes to collect features at different scales, it is intended to increase the efficiency and accuracy of image classification tasks.

Each module in the Inception v3 architecture is made up of a number of convolutional layers, pooling layers, and other types of layers. These modules are set up in a hierarchy, with lower-level modules concentrating on the capture of low-level features like edges and textures and higher-level modules concentrating on the capture of more complex features like object components and complete objects.The modules of this network is made up of parallel convolutional layers with various filter sizes and a max-pooling layer that helps in reducing the output's dimensionality. Each module's output is then combined and passed into the next module.Inception-v3 additionally employs batch normalisation and dropout regularisation.

Overall, it has been demonstrated that Inception v3 performs at the cutting edge on a variety of image recognition tasks, including object detection, image classification, and visual question answering.Additionally, this has been utilised to extract features that have proven to be highly effective for tasks related to transfer learning.

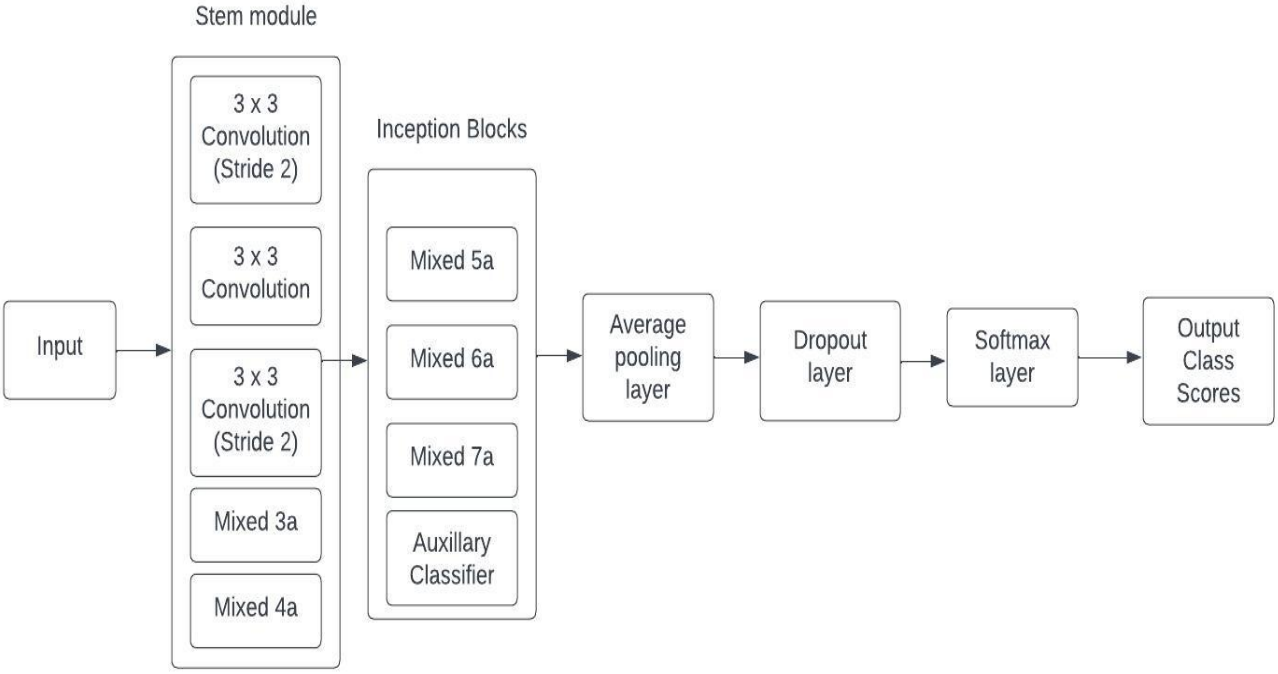


FIGURE 3.3: Architeture of InceptionV3

3.3.0.1 Architecture of InceptionV3

- Input layer: The layer that receives and transmits the raw image data to the following layer.

- Factorized convolutions: InceptionV3 substitutes more computationally effective factorised convolutions for conventional convolutions. The typical convolution is divided into two smaller convolutions by factorised convolutions, one along the input's width and the other along its height. As a result, the network's parameters are reduced, which speeds up network training and inference.

- Inception Modules:The foundation of the InceptionV3 architecture is made up of Inception modules. They include a number of 1x1 convolutional layers, pooling layers, and parallel convolutional layers of various sizes.

These modules are designed to collect features at various scales and merge them into a single layer.

- Better pooling: InceptionV3 substitutes ”global average pooling” for the more conventional ”max pooling” method. Global average pooling minimises overfitting and is more resistant to slight input fluctuations. Moreover,number od parameters in network is reduced.

- Auxiliary classifiers: InceptionV3 has two auxiliary classifiers that are coupled to the network’s intermediate layers. In order to increase overall accuracy and decrease overfitting, these auxiliary classifiers give the network more training signals during training.

- Batch normalisation: The batch normalisation layers in InceptionV3 serve to stabilise the learning process and lessen the network’s sensitivity to the initial weight values.

- stem network: The architecture of InceptionV3 includes a ”stem network” that is intended to preprocess input images and extract basic features. This aids in decreasing the amount of parameters and enhancing network performance.

- Reduction layers: InceptionV3 has ”reduction layers” that are used to increase the number of channels while decreasing the spatial dimensions of the feature maps. This aids in lowering the network’s computational complexity and raising performance.

- Fully connected layers: The final classification of the input image is carried out by the fully connected layers at the end of the network. Two fully connected layers, one with 2048 units and the other with 1000 units, are present in InceptionV3.

- Softmax layer: The softmax layer creates a probability distribution over the classes using the output of the final fully linked layer.

3.4 DenseNet

Dense Convolutional NetworkDenseNet (DenseNet) is a architecture used in deep learning that was first introduced by Huang et al. in 2017. It architecture of convolutional neural network (CNN) that is characterized by densely connected layers.

Many dense layers make up each of the dense blocks that make up DenseNet.Each layer in conventional CNNs only gets input from the layer before it. But in denseNet,each layer is a dense block is connected to all layers before it in a feed-forward manner,so DenseNet takes input from all layers that came before it. This allows for more effective parameter use and can help eliminate the vanishing gradient problem by reusing the feature maps created by earlier layers and lower the spatial dimensionality of the feature maps.

For a range of computer vision tasks, including image classification, object recognition, and semantic segmentation, DenseNet has been demonstrated to produce state-of-the-art results. It has also been applied in transfer learning for applications like as medical image analysis.

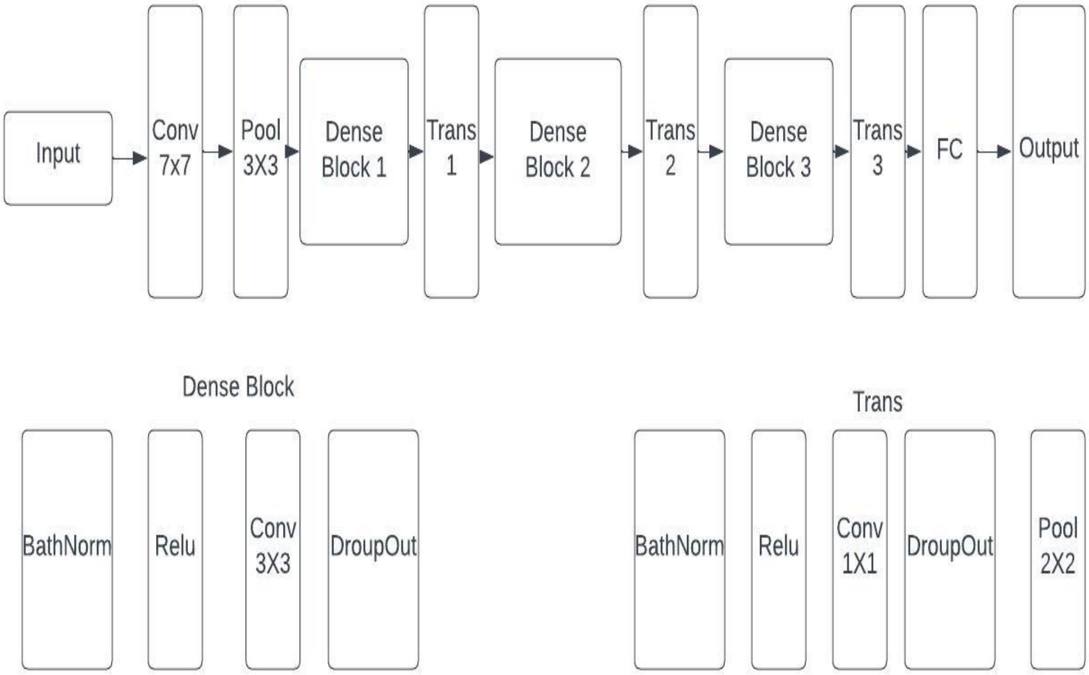


FIGURE 3.4: CNN ARCHITECTURE

3.4.0.1 Architecture of DenseNet

A number of dense blocks, each of which is made up of several dense layers, make up DenseNet. A batch normalisation layer, a Rectified Linear Unit (ReLU) activation function, and a convolutional layer sequence with the same number of filters make up each dense layer. All of the dense block's preceding layers' outputs are combined and sent into the current layer as input.

Each dense block's output is then transferred via a transition layer, which carries out a number of processes to reduce both the number and the spatial dimensions of the feature maps. A batch normalisation layer, a 1x1 convolutional layer with a reduction factor (often set to 0.5), and a 2x2 average pooling layer with stride 2 make up the transition layer.

The final classification output is created by applying a fully connected layer with softmax activation after the last dense block and a global average pooling layer to the feature maps.

- Input layer
- Initial convolutional layer with batch normalization and ReLU activation
- Dense blocks
    - Dense layers
        - Batch normalization layer
        - ReLU activation layer
        - Convolutional layer
    - Concatenation of all previous dense layer outputs
- Transition layer
    - Batch normalization layer
    - 1x1 convolutional layer with a reduction factor (usually 0.5)
    - 2x2 average pooling layer with stride 2
- Global average pooling layer
- Fully connected layer with softmax activation for classification output

3.5 MobileNet

MobileNet a CNN architecture creatred by Google researchers , which uses depthwise separable convolutions to lower the computing cost of the network without substantially reducing accuracy. The principle behind depthwise separable convolutions is to substitute two independent operations, a depthwise convolution and a pointwise convolution, for the normal convolution operation, which convolves the input tensor with a filter tensor. The depthwise convolution creates a set of output channels by performing a spatial convolution on each each input channel using a tiny filter kernel, often 3x3 or 5x5. The depthwise convolution's output is then subjected to a 1x1 convolution in the pointwise convolution, which combines the output channels using a linear filter. This network requires fewer calculations while still being able to learn complicated features because to the use of depthwise separable convolutions. This makes MobileNet an excellent option for devices with limited resources, including smartphones and embedded systems.

Overall, MobileNet is an effective CNN architecture that is small and lightweight, making it perfect for mobile and embedded devices. It may also be applied to other applications that have limited computational resources.

3.5.0.1 MobileNet Architecture

The primary concept underlying MobileNet is Using depthwise separable convolutional layers, which require fewer parameters and computations than conventional convolutional layers while maintaining high accuracy. Further

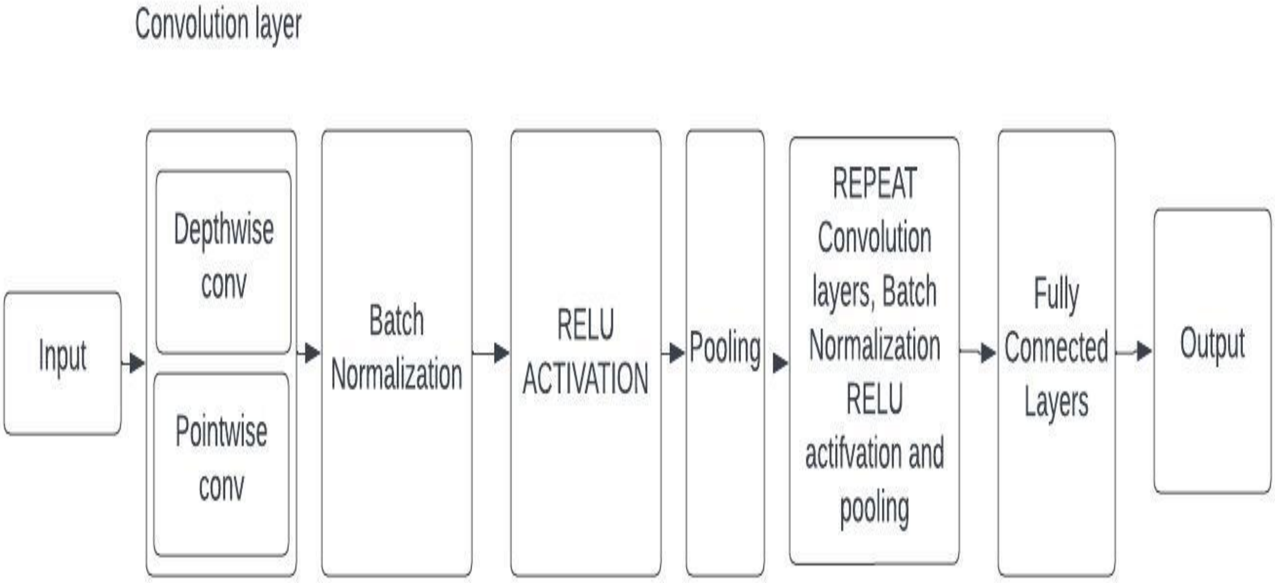


FIGURE 3.5: Architecture of MobileNet

enhancements including strided convolutions, batch normalisation, ReLU activation functions, and global average pooling are also included of MobileNet.

- Input Layer: An picture with a width and height of 224 pixels and three colour channels is used as the network's input (RGB).
- convolutional layer: The first layer has 32 filters, a 3x3 kernel size, and a stride of 2. This layer creates 32 feature maps after applying 32 different filters to the input image.
- Depthwise convolutional layer: The 32 feature maps created by the preceding layer are each subjected to a separate 3x3 depthwise convolution in this layer. 32 sets of feature maps with only one channel each are the end result.

- Pointwise convolutional layer:Each of the 32 sets of feature maps generated by the depthwise convolution is subjected to a 1x1 pointwise convolution in this layer. This enables the network to learn feature combinations across several channels.

- Depthwise convolutional layer followed by pointwise convolutional layer:By repeating depthwise and pointwise convolution, the number of channels are normally increased by a factor of two each time.

- Global average pooling layer:By averaging each feature map, this layer reduces the spatial dimensions of the feature maps to 1x1.

- Fully connected layer:The final layer of the network is a fully connected layer with a softmax activation function,which generates the probabilities for each class as the result.

3.6 Mask R-CNN

Mask R-CNN is a deep learning model for object detection and instance segmentation.By including a branch to forecast segmentation masks for each identified object, the Mask R-CNN model expands upon the Faster R-CNN model. As a result, the model is able to segment objects in a picture, i.e., identify which pixels are part of an object and which are not, in addition to detecting the objects in the image.

Mask R-CNN can be used in academic articles as a starting point or cutting-edge model for numerous computer vision problems, including object identification,

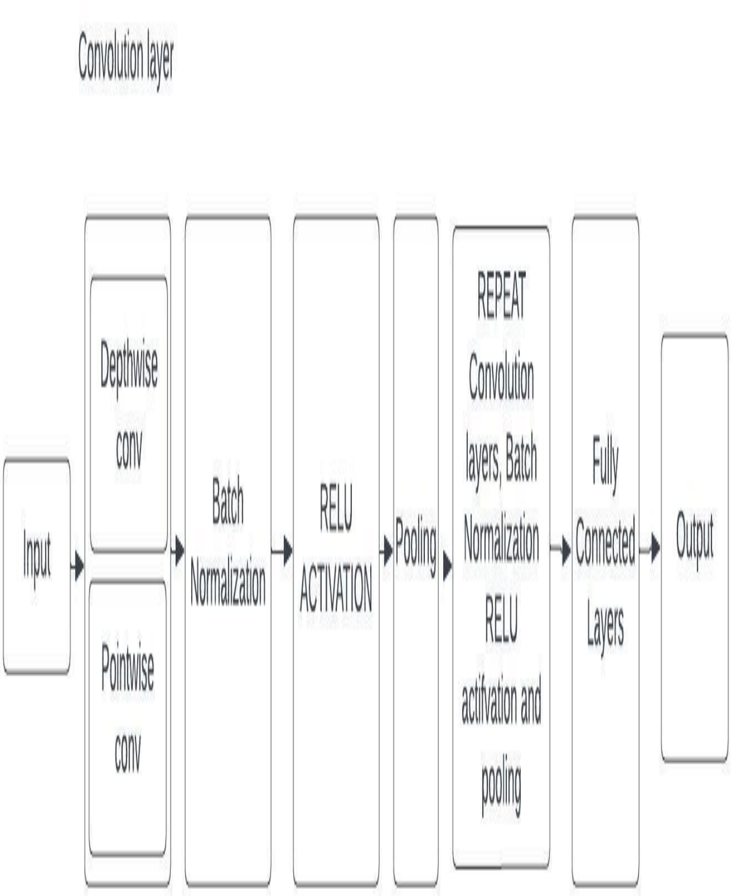


FIGURE 3.6: MobileNet Architecture

instance segmentation, and semantic segmentation. To assess the effectiveness of their suggested approaches or models, researchers can employ pre-trained Mask R-CNN models or train their own models on their own datasets. Moreover, researchers can use the pre-trained Mask R-CNN model to extract features from pictures that can be applied to other tasks or downstream models. This process is known as feature extraction or transfer learning.

Mask R-CNN has achieved state-of-the-art performance on a variety of object detection and segmentation benchmarks, including COCO and Cityscapes. It has been used in a wide range of applications, such as autonomous driving, medical imaging, and robotics.

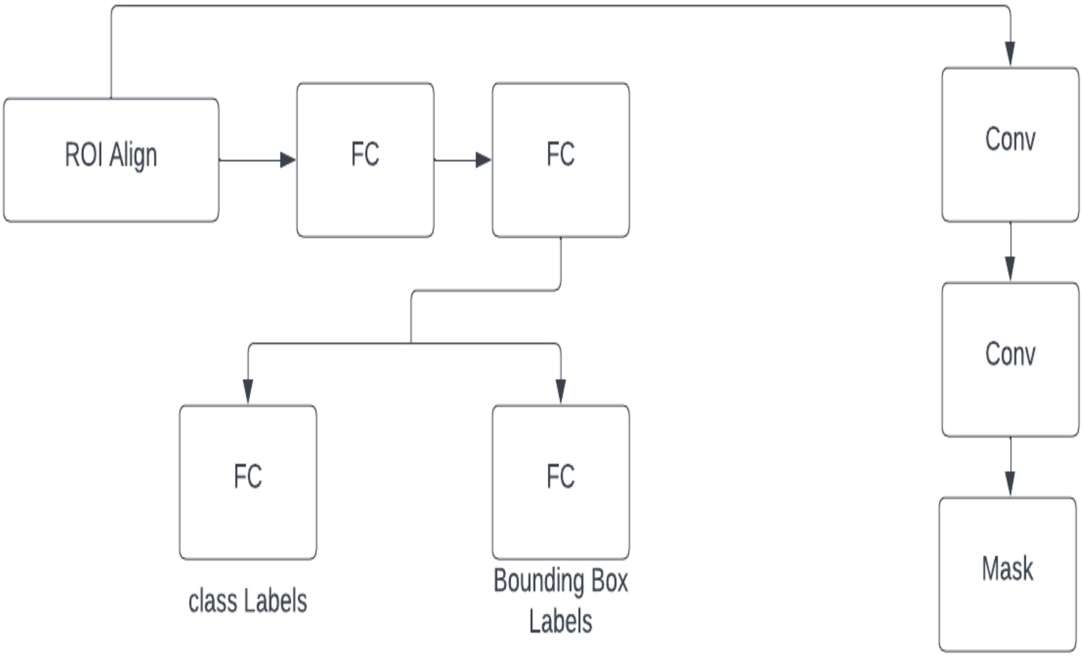


FIGURE 3.7: Architecture of Maskrcnn

- Input Layer:This layer receives the input image and sends it to the network.
- Backbone network: Convolutional neural networks (CNNs) are frequently employed as the backbone network because they are effective at extracting information from input images. The backbone network frequently uses ResNet, VGG, and Inception.
- Region Proposal Network (RPN) layers: The backbone network's feature maps are used by the RPN layers to produce a set of candidate object proposals, each of which is represented by a bounding box. The RPN layers

typically start with a convolutional layer and then have two sister layers that predict the bounding box's coordinates and the likelihood that an anchor box would include an object.

- Pointwise convolutional layer:Each of the 32 sets of feature maps generated by the depthwise convolution is subjected to a 1x1 pointwise convolution in this layer. This enables the network to learn feature combinations across several channels.

- RoI (Region of Interest) Pooling layer:For each object suggestion produced by the RPN layers, the RoI (Region of Interest) pooling layer extracts a fixed-size feature map.

- Mask Head layers: The Mask Head layers create a binary mask for each item proposition using the feature maps produced by the RoI pooling layer. Many convolutional layers and a fully linked layer make up the layers that make up the Mask Head.

- Bounding box refinement layer: This layer improves the bounding boxes' coordinates after they have been created by the RPN layers and the Mask Head layers.

- Detection layer: The detection layer creates the final object detections and instance segmentations by combining the bounding boxes and binary masks produced by the RPN layers and the Mask Head layers.

- output Layer: This layer provides the final segmentations of instances and object detections.

CHAPTER 4

Proposed System Design

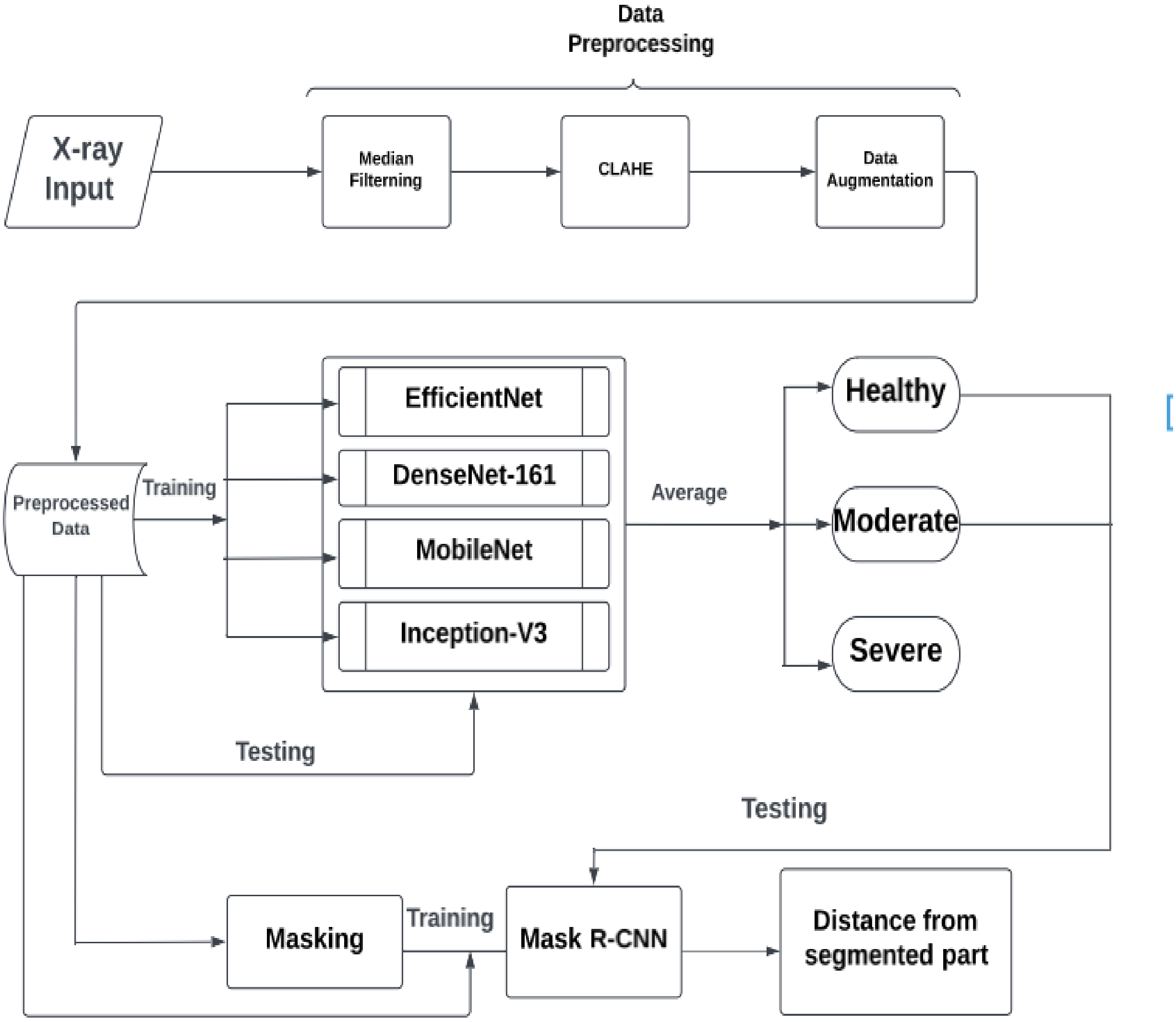


FIGURE 4.1: Overall System Architeture

The system that was utilised to carry out the task is described in detail in the section that follows, along with an explanation of each component. The system is mainly divided into two parts: the system needed for classification is in the first part, and the system needed for distance finding is in the second.Now lets look about all the components in detail in the upcoming parts.

4.1 System overall Algorithm

Algorithm 1 Classifying the Knee Oesteoarthritis Severity and distance

```
Input: X-ray Image.
Output: Grade of Disease                                        ▷ Grade 0/1/2/3/4
function CLASSIFYGRADE(image)
    for All the Images in the Dataset do
        $ImageWithLessNoise \leftarrow XrayImage$                 ▷ Median Filtering
        $HighContastImage \leftarrow ReducedNoiseImage$           ▷ CLAHE
    end for
    $BalancedDataset \leftarrow UnbalancedDataset$              ▷ Data Augmentation
    $TrainingSet + TestingSet \leftarrow Dataset$
    $Inception \leftarrow Create, Train, Test$
    $EfficientNet \leftarrow Create, Train, Test$
    $MobileNet \leftarrow Create, Train, Test$
    $DenseNet \leftarrow Create, Train, Test$
            ▷ emsemble Model is Created by taking the average of the probability
    $EnsembleModel \leftarrow EfficientNet + MobileNet + DenseNet + IncpetionV3$
    $Healthy/Moderate/Servere \leftarrow EnsembleModel$
    if Healthy or Modrate then
        Find Distance
    end if
end function
```

4.2 Design Overview

The input is first subjected to dataprepocessing; the methods employed are median filtering and CLACHE algorithms.Dataset is balanced with the right data augmentation.

In order to categorise the severity of disease based on the grades of healthy, moderate, and severe, the preprocessed data is now submitted via the ensemble model.MobileNet, IncpetionV3, DenseNet, and Efficient Net are the models utilised for ensemble.

Now the distance is found out for the images which are classified as healthy and moderate only.

4.3 Dataset used

X-ray scans of the knee joints from patients with knee osteoarthritis make up the dataset. The Kellgren and Lawrence grading system [10], a commonly recognised approach for determining the severity of knee osteoarthritis, is used to label the dataset [13].

826 photos are included in the dataset for validation, 1656 for testing, and 5778 for training. Each image has a label indicating the Kellgren and Lawrence grade that corresponds to it, which goes from 0 to 4. Grades 0 through 4 denote increasing degrees of osteoarthritic alterations in the knee joint, with grade 0 indicating no osteoarthritic characteristics.

The dataset can be used to create machine learning models that automatically categorise X-ray pictures of knee osteoarthritis into different grades and is freely accessible on Mendeley Data.

To avoid overfitting with this dataset, it is crucial to thoroughly preprocess the data and utilise the proper data augmentation strategies.

X-ray Images			
Disease Grade	Training	Testing	Validation
All Grades	5778	1656	826
Grade 0	2286	639	328
Grade 1	1046	256	153
Grade 2	1516	447	212
Grade 3	757	223	106
Grade 4	173	51	27

Further modules are explained below.

4.4 Median Filtering

A common method for processing photos that removes noise is median filtering. N. S. Young et al. provided a detailed explanation of the procedure [13].

The fundamental concept behind median filtering is to replace each pixel in an image with the median value of the pixels around it. To do this, a small window is moved over the image, and for each pixel in the window, the median of the pixel values in the window is computed.

The ability of median filtering to maintain image edges while eliminating noise is one of its key benefits. This is due to the fact that the median operation typically keeps the distinct image transitions while eliminating the smaller, noise-related spurious fluctuations.

The recursive median filter works by iteratively applying the median filter to a series of images, each of which is produced by taking the original image and deleting the most recent estimate of the noise.

Overall, median filtering is a widely used and straightforward technique for reducing noise from images.

4.5 CLAHE

The notion of Contrast Limited Adaptive Histogram Equalisation [14] (CLAHE) was studied by Karel Zuiderveld.

A typical method in image processing for enhancing contrast in an image is histogram equalisation. The technique involves rearranging an image's values of intensity in order to produce a more consistent histogram. Standard histogram equalisation, however, might result in the overamplification of noise in low contrast areas.

With CLAHE, histogram equalisation is modified such that it is selectively applied to certain areas of an image rather than the full image, limiting the amplification of noise. The image is separated into discrete areas in CLAHE called tiles, and histogram equalisation is applied to each tile individually. To prevent noise from being amplified, the histogram of each tile is clipped at a specific maximum value. The final image is created by stitching the resulting image back together.

One of CLAHE's key benefits is its capacity to boost contrast while retaining details and preventing noise overamplification in an image.

4.6 Data Augmentation

Data augmentation is done by intentionally producing additional, slightly altered versions of the original data, the machine learning technique allows one to enhance the size of a training dataset. The idea of data augmentation was initially presented in a study[15] by Alex Krizhevsky et al.

Data augmentation is predicated on the notion that a machine learning model would perform better on fresh, unexplored data the more training data it is exposed to. However, it can frequently be time-consuming and expensive to gather significant volumes of high-quality training data. This issue is solved by adding fresh training examples that are roughly identical to the original data but with minor differences.

Applying a variety of straightforward modifications to the source images, such as rotation, translation, scaling, flipping, and cropping, is the most popular method of data augmentation for image collections. From a small amount of source images, a significant number of new training examples can be produced by applying these changes randomly and in various combinations.

4.7 Classification Model

Preprocessed data is now used to train each model individually, including EfficientNet, InceptionV3, MobileNet, and DenseNet.After each model has been trained, new data from the dataset is selected as test data, and each model is then independently tested.

After the individual model training and testing, each model's gives a probability distribution for each image belonging to certain class. Finally, we develop an ensemble model that selects the class with the highest probability distribution as the predicted class by averaging [**?**] the probability distribution.The ensemble model must be trained and tested seperately.

Now distance that is the joint space width for the classes healthy and moderate are found.

4.8 Creating the masks

The technique of segmenting an image or a collection of images to pinpoint particular regions of interest is also referred to as creating masks. These areas are separated and extracted using masks, and they can subsequently be used for a variety of tasks like object detection, tracking, and measuring.

Then we create mask for all the x ray images. The mask should be created using a polygon Tool.

4.9 Training Mask R CNN

To train your Mask R-CNN model, you’ll need to compile and classify a dataset of images. The objects in the images that you wish to recognise is the space between the joints and segment must be annotated.

The correct parameters for your Mask R-CNN model, such as the number of classes you wish to detect, the learning rate, and the number of training epochs must be set.After the model has been trained, performance on a validation dataset to determine how well it does on photos that it has never seen before.

4.10 Segmenting and finding distance

Use the above trained model to perform segmentation on the images. The segmented part will be covering the region between two bones, the distance between the bones can be found by calculating the centre point and then the points above and below are chosen to find the distance. Distance is the important factor in determining the degree of injury. Measuring the distance between knee bones can help detect osteoarthritis at an early stage. With this distance we will tell how prone the patient if from going to the next stage.

CHAPTER 5

IMPLEMENTATION

5.1 Libraries and Technologies used

TensorFlow: An open-source framework for creating and deploying machine learning models. It was created by Google and is widely utilised in business and education. Users can build and train neural networks using TensorFlow, which also offers tools for deploying models in real-world settings.

Keras:A high-level neural network API called Keras was created in Python and could potentially used with TensorFlow, CNTK, or Theano. With its user-friendly interface for constructing models and training neural networks, Keras makes it simple to construct and experiment with neural networks.

NumPy: The Python package NumPy is used to manipulate arrays. It offers resources for performing addition, multiplication, and dot product operations on arrays. Data analysis, scientific computing, and machine learning all make extensive use of NumPy.

Pandas:A Python library for analysing and manipulating data. Users can execute operations like merging, grouping, and filtering using the tools provided for working with tabular data, such as data frames.

Matplotlib:A Python package for making visualisations. It offers resources for making many different kinds of graphs, including line plots, scatter plots, bar charts, and more. Scientific computing and data analysis both make extensive use of Matplotlib.

OpenCV: Open Source Computer Vision Library is an image and video processing computer vision library. It offers resources for operations like face recognition, image segmentation, and object detection.

SeaBorn:A Python module called Seaborn is used to produce statistical visualisations. It offers resources for making many different kinds of graphs, including histograms and heatmaps. Machine learning and data analysis both frequently employ Seaborn.

5.2 Detailed Implementation of Each Module

5.2.1 Data Preprocessing

A key stage in data analysis and machine learning, ata preprocessing is preparing and transforming raw data into a more useful form. Data cleaning, transformation, and normalisation are often involved in order to make the data more acceptable for analysis and modelling.

Data preprocessing's main objective is to make sure the data is precise, consistent, and appropriate for the particular analysis or machine learning activity at hand. This often involves identifying and modifying the most crucial elements or variables in the data, as well as dealing with problems including missing data, outliers, inconsistencies, and formatting errors.

Overall, because it can considerably affect the accuracy and quality of the data, data preparation is an important phase in the data analysis process.

The preprocessing techniques we employ are Median Filtering,CLAHE,Data AUgmentation.

5.2.1.1 Median Filtering

A median filter is a digital signal processing method that effectively reduces noise and maintains edges by replacing each pixel value in an image or signal with the median value of its neighbours.

A specific factor value is chosen that is how many neeighbouring pixels to look at.A Window used in the filtering process have a width and height of 5 pixels because the median filtering ”factor” is 5. This shows that the median filter calculates the median value of the pixels in the 5x5 neighbourhood surrounding each pixel in the image or signal to replace the original value at the neighborhood’s centre.

5.2.1.2 CLAHE

A computer image processing technique called CLAHE (Contrast Limited Adaptive Histogram Equalisation) divides an image into small, overlapping sub-regions and applies histogram equalisation separately to each sub-region in order to improve contrast while limiting the amplification of noise.

Clip limit describes a setting that regulates the maximum amount of contrast enhancement that can be applied to each individual pixel in an image.The clip limit chosen is 2.

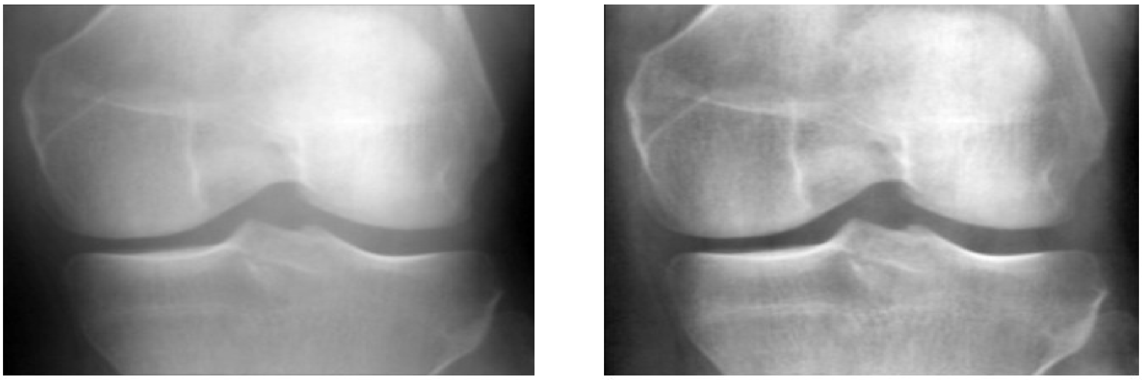


FIGURE 5.1: Before And After

Above is an example of a Xray image where before and after median filter and CLAHE is applyed.We can see that after the techniques image is enhanced more.

5.2.1.3 Data Augmentation

Data Augmentation: Data augmentation is a technique used to artificially increase the size of a training dataset by applying various transformations to the existing data. We applied horizontal flip and a rotation of 20 degrees on both side on the existing images for performing augmentation

Data Augmentation		
Grade	BEFORE	AFTER
HEALTHY	4848	500
MODERATE	757	500
SEVERE	173	500

5.2.2 Hyper Parameter

Hyperparameters are variables that need to be set before a machine learning algorithm can learn them during training. These variables affect the learning process in a variety of ways, including how many hidden layers there are in a neural network, how quickly the optimisation algorithm learns, how strong the regularisation is, and how many iterations or epochs there are.

5.2.2.1 Learning Rate

This hyperparameter that controls how much the model weights should be adjusted in response to the training error. A lower learning rate might cause sluggish convergence or being stuck in local minima, whereas a higher learning rate can lead to faster convergence but also to overshooting the ideal weights.

5.2.2.2 Number of Epoch

It determines how many times the entire dataset will be used to train the model

5.2.2.3 Batch Size

Determines how many samples are used in each training iteration. Although larger batch sizes may shorten training durations, they may also need more memory and may not generalize as well.

5.2.2.4 Activation Funcation

a neural network function that adds nonlinearity to each neuron's output and aids in the network's ability to recognize complicated patterns in the data. The activation methods ReLU, sigmoid, and softmax are frequently used.

5.2.2.5 Optimisation

A method for determining the ideal set of model parameters to minimize the training loss function. Algorithms for optimization including Adam, RMSProp, and gradient descent are frequently used.

5.2.2.6 Dropout Rate

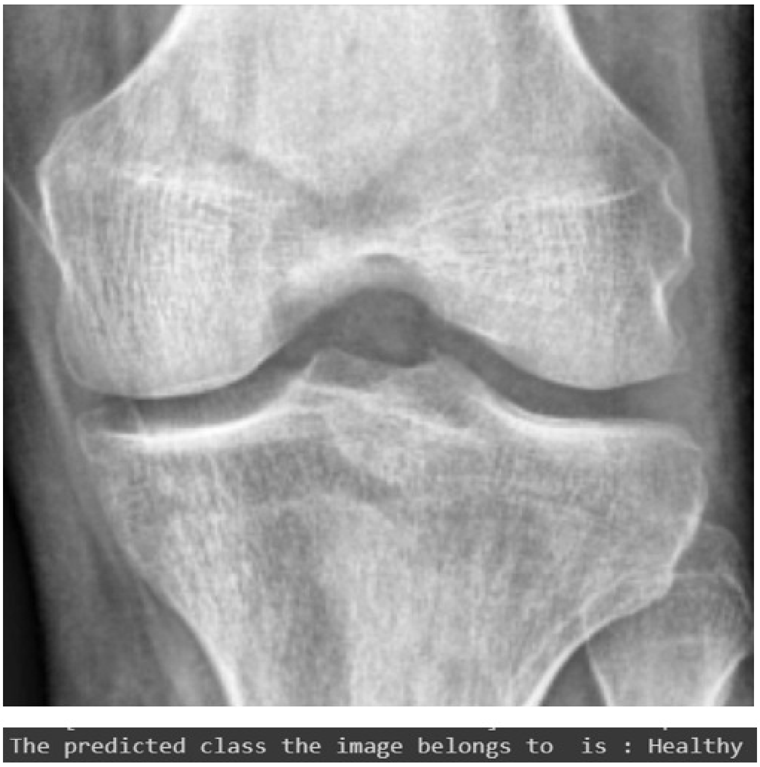
a regularization hyperparameter that, during training, randomly ”drops out” (sets to zero) a subset of the neurons in a layer in order to avoid overfitting.

5.2.2.7 Loss Function

A function that measures how well the model is performing on a given task during training. The objective is to reduce the gap between the expected and actual values, or the loss function. Mean squared error, cross-entropy, and hinge loss are typical loss functions.

Data Augmentation	
HyperParamter	Value
Learning Rate	0.001
Number of Epoch	40
Batch Size	32
Activation Funcation	SoftMax
Optimisation	Adam Optimiser
Dropout Rate	0.4
Loss Function	categorical crossentropy

5.2.3 Training and Ensemble



```
The predicted class the image belongs to  is : Healthy
```

FIGURE 5.2: Class Predicated by Ensemble

The preprocessed images are used for training all the models. Every image is of 224x224 size. The total train dataset is divided into batch size of 32 for each models. The training procedure starts with a learning rate of 0.001, which is gradually decreased to 0.0005 after 20 epochs. For training, the Adam optimizer is utilized. An ensemble model is then built from the individual models. For image classification, each model creates a probability distribution. The ensemble model is designed to combine the results from each model by averaging the probability distributions and choosing the class with the highest probability. The ensemble model is trained with a separate dataset and evaluated accordingly.

5.2.4 Finding Distance

5.2.4.1 Creating Mask

We manually create a mask for all the X-ray images using polygon tool such that the region of interest is identified.

5.2.4.2 Using Mask For training the data

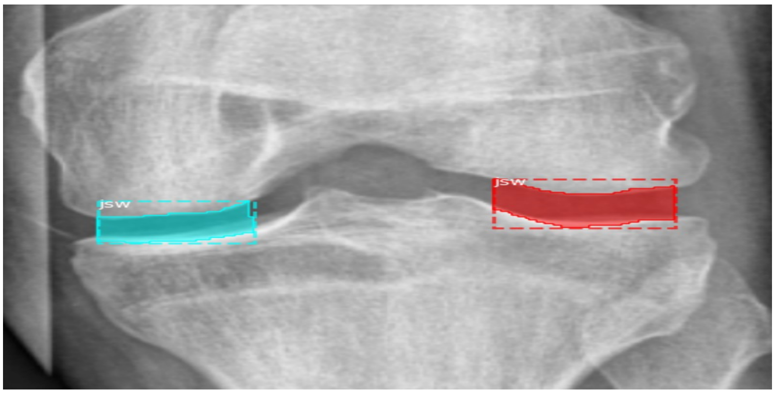


FIGURE 5.3: Regoin identified by maskrcnn

We use Mark rcnn architecture for segmentation purpose, the architecture is trained using the xray images with mask.

5.2.4.3 Finding Joint Space Width

First we identify the centroid of masked region and after finding this point we find the points upper and below the centroid.After finding the top and bottom points the distance is foundout.Now the left and right point of centroid are found and then again the processes is repeated.Now the minimum distance is chosen.

CHAPTER 6

Performance Evaluation

6.1 Metric Used

For evaluating the perforamance of the model built the following performance metrics are made use of Precision,recall,F1 score,Support and Accuracy for classification.Intersection over union is used as the performance metric for segmentation.

6.1.1 Precision

A metric that evaluates the percentage of samples that were accurately identified as positive out of all samples that were positive. In order to compute it, divide the total number of true positives by the total number of false positives, or TP / (TP + FP).

6.1.2 Recall

a metric which measures the percentage of samples that are truly positive out of all samples. It is determined using the formula TP / (TP + FN), where FN stands for false negatives.

6.1.3 F1 score

An evaluation that balances precision and recall by combining the two parameters into one score. The harmonic mean of precision and recall is determined as follows: 2 * (precision * recall) / (precision + recall).

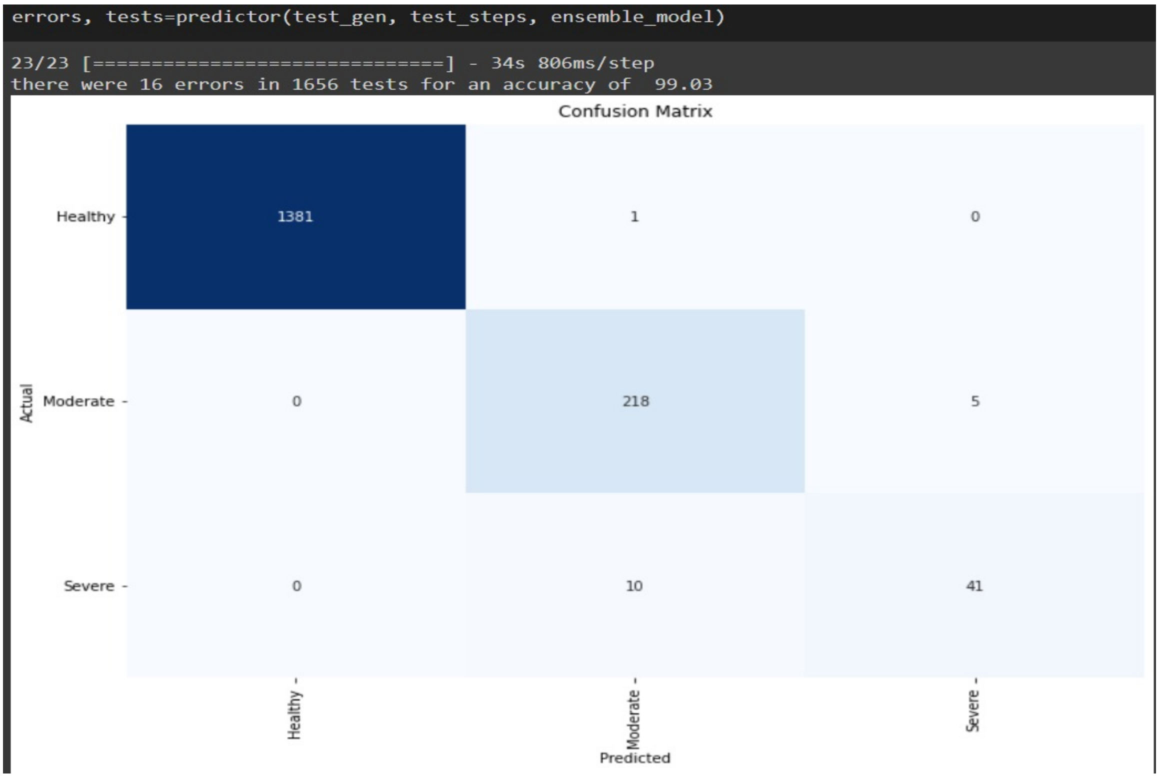
6.1.4 Support

The number of samples in each class. For problems involving multi-class classification, this is used to determine the weighted average of the performance indicators.

6.1.5 Accuracy

A metric which calculates the percentage of samples that were correctly identified out of all samples. The formula is (TP + TN) / (TP + TN + FP + FN), where FP stands for false positives and TN for true negatives.

Accuracy	
Model	Accuracy
EfficientNet	94.14
DenseNet	93.42
MobileNet	93.66
InceptionV3	94.08
Ensemble	99.03



```
errors, tests=predictor(test_gen, test_steps, ensemble_model)

23/23 [==============================] - 34s 806ms/step
there were 16 errors in 1656 tests for an accuracy of  99.03
```

FIGURE 6.1: Confusion Matrix

6.1.6 Intersection over Union

Intersection over Union (IoU) or also called as Jaccard Index is used to evaluate the accuracy of an object detector or segmentation algorithm. By computing the ratio of the areas of the intersection and union of the two, it determines the overlap between the predicted bounding box or segmentation mask and the ground truth.

We used IOU as the metrics for segmentation and we obtained an IOU score 0f 0.7321

CHAPTER 7

Conclusion

The study's finding is that the suggested deep learning model could potentially be able to precisely predict the degree of knee osteoarthritis from X-ray pictures. The model demonstrated great accuracy in classifying the severity of the disease using the Kellgren-Lawrence grading system after being trained on a sizable sample of X-ray images.

Further, our work will also be able to locate the region where the disease occurs and determine the distance between the knee bones. According to the study, this method might be applied in clinical settings to help radiologists and doctors correctly diagnose and treat knee osteoarthritis.

However, there are certain limitation that will be discussed in upcoming sections, based on those, additional study is required to confirm the model's precision in actual clinical situations and to increase its generalizability to various groups.

7.0.1 Limitations

To obtain accurate measurements of the distance in knee osteoarthritis patients, it is essential that the X-ray beam is projected parallel to the knee joint from a front view. Any changes in the viewpoint will affect the measured distance.

In instances where X-ray images are taken from a side angle or with the patient in a different position, the distance measurement may not be precise and can vary. For our project, we only include X-ray images of the knee joint taken from a straight front view.

Additionally, it is not possible to calculate the distance for grade 4 patients as their knee bones may be touching or even overlapping, resulting in a negative distance measurement. Therefore, we only calculate the distance for patients with grades 0, 1, 2, and 3.

7.0.2 Future Work

Automatic masking can be tried instead of manually creating the mask. Manually creating the mask can be tedious and time consuming

Various types of X Ray images can also be used to study the performance of masking and classification. There are several types of x rays taken to evaluate the osteo arthritis disease and calculating the distance in each image will be different

REFERENCES

1. Ahmed, S.M.; Mstafa, R.J. Identifying Severity Grading of Knee Osteoarthritis from X-ray Images Using an Efficient Mixture of Deep Learning and Machine Learning Models. Diagnostics 2022, 12, 2939. https://doi.org/10.3390/diagnostics12122939

2. Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv:1905.11946

3. G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, ”Densely Connected Convolutional Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243.

4. Howard, Andrew Zhu, Menglong Chen, Bo Kalenichenko, Dmitry Wang, Weijun Weyand, Tobias Andreetto, Marco Adam, Hartwig. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

5. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, ”Rethinking the Inception Architecture for Computer Vision,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.

6. Abedin, Jaynal Antony, Joseph McGuinness, Kevin Moran, Kieran O’Connor, Noel Rebholz-Schuhman, Dietrich Newell, John. (2019). Predicting knee osteoarthritis severity: comparative modeling based on patient’s data and plain X-ray images. Scientific Reports. 9. 10.1038/s41598-019-42215-9.

7. Thomas KA, Kidziński Ł, Halilaj E, Fleming SL, Venkataraman GR, Oei EHG, Gold GE, Delp SL. Automated Classification of Radiographic Knee Osteoarthritis Severity Using Deep Neural Networks. Radiol Artif Intell. 2020 Mar 18;2(2):e190065. doi: 10.1148/ryai.2020190065. PMID: 32280948; PMCID: PMC7104788.

8. Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. Sci Rep. 2018 Jan 29;8(1):1727. doi: 10.1038/s41598-018-20132-7. PMID: 29379060; PMCID: PMC5789045.

9. Ardila, D., Kiraly, A.P., Bharadwaj, S. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 25, 954–961 (2019). https://doi.org/10.1038/s41591-019-0447-x

10. Kohn MD, Sassoon AA, Fernando ND. Classifications in Brief: Kellgren-Lawrence Classification of Osteoarthritis. Clin Orthop Relat Res. 2016 Aug;474(8):1886-93. doi: 10.1007/s11999-016-4732-4. Epub 2016 Feb 12. PMID: 26872913; PMCID: PMC4925407.

11. Marques G, Agarwal D, de la Torre Díez I. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. Appl Soft Comput. 2020 Nov;96:106691. doi: 10.1016/j.asoc.2020.106691. Epub 2020 Aug 29. PMID: 33519327; PMCID: PMC7836808.

12. Mustapha Oloko-Oba and Serestina Viriri Ensemble of EfficientNets for the Diagnosis of Tuberculosis https://doi.org/10.1155/2021/9790894

13. Chen, Pingjun (2018), “Knee Osteoarthritis Severity Grading Dataset”, Mendeley Data, V1, doi: 10.17632/56rmx5bjcr.1

14. Russ, John C. ”The Digital Image Processing Handbook.” CRC Press, 1995.

15. Zuiderveld, Karel. ”Contrast Limited Adaptive Histogram Equalization.” Graphics Gems IV, Academic Press, 1994.

16. Krizhevsky, Alex, and Hinton, Geoffrey. ”Data Augmentation for Image Classification: Random Crop and Patch Techniques.” Proceedings of the 18th International Conference on Neural Information Processing Systems (NIPS), 2012.