Jimma University
Jimma Institute of Technology
Faculty of Computing and Informatics
Program of Information Technology

MSc in Information technology
MSc Thesis Research
On

Classification of Chronic Obstructive Pulmonary Diseases from Chest X-Ray Images Using Deep Learning

By
Amanuel Meseret

A Thesis Submitted to the School of Graduate Studies of Jimma University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Technology

Principal Advisor
Kula kekeba (PhD)

Co-advisor
Tafary Kababa (Msc)

November 2021
Jimma, Ethiopia

Declaration

This thesis is presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with proper citation of sources. I, the undersigned, declare that this thesis has not been presented for a degree in any other university

Amanuel Meseret

Jimma University
Jimma Institute of Technology
Faculty of Computing and Informatics

A Thesis Submitted to the School of Graduate Studies of Jimma University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information technology

Approval sheet

Thesis Title: Classification of Chronic obstructive pulmonary diseases from Chest X-ray Images using deep learning

Submitted by:

Mr Amanuel Meseret — Signature — Nov. 25, 2021 Date

Approved by:

Name and Signature of advisors

Kula Kekeba (PhD) Principal advisor	Signature	Nov. 25, 2021 Date
Tafary Kababa (MSc) Co-advisor	Signature	Nov. 25, 2021 Date

Name and Signature of the Examining Board

Mr Tesfu Mekonen Chairman, Examining Board	Signature	Date
Dr Getachew Mamo Internal Examinar	Signature	Date
Dr. Teklu Urgessa External Examinar	Signature	Dec 15,2021 Date

December 2021
Jimma, Ethiopia

ACKNOWLEDGMENT

First and foremost, I want to express my gratitude to the beloved God, who is the Source of Wisdom and the Origin of Knowledge. I praise and thank Him for all of His assistance during this work. My deepest gratitude is to my advisors, Dr. Kuulaa Qaqqabaa and Mr. Tafary Kababa, I am deeply thankful to them for their patience, motivation, enthusiasm, immense knowledge, willingness, and commitment. They were volunteered to invest their time to assist me in completing this research study, as well as providing their views and critical criticism.

I would also thank Dr. Mulusew Gerbaba, for his teaching me about Machine Learning and Depp Learning Concepts. He helped me a lot in motivating me to do my work with high quality. My sincere thanks also go to the staff of Addis Abeba Tikur Anbesa specialized Hospital, st paul's specialized Hospital, and Betele specialized Hospital, and MSF Holland Gambella Branch who have provided a collection of the possible x-ray images. A special thanks go to Dr. Samule Gizaw (Radiologist), Dr. Lamlem Terefe ((Radiologist), Dr. Mabratu Kassay for their willingness to sharing of the x-ray images and provide constructive ideas about the detection and classification of chronic obstructive pulmonary Disease.

I'd also want to thank my best friend Dr. Abel Tariku (Radiologist), Dr. Minitesinot Ferede (Radiologist), and others who have spent many hours discussing how to make things work out, and who are prepared to offer the possible x-ray images as soon as. Finally, I'd want to express my gratitude to my father, Mr. Meseret Bekele, and mother, Mrs. Genet Muleta and my Girl Friend Dr. Hawi Teshale, as well as the rest of my family and friends, who have all contributed in some way to my academic success. Without the love and patience of my family and friends, none of this would have been possible. Their encouragement and concern aided me in staying focused on my graduate studies.

Contents

List of Figures **Page Number**

List of Table **Page Number**

Abstract

To reduce the risk of Chronic Obstructive Pulmonary Disease and we have proposed the applications of digital image processing techniques. To accomplish our study, we have adopted a design science methodology and followed its scientific procedures starting from collecting the required data set to test the developed model. We have collected about 2248 images having 350 Images or more for each class. We have applied different image preprocessing tasks to enhance the image. And augmentation is applied to increase the number of images to a total of 2248 chest X-ray Images. Therefore, to overcome that problem, we applied zooming, rotation, and flipping at a different angle as augmentation techniques. Then Features are extracted from gray-level images using a CNN feature extraction and a classification model is built using 5 Different Pre-trained models called InceptionV3, VGG16, EffeceintNetB0, and Resnet50 including our own CNN model.

The convolutional neural network architecture with the sequential model is implemented with many layers such as convolutional, activation, max-pooling to extract important features from the Chronic Obstructive Pulmonary diseases x-ray image.

A total of 2248 COPD Chest X-ray datasets were collected from St. Paul Millennium Medical College Hospital Black Lion Specialized Hospital, Betele Specialized hospital, ReftyVally University Collage Specialized Hospital, MSF Holland Medical Center Gambella Branch, and Jimma University Medical Center. An adequate set of a report for labeling was not available and requires tremendous effort and time. We have used 80/20 by splitting the data into 80% for training and 20% for tests. Transfer learning and data augmentation techniques were applied. The proposed CNN classification model achieved an average accuracy of 81.1%. While the InceptionV3 with its filtering mechanism has achieved a better classification performance with an accuracy of 90.1% and it was reported as Highly Accuracy we obtained through the investigation.

The study could contribute to the medical profession by providing a system that supports experts to estimate chronological age

Keywords: VGG16, ResNet50, EffeceientNetB0, CNN, Neural network, Feature extraction Design Science, Transfer Learning, Deep Learning

CHAPTER ONE
INTRODUCTION

1.1 Background of the study

Respiratory diseases cause a huge worldwide health burden, hundreds of millions of people suffer and, more than 1 million persons suffer from chronic respiratory conditions. At least 2 billion people are exposed to the toxic effects of biomass fuel consumption, 1 billion are exposed to outdoor air pollution and 1 billion are exposed to tobacco smoke. Each year, 4 million people die prematurely from chronic respiratory disease. Infants and young children are particularly susceptible [1]. Nine million children under 5 years of age die annually and lung diseases are the most common causes of these deaths. COPD is the leading cause of death worldwide and the numbers are growing. Among the type of COPD Asthma is the most common chronic disease, affecting about 21% of children globally and rising and chronic bronchitis killing over 9.2 million worldwide also 3.5 million people have been diagnosed with emphysema with more than 90 percent of cases involving people over age 45.[3]

In lung disease diagnosis, clinicians integrate their medical knowledge and chest x-ray image to obtain the nature and pathological characteristics of lung diseases and to decide on treatment options. However, manual detection and classification of lung diseases in CXR, where a large number of CXR is taken for each patient, is tedious and subjected to inter-observer and intra observer detection and classification variability [2].

1.1.1 Lung anatomy

The lungs are an internal organ in the body and the only internal organ that is exposed constantly to the external environment. Everyone who breathes is vulnerable to the infectious and toxic agents in the air. While respiratory disease causes death in all regions of the globe and all social classes, certain people are more vulnerable to environmental exposures than others [1].

Lungs are a pair of air-filled organ that is connected to the trachea by the right and left bronchi; on the under surface, the lungs are bordered by the diaphragm. They are covered by the pleurae, which are attached to the mediastinum. The main function of the lung is the process of gas exchange known as respiration, which gives enough supply of oxygen to the different parts of the body and gets rid of carbon dioxide and other waste from the blood. Lung disease is any problem

In the lung that prevents proper function of the lung; there are different lung diseases due to different causes. Lung diseases are diagnosed by visualization or interpretation of chest radiography.

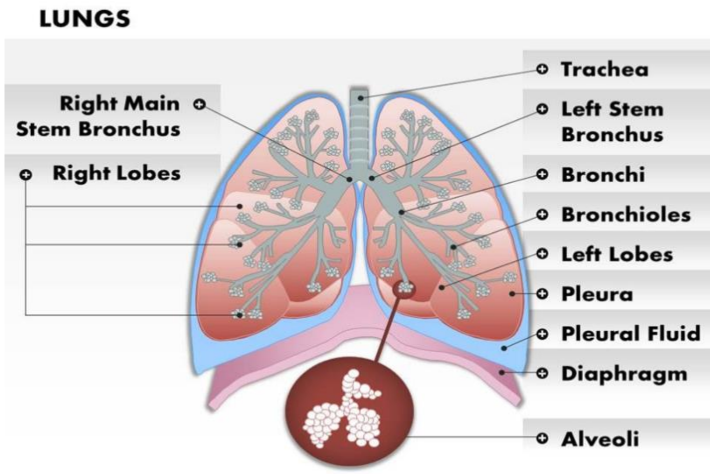


Figure 1 Anatomy of the lung [2]

Causes

The main cause of COPD in developed countries is tobacco smoking. In the developing world, COPD often occurs in people exposed to fumes from burning fuel for cooking and heating in poorly ventilated homes.

Only a few continual people who smoke expand clinically obvious COPD, even though many people who smoke with long smoking histories may additionally increase reduced lung characteristic. Some smokers increase less commonplace lung conditions. They'll be misdiagnosed as having COPD until a greater thorough evaluation is accomplished

1.1.2 Existing diagnosis methods

The main and commonly used diagnosis techniques are Spirometer, sputum test, CXR, and CT. According to an interview report held with radiologists, currently, in JUMC, suspicious and careful examination of chest X-ray image is the Gold standard way to diagnose the lung diseases

Effectively. The radiologist will inspect chest X-ray images visually. Visual inspection of chest X-ray images is a great challenge for the reason that:

- Requires a high level of expert knowledge
- The task is very time consuming
- Overlapping of the tissue structures leads to misdiagnosis.
- It also led to inter or intra observer variability again the result may be inaccurate and the result report will be highly subjective.
- It gives a false-positive result.

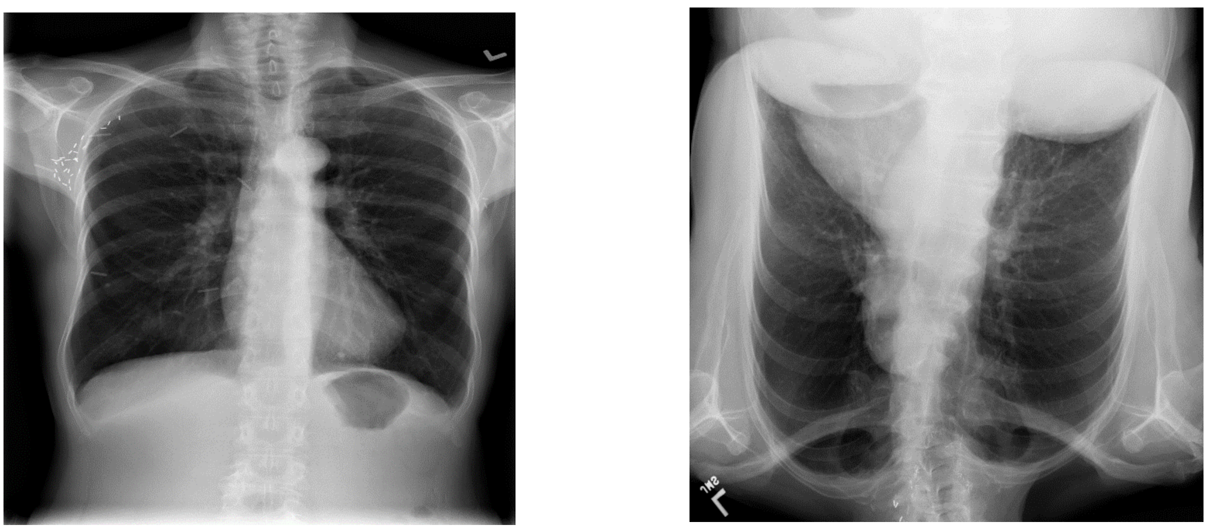


Figure 2 Sample Chest X-Ray Image [31]

1.2 Motivation

Chronic obstructive pulmonary disease (COPD) is one of the most causes of death among patients of all ages especially for those older in major medical centers and hospitals in Ethiopia, chest radiographs are widely used in the detection and diagnosis of lung diseases. The main reason for automating the manual diagnosis is to overcome critical challenges that the radiologists face during the manual diagnosis like misdiagnosis of diseases due to shortage of high-level expertise and the diagnosis procedure is time-consuming and tedious, Even in JUMC, there is only one expert radiologist to manage all imaging modalities. In addition to this, the occurrence of the disease is frequent enough in x-ray reports to provide a reasonable test set, and Chest radiography is an economical and easy-to-use medical imaging and diagnostic technique when we compare with other imaging equipment like MRI and CT scan.

The great advantages of chest X-rays include their low cost and easy operation. Even in underdeveloped areas like Ethiopia, modern digital radiography (DR) machines are affordable. Therefore that is why most hospitals and Health facilities use x-ray machines as common medical equipment to diagnose a disease.

1.3 Statement of the problem

Chronic obstructive pulmonary disease (COPD) is a problem in the lungs that prevents the lungs from working properly. The Global Burden of Disease Study reports a prevalence of 251 million cases of COPD globally in 2016 and Globally, it is estimated that 3.17 million deaths were caused by the disease in 2015 (that is, 5% of all deaths globally in that year) also More than 90% of COPD deaths occur in low- and middle-income countries since most of the cause is toxic effects of biomass fuel consumption, outdoor air pollution, tobacco smoke.[16] According to World Health Organization (WHO) estimated in 2011 that 34% of the Ethiopian population was affected by non-communicable diseases and dying by chronic obstructive pulmonary disease (COPD).

In Ethiopia, national data regarding burden and different aspects of acute exacerbations of chronic obstructive respiratory diseases are limited. However, there are some disaggregated studies conducted in different parts of the country. In one study, the prevalence of asthma was 3.5% -9.1% and that of COPD was 4% [66]. About 0.6% of deaths, both in urban and rural Ethiopia, were caused by asthma. COPD related mortality in the southern and central part of Ethiopia was 5.2% and 3%, respectively [67].

Similarly, Global Burden of Disease (GBD) studies estimated age-standardized death rates of 800 per 100,000 population for non-communicable diseases in Ethiopia, of which higher death rates (approximately 450 per 100,000) were attributed to cardiovascular disease and diabetes, 150 per 100,000 attributed to cancer, and 100 per 100,000 to chronic obstructive pulmonary disease [17]. These estimations were much higher than in many developed countries. Although these estimates of cardiovascular disease, cancer, diabetes mellitus, and chronic obstructive pulmonary disease look higher in Ethiopia, estimations by WHO and GBD studies are highly uncertain because the causes of deaths were predicted using cause-of-death models due to lack of information on the level of mortality or cause of death at the country level, which should be substantiated by national pieces of evidence [18].

The reason that people in developing countries are highly at risk of the disease is that they cannot get diagnosed at the early stage of the disease. Accurate and effective detection of disease has great benefits to control and treat disease as fast as possible. Unfortunately, there is still a severe shortage of radiologists in our country since we are far below the target for the number of radiographers, with only 87 radiographers for the entire country compared to the HSDP III goal of 620. The HSDP

III MTR does not provide specific data on radiologists. However, according to the AAU Radiology website, the nationwide radiologist to patient ratio is approximately 1:1,000,000. If this ratio is applied to the HSDP III MTR data, it implies a total of 60-80 radiologists for the entire country (Ethiopia) but there is only 25 senior radiology throughout the country. When we come to Jimma, there is only one radiology specialist for Jimma University Medical Center as well as private clinics. As a result, the diagnosis takes a long time and also difficult to differentiate the Chronic obstructive pulmonary disease (COPD). This keeps many people from having the diagnosis at the right time and the treatment is also delayed. The delay of treatment leads to death. [19]

The extreme shortage of radiologists and allied professionals highlights the importance of teleradiology, especially since almost all radiologists live in urban areas. However, the FMOH recognizes that telemedicine can only partially address its physician manpower problem. The need for "homegrown" talent has been recognized. [20] Similarly our chronic obstructive pulmonary disease classification model can support of the radiologist by helping them during the Chest X-ray Image reading and interpretation (diagnoses)

In addition to the above reason, Medical errors are a leading cause of morbidity and mortality in the medical field and are substantial contributors to medical costs. [21] Radiologists play an integral role in the diagnosis and care of patients and, given that those in this field interpret 113,204 examinations annually in Ethiopia, this may contribute to diagnostic errors [24]. Errors can be categorized as a "miss" when a primary or critical finding is not observed or as a "misinterpretation" when errors in interpretation lead to an incorrect diagnosis. Recognizing the cognitive processes that radiologists use while interpreting images should improve one's awareness of the inherent biases that can impact decision-making. [23] The automation will reduce common biases that impact clinical decisions, as well as strategies to counteract or minimize the potential for misdiagnosis. [22]

Previously most research has been conducted for many diseases from chest x-ray images using machine learning and Deep learning even for Chronic Obstructive pulmonary disease too. to solve the shortage and misdiagnosis of disease from chest x-ray but still, there wasn't any researcher have investigated to classify sub type of Chronic Obstructive pulmonary disease so in this study we have taken the classification of sub type of Chronic obstructive pulmonary disease as research Gap.

1.4 Research Question

The proposed Study aims to develop a deep learning model for multiclass classification of Chronic obstructive pulmonary diseases (COPD) from X-ray images to answer the following research questions (RQ.

(RQ1): What are the efficient parameters to get high accuracy in the COPD classification?

(RQ2): How to Build Our Own CNN Model that can perform almost the same as the Pre-Trained Model?

(RQ3): Which pre-trained model can give high accuracy in the COPD classification?

1.5 Objective of Study

1.5.1 General objective

The main objective of the research is to develop an automated system that can classify Types of COPD into asthma, Emphysema, or Chronic bronchitis from chest X-ray images.

1.5.2 Specific objectives

To achieve the general objective, the following specific objectives are addressed.

- Reviewing different image processing and deep learning related literatures.
- Collecting desired x-ray image data and preparing the dataset
- Construct our Own CNN Classification model and experiment with it.
- Train more 4 State of art Pre-trained CNN architectures; VGG16, InceptionV3, ResNet50 and EfeceintNetB0.
- Measuring the performance of the model which is developed.
- Compare performance of our CNN model with the Pre-Trained Models.
- Making Conclusion and recommendation

1.6 Significance of the study

Increasing healthcare costs have threatened many nations' financial health, and the effort needed to care for the ill and dying affects national productivity. It has become abundantly clear that the economic development of countries is tightly linked to the health of their citizens. Poor health, both individual and public, along with lack of education, is a major impediment to a country's development and is the root of poverty. Poor health impoverishes nations and poverty causes poor health, in part related to inadequate access to quality healthcare. Even more distressing is the enormous suffering that living with illness causes. Those who are most disadvantaged suffer most due to poor health. Those diseases highly affect HIV patients and can be caused by environmental pollution, this shows that the problem highly affects developing countries.

Automated Chronic obstructive pulmonary disease (COPD) detection and classification can be an appropriate solution to the problem. Diagnosis of lung diseases in Ethiopia is still manual by visualization of the chest x-ray. As a result, developing an automated chronic obstructive pulmonary disease (COPD) classification mechanism is vital to reduce disease misdiagnosis, reduce radiologists' workload, provide patients with faster and more accurate diagnoses, lower medical costs, and the model can serve as a laboratory for medical students to experiment with because they can't get enough radiologists. In addition to this, the model will reduce the issues of bias in the field [23].

1.7 Scope and Limitation of Study

1.7.1 Scope of the study

The goal of the study is to build a classification model that will be used by experts to Classify Sub type of Chronic Obstructive Pulmonary Disease: Emphysema, Chronic bronchitis, Asthma and Normal. Other respiratory disease are out of scope of the study. The model can be accessed via a web application that was developed for evaluation purposes in the study. The model mimics the reading method of the experts and it can be used with their absence. However, the study has no intention of replacing human radiologists or their radiologic interpretation. The study doesn't have the aim of assessing the reliability of conventional techniques for COPD determinations in Ethiopia. The aim is to make the Classification of COPD.

The study provides 5 different COPD classification models while 4 of them are based on the State of art Pre-trained CNN model. Image preprocessing was applied to improve the quality of X-ray images and to investigate effects on the performance of deep learning algorithms. Data Augmentation is applied to prevent overfitting. Among Deep Learning algorithms CNN is used with different Parameters Value.

1.7.2 Limitations of the study

The data was collected from St. Paul Millennium medical College Hospital, Black lion Specialized Hospital, Betel Specialized Hospital, Jimma Univesity Medical Center, and MSF Holland Medical center Gambella. All collected data did not have any label so the dataset was labeled with the guidance of Radiologists. It was hard to collect chest X-ray images and reports since the Hospital doesn’t have a system to export an image with specific pathology. Image exported with Patients' card number registers on the book. Searching and exporting were done for each on the remote archival system which was password protected this bottleneck limited the researcher to collect every single image in the archival system.

So to make this research manageable and to be more explanatory, this study will be limited in scope, time, and budget, and also lack of GPU is another limitation. Accordingly, to research the classification of lung disease from chest x-ray image using deep learning would be comprehensive and it needs a huge amount of money and a long process hence, prioritizing is a must only for the significant part to be done because of the above reasons this study limited to classification of chronic obstructive pulmonary diseases. Besides, possible efforts will be exerted to overcome the above constraints and to accomplish the desired work.

1.8 Organization of the Thesis

The thesis is organized into five chapters. Chapter one presents the introductory part of the thesis, which includes background, motivation, statement of the problem with research questions, objectives, scope and limitation of the study, and overview of research methodology. Chapter two shows a literature review on the theory of Medical Imaging, Chronic Obstructive pulmonary diseases Diagnosis methods. Brief description of Image processing techniques and Deep Learning techniques presented. An elaborated description of the state-of-the-art pre-trained feature extractors in the Classification framework in comparison with different CNN architectures is presented. A detailed review of works related to COPD using Image Processing, Computer Vision, Machine Learning, and Deep learning techniques are delivered. Finally, CNN architectures that have been used for COPD are presented.

Chapter three provides methods and approaches used to conduct the research. The high-level proposed model design is presented with a detailed description. Discussion includes techniques used for dataset preparation, Image Preprocessing, feature extraction, labeling and training, testing, or model evaluation procedure. A step-by-step procedure for the training and testing experiments on the model is explained.

Chapter four is all about the proposed Architecture for the entire study

Chapter Five presents the experiment result for each research question and lesson learned. It also includes analysis and interpretation of the experiment result along with the findings of the study. Chapter six is the last chapter which provides a Conclusion and Recommendation. In this chapter, the concluding remarks from the study result, challenges, and future work are presented.

CHAPTER TWO

LITERATURE REVIEW

2.1 Brief Discussion about Chronic Obstructive Pulmonary Disease

2.1.1 How lungs are affected

Air travels down the windpipe (trachea) and into the lungs via large tubes (bronchi). within the lungs, those tubes divide commonly like the branches of a tree into many smaller tubes (bronchioles) that lead to clusters of tiny air sacs referred to as alveoli. The air sacs have very thin partitions complete of tiny blood vessels (capillaries).

The air sacs have very thin walls full of tiny blood vessels (capillaries). The oxygen in the inhaled air passes into these blood vessels and enters the bloodstream. On the identical time, carbon dioxide a gasoline that may be a waste product of metabolism is exhaled. The lungs rely upon the natural elasticity of the bronchial tubes and air sacs to force air out of the body. COPD reasons

them to lose elasticity and over-extend, which leaves a few air trapped inside the lungs when we exhale.

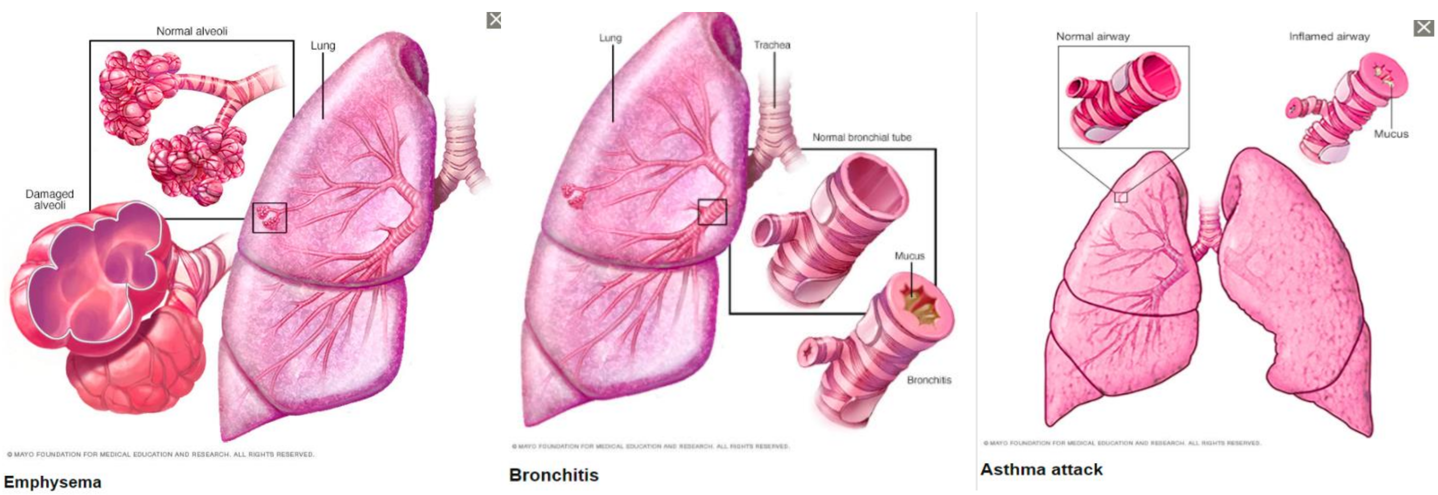


Figure 3. An illustration of COPD Disease (Emphysema Bronchitis and Asthma) [15]

2.1.3 Detection Error and discrepancy in radiology

It is estimated that a thousand million radiologic examinations are achieved worldwide annually, maximum of which might be interpreted through radiologists [59]. Most professional bodies would agree that all imaging procedures should include an expert radiologist's opinion, given employing a written report. This activity constitutes much of the daily work of practicing radiologists. However, radiologists don't always get it right. Although not always appreciated by the public, or indeed by referring doctors, radiologists' reports should not be expected to be definitive or incontrovertible. [60] they represent clinical consultations, resulting in opinions which are conclusions arrived at after weighing the evidence;" opinion" can be defined as "a view held about a particular subject or point; a judgment formed; a belief" [61].

Sometimes it is possible to be definitive in radiological diagnoses, but in most cases, radiological interpretation is heavily influenced by the clinical circumstances of the patient, relevant history and previous imaging, and myriad of other factors, including biases of which we may not be aware. Radiological studies do not come with inbuilt labels denoting the most significant abnormalities, and interpreting them is not a binary process (normal vs abnormal, cancer vs "all-clear").

The use of the term "error" implies that there is no potential for disagreement about what is "correct", and indicates that the reporting radiologist should have been able to make the correct diagnosis or report, but did not [60]. In real life, there is frequently room for legitimate differences of opinion about diagnoses or for "failure" to identify an abnormality that can be seen in retrospect. Expert opinion often forms the basis for deciding whether an error has been made [61]. . Any discrepancy in interpretation that deviates substantially from a consensus of one's peers is a reasonable and commonly accepted definition of interpretive radiological error [59], but even this is a loose description of a complex process, and may be subject to debate in individual circumstances. Certainly, in some circumstances, diagnoses are proven by pathologic examination of surgical or autopsy material, and this proof can be used to evaluate prior radiological diagnoses [59], but this is not a common basis for determining whether the error has occurred. Many cases of supposed error fall within the realm of reasonable differences of opinions between conscientious practitioners. "Discrepancy" is a better term to describe what happens in many such cases. This is not to suggest that radiological error does not occur; it does, and frequently.

2.3 Digital Image Processing

An image is defined as a two-dimensional function, f (a, b), where a and b are spatial or geographical coordinates. The amplitude of f at any pair of coordinates (a, b) is also called the intensity of the image at that point [87]. When a, b, and the amplitude values of an image f are all finite, discrete quantities, the image is a digital image. In general, as noted by Gonzalez [55], a digital image is an array of numbers representing the spatial distribution of certain appearance parameters such as reflectivity of electromagnetic radiation, emissivity, temperature, or some geophysical or topographical elevation. A digital image is composed of a finite number of elements, each of which has a particular location and value [56]. A digital image consists of discrete picture elements called pixels [57]. related to every pixel is a number of represented as DN (digital quantity) that shows the average radiance of a highly small vicinity inside a scene, with DN values commonly starting from zero to 255 in three elements or 0 to 224-1 in single numbering. The dimensions of this vicinity influences the reproduction of details within the scene. As the pixel length is decreased more scene detail is preserved in digital illustration. The sector of virtual picture processing refers to processing digital pics utilizing a virtual pc. Digital image processing involves efficient techniques of data acquisition and retrieval through sound image representation, display, pre-processing, and segmentation approaches [56]

2.4 Machine Learning

2.4.1 Classical Machine Learning

2.4.1.1 K-nearest Neighbors

The KNN algorithm assumes that similar matters exist nearby. In other phrases, comparable things are close to each different. / determine Figures 4 showing how comparable facts factors usually exist close to each other note in determine four above that most of the time, comparable information factors are close to every different. The KNN algorithm hinges in this assumption being real enough for the algorithm to be useful. KNN captures the concept of similarity (sometimes known as distance, proximity, or closeness) with some arithmetic we'd recognize from our formative years. k Nearest Neighbor set of rules falls beneath the Supervised getting to know category and is used for category and regression. It's far a flexible algorithm that also can be used for imputing lacking values and resampling datasets.

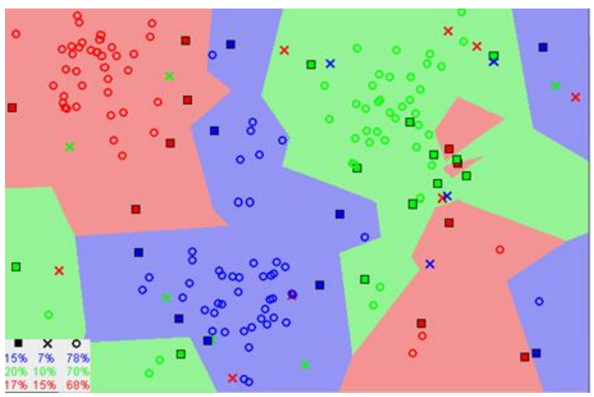


Figure 4 Image showing how similar data points typically exist close to each other

Because the call suggests, it considers okay nearest associates (records factors) to be expecting the magnificence or continuous cost for the new statistics point. The set of rule's gaining knowledge of is

1. example-based studying: right here we do no longer learn weights from schooling records to expect output (as in model-based algorithms) however use whole education instances to predict output for unseen information.

2. Lazy Learning: Model is not learned using training data prior and the learning process is postponed to a time when prediction is requested on the new instance.

3. Non -Parametric: In KNN, there is no predefined form of the mapping function.

How does KNN Work?

1. Principle:

Consider the following figure 5. Let us say we have plotted data points from our training set on a two-dimensional feature space. As shown, we have a total of 6 data points (3 red and 3 blue). Red data points belong to 'class1' and blue data points belong to 'class2'. And yellow data point in a feature space represents the new point for which a class is to be predicted. We say it belongs to 'class1' (red points).

This is because its nearest neighbors belong to that class!

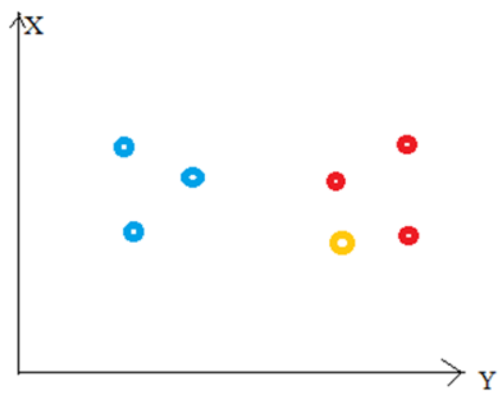


Figure 5 Classification in KNN algorithm [26]

That is the principle at the back of k nearest friends. Here, nearest acquaintances are those facts factors which have minimum distance in function space from our new statistics factor. And ok is the quantity of such records points we recall in our implementation of the set of rules.

Therefore, distance metric and okay value are two important issues even as the usage of the KNN set of rules. Euclidean distance is the most famous distance metric. You can also use Hamming distance, big apple distance, Minkowski distance as in line with your want. For predicting magnificence/ continuous fee for a new records point, it considers all the statistics points inside the training dataset. Reveals new information point's 'okay' nearest acquaintances (data points) from characteristic space and their class labels or continuous values.

For classification,   a class label assigned to most of the people of ok nearest acquaintances from the education dataset is taken into consideration as an expected class for the new information point.

For regression: suggest or median of continuous values assigned to okay nearest friends from training dataset is an anticipated continuous fee for our new facts factor.

2.4.1.2 Decision Tree Algorithm

Class is a -step technique, the gaining knowledge of step and prediction step, in system getting to know. Within the learning step, the model is advanced primarily based on given training records. Within the prediction step, the model is used to expect the reaction for given statistics. A selection Tree is one of the simplest and popular class algorithms to apprehend and interpret.

The decision Tree algorithm belongs to the circle of relatives of supervised studying algorithms. Unlike other supervised studying algorithms, the decision tree algorithm can be used for fixing regression and Classification issues too.

The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data (training data).

In choice timber, for predicting a category label for a record we begin from the root of the tree. We compare the values of the foundation attribute with the document's characteristic. Based totally at the contrast, we observe the branch similar to that value and jump to the following node

Types of Decision Trees

Forms of choice bushes are primarily based at the sort of target variable we've got. It could be of sorts:

1. Categorical Variable Decision Tree: The decision Tree has a categorical target variable then it is called a Categorical variable decision tree.

2. Continuous Variable Decision Tree: A decision Tree has a continuous target variable then it is called a Continuous Variable Decision Tree.

Example: let’s say we've a trouble predicting whether or not a client pays his renewal premium with an insurance organization (yes/ no). Right here we recognize that the earnings of customers is a considerable variable but the coverage agency does now not have earnings info for all clients.

Now, as we realize that is an important variable, so we can build a decision tree to predict client income based totally on career, product, and diverse other variables. In this example, we're predicting values for the continuous variables.

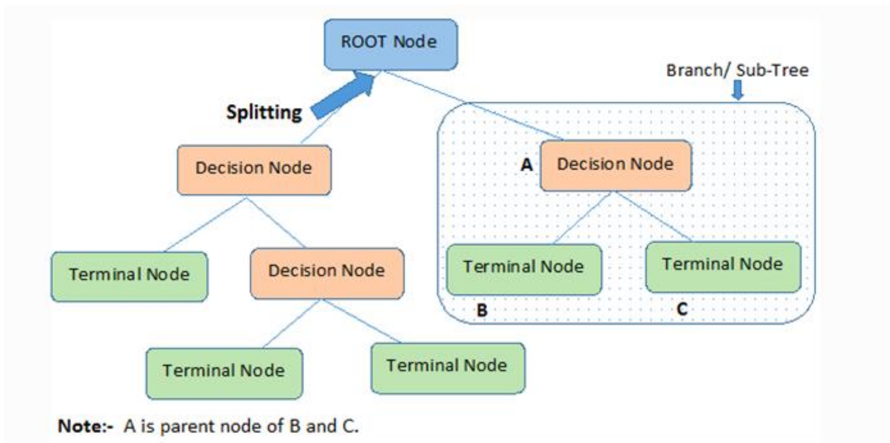


Figure 6 A is Parent node of B and C[31]

Decision trees classify the examples by sorting them down the tree from the root to some leaf/terminal node, with the leaf/terminal node providing the classification of the example. each node within the tree acts as a test case for some attribute, and every facet descending from the node corresponds to the viable solutions to the test case. This system is recursive and is repeated for each subtree rooted at the new node.

2.4.1.3 Support Vector Machines (SVMs)

Given a fixed of schooling examples, each marked as belonging to 1 or the opposite of two training, an SVM set of guidelines builds a version that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

The goal of applying SVMs is to discover the best line in dimensions or the great hyperplane in extra than two dimensions to help us separate our area into training. The hyperplane (line) is located through the most margin, i.e., the most distance between statistics points of both training Imagine the labeled training set are two classes of data points (two dimensions):

Man and Woman. To separate the two classes, there are so many possible options of hyperplanes that separate correctly. Shown within the graph beneath, we can gain the identical end result using extraordinary hyperplanes (L1, L2, and L3). However, if we upload new facts points, the result of the use of diverse hyperplanes will be very one-of-a-kind in terms of classifying new facts point into the proper institution of sophistication.

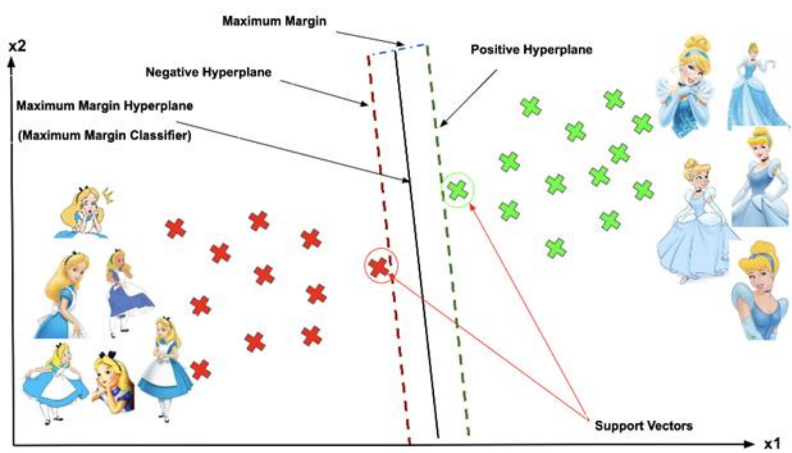


Figure 7 Different hyperplanes (L1, L2, and L3) [26]

Support Vector, Hyperplane, and Margin

The vector points closest to the hyperplane are referred to as the guide vector points because only these factors are contributing to the end result of the set of rules, and different factors aren't. If an information point is not an assist vector, getting rid of it does not have an effect on the version. On the other hand, deleting the assist vectors will then alternate the placement of the hyperplane.

Size of the hyperplane relies upon upon the wide variety of capabilities. If the number of input features is two, then the hyperplane is just a line. If the form of input abilties is 3, then the hyperplane turns into a two-dimensional aircraft. It turns into difficult to consider when the range of features exceeds 3. The distance of the vectors from the hyperplane is called the margin, this is a separation of a line to the nearest elegance factors we would love to pick out a hyperplane that maximizes the margin between training.

The graph below shows what good margins and bad margins are.

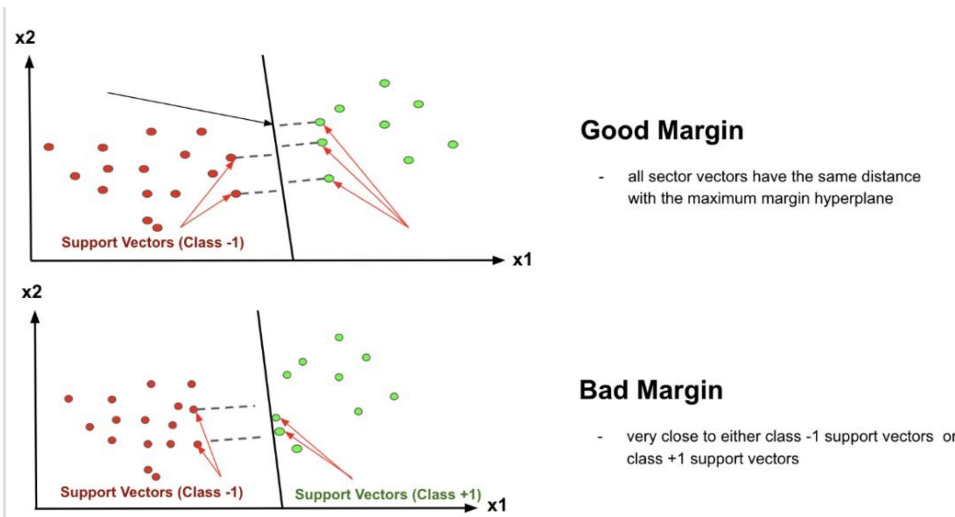


Figure 8 Good and Bad Margin [21]

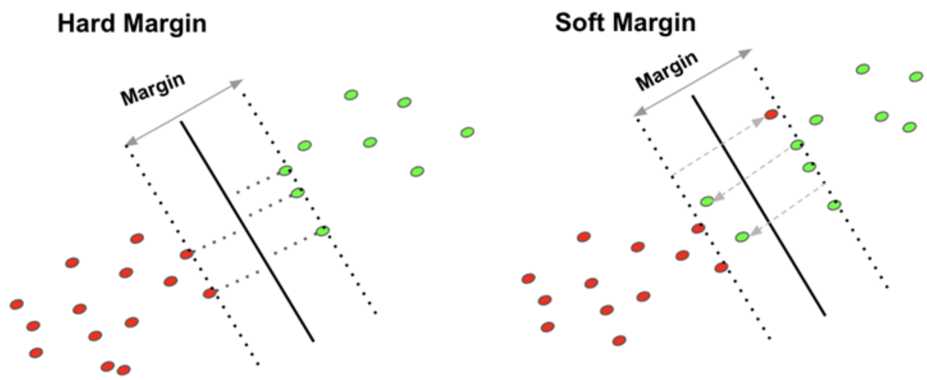


Figure 9 Hard and soft Margin [21]

Hard Margin

Based on the above figure 9 illustration if the education records is linearly separable, we are able to pick out two parallel hyperplanes that separate the two training of facts, so that the gap between them is as massive as viable.

Soft Margin

As most of the real-global information aren't fully linearly separable, we are able to allow a few margin violation to arise, that's called gentle margin class. It is higher to have a huge margin, even though some constraints are violated violation approach deciding on a hyperplane, which can permit some data factors to stay on both the wrong side of the hyperplane and among the margin and the correct aspect of the hyperplane

To find the maximal margin, we want to maximize the margin between the facts points and the hyperplane. Within the following session, i can proportion the mathematical ideas in the back of this set of rules

Maximizing the Margin

You probably found out that an equation of a line is y=ax+b.However, you will often find that the equation of a hyperplane is defined by:

Note that

$$y = ax + b$$

is the same thing as

$$y - ax - b = 0$$

Given two vectors $\mathbf{w}\begin{pmatrix} -b \\ -a \\ 1 \end{pmatrix}$ and $\mathbf{x}\begin{pmatrix} 1 \\ x \\ y \end{pmatrix}$

$$\mathbf{w}^T\mathbf{x} = -b \times (1) + (-a) \times x + 1 \times y$$

$$\mathbf{w}^T\mathbf{x} = y - ax - b$$

Figure 10 equation of a hyperplane [21]

The two equations are just two different ways of expressing the same thing.

For Support Vector Classifier (SVC), we use **w**T**x**+*b* where **w** is the weight vector, and *b* is the bias.

$$\mathbf{w}^T\mathbf{x} + b = 0$$

You can see that the name of the variables in the hyperplane equation are w and x, which means they are vectors! A vector has magnitude (size) and direction, which works perfectly well in 3 or more dimensions. Therefore, the application of "vector" is used in the SVMs algorithm.

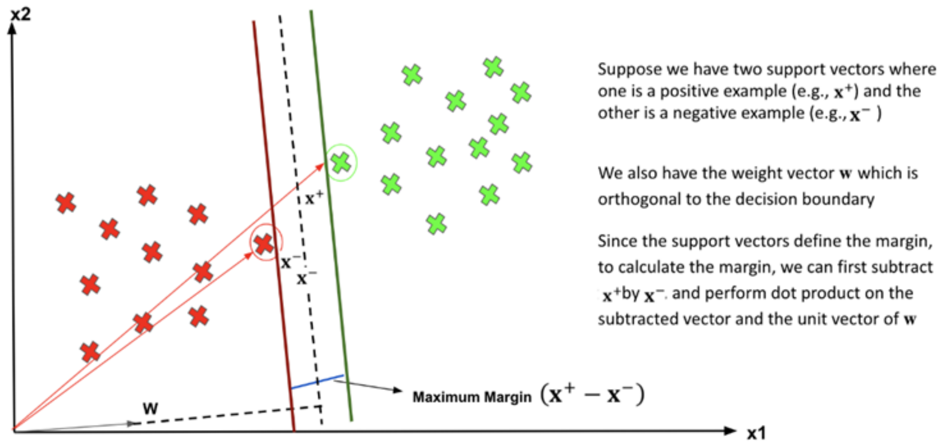


Figure 11 application of "vector" is used in the SVMs algorithm. [22]

$$h(\boldsymbol{x}_i) = sign\left(\sum_{j=1}^{s} \alpha_j \; y_j \, K(\boldsymbol{x}_j, \boldsymbol{x}_i) + b \right)$$

$$K(\boldsymbol{v}, \boldsymbol{v}') = \exp\left(\frac{||\boldsymbol{v} - \boldsymbol{v}'||^2}{2\gamma^2}\right)$$

Figure 12 the equation of calculating the Margin [22]

Support Vector Machine for Detecting Brain Tumor

The support Vector machine (SVM) turned into first proposed by using Vapnik and has when you consider that attracted a high diploma of interest within the machine learning studies community. Numerous recent studies have suggested that the SVM (assist vector machines) generally are able to delivering higher overall performance in phrases of class accuracy than different records classification algorithms.

SVM is a binary classifier based on supervised learning which gives better performance than other classifiers. SVM classifies between two classes by constructing a hyperplane in high-dimensional feature space which can be used for classification. Hyperplane can be represented by the equation- **W.X+B=0**

W is a weight vector and is normal to the hyperplane. **B** is bias or threshold.

2.4.2 Neural Network

Neural Nets have become pretty popular today

2.4.2.1 The Perceptron

Rosenblatt's perceptron, commonly known as the perceptron, is the most effective form of a neural community used for binary type of styles stated to be linearly separable. It consists of a single neuron with adjustable synaptic weights and bias. The set of rules used to adjust the unfastened parameters of this neural community first seemed in a gaining knowledge of system evolved by using Rosenblatt in 1958 for his perceptron brain version [25]

Indeed, Rosenblatt proved that if the styles used to train the perceptron are drawn from two linearly separable lessons, then the perceptron algorithm converges and positions the decision surface inside the form of a hyperplane among the two lessons. Therefore, the perceptron is optimal only when the classification problem can be linearly separated in that space. The synaptic weights of the perceptron can be adapted on an iteration-by-iteration basis. For the adaptation, we may use an error-correction rule known as the perceptron convergence algorithm [26].

The perceptron constructed round a single neuron is restricted to acting pattern classification with simplest two training. By way of increasing the output layer of the perceptron to encompass multiple neuron, we may additionally correspondingly perform category with greater than two lessons. In a layered neural network, the neurons are organized inside the shape of layers. Within the best shape of a layered community, we've an enter layer of source nodes that projects directly onto an output layer of computation neurons, but not vice versa.

In single-layer perceptrons, 'single-layer refers to the output layer of computation neurons. We do not count the input layer of source nodes because no computation is performed there.

The perceptron consists of 4 parts.

- Input values or One input layer
- Weights and Bias
- Net sum
- Activation Function

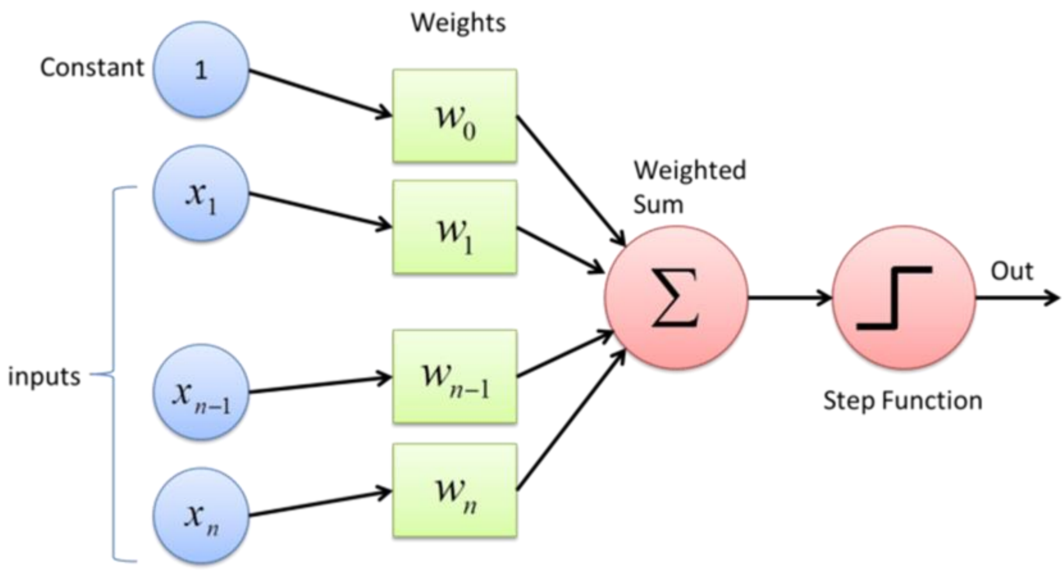


Figure 13 : Perceptron [34

The perceptron works on these simple steps

A. All the inputs x are multiplied with their weights w. Let's call it k.

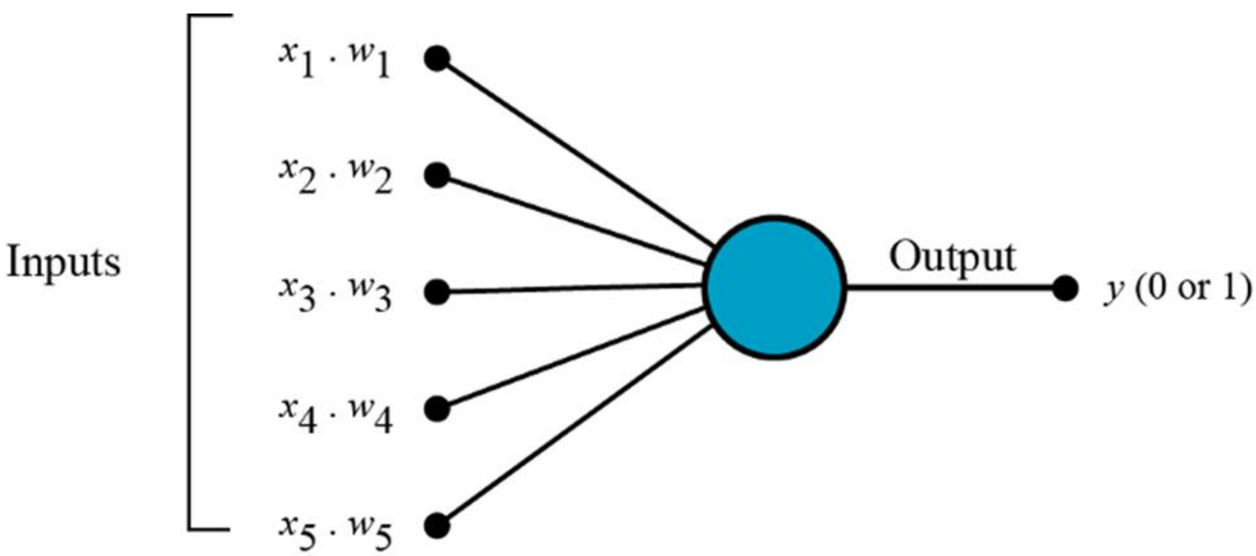


Figure 14 : Multiplying inputs with weights for 5 inputs [34]

output size = (input size - 1)*stride - 2*padding + (kernel size - 1) +1

B. Add all the multiplied values and call them Weighted Sum.

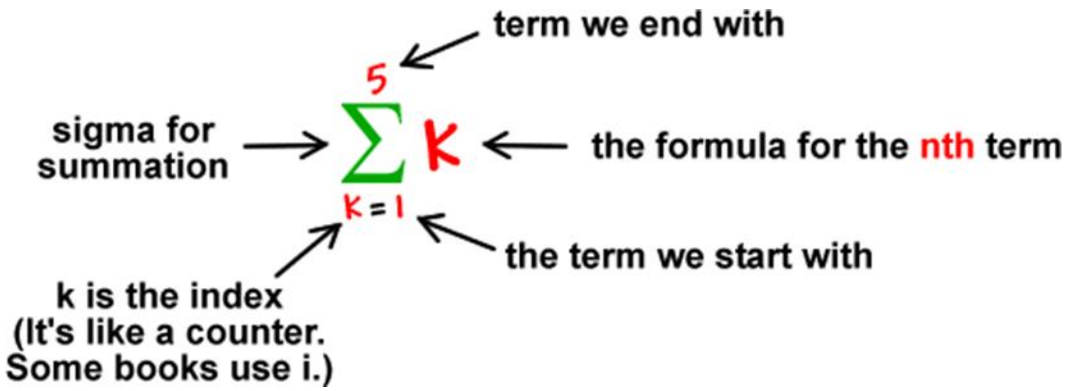


Figure 15 Sigma for summation [34]

C. Apply that weighted sum to the correct Activation Function.

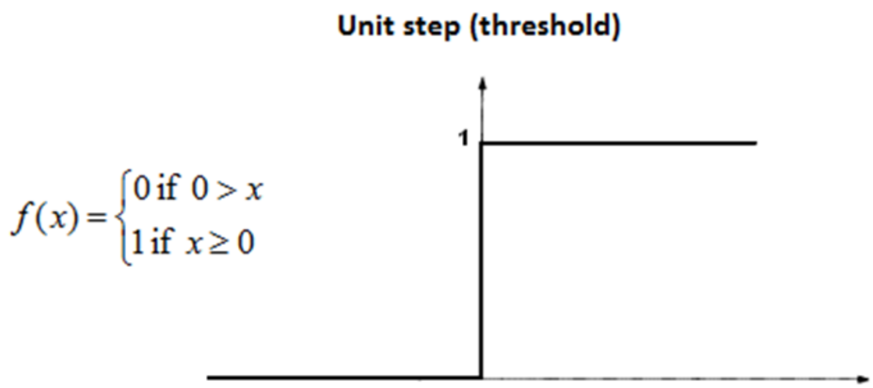


Figure 16 Unit Step Activation Function [31]

Why do we need Weights and Bias? Weights show the strength of the particular node. A bias value allows you to shift the activation function curve up or down.

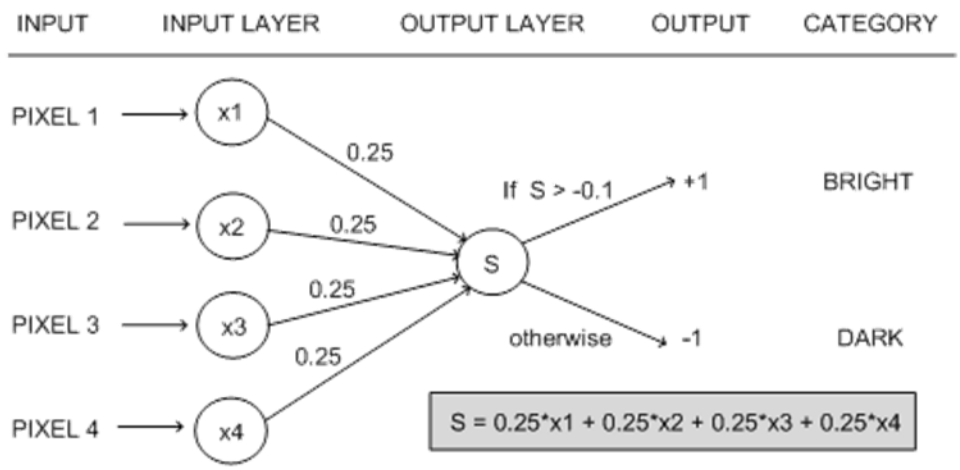


Figure 17 Weight and bias [34]

Why do we need an Activation Function? In short, the activation functions are used to map the input between the required values like (0, 1) or (-1, 1).

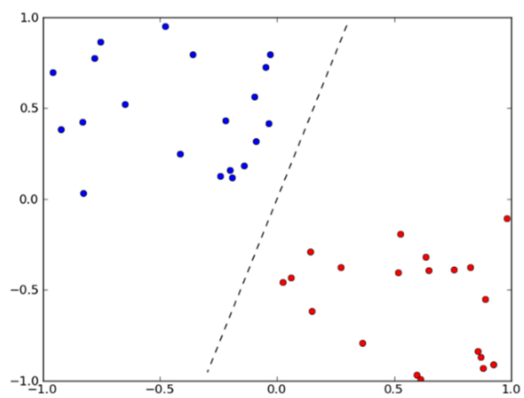


Figure 18 shows the effect of activation function on Neural network [32]

2.4.2.2 Multi-Layer Perceptron's (MLP)

An MLP is a network that consists of at least three layers of nodes: an input layer, a hidden layer, and an output layer.

For the input nodes, each node is a neuron that uses a nonlinear activation function. MLP makes use of a supervised mastering technique called backpropagation for schooling. [28][29] Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It may distinguish information that isn't linearly separable. [30]

Multi-Layer perceptrons are an example of feed forward neural networks, where single-layer perceptrons are connected. An MLP contains neurons organized in layers (see figure 14). Instead of the single perceptron, several neurons are connected to the same inputs $x1...xn$, with a different set of weights. The outputs of all these neurons are inputs for a new layer of neurons. Considered altogether, the weights of each neuron k ((k) (i)), define a weight matrix from layer Li−1 to layer Li: (i). Thus the output (vector) of a given layer Li can be computed as the multiplication of the input vector y (i−1) by the weight matrix **W**(i), the addition of a bias vector **b** (i), and the element-wise application of a non-linear function fi [31].

The Perceptron consists of an enter layer and an output layer which are fully linked. MLPs have the identical input and output layers however may have more than one hidden layers in among the aforementioned layers, as shown below.

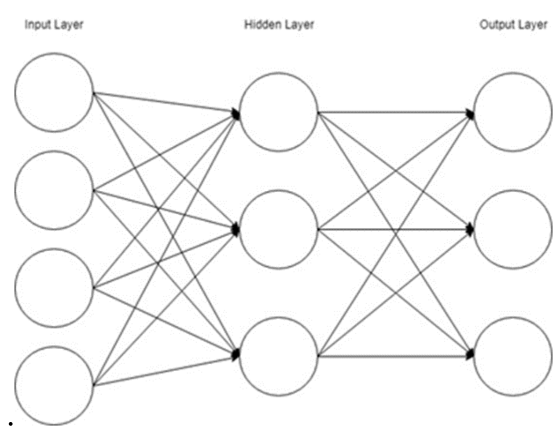


Figure 19 Multi-Layer Perceptron's [37]

Learning occurs in the perceptron by changing connection weights after each piece of data is processed, based totally on the amount of errors in the output compared to the predicted end result. That is an example of supervised gaining knowledge of and is achieved via backpropagation, a generalization of the least suggest squares set of rules within the linear perceptron.

We can represent the degree of error in an output node j in the nth data point (training example) by ei(n)=d(n)-yi(n) where d is the target value and y is the value produced by the perceptron. The node weights can then be adjusted based on corrections that minimize the error in the entire output, given by

$$\mathcal{E}(n) = \frac{1}{2}\sum_j e_j^2(n).$$

Using gradient descent, the change in each weight is

$$\Delta w_{ji}(n) = -\eta \frac{\partial \mathcal{E}(n)}{\partial v_j(n)} y_i(n)$$

Where yi is the output of the previous neuron and n is the learning rate, which is selected to ensure that the weights quickly converge to a response, without oscillations Vj. which itself varies. It is easy to show that for an output node this by-product may be simplified to by-product is

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = e_j(n)\phi'(v_j(n))$$

Where ø is the by-product of the activation characteristic defined above, which itself does not vary the analysis is extra difficult for the change in weights to a hidden node, but it may be shown that the relevant

$$-\frac{\partial \mathcal{E}(n)}{\partial v_j(n)} = \phi'(v_j(n)) \sum_k -\frac{\partial \mathcal{E}(n)}{\partial v_k(n)} w_{kj}(n).$$

The derivative to be calculated depends on the induced local field

The algorithm for the MLP is as follows:

- simply as with the perceptron, the inputs are pushed ahead through the MLP by using taking the dot made from the input with the weights that exist between the enter layer and the hidden layer (WH). This dot product yields a price at the hidden layer. We do not push this price forward as we would with a perceptron even though.
- MLPs make use of activation features at each of their calculated layers. There are many activation functions to speak about: rectified linear gadgets (ReLU), sigmoid feature, tanh. Push the calculated output at the modern-day layer thru any of these activation features. Once the calculated output at the hidden layer has been pushed via the activation feature, push it to the next layer in the MLP by means of taking the dot product with the corresponding weights. Repeat steps and 3 till the output layer is reached.
- at the output layer, the calculations will either be used for a backpropagation algorithm that corresponds to the activation function that turned into decided on for the MLP (inside the case of training) or a selection could be made based at the output (inside the case of testing). MLPs shape the premise for all neural networks and have substantially advanced the electricity of computer systems while carried out to classification and regression problems. Computer systems are not confined by using XOR cases and may research rich and complicated models thanks to the multilayer perceptron.

2.4.2.3 Feed-Forward Neural Network

This is the most basic type of neural network that came about in large part to technological advancements which allowed us to add many more hidden layers without worrying too much about computational time. It also became popular thanks to the discovery of the backpropagation algorithm by Geoff Hinton in 1990.

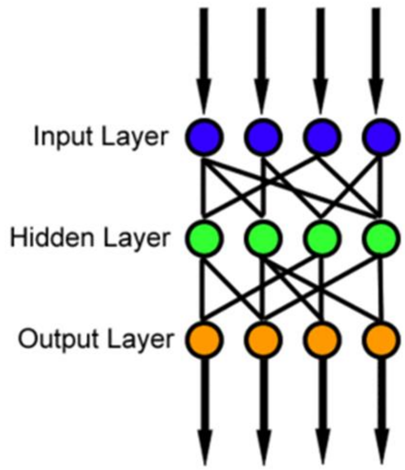


Figure 20 backpropagation algorithm by Geoff Hinton in 1990. [37]

This kind of neural network basically includes an enter layer, a couple of hidden layers, and an output layer. There's no loop and information only flows forward.

Feed-forward neural networks are normally perfect for supervised gaining knowledge of with numerical information, although it has their disadvantages too: 1) It can't be used with sequential statistics; 2) It doesn't paintings too properly with photograph facts because the overall performance of this model is heavily reliant on capabilities, and finding the features for an image or textual content information manually is a quite tough exercise on its own.

This brings us to the next two classes of neural networks: Convolutional Neural Networks and Recurrent Neural Networks.

2.4.2.4 Recurrent Neural Networks (LSTM/GRU/Attention)

RNNs can assist us study the sequential shape of text wherein every phrase is depending on the preceding phrase, or a phrase inside the preceding sentence.

For a easy clarification of an RNN, think of an RNN mobile as a black container taking as enter a hidden state (a vector) and a phrase vector and giving out an output vector and the subsequent hidden nation. This field has some weights which need to be tuned using backpropagation of the losses. Additionally, the identical cell is implemented to all the words so that the weights are shared throughout the phrases within the sentence. This phenomenon is called weight-sharing.

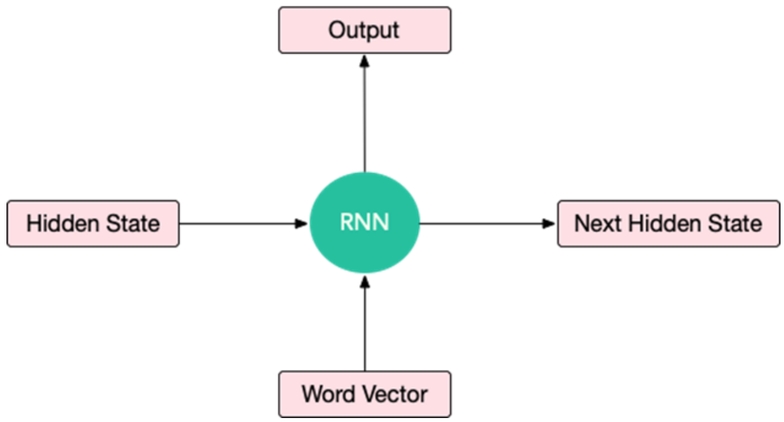


Figure 21 weight-sharing in RNN [31]

Below is the expanded version of the same RNN cell where each RNN cell runs on each word token and passes a hidden state to the next cell. For a sequence of length 4 like "the quick brown fox", the RNN cell finally gives 4 output vectors, which can be concatenated and then used as part of a dense feed-forward architecture like below to solve the final task Language Modeling or classification task:

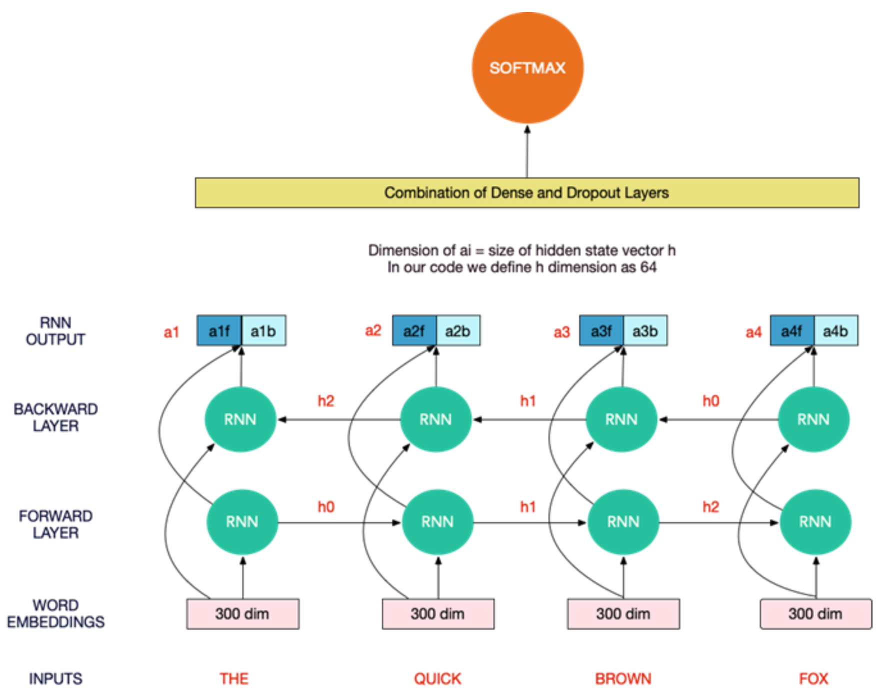


Figure 22 RNN [30]

Short Term Memory networks (LSTM) and Gated Recurrent units (GRU) are a subclass of RNN, specialized in remembering information for prolonged durations by means of introducing numerous gates which alter the cell state through adding or putting off information from it. From a very high point, you can apprehend LSTM/GRU as a play on RNN cells to study long-time period dependencies.

RNNs/LSTM/GRU have been predominantly used for various Language modeling tasks where the objective is to predict the next word given a stream of input Word or for tasks that have a sequential pattern to them.

2.4.2.5 Convolutional Neural Networks

Convolutional neural networks (CNNs), also known as convolutional networks, are specialized kinds of neural networks for processing data that has a grid-like topology. Examples include image data which can be thought of as a two-dimensional grid of pixels. CNN's can be simply defined as neural networks that use a mathematical operation, termed as convolution, in at least one of their layers [32].

Typical architecture or building blocks of CNNs have two components: The feature extraction part and classification part. In the feature extraction part, the network performs a series of convolutions, pooling (subsampling), and non-linear transformations throughout which the capabilities are detected. The convolution is performed on the enter statistics with using a kernel to then produce a feature map. The kernel slides throughout the input feature map.

The kernel slides across the input feature map. At each place, the product between each element of the kernel and the input element it overlaps is computed and the outcomes are summed as much as obtain the output inside the modern vicinity. The method may be repeated using extraordinary kernels to shape as many output characteristic maps as desired. The very last outputs of this technique are referred to as output function maps. The convolution extracts exclusive functions of enter.

The primary convolution layer extracts low-degree functions like edges, shape, texture, and corners. Better-stage layers extract better-level features. In comparison, pooling shrinks the measurement of enter by way of an integer component.

A deep learning algorithm which is CNN will be chosen in different literature that will be conducted in computer vision especially in image classification. CNN's represent an interesting method for adaptive image processing. The algorithm will be used for feature extraction, classification, training, testing as well as evaluating the accuracy of the model. CNN's take data, without the need for a separate pre-processing or feature extraction stage and with the preprocessing data. In addition to these, feature extraction and classification stages occur naturally within a single framework.

As the main advantage of using the CNN algorithm for the classification of chronic obstructive pulmonary diseases, it is more automated than classical machine learning algorithms. In the classical machine learning algorithm, there is a need to develop different algorithms for different problems, therefore it uses more handcrafted algorithms but in our case, we were used a deep learning algorithm called Convolutional Neural Network to extract features of chronic obstructive pulmonary disease and Classify them based on their types [26].

Pooling is also known as subsampling and is widely used in deep learning. Pooling layers reduce the dimension and resolution of input while preserving the most important information [32]. The CNNs also include other architectural features, namely: the stride and zero padding. Stride is the distance between two consecutive positions of the kernel along an axis, whereas zero padding is the number of zeroes concatenated at the beginning and the end of the axis. A convolutional layer's output form is suffering from the form of its input as well as the choice of kernel shape, strides, and 0 padding. Moreover, the relationship among these properties isn't always trivial to deduce. This contrasts with the fully connected layers, whose output size is independent of the input size [34].

In recent years, following the breakthroughs that have been gained using deep learning techniques, deep CNNs have become the de facto standard for complex computer vision tasks. Some other successful application areas of deep CNN include image classification, object detection, video processing, natural language processing (NLP), handwriting recognition, and speech recognition. The effective computerized characteristic extraction potential of deep CNN reduces the want for a separate handmade function extraction system. This capability is typically due to using a couple of characteristic extraction degrees that may robotically examine representations from raw records.

The vast development in the representational capacity of the deep CNNs is carried out thru architectural innovations. For a survey at the distinct architectures of deep CNNs, check with [33].

2.4.3 Transfer learning

Transfer learning is the reuse of a pre-trained model on a new problem. It's currently very popular in deep learning because it can train deep neural networks with comparatively little data. This is very useful in the artificial intelligence and data science field since most real-world problems typically do not have millions of labeled data points to train such complex models. [51]

2.4.4 Neural Network Architectures

2.4.4.1 LeNet5

LeNet5 is a neural network architecture that was created by Yann LeCun in the year 1994. LeNet5 propelled the deep Learning field. It may be stated that LeNet5 became the very first convolutional neural community that has the leading function at the beginning of the Deep studying subject.

LeNet5 has a very fundamental architecture. Across the entire image will be distributed with image features. Similar features can be extracted in a very effective way by using learnable parameters with convolutions. When the LeNet5 was created, the CPUs were very slow, and No GPU can be used to help the training.

The main advantage of this architecture is the saving of computation and parameters. In an extensive multi-layer neural network, each pixel was used as a separate input, and LeNet5 contrasted this. There are high spatially correlations between the images and using the single-pixel as different input features would be a disadvantage of these correlations and not be used in the first layer.

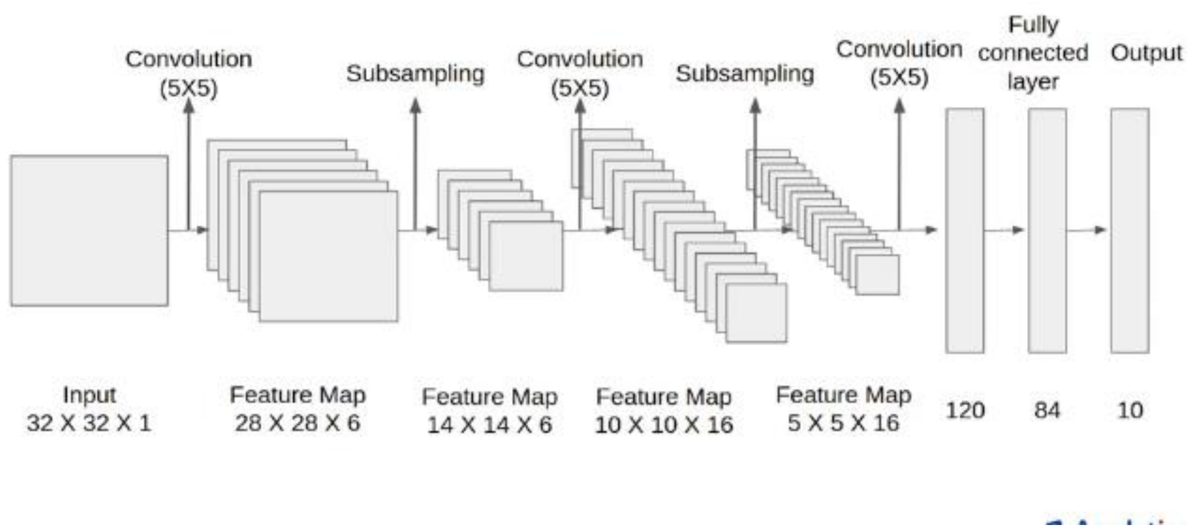


Figure 23 Diagram of LetNet5 [41]

Features of LeNet5:

- The cost of Large Computations can be avoided by sparsing the connection matrix between layers.
- The final classifier will be a multi-layer neural network
- In the form of sigmoids or tanh, there will be non-linearity
- The spatial average of maps are used in the subsample
- Extraction of spatial features are done by using convolution
- Non-linearity, Pooling, and Convolution are the three sequence layers used in a convolutional neural network

In a few words, it can be said that LeNet5 Neural Network Architecture has inspired many people and architectures in the field of Deep Learning.

2.4.4.2 Dan Ciresan Net

The neural network did not progress much from the year 1998 to 2010. Many researchers were slowly improving, and many people did not notice their increasing power. With the rise of cheap digital and cell-phone cameras, data availability increased. GPU has now become a general-purpose computing tool, and CPUs have also become faster with the increase of computing power. In those years, the progress rate of the neural network was prolonged, but slowly people started noticing the increasing power of the neural network.

The very first implementation of GPU Neural nets was published by Jurgen Schmidhuber and Dan Claudiu Ciresan in 2010. There were up to 9 layers of the neural network. It was implemented on an NVIDIA GTX 280 graphics processor, and it had both backward and forward.

2.4.4.3 AlexNet

This neural network architecture has won the challenging competition of ImageNet by a considerable margin. It is a much broader and more in-depth version of LeNet. Alex Krizhevsky released it in 2012.

Complex hierarchies and objects can be learned using this architecture. The much more extensive neural network was created by scaling the insights of LeNet in AlexNet Architecture

The work contributions are as follows:

- Training time was reduced by using GPUs NVIDIA GTX 580.
- Averaging effects of average pooling are avoided, and max pooling is overlapped.
- Overfitting of the model is avoided by selectively ignoring the single neurons by using the technique of dropout.
- Rectified linear units are used as non-linearities

Bigger images and more massive datasets were allowed to use because training time was 10x faster and GPU offered a more considerable number of cores than the CPUs. The success of AlexNet led to a revolution in the Neural Network Sciences. Useful tasks were solved by large neural networks, namely convolutional neural networks. It has now become the workhorse of Deep Learning.

One thing to note here, due to the fact Alexnet is a deep structure, the authors added padding to save you the size of the function maps from decreasing notably. The input to this model is the images of size 227X227X3.

Table 1 Architectiure of AlexNet [38]

Layer	# filters / neurons	Filter size	Stride	Padding	Size of feature map	Activation function
Input	-	-	-	-	227 x 227 x 3	-
Conv 1	96	11 x 11	4	-	55 x 55 x 96	ReLU
Max Pool 1	-	3 x 3	2	-	27 x 27 x 96	-
Conv 2	256	5 x 5	1	2	27 x 27 x 256	ReLU
Max Pool 2	-	3 x 3	2	-	13 x 13 x 256	-
Conv 3	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 4	384	3 x 3	1	1	13 x 13 x 384	ReLU
Conv 5	256	3 x 3	1	1	13 x 13 x 256	ReLU
Max Pool 3	-	3 x 3	2	-	6 x 6 x 256	-
Dropout 1	rate = 0.5	-	-	-	6 x 6 x 256	-

Convolution and Maxpooling Layers

Then we apply the first convolution layer with 96 filters of size 11X11 with stride 4. The activation function used in this layer is relu. The output feature map is 55X55X96.

2.4.4.3 VGG

The ImageNet large Scale visible recognition project (ILSVRC) is an annual pc vision competition. Every year, teams compete on two obligations. The primary is to come across gadgets inside a picture coming from 2 hundred training, that's called object localization. The second is to classify pics, every categorized with certainly one of one thousand categories, that's known as picture classification. VGG 16 became proposed by means of Karen Simonyan and Andrew Zisserman of the visible Geometry organization Lab of Oxford University in 2014 this model won 1st and 2nd place in the above categories in the 2014 ILSVRC challenge.

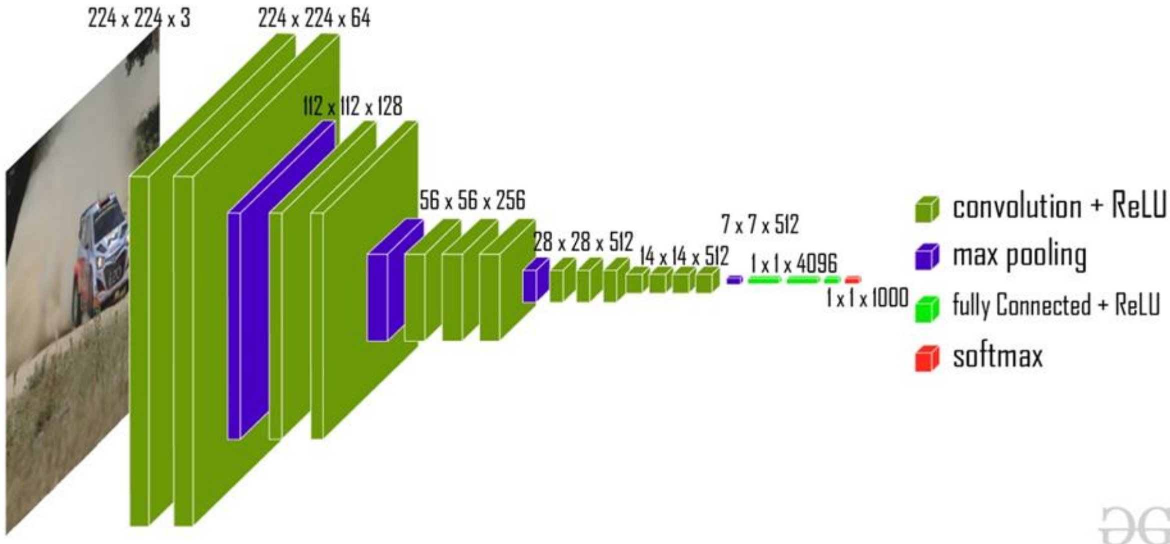


Figure 24 Architecture of Visual geometry group 16 (VGG16) [41]

This model achieves 92.7% top-5 test accuracy on the ImageNet dataset which contains 14 million images belonging to 1000 classes.

It was one of the famous models submitted to ILSVRC-2014. As you can see from below figure 25 it improves AlexNet by replacing large kernel-sized filters (11 and 5 in the first and second convolutional layer, respectively) with multiple 3×3 kernel-sized filters one after another with 3 dense layers VGG16 was trained for weeks and was using NVIDIA Titan Black GPU's.

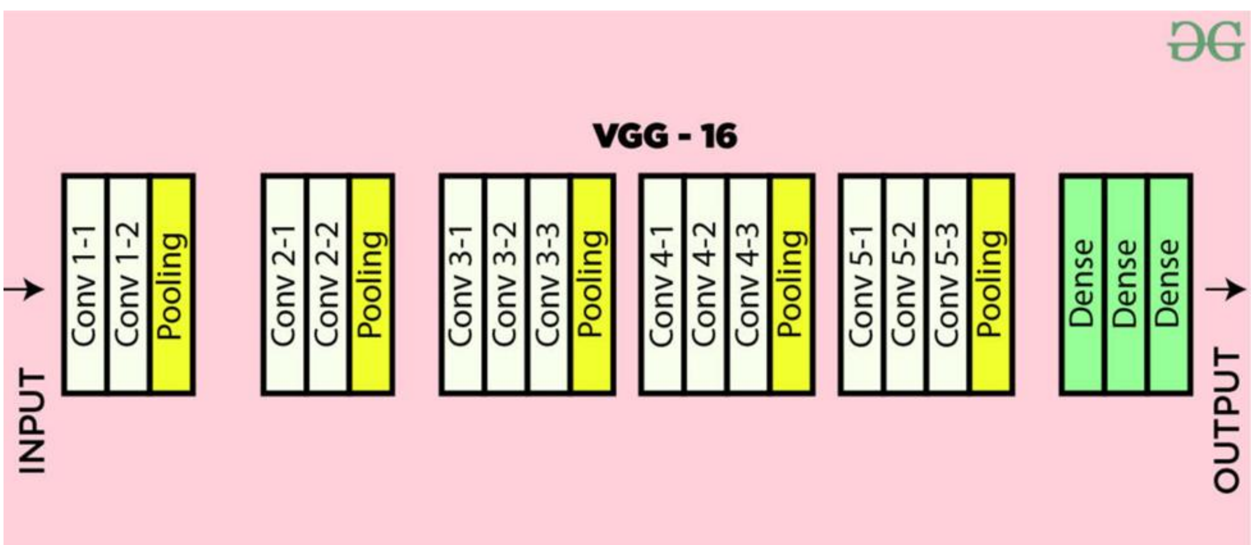


Figure 25 Architecture of Visual geometry group 16 (VGG16) [41]

All the hidden layers use ReLU as its activation function. ReLU is extra computationally green as it results in faster studying and it also decreases the chance of vanishing gradient troubles. **Challenges of VGG 16**:

It's far very gradual to educate (the unique VGG model changed into skilled on Nvidia Titan GPU for two-three weeks). The scale of VGG-16 skilled imageNet weights is 528 MB. So, it takes quite lots of disk area and bandwidth that makes it inefficient.

2.4.4.4 GoogLeNet and Inception

GoogLeNet is the primary inception structure that objectives at reducing the load of computation of deep neural networks. The categorization of video frames and snap shots content become carried out by using deep Learning model. Large deployments and performance of architectures on the server farms became the primary hobby of huge internet giants inclusive of Google In 2014, many individuals agreed that neural networks and deep learning are here to stay.

GoogLeNet put forward a leap forward overall performance on the ImageNet visual reputation project (in 2014), which is a reputed platform for benchmarking image popularity and detection algorithms. In conjunction with this, it prompt a ton of research within the advent of new deep learning architectures with modern and impactful ideas.

Figure 26 Architecture of GoogleNet[38]

Architectural Details

A new type of architecture – GoogLeNet or Inception v1. It is a convolutional neural network (CNN) that is 27 layers deep. Below is the model summary:

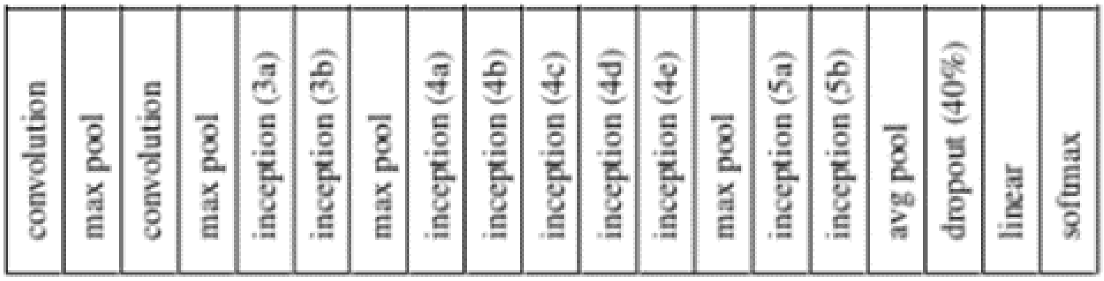


Figure 27 GoogleNet Model Summery [38]

Notice in the above image that there is a layer called the inception layer. This is the main idea behind the approach. The inception layer is the center concept of a moderately linked architecture.

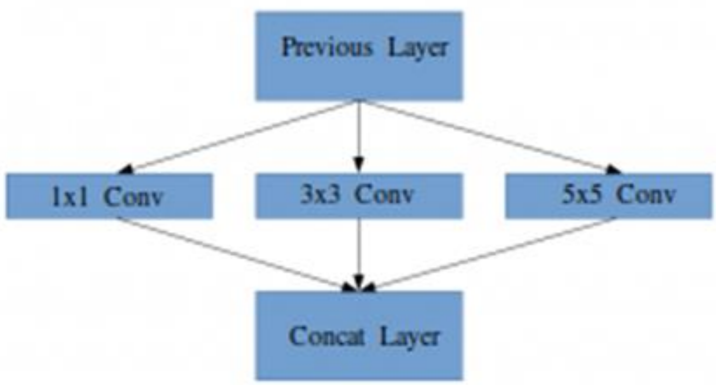


Figure 28 Idea of an Inception module [38]

Along with the above-mentioned layers, there are two major add-ons in the original inception layer:

- 1×1 Convolutional layer before applying another layer, which is mainly used for dimensionality reduction
- A parallel Max Pooling layer, which provides another option to the inception layer

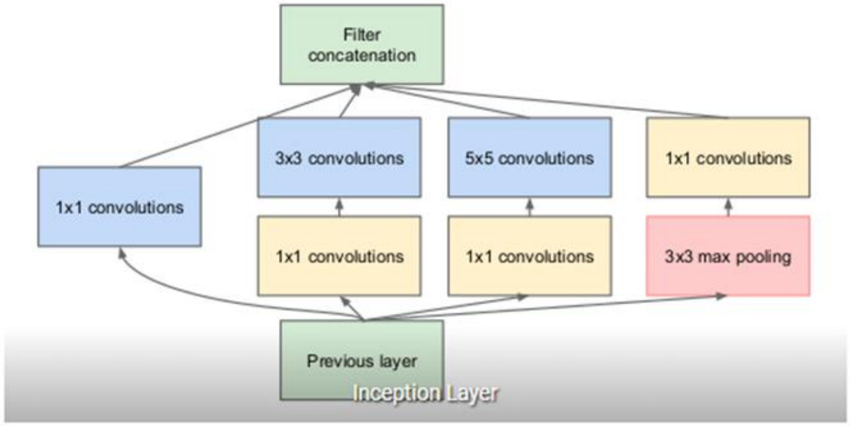


Figure 29 Inception layer [38]

2.4.4.5 ResNet

The idea of ResNet is straightforward, and that is to bypass the input to the next layers and also to feed the output of two successive convolutional layers. More than a thousand layers of the network were trained for the first time in ResNet.

After the prestigious victory of AlexNet at the LSVRC2012 category contest, deep Residual network [2] changed into arguably the maximum groundbreaking work within the computer imaginative and prescient/deep learning network within the previous couple of years. ResNet makes it possible to train up to masses or even hundreds of layers and still achieves compelling performance

Taking benefit of its effective representational capability, the overall performance of many laptop imaginative and prescient packages aside from photo type had been boosted, consisting of item detection and face recognition.

Since AlexNet, the state-of-the-art CNN architecture is going deeper and deeper. While AlexNet had only 5 convolutional layers, the VGG network [3] and GoogleNet (also codenamed Inception_v1) [4] had 19 and 22 layers respectively.

But, growing network depth does no longer work by certainly stacking layers collectively. Deep networks are difficult to educate due to the vanishing gradient trouble because the gradient is returned-propagated to earlier layers, repeated multiplication may additionally make the gradient infinitely small. As end result, as the network goes deeper, its performance gets saturated or maybe starts offevolved degrading swiftly

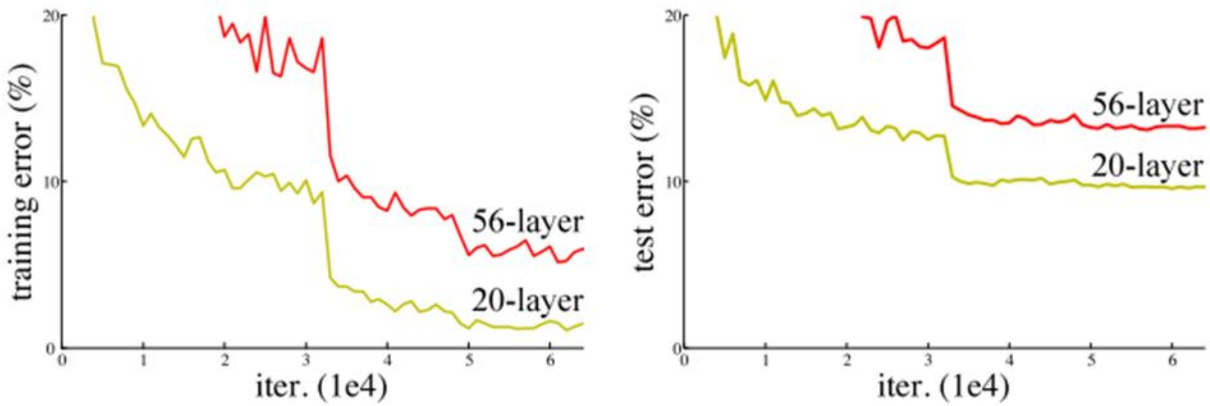


Figure 30 Sample Training using ResNet[39]

The core idea of ResNet is introducing a so-called "identity shortcut connection" that skips one or more layers, as shown in the following figure:

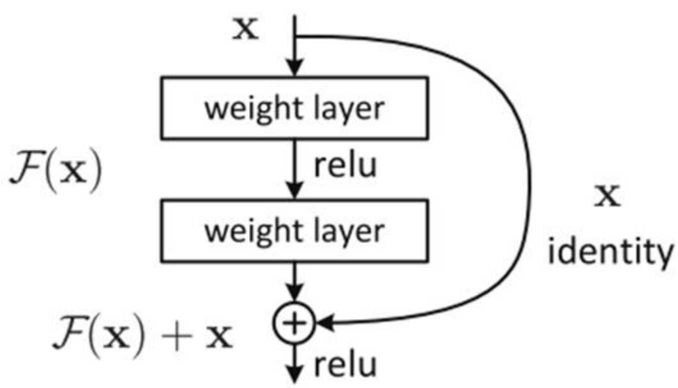


Figure 31 A residual block [41]

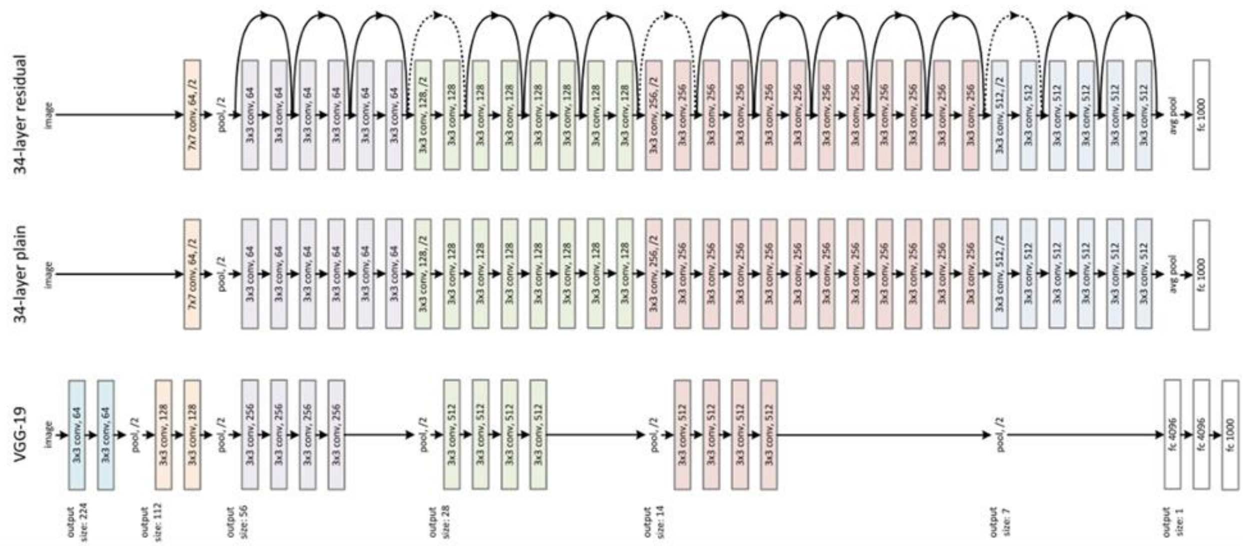


Figure 32 Full Architecture of ResNet [42]

The authors argue that stacking layers shouldn't degrade the community overall performance, because we ought to without a doubt stack identity Mappings (a layer that doesn't do something) upon the modern-day network, and the resulting structure might carry out the equal. This shows that the deeper version should no longer produce a Training error better than its shallower opposite numbers.

They hypothesize that letting the stacked layers fit a residual mapping is less difficult than allowing them to without delay match the preferred underlaying mapping. And the residual block above explicitly allows it to do precisely that.

2.4.4.6 Densely Connected CNN

Huang et al. [9] proposed a unique structure known as DenseNet that similarly exploits the consequences of shortcut connections. It connects all layers immediately with each different. On this novel architecture, the center of each layer consists of the function maps of all in advance layers, and its output is passed to each next layer. The feature maps are aggregated with depth-concatenation.

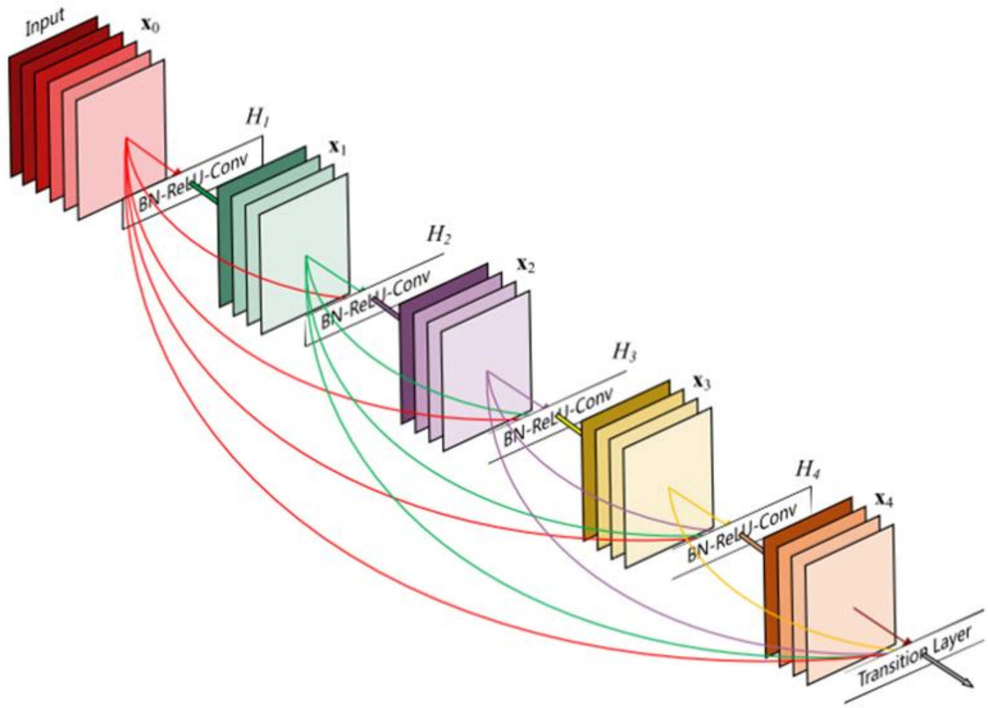


Figure 33 Illustration of DenseNet [43]

Other than tackling the vanishing gradients problem, the authors argue that this architecture also encourages feature reuse, making the network highly parameter-efficient. One easy interpretation of that is that the output of the identification mapping turned into introduced to the next block, which may obstruct statistics float if the characteristic maps of layers have very exclusive distributions. Therefore, concatenating feature maps can keep all of them and growth the variance of the outputs, encouraging characteristic reuse.

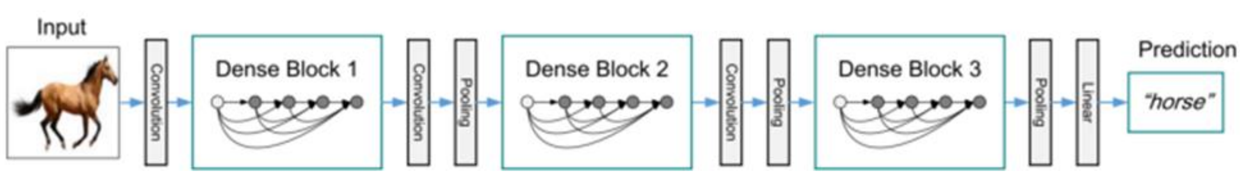


Figure 34 DenseNet Full architecture [46]

Following this paradigm, we know that the l_th layer will have okay * (l-1) + k_0 input characteristic maps, wherein k_0 is the number of channels in the enter photograph. The authors used a hyper-parameter referred to as growth rate (ok) to save you the community from growing too extensive, additionally they used a 1x1 convolutional bottleneck layer to lessen the range of function maps before the high-priced 3x3 convolution. The overall architecture is shown in the below table:

2.4.4.7 Fast AI

Fastai is prepared round important design desires: to be approachable and rapidly efficient, at the same time as additionally being deeply hackable and configurable.

Other libraries have tended to pressure a desire between conciseness and pace of improvement, or flexibility and expressivity, but no longer each. We wanted to get the clarity and development speed of Keras [2] and the customizability of PyTorch.

This reason of getting the exquisite of both worlds has inspired the layout of layered architecture. A high-level API powers ready-to-use capabilities to educate models in numerous programs, offering customizable models with sensible defaults. It is built on pinnacle of a hierarchy of decrease-level APIs which give composable building blocks. This way, a consumer looking to rewrite a part of the high-stage API or upload particular conduct to suit their desires doesn't need to discover ways to use the lowest degree.

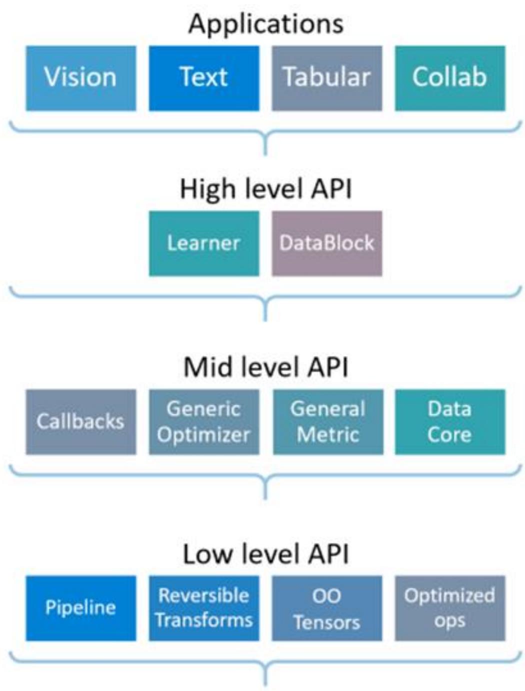


Figure 35 the layered API from fastai [51]

The excessive-stage of the API is maximum in all likelihood to be useful to beginners and to practitioners who are mainly inquisitive about making use of pre-existing deep learning strategies. It offers concise APIs over four primary utility regions: vision, text, tabular and time-series evaluation, and collaborative filtering. These APIs pick out wise default values and behaviors primarily based on all to be had statistics. As an instance, fastai provides an unmarried Learner magnificence that brings collectively architecture, optimizer, and facts, and automatically chooses the perfect loss function in which possible.

Integrating these concerns into a single elegance enables fastai to curate suitable default selections. to present any other instance, generally, a schooling set must be shuffled, and validation does not. So fastai provides a single DataLoaders class that automatically constructs validation and education information loaders with those details already treated. This enables practitioners make sure that they don't make errors such as failing to consist of a validation set.

Training, layer freezing, and discriminative getting to know charges [4]. In widespread, the library's use of integrated defaults method it requires fewer traces of code from the person to re-specify records or merely to connect components. As a end result, each line of user code has a tendency to be much more likely to be meaningful, and less complicated to examine.

The mid-level API provides the center deep gaining knowledge of and information-processing methods for every of those applications, and occasional-level APIs offer a library of optimized primitives and practical and object-orientated foundations, which lets in the mid-level to be Advanced and customized. The library itself is constructed on top of PyTorch [5], NumPy [6], PIL [7], pandas [8], and various other libraries. To achieve its purpose of hackability, the library does not intention to supplant or disguise those decrease stages or this foundation?

Inside a fastai version, you can actually engage at once with the underlying PyTorch primitives; and inside a PyTorch model, you can still incrementally undertake components from the fastai library as conveniences instead of as an incorporated package. They believe that fastai meets its design goals.

2.4.4.8 MobileNet

As the name suggests, the MobileNet version is designed for use in mobile applications, and it is TensorFlow's first mobile Personal Computer vision. MobileNet uses depthwise separable convolutions.It notably reduces the number of parameters whilst in comparison to the community with everyday convolutions with the same intensity in the nets. This consequences in lightweight deep neural networks.

A depthwise separable convolution is made from two operations.

1. Depthwise convolution.
2. Pointwise convolution.

MobileNet changed into open-sourced through Google, and consequently, this offers us a terrific start line for learning our classifiers that are insanely small and insanely rapid. The velocity and energy intake of the community is proportional to the variety of MACs (Multiply-Accumulates) that is a degree of the wide variety of fused Multiplication and Addition operations.

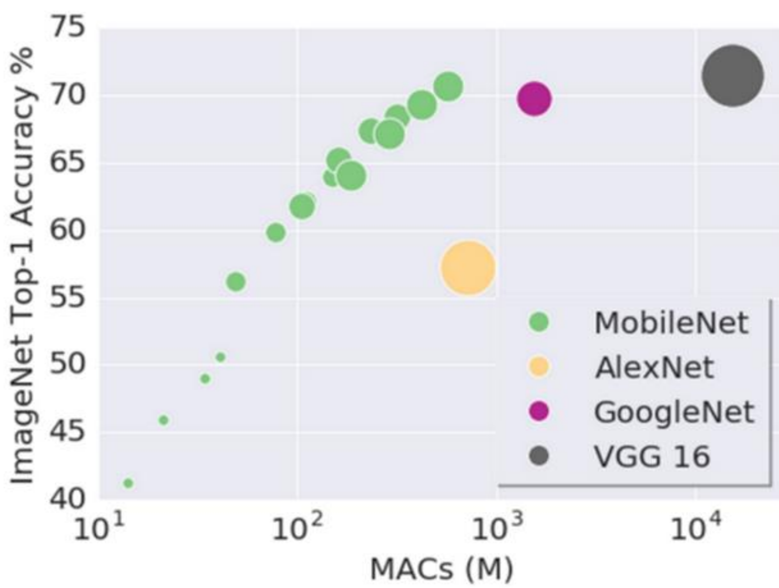


Figure 36 ImageNet Top 1 Accuracy [51]

Table 1. MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5× Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
5× Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

Figure 37 Architecture of MobileNet [53]

MobileNet makes use of Depthwise Separable Convolution. This convolution originated from the concept that a clear out's intensity and spatial dimension may be separated thus, the name is separable. Let us take the example of the Sobel filter, used in image processing to detect edges.

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Figure 38 Sobel Filter. Go for the vertical edge, Gy for horizontal edge detection [37]

We can separate the peak and width dimensions of those filters. Gx filter out can be regarded as a matrix made from [1 2 1] transpose with [-1 0 1]. We note that the filter had disguised itself. It suggests it had nine parameters, however it has 6. This has been viable due to the separation of its peak and width dimensions. The equal idea applied to separate intensity size from horizontal (width*peak) gives us depth-smart separable convolution in which we carry out intensity-clever convolution. After that, we use a 1*1 clear out to cowl the intensity dimension

One factor to word is how many parameters are reduced by means of this convolution to output the same number of channels. To supply one channel, we want 3*3*3 parameters to carry out intensity-clever convolution and 1*three parameters to perform further convolution in-intensity size. however If we want three output channels, we handiest need 31*3 depth filters, giving us a complete of 36 (= 27 +9) parameters even as for the identical number of output channels in everyday convolution, we want 33*3*3 filters giving us a complete of 81

parameters.Depthwise separable convolution is a depthwise convolution followed by a pointwise convolution as follows:

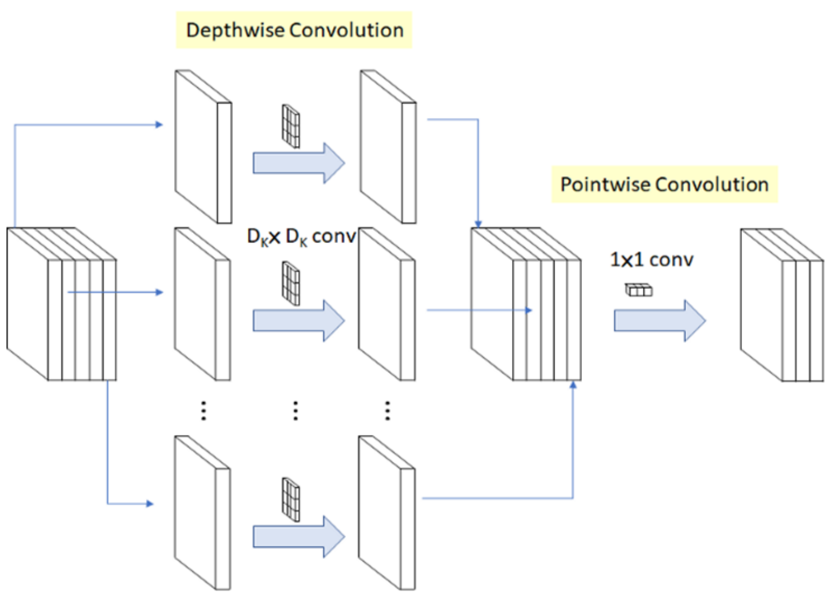


Figure 39 Depthwise Separable Convolution [36]

2.4.5 Neural Network Optimization

When training a neural network, two key things need to be considered. These are Hyper-parameters and parameters. Hyper-parameters are parameters that are manually set before training a system to get the finest result and once they are set they cannot be changed during training. [31] A typical set of hyper-parameters for a neural network includes the number of hidden layers, initialization weight values, learning rate, decay rate, dropouts, and Activation functions.

Parameters are those which would be learned by the machine-like weights and biases. These values differ for each experiment and epoch. Moreover, they are highly dependent on the type of data and the task at hand. Parameters are the ones that the "model" uses to make predictions. [32]

2.5 Related Work

Due to the complexity of formalizing the judgment for automatic detection of lung pathologies based on images, many researchers attempted to develop an automatic detection system using e machine learning and deep learning methods. This requires the creation of a reasonable training set that consists of images with labeled lung contours [3]. The wide application of chest x-rays and the complexity of reading them make computer-aided (CAD) systems and computer vision a hot research topic.

Studies have been done using machine learning on two cascaded SVM classifiers [4] to classify the CXR image as TB and non-TB In this method, a small dataset has been used, to detect only TB from multiple lung diseases The result has relatively low accuracy (about 84%). Moreover, lung cancer and pneumonia detection using ANN were also proposed [5].

Even though the method detects both lung cancer and pneumonia diseases, with improved accuracy, it lacks further classification and the source of the data was not specified this will be less acceptable. And also, another method using Feedforward neural network [6] was also proposed to improve the accuracy and increase the classification of lung cancer, pneumonia, and TB, and they have used 1450 x-ray images but the source was not specified.

Recent state-of-the-art deep neural network (DNN) models have been used in many works for breast cancer diagnosis [38, 39]. To classify breast tissues biopsy images as normal, benign, malignant, and invasive carcinoma, a deep CNN-like patch level voting model and merging model was used and an accuracy of 87.5 % was achieved [70]. Likewise, DCNN and gradient boosted tree method was used to classify breast cancer into the basic 4 types [40] and an accuracy of 87.2 + 2.6% was reported.

CNN as a feature extractor and support vector machine as a classifier has been implemented in another study [41] by retrieving image information at different scales, including both nuclei and overall tissue organization. An accuracy of 77.8% for the four classes and 83.3% for carcinoma (in situ and invasive) or non-carcinoma (normal and benign) was claimed. The inception-v3 convolutional neural network has also been adapted and fine-tuned to make patch classification and majority voting was considered [42] for the whole slide classification yielding an accuracy of 85% over the four classes and 93% for non-cancer (normal and benign) versus malignant (in situ or invasive carcinoma) was reported. Similarly, SVM-based classification with texture features [7] was proposed to solve the problem related to the accuracy, it achieved 96%, the however further classification problems are still existing.

The problem was solved by doing PCA-SVM classification [8]. It was able to classify 14 thoracic diseases, nevertheless, the classified diseases are not common, and the accuracy was 92% on training dataset.

From Jimm University also Multiple Lung Diseases Classification from Chest X-Ray Images using Deep Learning approach was proposed to classify Lung cancer, Chronic Obstructive Pulmonary Disease (COPD), pneumonia, and tuberculosis (TB) and there dataset was 11,716 among only 443 have been collected from jimm university medical center while the other 11,273 x-ray images collected from National Institute of Health (NIH) online dataset repository during their investigation sensitivity, and specificity of 97.3%, 97.2%, and 99.4%, Accuracy respectively have been achieved for multi-class classification.[65]

In addition to the above machine learning methods, several studies have been carried out using the recent state-of-the-art, deep learning. Deep Learning for abnormality detection in Chest X-Ray images was introduced [9] by using the three neural network models GoogLeNet, InceptionNet, and ResNet to diagnose lung diseases, but it was a binary classification task either it is normal or abnormal, or the accuracy was not specified.

In another work [10], the Residual Inception module has been used to classify TB and pneumonia [10]. The system has an accuracy of 96% for both classes. , But, it still lacks further classification, the amount of data was not exactly specified, and the method is not clear enough. Another study has also been done on Chest Pathology identification using Deep Feature Selection with Non-Medical Training using Convolutional Neural Network (CNN) that was trained with ImageNet to classify the pattern of the diseases [11]., But, there is a problem which is the system detecting the pattern but not the diseases, the dataset was very small (93 chest x-ray images), and the accuracy was not indicated on the paper.

Furthermore, CNN has been used to classify general opacity, diffuse lung opacity, cardiomegaly, abnormal hilar with the accuracy of 86.7% [12]. But similarly, it lacks further classification, detects the Lung, not the diseases, and has relatively low accuracy.

In addition to this, DCNN has been used to classify 8 Features in disease [13]. And still, the mixing of patterns with pathology and the accuracy was not indicated.

The chest Net model that consists of a classification branch and an attention branch has been introduced to classify 14 thoracic diseases [14]. Even though the problem on further classification is solved, there is a problem on the accuracy, the mixing of pattern and pathology, the pathologies detected are not the common ones, there is a strong class imbalance, and random image split (one patient appear at the same time for training and test).

Table 2 Summery of Related Works

Cite	Problem	Methods	Class	Result
[4]	TB Classification	two cascaded SVM classifier	2 class	84%
[38,39].	Breast Disease Classification	CNN	4 Class	87.5 %
[41]	Breast cancer Classification	DCNN	4 Classes	93.8+2.3%
[42]	Classification of carcinoma (breast Disease)	CNN, Inception-v3	2 Classes	83.3%
[7]	CXR Pathogens	SVM	4 Classes	96%
[10]	Classification of TB and pneumonia	GoogLeNet, InceptionNet, and ResNet	2 classes	96%
[12]	Classification of general opacity, diffuse lung opacity, cardiomegaly, abnormal hilar	CNN	4 Classes	86.7%
[65]	Multiple Lung Diseases Classification	DCNN	4 Classes	93.7%

2.6 Research Gap

We have reviewed a lot of literatures that related to our works and all of them are based on the Classification of generic/base disease and even COPD is also investigated as one class it will classify if the given x-ray images if COPD or Normal not more than so based on the above reason no research investigated or conducted to classify subtype of chronic obstructive pulmonary disease and we are the first one working on this.

2.7 Summary

In this chapter, the researcher has discussed research works related to Chest x-ray Disease Classification systems. Literature works which employed Deep learning, Data augmentation, Transfer learning, and Image processing were reviewed.

An artificial intelligence algorithm could reduce the radiologist's effort and provide alternative technology. Each literature works which are reviewed in this chapter has its contributions and limitations. This study is motivated by literature works that applied transfer learning for the small medical dataset and the successful application of Deep Neural Network architectures. The researcher attempted to fill the gaps found in the related works in the next chapters.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Research Design

To conduct this research, a Design science Methodology will be used. The research paradigm focuses on the development and validation of prescriptive knowledge and the researcher manipulated and controlled testing to understand the causal process. It is used for the researchers for identifying the cause and observe which powerful tool the diagnostic of the results is. In this study, the researcher will be used the Design science method for model building, analysis, and prototype development and testing, whereas a non-experimental method will be used for knowledge elicitation through discussion with experts and document review.

3.2 Ethical Considerations

The researcher learned enough about the culture of informants to ensure it is respected during the data collection process and ensure the confidentiality of the data obtained. To do that, an Approval letter of ethical clearance was obtained from the Research and Ethical Review Board (RERB/IRB) of Jimma Institute of Technology, Jimma University for Jimma Medical Center, MSF Holland Gambella Branch, St. Paul Specialized Hospital, Black Lion Specialized Hospital, and Betel Specialized Hospital.

Confidentiality was ensured during the data preparation and data collection and interview of domain experts; thus, the name and address of the patient were not recorded and the data stained X-ray images were used only for the research purpose. The work of others was properly cited or credited, whether published or unpublished and whether or not it became written work, an oral presentation, or material on an internet site. It’s far a primary duty of a researcher to avoid either a fake declaration or an omission that distorts the research file. A researcher need to not file anticipated research consequences that had no longer but been discovered at the time of submission of the file.

3.3 Data Quality assurance

To increase the quality of the data, the researcher was prepared to work manually to check every day's progress. And also, assistants are selected and trained to handle the data carefully. The researcher has checked the reliability and accuracy of the data as well.

3.4 Data preparation

Data was collected from chest x-ray of the patients with the help of the radiology department coordinator of JUMC, St Paul's Millennium Medical College, Black Lion Specialized Hospital, Hayat Medical College, and the other data, which is the online dataset of chest X-ray image collected from the source Padchest. A total of around 3600 CXR images was collected for the three Type of COPD and Normal (non-infected) image [25].

The reason why we have prioritize this 6 health facility is they are the top huge and old health facility in the Ethiopia so that they has more data than others in addition to the above reason we have a peoples in this 6 health facility that can collaborate with us in the data collection and annotating.

The images are saved in Joint Photograph Experts Group (.JPG) file format, with 24 bits per pixel. The images saved can then be stored in the computer and have processed in real-time mode using deep learning techniques. These image datasets are the result of a real patient's X-ray that was prepared through the process using examination by domain experts.

3.5 Data Annotating /Labeling

In machine learning, data labeling is the process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to provide context so that a machine learning model can learn from it. [43] It is an important part of data preprocessing for ML, in which both input and output data are labeled for classification to provide a learning basis 44] Errors in data labeling impair the quality of the training dataset and the performance of any predictive models it's used for. To mitigate this, many organizations take a Human-in-the-Loop (HITL) approach, maintaining human involvement in training and testing data models throughout their iterative growth. [43]

As I discussed above, the Dataset was collected from 6 Hospitals Located in Ethiopia, Namely Jimma University Medical Center, MSF Holland Gambella Branch, St paul's specialized Hospital, Balck Lion Specialized Hospital, Reftyvally University collage specialized hospital Adama Branch and Betele specialized Hospital and online datasets.

Also, professionals or experts too from the 6 Hospitals named as Radiologist 5 (Sub Specialist radiologist) from Jimma Medical Center, Radiologist 4 (Sub Specialist) from MSF Holland Gambella Branch, Radiologist 1 (Sub Specialist Radiologist) from St.paul Specialized Hospital, Radiologist 2 (Sub Specialist) from Balck Lion Specialized Hospital, Radiologist 3 (Sub Specialist) from Rift valley University Collage specialized Hospital Adama and Radiologist 6 (Sub Specialist Radiologist) from Betel Specialized Hospital Addis Ababa were involved in data annotation.

To minimize the rate of error and bias, I have Distributed X-ray images according to the below table (for example it's assumed that dataset or X-ray comes from St paul's specialized Hospital have been labeled by Radiologist 1 since Radiologist 1 is an Expert of St paul's specialized Hospital, so I have given that data to Radiologist 2 who is expert of a Black Lion specialized Hospital so if the Label of Radiologist 2 is the same with Radiologist 1 I will take that X-ray image for training but if it’s not the same I will give that image to other experts to Radiologist 3 or Radiologist 4 , so on). I will do this process to get an error-free labeled dataset to train the network.

Table 3 X-ray Image Distribution for Labeling

	Jimma University Medical Center	**MSF Holland Gambella branch**	**St paul's specialized Hospital**	**Black Lion specialized Hospital**	**Reftyvally University collage specialized hospital**	**Betele specialized Hospital**
Radiologist 1	2	1			3	2
Radiologist 2	1	2		3		
Radiologist 3			1			3
Radiologist 4			2	1	2	
Radiologist 5	3		3		1	
Radiologist 6		3		2		1

3.6 Image (data) preparation

Before feeding the data to the network, the dataset was split into training and test set. The data was split into training and: test set randomly; the well-known rule of splitting the data is 80–20 percent training and testing sets respectively. 80% will be used for Training while 20% was used for test. A training set will be used to train the network while a Test set will be used to Test the model performance after the training process, fine-tune hyper-parameters, and perform model selection.

3.7 Preprocessing

Data preprocessing is a required task for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

3.7.1 Data Format Conversion

The main purpose of preprocessing is to convert the DICOM file (default extension for x-ray images) to a common image file (JPG). In CXR image analysis, Preprocessing will be done to enhance and remove noises from the image and to extract essential information for further image analysis. It includes Blur and focuses corrections, Enhancements, Lighting corrections, filtering, noise removal, Thresholding. Edge sharpening, and noise suppression.

3.7.2 Image Augmentation

Data augmentation is a process of increasing the number of training data points in a dataset by generating more data from the original dataset [26]. It is important to increase the number of datasets. It helps the network to learn more complex features from the data and prevent the problem of Over-fitting. In this study, various data augmentation techniques were performed on the original images dataset.

3.8 Feature Extraction

Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing.

In Classical machine learning, features are crafted by a human being by hand or manually but recent algorithms like CNN, DCNN, and VDCNN extract features automatically by learning deeply the nature of data and which one is an important part of data and which one is not. And it's very useful when you need to reduce the number of resources needed for processing without losing important or relevant information. Not only can that but feature extraction also reduces the amount of redundant data for a given analysis.

3.9 Classification

Image classification is the process of predicting a specific class, or label, for something that is defined by a set of data points. Image classification is a subset of the classification problem, where an entire image is assigned a label. Perhaps a picture will be classified as a daytime or nighttime shot.

3.10 Implementation Tools

Investigation of available software tools with their libraries was conducted to select the appropriate tool for the implementation of the CNN algorithm for COPD using X-ray image classification. During the investigation, we have seen that there are tools that are general for deep learning and image processing algorithms and specific only for one of them. Before selecting the tools, we considered some criteria which are helpful to select the appropriate software tools with their corresponding libraries. Software tools that we used to implement the CNN algorithm are python as a programing language with Tensor-Flow and Keras libraries on an anaconda environment. These tools fulfill all the consideration criteria's and they were used in python which is familiar to us.

3.10.1 Python Programming language

I used python programming language because it offers concise and readable code. Whilst complex algorithms and flexible workflows stand in the back of deep learning and AI, Python's simplicity lets in developers to write dependable structures. Not only is this, Python code understandable by humans, but this also makes it easier to build models for machine learning.

3.10.2 Tensorflow and Keras

Keras is a high-level neural network API that is written in python which runs on top of TensorFlow or Theano or Microsoft Cognitive Toolkit (CNTK). It is very simple to develop a Model, user-friendly, easily extensible with python, and most importantly it contains pre-trained CNN models such as VGG16, Inception, EffeceintNetB0, and Resnet50 that we used during the experiment. It allows easy and fast prototyping and supports both CNN and ENN or the combination of the two [26].

3.10.3 Anaconda Environment

Anaconda will be used for the implementation of the model and it is a free and open-source distribution of the python for image processing, data science, machine learning, and related applications that aim to add and simplify package management and deployment. I have used Gradio to develop a User Interface for Models I have developed and I have used Google Collaborator notebook to implement the coding part. Which has a Graphical Processing Unit (GPU) for free and is easy to use and run in a web browser.

3.10.4 DICOM viewer

The Digital Imaging and Communications in Medicine (DICOM) standard was created by the National Electrical Manufacturers Association (NEMA) to aid the distribution and viewing of medical images, such as CT scans, MRIs, and ultrasound. ... This format is an extension of the older NEMA standard. All X-ray images we collected were in DICOM format and to view the images, a Dicom viewer was needed.

3.10.5 OpenCV

I used an open computer vision module for the data (image) preprocessing (image conversion, contrast enhancement, edge sharpening, and noise suppression) because OpenCV is an open-source computer vision and machine learning software library.

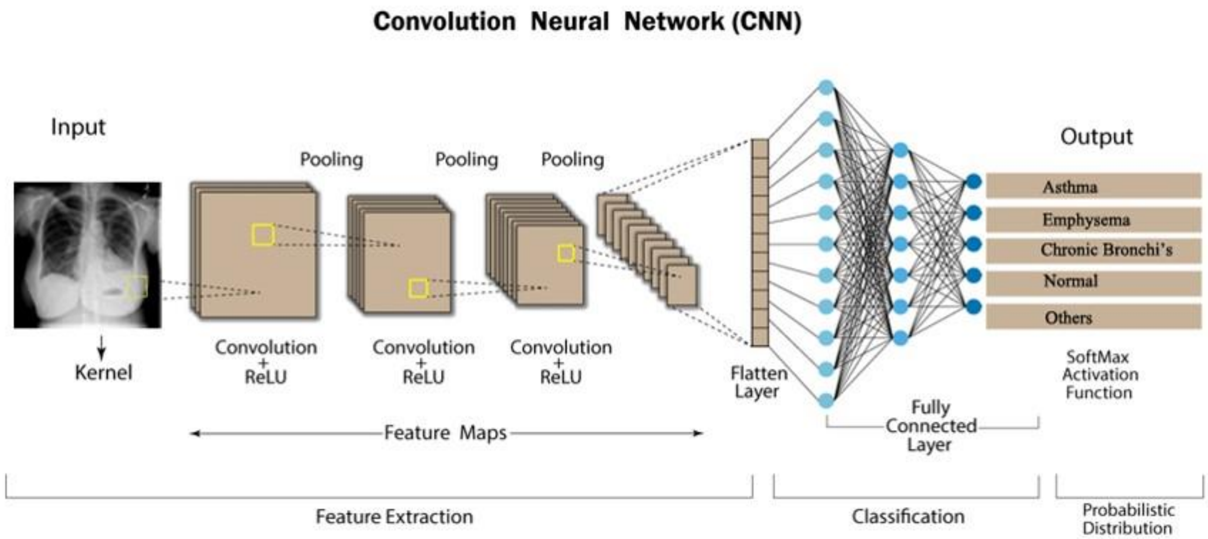


Figure 40 above shows the architecture of CNN [28].

3.10.6 TPU or GPU

I was tried to get TPU as a standalone machine Unfortunate it wasn't easy so we used GPU (Graphical Processing Unit) on Google Collaborator.

3.15 Pre-Trained Model.

I used, VGG16, InceptionV3, EffeceinNetB0, and Resnet50 Pre-Trained Model to boost my Model Accuracy and to implement Transfer learning because the pre-trained models contain trained weights for the network. Hence, if a network pre-trained for some classification task is used. The features extracted will be similar. Instead of initializing the model with random weights, initializing it with the pre-trained weights reduces the training time and hence is more efficient.

3.16 Methodology Architecture

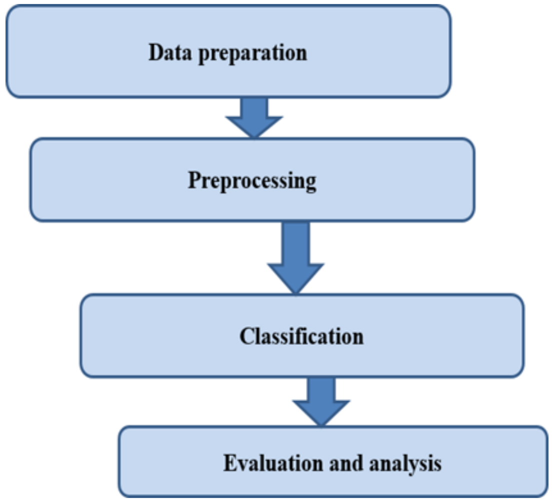


Figure 41 General block diagram of the method [7]

3.17 Evaluation of Model

This experiment uses the F1 score and the accuracy to evaluate the performance of the model.

The metrics used to evaluate the model in this classification task are accuracy of classification accuracy (CA), precision, recall, and F1 score. Accuracy: - the model accuracy was calculated as the percentage of correct prediction of the top class (the class having the highest probability as indicated by the CNN model) and the target class will be assigned by the author beforehand is the same. It will be represented by the formula below,

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Where,

TP (True positive) represents the positive instances that are correctly classified as positive, TN (True Negative) represents the negative instance that is correctly classified as negative, FP (False positive) represents the negative instance that is wrongly classified as positive, FN (False Negative) represents the positive that is wrongly classified as negative.

Precision: is calculated as the fraction of true positives (TP) from the sum of the relevant classes, i.e. the sum of the true positives and the false positives. It will be represented by the following formula.

$$Precision = \frac{TP}{TP + FN}$$

Recall: is calculated as the fraction of true positives from the sum of True positives and False Negatives. It can be represented by the following formula

$$Recall = \frac{TP}{TP + FN}$$

F1 score:-will be used in this experiment as the dataset is imbalanced. F1 score is interpreted as the harmonic mean of precision and Recall. It will be represented by the below.

$$F1\ Score = \frac{2 * precsion * Recall}{Precisoin + Recall}$$

CHAPTER FOUR

SYSTEM ARCHITECTURE

4. Introduction

A system architecture is a conceptual model that defines the structure, behavior, and more views of a system. [1] An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

A system architecture can consist of system components and the sub-systems developed, that will work together to implement the overall system.

4.1 Proposed architectures

This chapter presents the proposed Architecture for the Classification of Chronic obstructive pulmonary diseases from Chest X-ray Images using deep learning. I have used Pre-Trained Models Called VGG16, InceptionV3, EffeceintNetB0, and ResNet50 all of them are tested on a large image dataset called ImageNet.

The developed system is fine-tuned and mainly trained to classify the 3 sub-types of chronic obstructive pulmonary diseases and the other 2 classes namely Normal and Other. After the development of the model, there will be compare and contrast of each model architecture and its accuracy. To do so, data has been collected from one Publicly Available Online dataset and 6 Local Hospitals. The overall architecture or illustration of the proposed system is discussed below.

4.1.1 General description

Chest –Ray images will be divided into five-part. After the annotation then pre-processing techniques like Re-Scaling, Normalizing the image, and Augmentation will be performed. After that the network will be developed based upon Pre-trained models then the weight inside the pre-trained model and feature extracted from a new dataset (COPD Images) will be gathered and transferred to a classifier then the classifier will classify based upon the features (patterns).

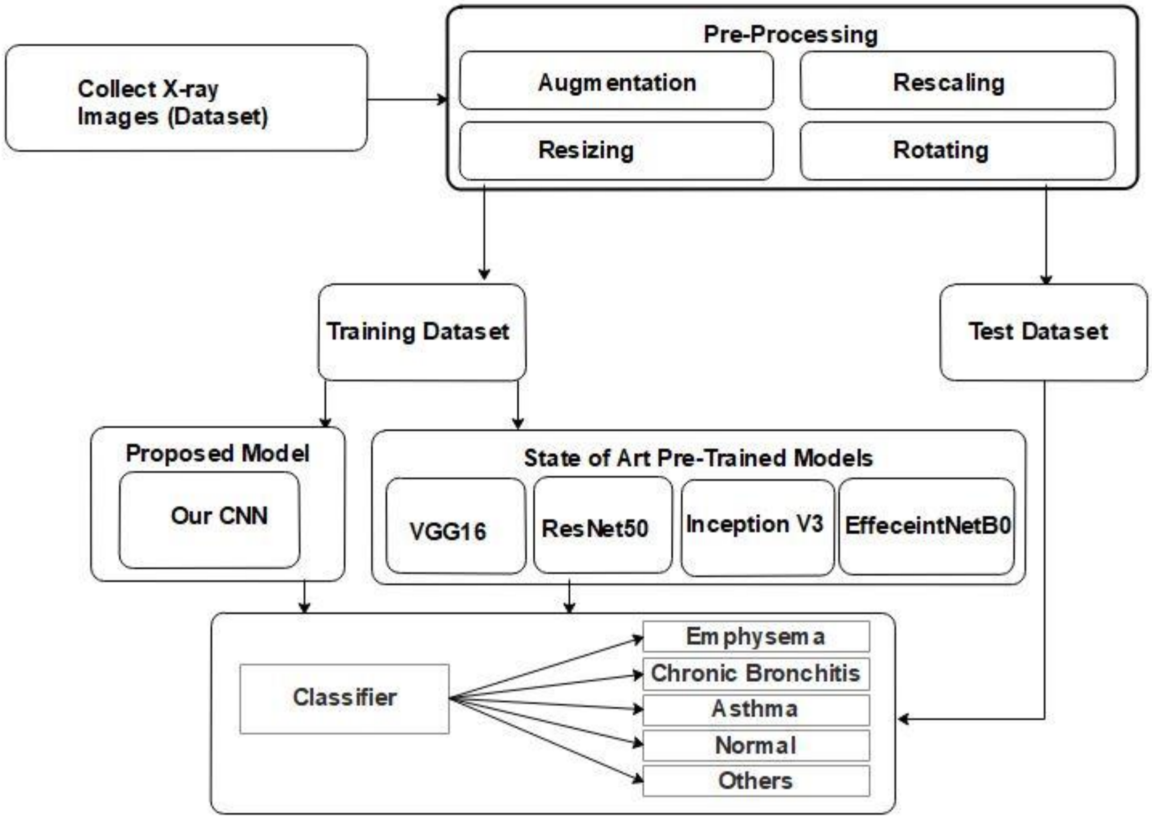


Figure 42 Proposed architecture

4.1.1.1 Description of the proposed architecture

The above model has both our proposed CNN model and state of art pre-trained model a CNN that we have developed on Google Collaborator Notebook .as you can see in above figure 42 there are Four components to construct and train the dataset, firstly the data were collected and labeled then after the labeling Process the data were pre-processed in this stage we have Augmented to classify images without worries about the Rotation of images, Re-scaled this to have some common image size and Normalized all images then in the third step the CNN network will extract features from the image automatically and for the pre-trained they will use their pre-existed features from which they acquired from ImageNet after the feature extraction the Network was learned features Finally the network classified the images to correspondent classes or Diseases.

The Pre-Trained model was adapted and fine-tuned to classify COPD images. The model was used as a feature extractor. That is, both its Architecture and weight value (from ImageNet) were adapted. Then the top layers were frozen, and some parameters like the learning rate, optimizer type, loss function, and decay rate were used to optimize the model. Finally, by using a soft-max classifier the probability distribution of each possible class was found. And based on the most probable class, the images were classified accordingly. Unlike the previous model this model has 5 main components while only one component called the Pre-trained model is unique.

CHAPTER FOUR:

RESULTS AND DISCUSSION

5.1 INTRODUCTION

In this section, the experimental evaluation that has been conducted for the detection and classification of Chronic Obstructive Pulmonary Disease is discussed. And, the experimental evaluation results show the reality of our proposed model. In this case, we test the accuracy of an end-to-end (CNN model and 4 different Pre-Trained models independently. The experimental evaluations enable to check the performance, efficiency, and accuracy of the proposed models. To this end, the decision/confusion matrix, loss/accuracy curve, and evaluation metrics results are used to illustrate the accuracy of the model.

5.1.1 Dataset preparation

We collected the data from four hospitals, namely Addis Ababa Tikur Anbesa specialized Hospital, st paul's specialized Hospital, Betele specialized Hospital, Jimma University Medical Center, Reftyvally University collage specialized hospital and MSF Holland Gambella Branch. , After collecting those data from the selected hospital's data augmentation was applied to the collected data to minimize the over-fitting problem and to increase the dataset size since CNN needs more data. The format of the collected images is DICOM (Digital Imaging and Communications in Medicine) which is .dcm files, but it is really difficult to work with Dicom image formats. Since it has low compression, it needs more memory to process and more computing time, so must be converted to JPEG (Joint Photographic Experts Group) format jpeg has great compression, for example, we try to convert the images from DICOM format to JPEG, we used one of the .dcm images and realize that the size of (.dcm) is 15.5MB and the (.jpeg) is only 330KB, therefore jpeg has great compression. From the 2248 image data, 80% of the dataset was used for training and 20% was used for testing in our CNN Model also for the Pre-trained Model too.

Table 4 Source of Data

Sr No	Source of Images	No of Images	Type of Image	Image Format
1	Jimma University Medical Center	19	Emphysema and Asthma	JPEG
2	Reftyvally University collage specialized hospital	310	Asthma ,Chronic Bronchi's and Emphysema	JPEG
3	St paul's specialized Hospital	422	Asthma ,Chronic Bronchi's and Emphysema	JPEG
4	Addis Ababa Tikur Anbesa specialized Hospital	521	Asthma ,Chronic Bronchi's and Emphysema	JPEG
5	Betele specialized Hospital	72	Asthma ,Chronic Bronchi's and Emphysema	JPEG
6	MSF Holland Gambella Branch.	33	Asthma ,Chronic Bronchi's and Emphysema	JPEG
	Online Dataset (Padchest)	871	Normal and Others	JPEG

Table 5 Data preparation

Class	Original collected images	Image format
Asthma	489	JPEG
Chronic bronchitis	338	JPEG
Emphysema	550	JPEG
Normal	510	JPEG
Others	361	JPEG

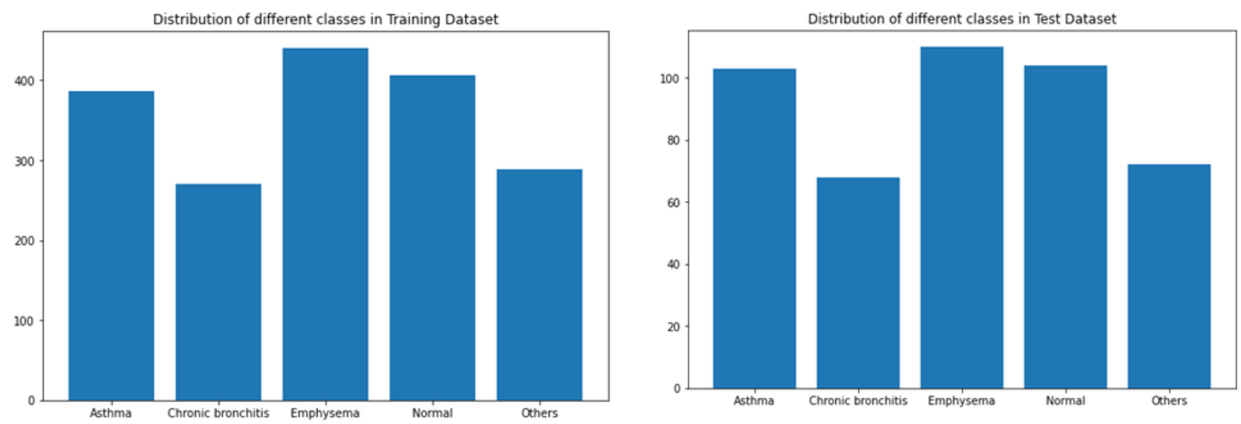


Figure 43 Data preparation

To implement the Chronic Obstructive Pulmonary Disease classification model, we make use of several open-source libraries. We used Google Collaborator 2021 with GPU for work with many open-source packages and libraries Google Collaborator 2021 is a distribution of the python programming languages for scientific computing (data science, machine learning application, large scale data processing, predictive analysis, etc.), that aims to simplify package management and development. And we used tensor flow, for numerical computation and large-scale machine learning, we used PIL from the Images to take images from the disk, Matplotlib, for constructing figures. Python programming language with version 3.9 for scientific computing on the hardware of 500 GB HDD, CPU of 3.2 GHz on 8GB RAM core i5 Lenovo laptop was used., Scikit-learn is another machine learning library used in this research. It consists of various classifications, clustering, and regression algorithms. We used a free source Keras (using TensorFlow as backend) and Tensor Flow library in this environment on the Windows 10 operating system. Keras is a free source deep learning library written in Python.

It is easily running on top of Tensor Flow. Tensor Flow is a representative math library and is also used for scientific computing tenders such as neural networks. And it is a free software library for deep learning and machine learning dataflow programming. TensorFlow was created by the Google brain team for use and analysis by Google users all over the world. It's a base library that can be used to create deep learning models right away. And OpenCV-Python is a Python linking library for solving computer vision problems. CV2 module is inside this library which is used to read the image, resize an image, change RGB to grayscale, etc.

5.1.2 Hyper Parameters in the model

When training with our model, some hyperparameters determine the network structure for an optimized result of training. Those groups of hyperparameters are determined for the network according to several training data set by choosing the batch size. Those are the number of an epoch: it determines how many times the model reads all the data set, batch size: parts from all datasets to be highlighted at a time, number of iterations: the number of batches to finish the dataset in one epoch. And, the other group of hyperparameters is also the following that we discussed in chapter two, which may differ for different convolution layers to choose the best value of Parameters our data we have experiment little bit and the report is below.

Table 6 Experimental Result of parameters Value

Sr No	Batch Size	No of epoch	Learning Rate	Dropout Value	Number of Hidden Layer	Accuracy	Loss
1	16	20	0.01	0.20	5	71.9%	0.655
2	16	20	0.01	0.40	5	66.5%	0.75
3	16	20	0.001	0.20	5	64.7%	0.81
4	16	20	0.001	0.40	5	66.0%	0.77
5	16	20	0.001	0.60	5	65.3%	0.79
6	16	20	0.0001	0.20	5	68.2%	0.75
7	16	20	0.0001	0.40	5	66.3%	0.80
8	32	20	0.01	0.20	5	68.6%	0.71
9	32	40	0.01	0.20	5	70.9%	0.66

After the experiment, we have got the hyper-parameters value that can give us the best performance for our model and it's described as below.

Table 7 Chosen CNN hyper-parameter Values

Hyper Parameter	Value
Activation	Relu
Striding	2
Padding	Same
Kernel size	3x3
Batch size	32
Learning rate	0.01
Dropout	0.2
Epoch	100
Optimizer	Adam

5.2 Result

In this Work, we have used our own CNN Architecture and the four transfer learning models were used. N namely InceptionV3, VGG16, EffecientNetB0 and ResNet50. The models were developed using Keras (TensorFlow of Python 3.7 as a backend) the model is trained for 100 epochs, a batch size of 32, a Dropout of 0.20, and a starting or initial learning rate of 0.01 (1e-3). The data is partitioned into training and testing datasets such that 80 percent of the data is allotted for training the model and 20 percent of the data is allotted for testing.

5.2.1 Evaluation result of Our Own CNN Model

In convolutional neural network model preparation, various parameters can influence the model performance. Therefore, we have to try to make optimum by assigning different values for those parameters, and finally, we have used a batch size of 32 because when the batch size is small the training time of the epoch is increased, epoch 100, optimizer “Adam” that is because adam is the best and suitable optimizer since it is the combination of Adgrad and RMSProp and we used grid search mechanisms to select this optimizer with 0.01 learning rate, dropout 0.2 to reduce model overfitting in the output 3 dense layers loss “categorical cross-validation” with the following model summary.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 222, 222, 16)      448
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 111, 111, 16)      0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 109, 109, 32)      4640
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 54, 54, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 52, 52, 64)        18496
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 50, 50, 128)       73856
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 25, 25, 128)       0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 23, 23, 512)       590336
_________________________________________________________________
dropout (Dropout)            (None, 23, 23, 512)       0
_________________________________________________________________
flatten (Flatten)            (None, 270848)            0
_________________________________________________________________
dense (Dense)                (None, 512)               138674688
_________________________________________________________________
dense_1 (Dense)              (None, 512)               262656
_________________________________________________________________
dense_2 (Dense)              (None, 5)                 2565
=================================================================
Total params: 139,627,685
Trainable params: 139,627,685
Non-trainable params: 0
```

Figure 44: Summary of the model

Our CNN model trained using the different activation functions like sigmoid, Tanager, and Relu, but among those functions, Relu performs better accuracy. Also, we preprocessed the data using 224x224 image sizes by using BICUBIC interpolation resizing techniques.

In this experiment, the model was trained at the Test and training phase using a CNN filter with 224X224 image size and with Relu as an activation function. From this experiment, we have achieved 81.1% for training accuracy. In the training and Test curve, there is some gap between them; this is because the major drawback of the median filter is that all pixels are substituted by the median of the window even if the pixel under concern is uncorrupted.

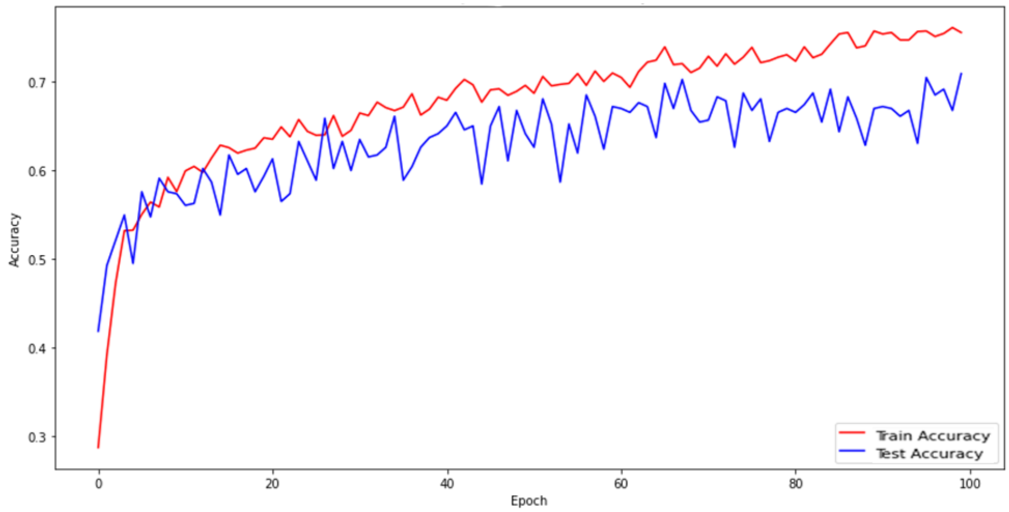


Figure 45 Training and Test Accuracy

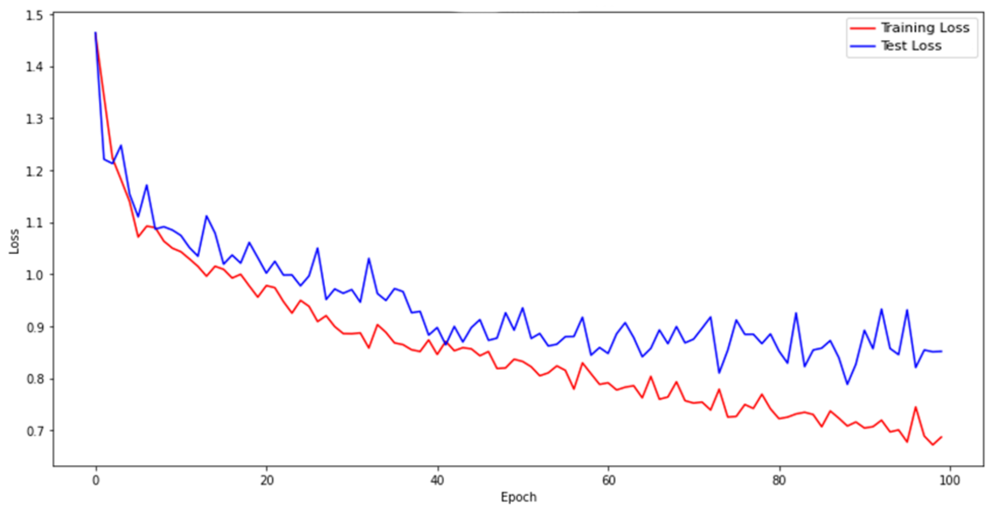


Figure 46 Training and Test Loss

The confusion matrix table of the softmax classifier in a convolutional neural network is shown in diagrams 46 and 47 we inferred from this testing that five classes (obstructive, non-obstructive, and others) were a little bit confused. That is, the best-verified outcome is postulated using a confusion matrix to show how our model is a little bit confused. As clearly depicted in figure 47 our model obtained 81.1% testing accuracy.

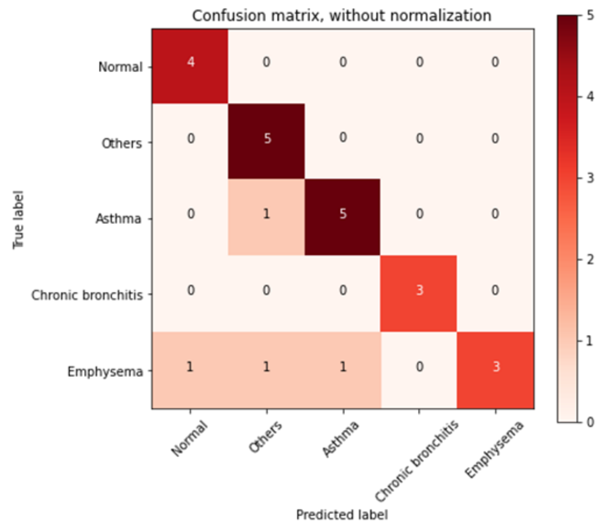


Figure 47 Confusion matrix

5.2.2 Evaluation result of InceptionV3 Pre-Trained Model

InceptionV3 is a pre-Trained Model of Google Brain and it has around 42 layers to perform the convolutions with different filter sizes on the input, performs Max Pooling, and concatenates the result for the next Inception module. Not only that, it introduced the 1 * 1 convolution operation to reduce parameters drastically and the model has only 7 million parameters. It was much smaller than the then prevalent models like VGG and Others Due to this, it takes only 3 hours and 32 minutes to train. In this experiment, the model was trained at the Test and training phase using Google Pre-Trained Model called Inception version 3 (InceptionV3) with 224X224 image size and with Relu as activation function. We have trained the model using 100 epochs and we have achieved 90.1% for Test accuracy respectively. This inceptionv3 model has high accuracy results on our test data. The Plot figure will show us the accuracy and loss result of the InceptionV3 model.

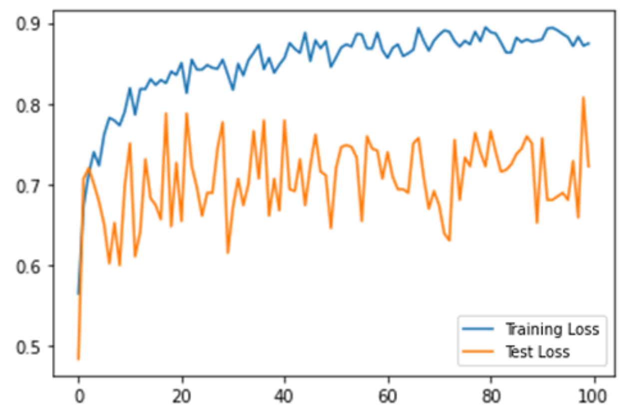


Figure 48 Loss of Training and Test

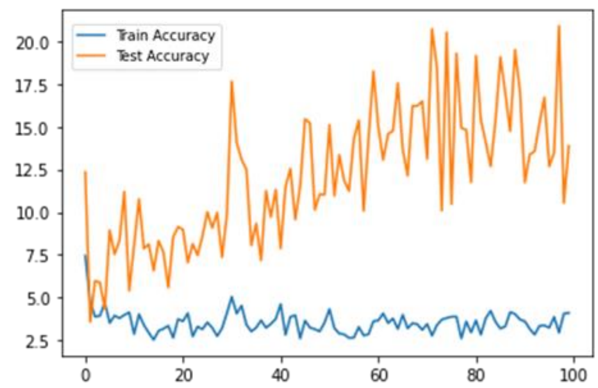


Figure 49 Training and Test Accuracy

5.2.3 Evaluation result of VGG16 Pre-Trained Model

VGG16 is one of the DCNN we have used in this work the architecture of VGG-16 is shown in Figure 51; it has 13 convolutional layers and 3 fully connected layers. The convolutional layers in VGG-16 are all 3×3 convolutional layers with a stride size of 1 and the same padding, and the pooling layers are all 2×2 pooling layers with a stride size of 2. The default input image size of VGG-16 is 224×224. After each pooling layer, the size of the feature map is reduced by half.

The last feature map before the fully connected layers is 7×7 with 512 channels and it is expanded into a vector with 25,088 (7×7×512) channels.

The default image size of VGG-16 was 224×224, which was the same as the size of the images in the dataset. We trained the model on the training data and tested the model on the test data (20% of the entire dataset) and the models' diagnostic efficiency of the test data. After 100 epochs (which took 4 hours and 45 minutes), we obtained 87.6% Test accuracy The Plot figure shows the accuracy and loss result of the VGG16 model.

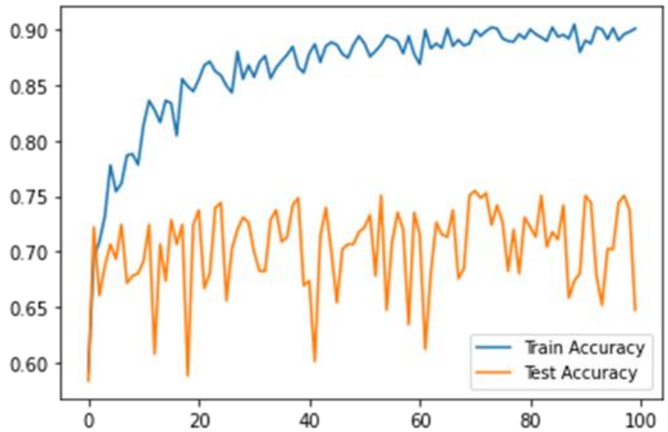


Figure 50 Training and Test Accuracy

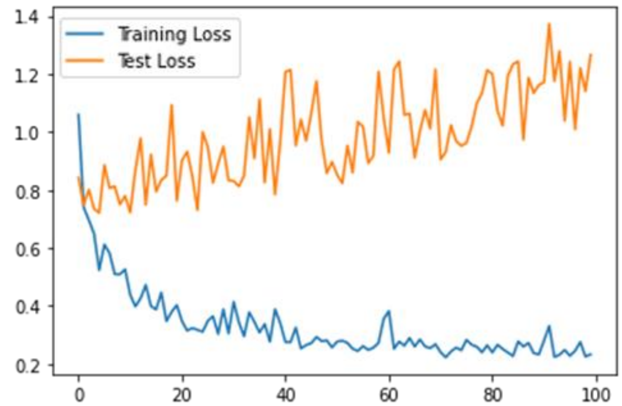


Figure 51 Loss of Training and Test

5.2.4 Evaluation result of EffeceintNetB0 Pre-Trained Model

EfeceintNetB0 is the latest Model from In EfficientNetB0, Google proposes a new Scaling method called Compound Scaling. We achieve much better performance and we have gotten around 85.7% accuracy using constant parameters we have used in the previous models and it took around 3 hours and 55 Minutes for 100 epochs using Google GPU.

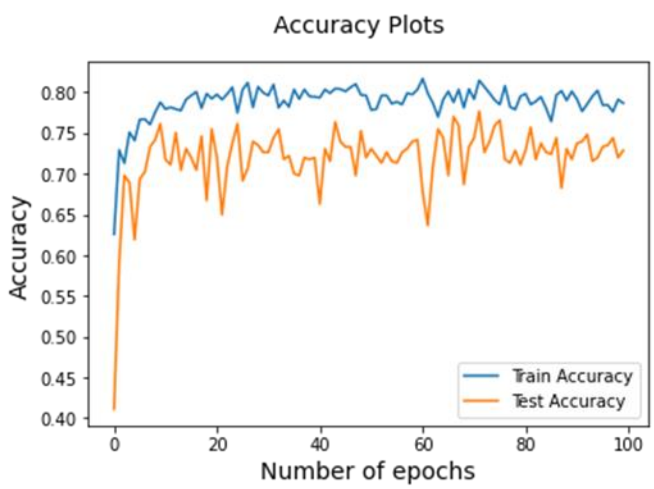


Figure 53 Training and Test Accuracy

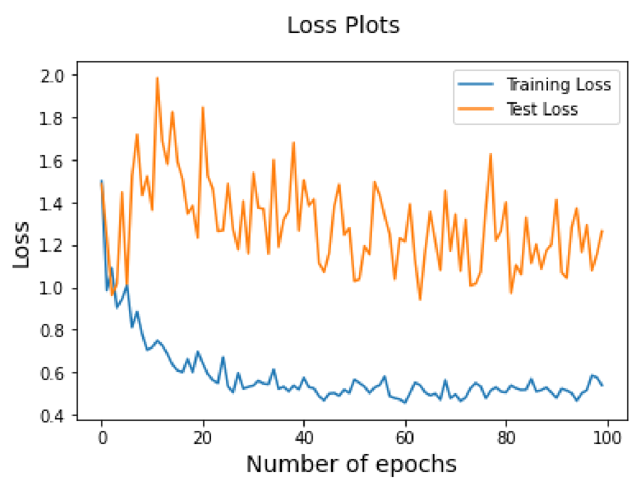


Figure 52 Loss of Training and Test

5.2.5 Evaluation result of ResNet50 Pre-Trained Model

The fourth DCCN Model is ResNet50 which has 50 Layers and 143,667,240 parameters with 224X224 image size and with Relu as an activation function. We have trained the model using 100 epochs.

Because of the smaller dataset, we were unable to see the exact power of the Resnet50 model and the accuracy we have got is around 67.5% and it was reported as the smallest accuracy in this research. Compared with another DCCN model we have used in this work, we have got the smallest accuracy because Resnet50 is very vast and has bulky layers due to these issues it needs a big dataset but in our case, we have only 2,248 datasets for both training and test respectively.

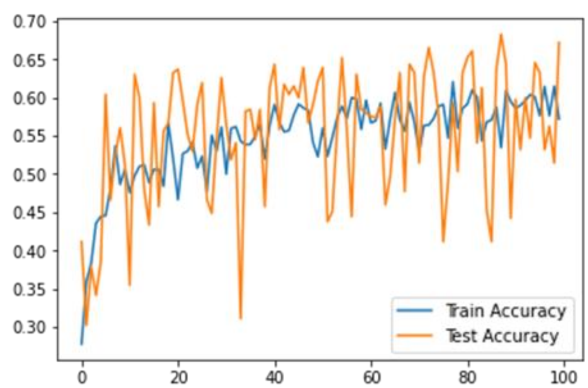


Figure 55 Training and Test Accuracy

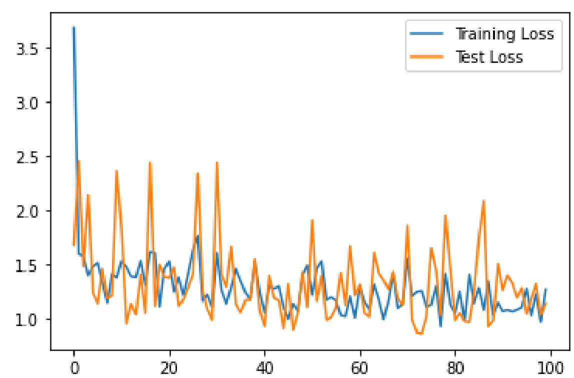


Figure 54 Loss of Training and Test

5.3 Discussion

We have collected 2248 images from different 4 health facilities and Online Dataset called Padchest we have distributed them to 8 health professionals (radiologists and medical doctors) to label images into 5 different classes namely Chronic Bronchi's, Emphysema, Asthma, Normal and Others among this 3 of them are a type of Chronic Obstructive pulmonary disease (COPD) we have used Normal Images (no Finding) to classify if the images are Normal then also we have collected non-COPD diseased images and named them as others so our model can classify if the x-ray images are COPD, non-COPD and even its Normal.

Data annotating part was very difficult and time-consuming it took about 2 months and a half to annotate 2248 x-ray images after the annotating step we have classified 489 images for Asthma, 338 images for Chronic bronchitis, 550 images for Emphysema, 510image for Normal, and 361images Others classes.

As Implementation environment, I have Google Collaborator because to train the images I need a Processor called Graphical Processing Unit which is available only on Servers and some few laptops called Gaming Computers so the only was using free only GPU services like Kaggle, Google Colab among them I have chosen the Google Collaborator because it has a very large community, tutorials and available free scripts, as main tensor flow open-source platform to build a CNN Network using its very powerful backend Keras and other libraries that makes the process very easy.

The common structure of a CNN for image classification has two main parts a long chain of convolutional layers for feature extraction and a few layers of the fully connected neural network for classification. For the feature extraction step, CNN with a sequential model with multiple layers is used. It consists of seven layers of Conv2D, with ReLu, Max-pooling2D, and fully connected layers. Then to classify Chronic Obstructive Pulmonary diseases, the extracted features are fed to dense layers. Features in the preprocessed dataset are extracted to reduce their dimension and use other Classification models like VGG16, Resnet50, InceptionV3, and EffeceintNetB0 on it. The convolutional neural network architecture with the sequential model is implemented with many layers such as convolutional, activation, max-pooling to extract important features from the Chronic Obstructive Pulmonary diseases x-ray image. Keras with the TensorFlow backend system helps to construct this model layer by layer. The Conv2D layer is used for input images as 2-

dimensional matrices and a dense layer is used for output images. As a filter matrix, the kernel size was 3 x 3. As an activation function for this model ReLU was used. The size of the input images is (224, 224) for all and (299,299) only for Inception =V3. That means the height and weight of the images should be 224.

This research aims to create a different pre-trained model and compare it with my own CNN Model to do that I have chosen 4 pre-trained models one is the Google Pre-trained Model called InceptionV3, EffeceintNetB0, VGG16, and ResNet50 all of them has weighted the obtained from the imageNet large image dataset. First I have tried to train our Own CNN model but the accuracy was very small around 14% then I have experimented on hyperparameters values after many tries we have got the best value of parameters as Batch size 32, Dropout 0.20, learning rate 0.01, activation function Relu, kernel size 3X3, optimizer Adam epoch 100, there is one true while the number of epoch increase the accuracy also increase infect it time taking. Pooling layers reduce the number of parameters when the images are too large. Here MaxPooling2D of feature map matrix with stride value of 2 is applied to downsample the convolved image. We used MaxPooling2D because in practice this has been found to perform better than average pooling for image classification. It results in a downsampled or pooled feature map highlighting the most relevant feature in the patch. To find the probabilistic value, a flattening layer is used to convert the three-dimensionalities of an image to a single one, followed by two fully connected dense layers containing a SoftMax activation function for the highest likelihood classification.

The parameters used for compiling the CNN model are optimizer, loss, and metrics. As an optimizer, 'adam' has been used. It is a successful optimizer during the entire training to adjust the learning rate. For the loss function,' Categorical cross-entropy is used. The 'accuracy' metric is used to see the accuracy score on the validation set during the training of the model to make it even easier to interpret.

The 'fit ()' feature was used to train the CNN model. Since it is the most widely used in CNN applications, as we have described earlier we have used the 80/20 training/test ratio. The time () function is used to measure the time needed for the training. We have set the batch size and number of epochs in the fit function, which will loop the model through the data. The testing was processed after the training process had been completed to check the effectiveness of the trained CNN model.

We have used our own CNN Architecture and four pre-trained models were used. Namely InceptionV3, VGG16, EffecientNetB0 and ResNet50.

In the beginning, I have divided my dataset into 80% training and 20% test set then I have uploaded it into Google Drive to Fitch them easily then I have created the first CNN model and the accuracy was just only 14%.

Based on the value of the above parameter I have trained again the model then I obtained a very good result around 81.1% training accuracy. For the pre-trained models most of the parameters are already fine-tuned so I don't need to alter or change values because I am of using this pre-trained model to use their fined tuned layer, parameters value with weights they have obtained from the ImageNet.

The first pre-trained model we have trained is VGG16 which has around a 16 layer convolutional neural network to carry out the task. Image for the VGG16 transfer learning has been given as input to predict the object through Google drive in Google Colab.

To train this VGG16 Model I have used all trained layers and just train the last 3 dense layers from the network architecture also among 14,840,133 Total Parameters I have trained only 125,445 which is 14,714,688 parameters are not trained for two reasons

- To use the transfer learning approaches
- We don't have this many big machines to train these all parameters.

And the trained took around 4 hours and 11 minutes then we obtained 87.6% training accuracy.

The InceptionV3 uses 42 Layers and the concept For the InceptionV3 also followed the same approaches we have trained last 3 dense layers among 42 Total layers and the concept of depthwise convolution method followed by a pointwise convolutional method to carry out the process. InceptionV3 model was invoked using Keras framework Image for the Inception transfer learning has been given as input to predict the object through Google drive in Google Colab. And among 22,458,149 Total Parameters, we have trained only 655,365 which is 21,802,784 to implement the concept of Transfer learning and the trained took a little bit long time relatively with the VGG16 because this model has a large size after the 5 hours and 32 minutes trained time we have obtained Very Good accuracy around 90.1%.

As a third model, we have used the EffeceintNetB0 Pre-trained model which was created by Google Company like Inception version 3 so the rest of the process is the same which VGG16 and Inception V3 but have trained few of the last dense layers and we obtained 85.7% training accuracy.

Finally, I have used the ResNet50 which is the most very large pre-trained I have used in this work the why I need to experiment on is to see the effect of large CNN architecture on small dataset like us so the rest of things keeps uniformly with others pre-trained model I have used unless we have used 23,534,592 parameters to train our data among 23,587,712 total of parameters ResNet50 has already, the result is so amazing we obtained only 67.5% training accuracy and it reported as the smallest accuracy in this work.

Table 8 Summery Table of Model Performance

Sr No	Models	Trainable Parameters	Accuracy
1	Our Own CNN Architecture	139,627,685	81.1%
	InceptionV3	655,365	90.1%
3	VGG16	125,445	87.6%
	EffeceintNetB0	325,445,21	85.7%
5	ResNet50	23,534,592	67.5%

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

6.1 Conclusion

Respiratory diseases cause a huge worldwide health burden, hundreds of millions of people suffer and, more than 1 million persons suffer from chronic respiratory conditions. At least 2 billion people are exposed to the toxic effects of biomass fuel consumption, 1 billion are exposed to outdoor air pollution and 1 billion are exposed to tobacco smoke. Each year, 4 million people die prematurely from chronic respiratory disease. As we know the risk of Chronic Obstructive Pulmonary Disease and applications of digital image processing, we proposed a Chronic Obstructive Pulmonary Disease classification model using different 4 Types of the art of science Pre-Trained Model which will use the Deep Convolutional Neural Network and image processing algorithms. In our work, we followed design science methodology, which follows its scientific procedures as starting from collecting the required data set to test the developed model. We have collected about 2248 images having 350 Images or more for each class. We have applied different image preprocessing tasks to enhance the image. And augmentation is applied to increase the number of images to a total of 2248 with approx. Therefore, to overcome that problem, we applied zooming, rotation, and flipping at a different angle as augmentation techniques. Then Features are extracted from gray-level images using a CNN feature extraction model then after we have extracted the features, the classification model is built using 4 Different Pre-trained models called InceptionV3, VGG16, EffeceintNetB0, and Resnet50 including our own CNN model.

The common structure of a CNN for image classification has two main parts: a long chain of convolutional layers for feature extraction, and a few layers of the fully connected neural network for classification. For the feature extraction step, CNN with a sequential model with multiple layers is used. It consists of seven layers of Conv2D with ReLu, Max-pooling2D, and fully connected layers. Then to classify Chronic Obstructive Pulmonary diseases, the extracted features are fed to dense layers.

Features in the preprocessed dataset are extracted to reduce their dimension and use other Classification models like VGG16, Resnet50, InceptionV3, and EffeceintNetB0 on it. The convolutional neural network architecture with the sequential model is implemented with many layers such as convolutional, activation, max-pooling to extract important features from the

Chronic Obstructive Pulmonary diseases x-ray image. Keras with the TensorFlow backend system helps to construct this model layer by layer. The Conv2D layer is used for input images as 2-dimensional matrices and a dense layer is used for output images. As a filter matrix, the kernel size was 3 x 3. As an activation function for this model, Tanh, ReLU, Leaky ReLU, and swish were used. The size of the input images is (224, 224). That means the height and weight of the images should be 224.

Pooling layers reduce the number of parameters when the images are too large. Here MaxPooling2D of feature map matrix with stride value of 2 is applied to downsample the convolved image. We used MaxPooling2D because in practice this has been found to perform better than average pooling for image classification. It results in a downsampled or pooled feature map highlighting the most relevant feature in the patch. To find the probabilistic value, a flattening layer is used to convert the three-dimensionalities of an image to a single one, followed by two fully connected dense layers containing a SoftMax activation function for the highest likelihood classification.

The parameters used for compiling the CNN model are optimizer, loss, and metrics. As an optimizer, 'adam' has been used. It is a successful optimizer during the entire training to adjust the learning rate. For the loss function,' Categorical cross-entropy is used. The 'accuracy' metric is used to see the accuracy score on the validation set during the training of the model to make it even easier to interpret.

We have used our own CNN Architecture and additional four pre-trained models were used. Namely InceptionV3, VGG16, EffecientNetB0 and ResNet50. The InceptionV3 uses 42 Layers and the concept of depthwise convolution method followed by a pointwise convolutional method to carry out the process. InceptionV3 model was invoked using the Keras framework. Image for the Inception transfer learning has been given as input to predict the object through Google drive in Google Colab. InceptionV3 uses only 16MB of disk space and accuracy varies between 88.5 percent for the InceptionV3 data set.

The VGG16 uses a 16 layer convolutional neural network to carry out the task. VGG16 model uses the input target size 224x224. Image for the VGG16 transfer learning has been given as input to predict the object through Google drive in Google Colab. VGG16 uses only 528MB of disk space and the accuracy was 87 percent.

Experimental results show that the InceptionV3 with its filtering mechanism has achieved a better classification performance with an accuracy of 90.1%.

6.2 Contribution

The following is the research's contribution:

Technical Contribution: we have investigated and developed 5 COPD classification models while 4 of them are based upon pre-trained models namely Visual geometry group (VGG16), Residual neural network (Resnet50), InceptionV3, EffeceintNetB0, and our own CNN Model and the implementation source code to replicate the experiments are available on GitHub.

Dataset: during the investigation, we have collected 2248 chest x-ray images from 5 different health facilities that we made publicly available. This dataset is very useful due to the lack of directories for medical images.

Scientific Contribution: as a scientific contribution, there was no previous research that is done on the classification of subtypes of Chronic Obstructive Pulmonary Disease (COPD) and we are the first one working on this.

6.3 Challenges

The major challenge for this study was the fact that algorithms learn specific patterns based on decision reports made by radiologists. The algorithm is expected to reach an accuracy of a hundred percent (100%) compared with the radiologists. Whether the model can perform well alone still needs to be investigated in the future.

Significant amounts of radiological images are being generated at hospitals. Even then, many of these images are not used for further training of machine learning or deep learning algorithms, as the training process restricts the resources available. It was challenging, therefore, to find a pre-trained model on a medical image. Moreover, X-ray images contain rich features that are clinically important. It is challenging to apply directly the x-ray raw data (DICOM) to the DCNN.

Even though transfer learning has immense potential and is a commonly required enhancement for existing learning algorithms. There are issues related to transfer learning that needs more research and exploration. New architectures need to be implemented to build networks that are invariant to these properties. While the central innovation of deep learning is its ability to learn useful features directly from data. However, its accuracy and efficiency are highly restricted by the data size. One potential solution to achieve this superior performance is to develop better models of deep learning. Either to collect more representative data that can be used continuously to enhance algorithms.

6.4 Recommendation

There is still a gap in this proposed system. To increase the performance, the proposed work can be expanded. As a result, when doing this research, the following are some noteworthy future work recommendations:

- The amount of data used in this study is not much enough. If researchers are solely working with large size deep learning architectures like effeceintNetB7 and Resnet150, it is preferable to expand the dataset because this kind of deep learning requires a large amount of data.
- Develop a system that can classify all chest x-rays.
- To improve the performance of the given method, use different Optimization approaches. Because Optimizing (Mats behind the model) them model is can be novel since I's new way and improve the performance of a given model.
- To have high Speed we advise that using GPU (graphical Processing Unit) minimum of 16 GB and TPU (tensor processing unit) from Google or Azure to train the model using more than epoch we have used.
- Applying Re-enforcement Techniques.
- Optimizing the Logic or Math's Behind The CNN model

6.4. Reference

[1] S. European Respiratory Society, Forum of International Respiratory Societies. Respiratory Diseases in the World. Realities of Today – Opportunities for Tomorrow. 2013.

[2] Helmberger M., Urschler M., Pienn M., Balint Z. “anatomy of lung.”, ResearchGate, 2013.

[3] A. N. Zakirov, R. F. Kuleev, A. S. Timoshenko, and A. V Vladimirov, Advanced Approaches to Computer-Aided Detection of Thoracic Diseases on Chest X-Rays. vol. 9, no. 88, pp. 4361–4369, 2015.

[4] A. Karargyris et al., Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays. Int. J. Comput. Assist. Radiol. Surg., vol. 11, no. 1, pp. 99–106, 2016.

[5] S. Pattar, Detection and Classification of Lung Disease – Pneumonia and Lung Cancer in Chest Radiology Using Artificial Neural Network, Int. J. Sci. Res. Publ., vol. 5, no. 1, pp. 2250–3153, 2015.

[6] E. Pune and V. I. T. Pune, Automatie Deteetion ofMajor Lung Diseases U sing Chest Radiographs and Classifieation by Feed-forward Artifieial Neural Network, pp. 1–5, 2016.

[7] G. P. Melendez and M. Cordel, Texture-based detection of lung pathology in chest radiographs using local binary patterns, BMEiCON 2015 - 8th Biomed. Eng. Int. Conf., 2016.

[8] L. Shen and R. Song, Semi-supervised Learning for Multi-label Classification.

[9] C. Tataru, D. Yi, A. Shenoyas, and A. Ma, Deep Learning for abnormality detection in Chest X-Ray images, 2017.

[10] M. Haloi, K. R. Rajalakshmi, and P. Walia, Towards Radiologist-Level Accurate Deep Learning System for Pulmonary Screening, 2018.

[11] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, Chest pathology identification using deep feature selection with non-medical training,Comput. Methods Biomech. Biomed. Eng. Imaging Vis., vol. 6, no. 3, pp. 259–263, 2018.

[12] C. Brestel and M. Cohen-sfaty, RadBot-CXR : Classification of Four Clinical Finding Categories in Chest X-Ray Using Deep Learning,Med. Imaging with Deep Learn., no. Midl 2018, pp. 1–9, 2018.

[13] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017–Janua, pp. 3462–3471, 2017.

[14] H. Wang and Y. Xia, ChestNet: A Deep Neural Network for Classification of Thoracic Diseases on Chest Radiography, pp. 1–8, 2018.

[15], Mayo Clinic Staff, June 2021, https://www.mayoclinic.org/diseases-conditions/copd/symptoms-causes/syc-20353679

[16] WHO Global Health Estimates 21 June 2021, https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)

[17]. Abegunde DO, Mathers CD, Adam T, Ortegon M, Strong K. The burden and costs of chronic diseases in low-income and middle-income countries. Lancet. 2007;370:1929–38. [PubMed] [Google Scholar]

[18] Tesfaye F, Byass P, Wall S. Population based prevalence of high blood pressure among adults in Addis Ababa: uncovering a silent epidemic. BMC Cardiovasc Disord. 2009;9:39. doi: 10.1186/1471-2261-9-39. [PMC free article] [PubMed] [Google Scholar]

[19] HSDP III mid-term review [homepage on the Internet]. [Cited 9/26/2009]. Available from: http://www.moh.gov.et/index.php?option=com_remository&Itemid=0&func=fileinfo&id =5

[20] Center for national health development in Ethiopia [homepage on the Internet]. [Cited 9/26/2009]. Available from: http://cnhde.ei.columbia.edu/programs/hep/index.html

[21] Ray Sipherd, June 10, 2021 https://www.cnbc.com/2018/02/22/medical-errors-third-leading-cause-of-death-in-america.html#:~:text=A%20recent%20Johns%20Hopkins%20study,after%20heart%20disease%20and%20cancer.

[22] Brady A. P. (2017). Error and discrepancy in radiology: inevitable or avoidable?. Insights into imaging, 8(1), 171–182. https://doi.org/10.1007/s13244-016-0534-1

[23] Busby LP, Courtier JL, Glastonbury CM. Bias in Radiology: The How and Why of Misses and Misinterpretations. Radiographics. 2018 Jan-Feb; 38(1):236-247. doi: 10.1148/rg.2018170107. Epub 2017 Dec 1. PMID: 29194009; PMCID: PMC5790309.

[24] Shimelis D, Atnafu A. Status of radiological services in Addis Ababa public hospitals. Ethiop Med J. 2011 Jul; 49(3):257-63. Erratum in: Ethiop Med J. 2011 Oct; 49(4):377. Tsige, Mesfin [corrected to Shimelis, Dagmawit]. PMID: 21991759.

[25] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6), 386–408. Doi: 10.1037/h0042519

[26] Haykin, S. (2009). Neural Networks and Learning Machines (3rd Ed.). Pearson Education, Inc.

[27]Hastie, Trevor. Tibshirani, Robert. Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, NY, 2009.

[28] Rosenblatt, Frank. x. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Spartan Books, Washington DC, 1961

[29] Rumelhart, David E., Geoffrey E. Hinton, and R. J. Williams. "Learning Internal Representations by Error Propagation". David E. Rumelhart, James L. McClelland, and the PDP research group. (editors), Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundation. MIT Press, 1986.

[30] Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function Mathematics of Control, Signals, and Systems, 2(4), 303–314.

[31] Bluche, T. (2015). Deep Neural Networks for Large Vocabulary Handwritten Text Recognition. PhD thesis, Université Paris-Sud. Retrieved from https://tel.archives-ouvertes.fr/tel-01249405/document

[32] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press

[33] Khan et al. (2020). A Survey of the Recent Architectures of Deep Convolutional Neural Networks. arXiv preprint. Retrieved from https://arxiv.org/abs/1901.06032v7

[34] Dumoulin, V., & Visin, F. (2018). A guide to convolution arithmetic for deep learning. arXiv preprint.

Retrieved from https://arxiv.org/abs/1603.07285v2

[35] C Santosh, S. Antani, "Automated chest x-ray screening: Can lung region symmetry help detect pulmonary abnormalities",IEEE Transactions on Medical Imaging, 37(5):1168–1177 (2018).

[36] Jaiswal, A.K.; Tiwari, P.; Kumar, S.; Gupta, D.; Khanna, A.; Rodrigues, J.J. Identifying pneumonia in chest X-rays: A deep learning approach. Measurement 2019, 145, 511–518.

[37] Jung, H.; Kim, B.; Lee, I. Classification of lung nodules in CT scans using three-dimensional deep convolutional neural networks with a checkpoint ensemble method. BMC Med. Imaging 2018, 18, 48.

[38], Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model," Sci. Rep., vol. 7, no. 1, pp. 1–10, 2017

[39] Y. Guo, H. Dong, F. Song, C. Zhu, and J. Liu, "Breast Cancer Histology Image Classification Based on Deep Neural Networks," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, vol. 10882 LNCS, pp. 827–836.

[40] A. Rakhlin, A. Shvets, V. Iglovikov, and A. A. Kalinin, "Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10882 LNCS, pp. 737–744, 2018.

[41] T. Araujo et al., "Classification of breast cancer histology images using convolutional neural networks," PLoS One, vol. 12, no. 6, pp. 1–14, 2017.

[42] A. Golatkar, D. Anand, and A. Sethi, "Classification of Breast Cancer Histology Using Deep Learning"," Santa One, vol. 8, no. 18, pp. 90–130, 2019.

[43] Ivy Wigmore, June 12, 2021, https://whatis.techtarget.com/definition/data-labeling

[44] AWS of Amazon, June 11, 2021, https://aws.amazon.com/sagemaker/groundtruth/what-is-data-labeling/

[45] O. Chapelle, B. Schlkopf, and A. Zien, Semi-supervised Learning. Cambridge: MIT Press, 2010.

[46] S. Sun, "A survey of multi-view machine learning," Neural Comput. Appl., vol. 23, no. 7–8, pp. 2031–2038, Dec. 2013.

[47] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," 2013, arXiv: 1304.5634v1.

[48] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," Inf. Fusion, vol. 38, pp. 43–54, Nov. 2017.

[50] D. Zhang, J. He, Y. Liu, L. Si, and R. Lawrence, "Multi-view transfer learning with a large margin approach," in Proc. 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, Aug. 2011, pp. 1208–1216.

[51] P. Yang and W. GAO, "Multi-view discriminant transfer learning," in Proc. 23rd International Joint Conference on Artificial Intelligence, Beijing, Aug. 2013, pp. 1848–1854

[52] K.D. Feuz and D.J. Cook, "Collegial activity learning between heterogeneous sensors," Knowl. Inf. Syst., vol. 53, pp. 337–364, Mar. 2017.

[53] Y. Zhang and Q. Yang, "An overview of multi-task learning," Natl. Sci. Rev., vol. 5, no. 1, pp. 30–43, Jan. 2018.

[54] W. Zhang, R. Li, T. Zeng, Q. Sun, S. Kumar, J. Ye, and S. Ji, "Deep model based transfer and multi-task learning for biological image analysis," in Proc. 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Aug. 2015, pp. 1475– 1484.

[55] A. Liu, N. Xu, W. Nie, Y. Su, and Y. Zhang, "Multi-domain and multi-task learning for human action recognition," IEEE Trans. Image Process., vol. 28, no. 2, pp. 853–867, Feb. 2019

[56] H.A. Minassie, "Image analysis for Ethiopian coffee classification," Addis Ababa University, Addis Ababa,Etthiopia, 2008.

[57] R. Gonzalez and R. Woods, "Digital Image Processing, Second Edition, Pearson Education," 2002.

[58] Siebenmorgen and L. Tao, "Digital Image Analysis Method for Rapid Measurement of Rice Degree of Milling," 2000

[59] Bruno MA, Walker EA, Abujudeh HH (2015) Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. Radiographics 35:1668–1676 2.

[60] Royal College of Radiologists (2006) Standards for the reporting and interpretation of imaging investigations. RCR, London 3.

[61] Robinson PJA (1997) Radiology's Achilles' heel: error and variation in the interpretation of the Röntgen image. BJR 70:1085–1098

[62], DeepAI, June 12, 2021 https://deepai.org/machine-learning-glossary-and-terms/epoch

[63], Jason Brownleeon July 20, 2021, https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/

[64], Jason Brownlee, June 21, 2021 https://machinelearningmastery.com/how-to-control-the-speed-and-stability-of-training-neural-networks-with-gradient-descent-batch-size/

[65], Yimer, Fethyaseid & Tessema , Abel & Simegn , Gizeaddis. (2021). Multiple Lung Diseases Classification from Chest X-Ray Images using Deep Learning approach. International Journal of Advanced Trends in Computer Science and Engineering.

[66], Adeloye D, Basquill C, Papana A, Chan KY, Rudan I, Campbell H. An estimate of the prevalence of COPD in Africa: a systematic analysis. Journal of Chronic Obstructive Pulmonary Disease. 2015; 12(1):71–81. pmid:24946179

[67], Misganaw A, Mariam DH, Ali A, Araya T. Epidemiology of major non-communicable diseases in Ethiopia: a systematic review. J Health Popul Nutr. 2014; 32(1):1–13. pmid:24847587

Appendix A: Sample Dataset (Chest X-ray)

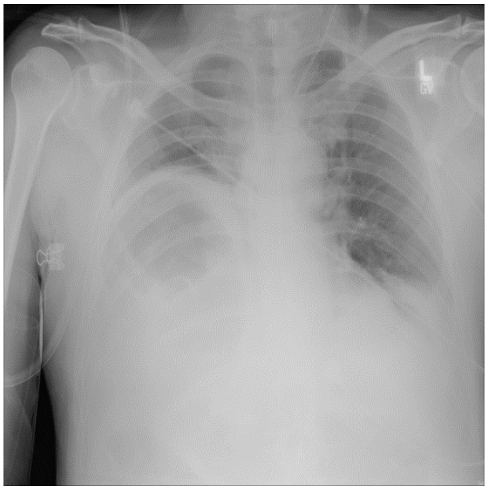


Figure 56 Sample of Asthma Chest X-ray Image

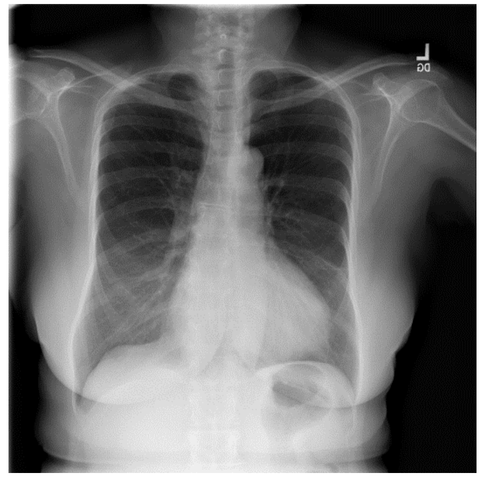


Figure 57Sample of emphysema Chest x-ray image

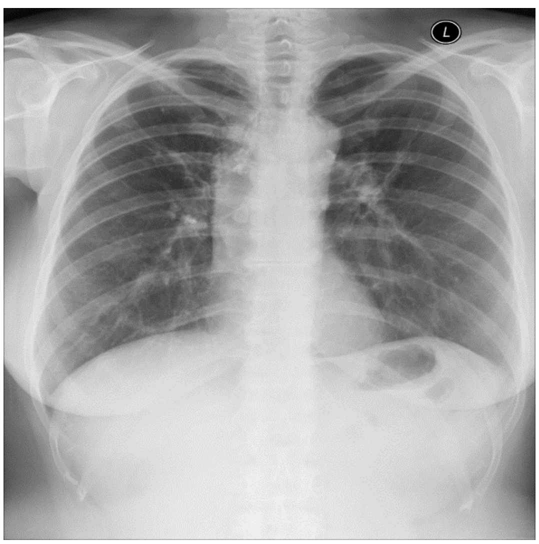


Figure 59Sample of chronic bronchitis Chest X-ray Image

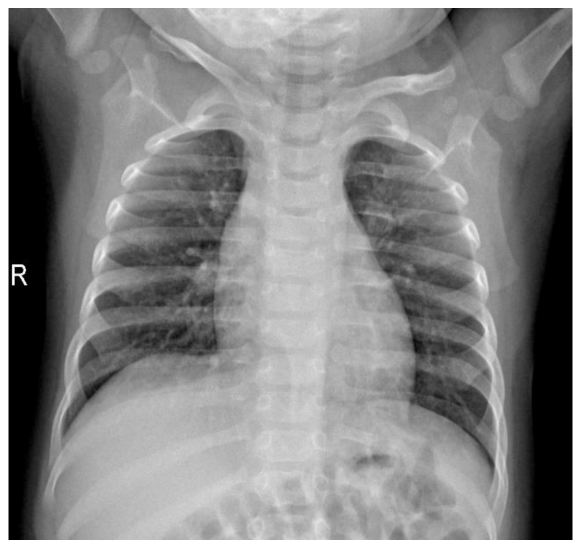


Figure 58Sample of Normal Chest X-ray Image

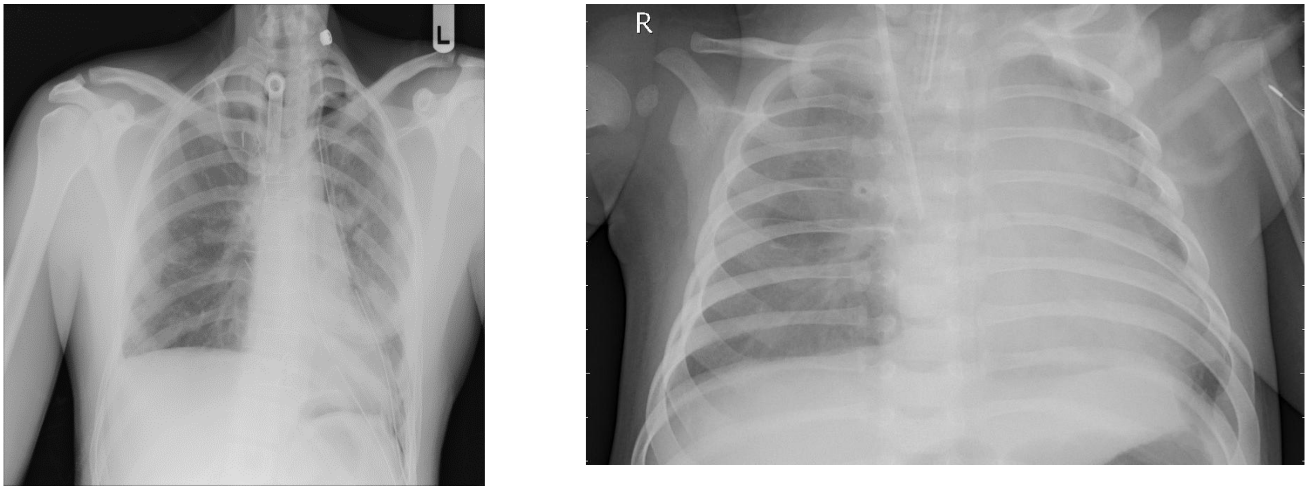


Figure 60Sample of Others (Non COPD) Disease Chest X-ray Images

Appendix B: Sample User Interface

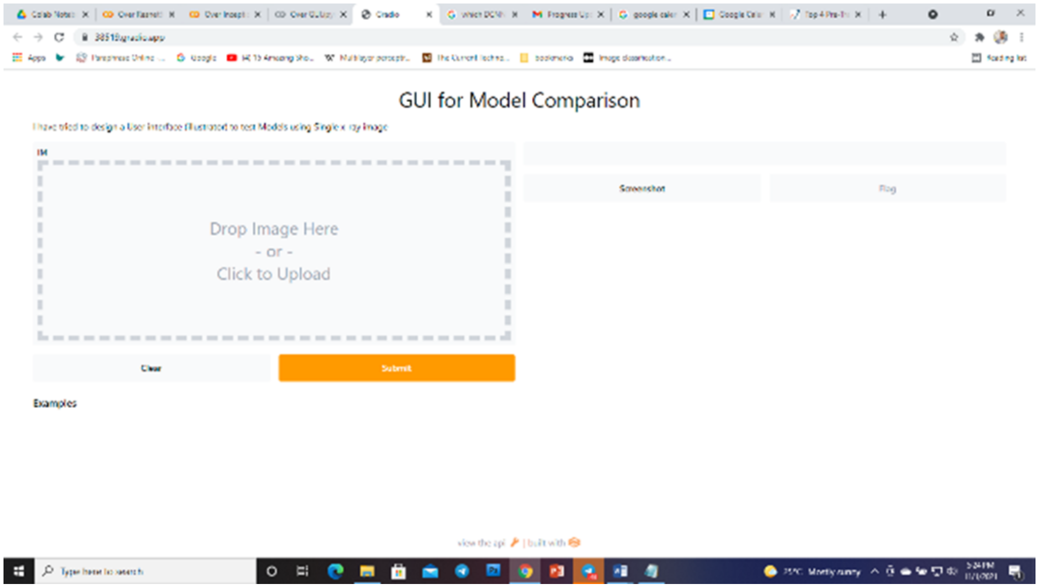


Figure 61 Front End of the System

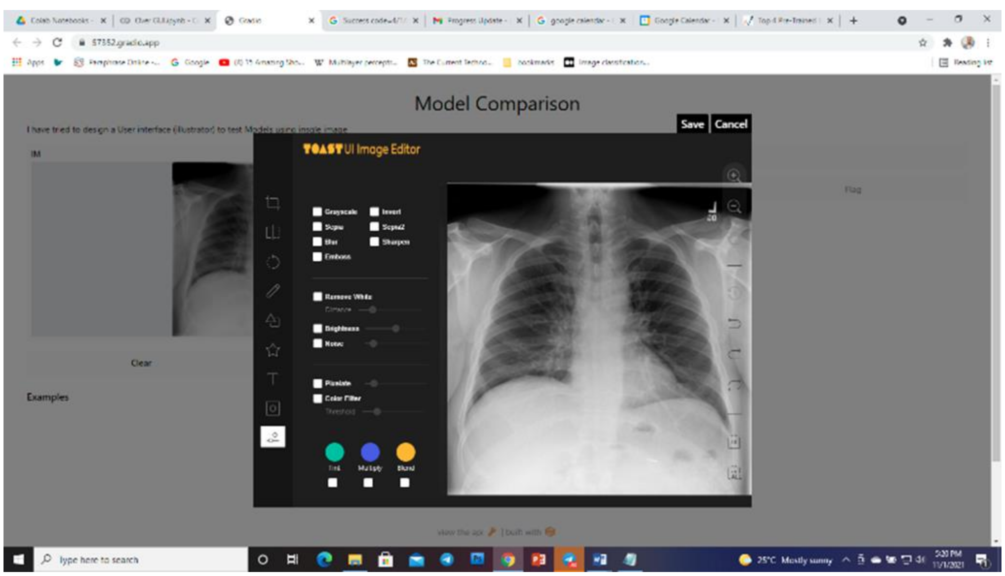
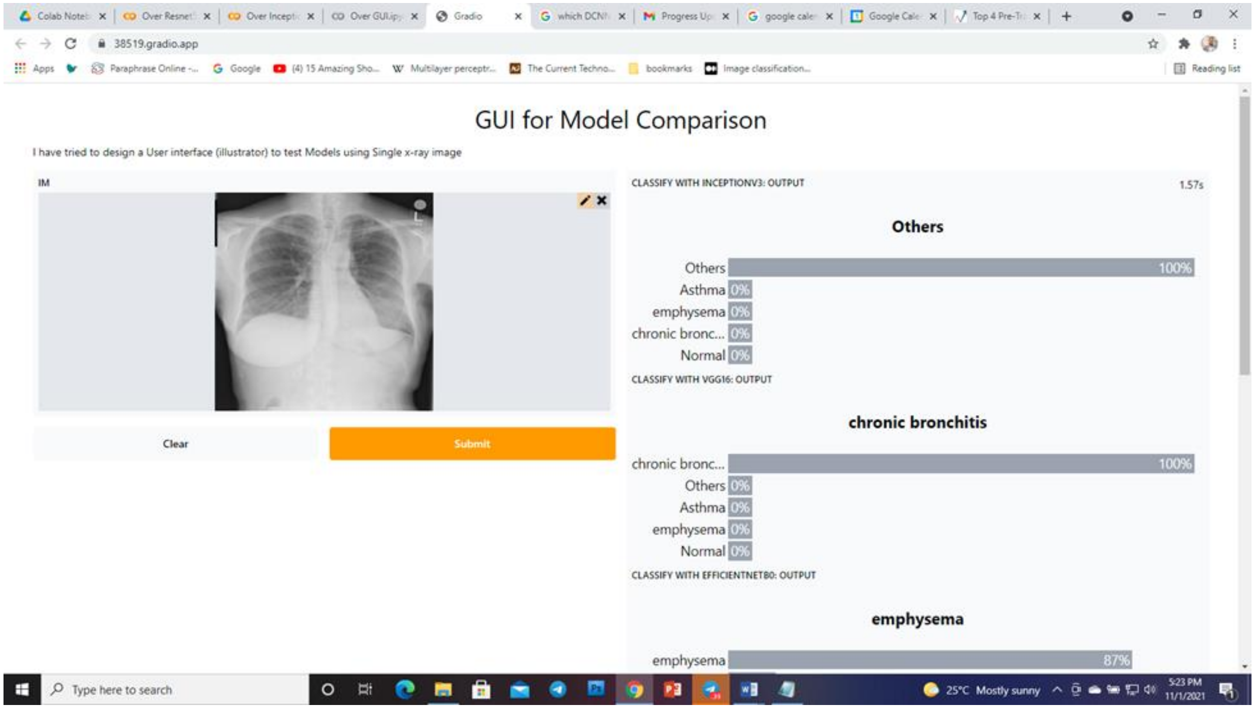


Figure 62 Pre-Built In Tools



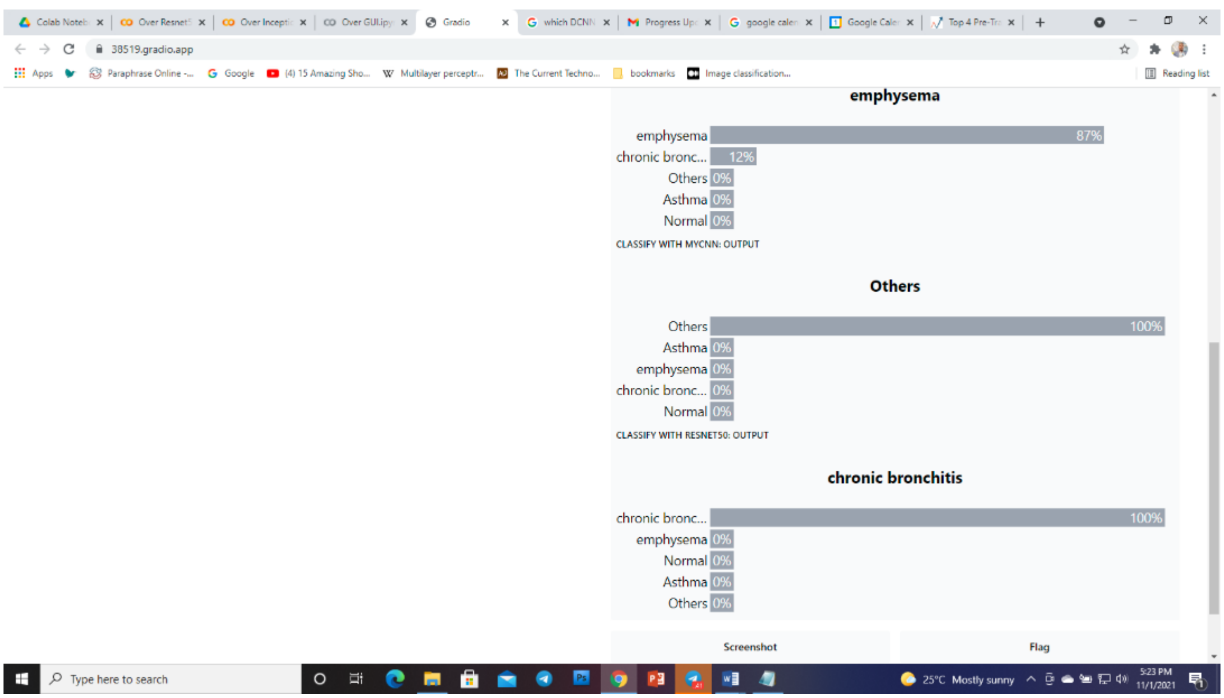


Figure 63 Model Prediction and Result comparison

Appendix C: Plagiarism Report

Plagiarism Checker X Originality Report

Plagiarism Quantity: 14% Duplicate

Date	Monday, November 22, 2021
Words	3231 Plagiarized Words / Total 22376 Words
Sources	More than 214 Sources Identified.
Remarks	Low Plagiarism Detected - Your Document needs Optional Improvement.

Appendix D: Source Code and Dataset Access Link

Our Own CNN Model Source Code Link: https://github.com/Amanuel-Meseret/OHBD-Project/blob/master/Copy_of_Over_My_CNN_Arch.ipynb

VGG16 Model Source Code Link: https://github.com/Amanuel-Meseret/OHBD-Project/blob/master/Over_RESENET50.ipynb

InceptionV3 Model Source Code Link: https://github.com/Amanuel-Meseret/OHBD-Project/blob/master/New_Over_Inception_V3.ipynb

EffeceinNetB0 Model source Code Link: https://github.com/Amanuel-Meseret/OHBD-Project/blob/master/Over_EfficientNetB0.ipynb

ResNet50 Model Source Code Link: https://github.com/Amanuel-Meseret/OHBD-Project/blob/master/Over_VGG16.ipynb

User Interface Source Code Link: https://github.com/Amanuel-Meseret/OHBD-Project/blob/master/Over_GUI.ipynb

Dataset access Link: https://drive.google.com/drive/folders/1f-oLuMJTIWzTKD5y8jbZ99PH14aOo5m0?usp=sharing

The End