

DETECTION OF COVID-19 FROM CHEST X-RAY IMAGES

Submitted in partial fulfilment of the requirements for the award of Bachelor of Engineering in Electronics and Communication.

By

CHONGA VEERA RAMA LAKSHMI,

38130048.



DEPARTMENT ELECTRONIC AND COMMUNICATION ENGINEERING

SCHOOL OF ELECTRICAL AND ELECTRONIC ENGINEERING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI – 600 119.

April – 2022.



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **CHONGA VEERA RAMA LAKSHMI (Reg. No: 38130048)** who carried out the project entitled "**DETECTION OF COVID-19 FROM CHEST X-RAY IMAGES**" under my supervision from June 2021 to November 2021.

Internal Guide

Dr.P. Kavi Priya,M.E.,Ph.d.,

Head of the Department

Dr. T. RAVI, M.E., Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I **Chonga Veera Rama Lakshmi** hereby declare that the Project Report entitled **DETECTION OF COVID-19 FROM CHEST X-RAY IMAGES** done by me under the guidance of **Dr.P. Kavi Priya** (Internal) Department of ECE and **Dr.Emalda Rosilin** (External) Department of ECE at **Sathyabama Institute of Science and Technology** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Sathyabama Institute of Science and Technology**.

Date:09-03-2022,

Place: Korukonda.

Ch.V.Ramalakshmi

Signature

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of SATHYABAMA for their kind encouragement in doing this project and for completing it successfully. I am grateful to them. I convey my thanks to **Dr. M.Nandhitha.**, Dean, School of Electrical and Electronics Engineering and **Dr. T.Ravi**, Head of the Department, Dept. of Electronics and Communication Engineering for providing me necessary support and details at the right time during the progressive reviews. I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.P.Kavi Priya** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work. I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Electronics and Communication Engineering who were helpful in many ways for the completion of the project.

ABSTRACT

COVID-19 global pandemic affects health care and lifestyle worldwide, and its early detection is critical to control cases' spreading and mortality. The actual leader diagnosis test is the Reverse Transcription Polymerase chain reaction (RT-PCR), result times and cost of these tests are high, so other fast and accessible diagnostic tools are needed. Inspired by recent research that correlates the presence of COVID-19 to findings in Chest X-ray images, this papers' approach uses existing deep learning models to process these images and classify them as positive or negative for COVID-19. The proposed system involves a pre-processing stage with lung segmentation, removing the surroundings which does not offer relevant information for the task and may produce biased results; after this initial stage comes the classification model trained under the transfer learning scheme; and finally, results analysis and interpretation via heat maps visualization. This best model achieved a detection accuracy of COVID-19 around 95%.

CONTENTS

CHAPTER NO.	PARTICULAR	PAGE NO.
Chapter 1	Introduction	1-3
Chapter 2	Literature survey	4-6
	Aim and Scope of present investigation	
3.1	Project proposal	
3.2	Mission	
3.3	Goal	
	Existing system	
3.4	3.4.1 Methods	7-11
	3.4.2 Disadvantages	
Chapter 3	3.5 Preparing the dataset	
	Methodology	
	3.6.1 Exploratory Data Analysis of X-Ray Images	
	3.6.2 Data Wrangling	
3.6	3.6.3 Data collection	
	3.6.4 Building the classification model	
	3.6.5 Advantages	
	Experimental methods and algorithms used	
	System study	
4.1	4.1.1 System requirement specifications	
	System Specifications	
4.2	4.2.1 Machine learning overview	
	4.2.2 Gradio Overview	
Chapter 4	4.3 Steps to download and install Python	
	4.3.1 IDE Installation for python	
	4.3.2 Python file creation	
	Python libraries needed	12-35
	4.4.1	
4.4	Numpy	
	4.4.2 Pandas	
	4.4.3 Matplotlib	

	4.4.4	Tensorflow	
	4.4.5	Gradio	
		Module details	
	4.6	4.6.1 Data Pre-processing	
		4.6.2 Data Validation/ Cleaning/Preparing Process	
		4.6.3 Exploration data analysis of visualization	
Chapter-4	4.6.4	Comparing Algorithm with prediction in the form of best accuracy result	12-35
		4.6.5 Algorithms and techniques	
		CNN-Convolution Neural Networks	
	4.7	4.7.1 Convolution layer-the kernel	
		4.7.2 Classification-Fully connected layer	
		Results and discussions, Performance analysis	
Chapter -5	5.1	Results	36-40
	5.2	Discussions	
	5.3	Performance analysis and results	
		Conclusion	
Chapter-6	6.1	Conclusion	41
	6.2	Future Work	
		References	42

LIST OF FIGURES

FIGURE NO	TITLE	PAGE NO
3.1	Existing system block diagram	8
3.2	Covid Images	10
3.3	Normal Images	10
3.4	Proposed system block diagram	11
4.1	Module Diagram 1	18
4.2	Module Diagram 2	19
4.3	Class Diagram	22
4.4	Convolution neural network systematic diagram	25
4.5	4x4x3 RGB image	26
4.6	CNN Kernel	27
4.7	Movement of Kernel	28
4.8	Convolution operation on a MxNx3 image matrix with a 3x3x3 Kernel	29
4.9	Convolution Operation with Stride Length = 2	30
4.10	SAME padding: 5x5x1 image is padded with 0s to create a 6x6x1 image	30
4.11	Types of Pooling	32
4.12	Fully connected layer in CNN	33
4.13	CNN Threshold	34
4.14	Detection mask through the trained model in an image.	35
5.1	Snapshot of Accuracy	37
5.2	Output Snapshots	39&40

List of Tables

5.1	Accuracy table	38
-----	----------------	----

List of Abbreviations:

- AI-Artificial Intelligence.
- Relu-Rectified Linear Unit.
- CNNs-Convolution Neural Networks.
- RTPCR-Real Time Polymerase Chain Reaction.
- MCDA-Multi Criteria Decision Analysis.

CHAPTER 1

INTRODUCTION

The World Health Organization (WHO) declared the spread of the coronavirus infection a pandemic in March 2020, which is called the coronavirus pandemic or COVID-19 pandemic. The coronavirus pandemic is caused by severe acute respiratory syndrome coronavirus. The outbreak originally started in Wuhan, China, and later spread to every country in the world. The coronavirus spreads through respiratory droplets of the infected person that are produced through cough or sneeze.

These droplets can further contaminate the surfaces increasing the spread. Coronavirus-infected persons may suffer from mild to severe respiratory illness and may require ventilation support. Older people and people with chronological disorders are easily prone to coronavirus infection. Thus, many governments have closed their borders and locked down people to break the cycle and prevent the spread of the pandemic.

With the sequencing of ribonucleic acid (RNA) from the coronavirus, many vaccines are being developed worldwide. The developed vaccines use both traditional and next-generation technology with six vaccine platforms, namely, live attenuated virus, inactivated virus, protein or subunit, viral vector-based, messenger RNA (mRNA), and deoxyribonucleic acid (DNA).

Although vaccines can reduce the rapid spread and facilitate the development of immunity via the production of suitable antibodies, the efficacy of the vaccines is still 95%. Many issues are encountered in administering the vaccine, such as supply chain logistical challenges, vaccine hesitancy, and vaccine complacency. A vaccine is a prevention measure rather than a cure.

Even with the availability of the vaccine, early detection of the coronavirus is important, as it can facilitate tracing of the people who were in contact directly and indirectly. By

tracing these people, further spread of the pandemic can be avoided. COVID-19 infection manifests as lung infection, and computed tomography (CT) and chest X-ray (CXR) images are primarily used in the detection of lung infection of any type.

Along with doctors and clinical personnel, researchers and technologists are focusing their efforts on early detection of coronavirus infections. According to PubMed, 755 academic articles were published with the search term “coronavirus” in 2019, and this number rose to 1245 in the first 80 days of 2020.

Artificial intelligence and deep learning methods are the most commonly used methods by researchers for the detection of coronavirus infection from CT and CXR images. Deep learning methods have shown significant performance in many research applications, such as computer vision, object tracking, gesture recognition, face recognition, and steganography.

Deep learning methods are widely used because of their improved performance compared to traditional methods.

In contrast to traditional methods and machine learning methods, the features need not be hand-picked. By changing the parameters and configurations of the deep learning convolutional neural network (CNN) architecture, a model can be trained to learn the best possible features for the dataset in use.

Researchers have used deep learning methods to explore the field of medical imaging even before the coronavirus pandemic. With the recent pandemic, the use of deep learning methods for the detection of coronavirus infection from images has increased tremendously.

A detailed survey of the available deep learning approaches for the detection of coronavirus infection from images such as CT scans or CXR images is conducted in this report.

Although other surveys are available in the literature, most of them cover a wider scope. For example, Ulhaq et al. surveyed all methods that address coronaviruses, such as medical image processing, data science methods for pandemic modelling, AI and the Internet of things (IoT), AI for text mining and natural language processing (NLP), and AI in computational biology and medicine.

This provides an overall view of what is happening in the research world. A survey on the application of computer vision methods for COVID-19 described the segmentation of lung images.

This report aims to exclusively describe coronavirus detection methods using deep learning methods. In the hope of helping researchers develop better coronavirus detection methods, this report summarizes all the methods that have been reported in the literature.

Along with the methods, the used datasets, commonly used metrics for evaluation and comparison are discussed and future direction are elaborated in this report.

CHAPTER 2

LITERATURE SURVEY

In recent months, researchers have investigated and analyzed chest X-ray images using deep learning algorithms to detect COVID-19. First, the images are pre-processed using the CNN technique for extracting better features, which are fed in deep learning algorithms for image classification.

- [1] Ahamed et al. proposed a deep neural network-based system where CNN provided high accuracy (94.03%). The authors trained the system with normal, pneumonia and COVID-19 patient's chest X-ray images. The limitation of the work was that a dataset with only 285 images was used for developing the system, and this small number of data was not perfect for training a deep learning-based system for the COVID-19 prediction.

- [2] Chowdhury et al. worked with chest X-ray images to develop a novel framework named PDCOVIDNet based on parallel-dilated CNN. In the proposed method, the authors used a dilated convolution in the parallel stack that could capture and stretch necessary features for obtaining a detection accuracy of 96.58%.

- [3] Azemin et al. used a deep learning method based on the ResNet-101 CNN model. In their proposed method, thousands of images were used in the pre-trained phase to recognize meaningful objects and retrained to detect abnormality in the chest X-ray images. The accuracy of this method was only 71.9%.

- [4] El-Rashidy et al. introduced a framework consisted of three layers: patient layer, cloud layer and hospital layer. A set of data was collected from the patient layer using some wearable sensors and a mobile app. A neural network-based deep learning model was used to detect COVID-19 using the patient X-ray images. The proposed model achieved 97.9% accuracy and 98.85% specificity.

- [5] Khan et al. developed a new architecture for the diagnosis of X-ray images as the COVID-19 or normal using pre-trained deep learning models like ResNet50, VGG16, VGG19 and DensNet121, where VGG16 and VGG19 showed the best accuracies. The proposed model consisted of two phases like pre-processing and data augmentation, and transfer learning, and finally showed 99.3% accuracy.

- [6] In the proposed model by Loey et al. three deep transfer models like Alex Net, Google Net and ResNet18 were employed on a dataset of 307 images with four different types of classes: COVID-19, normal, pneumonia bacterial and pneumonia virus. The research work was distributed into three scenarios to reduce memory consumption and execution time. At the last deep transfer model, achieved 100% testing accuracy and 99.9% validation accuracy.
- [7] Minaee et al. reported a deep learning-based framework to detect COVID-19 from chest X-ray images using four tuning models like ResNet18, ResNet50, Squeeze Net and DensNet-121. The proposed method took advantage of data augmentation to create a transformed version of the COVID-19 images, which increased the number of samples and finally achieved 98% sensitivity and 90% specificity.
- [8] Sekeroglu et al. developed a model using deep learning and machine learning classifiers where a total of 38 experiments was conducted by CNN for the detection of the COVID-19 using the chest X-ray images with high accuracy. Among them, 10 experiments were performed using 5 different machine-learning algorithms, and 14 experiments were carried out by the state-of-the-art pre-trained network for transfer learning.
The system demonstrated 98.50% accuracy, 99.18% specificity and 93.84% sensitivity. They concluded that the system developed by CNN was capable of achieving COVID-19 detection from a limited number of images without any pre-processing and with minimized layers.
- [9] Wang et al. developed a model using ResNet-101 and ResNet-151 with fusion effects to enhance their weight ratio dynamically. Classification of the chest X-ray images were carried out based on three classes, such as normal, COVID-19 and viral pneumonia. Performance accuracy of 96.1% was achieved during the testing phase.
- [10] Yoo et al. applied chest X-ray radiography (CXR) images to classify using a deep learning-based decision-tree classifier for detecting COVID-19. This classifier compared three binary decision trees based on the PyTorch frame. The decision tree classified CXR images as normal or abnormal, where the third decision tree achieved an average accuracy of 95%.
- [11] Khalifa et al. developed a classification approach for the treatment purposes of coronavirus on a single human cell-based on treatment type and treatment concentration level using deep learning and machine learning (ML) methods.

Numerical features of the data sets were converted to images for building the DCNN model.

The testing accuracy of treatment classification obtained by the model was as high as 98.05% compared to the other traditional ML methods, including support vector machine (SVM) and decision tree (DT).

However, the proposed DCNN model showed less testing accuracy (98.2%) compared to the DT (98.5%) for the prediction of treatment concentration level. Deep transfer models (i.e., Alexnet) have not been employed in their study.

[12] Wang et al. have developed a transfer learning method (Xception model) using deep learning models for diagnosing COVID-19. The proposed method showed 96.75% diagnostics accuracy.

Furthermore, Deep features and machine learning classification (Exception + SVM) were also employed to develop an efficient diagnostic method for improving the accuracy of the Exception model by 2.58%. From the result, the authors claimed that their proposed method attained higher classification accuracy and efficient diagnostic performance of the COVID-19. However, the authors have not compared them results with the existing similar works.

[13] Sahlol et al. proposed an improved hybrid classification approach using CNNs and marine predators' algorithm for classifying COVID-19 images, which were obtained from international cardiothoracic radiologists. Inception architecture of

CNNs was employed to extract features, and a swarm-based marine predators' algorithm was used to select the most relevant features from the images.

However, the research work did not consider any fusion approach to improve the classification and feature extraction of the COVID-19 images.

Most of the reported work in the literature has used chest X-ray images to diagnose COVID-19, and this highlights the importance of chest X-ray image analysis as an indisputable tool for doctors and radiographers. However, imbalance in data manipulation and lack of necessary extracted features from the images sometimes cannot provide expected accuracy in the classification result. To overcome these limitations, this work proposed fusion of features extracted by HOG and CNN and classify using CNN for improving the detection accuracy of the COVID-19.

CHAPTER 3

AIM AND SCOPE OF PRESENT INVESTIGATION

3.1 PROJECT PROPOSAL:

The project proposal is the term of documents. A project can describe the project proposal. It is the set of all plans of a project. Like, how the software works, what are the steps to complete the entire projects, and what are the software requirements and analysis for this project. In my project, I am doing all the steps and also risk and reward and other project dependencies in the project proposal.

3.2 Mission:

Using X-ray images for the detection of COVID-19 will be helpful especially for countries that are unable to purchase the kits for testing and to the countries those who don't have CT scanners.

3.3 Goal:

The goal is to develop a machine learning model for predicting the COVID-19 disease.

3.4 EXISTING SYSTEM:

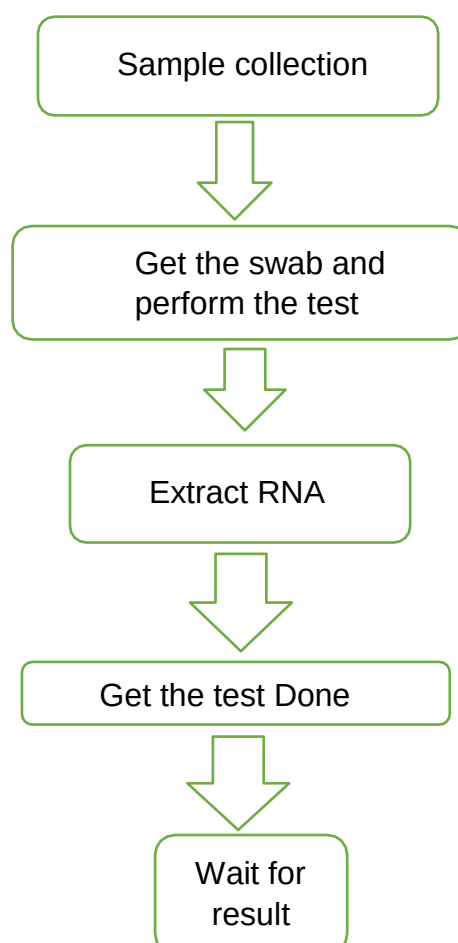
3.4.1 Methods:

There are three categories of testing, if I would want to simplify them. The first one is to identify whether the actual COVID virus genetic material exists, and that's called a NAAT test, N-A-A-T. And it's the PCR testing where you would have a nasal pharyngeal swab or a pharyngeal swab taken. And then they look for the genetic material of the virus itself. The second type of testing is when they try to identify one of the outer proteins of the viral shell or envelope, if you will. And that's called antigen testing. So, they try to detect the outer protein of the virus. And the third type is to detect within the human body, whether they've developed antibodies. So, it looks for antibodies that are specific to the outer portion of the virus itself. So, it shows whether the individual has mounted an immune response or developed immunity towards that specific virus or to COVID. So, those are the three big categories of testing that exist. The PCR testing, which is the NAAT testing or that tests the nucleic acid of the virus itself, are the most sensitive ones. And those, actually, to complete the testing process, they need to be done in quite a sophisticated laboratory setting. And, that's why the

turnaround time for these tests can take several days. And, if there's an outbreak and there's lots of samples, it will take several days longer than what we would want to or hope for. The antigen testing that exists now in the market are what we call the antigen rapid diagnostic tests. Those look for the antigen on the outer surface of the virus itself. And those have been developed in a way that they can be done at the bedside or in the field so they do not need the sophisticated laboratory setting to conduct them. They are not as accurate as the PCR testing, but they have a very important value as one of the tools to address the COVID pandemic.

There are several aspects of that. The first one is to identify whether a person has the disease or not. And that's important to be able to provide therapy, treatment, but also to specify who do we need to go and test further of the contacts of that specific person. So that's a big value to address the pandemic. When we talk about the value of the antibody testing, they become very important to know whether the person was exposed in the past. And, when you hear about understanding the immunity of the population, if you will, that is done through antibody testing.

Fig 3.1 Existing system block diagram



3.4.2 Disadvantages:

- The biggest disadvantage of this approach is that it takes long time to give the results.
- Since it is time Consuming, on an average it requires 48 hrs of time.
- Testing kits are more expensive.

3.6 METHODOLOGY:

The experiment conducted in this research consists of data collection, data analysis, model architecture, and consequent experimental results evaluated in terms of various performance parameters. These parts are described in the following sections.

Experiments are performed on the dataset collected from different sources for binary classification (COVID-19 vs. non-COVID). The following classes are subjected to classification: COVID-19, non-COVID. Non-COVID represents the normal categories of images.

3.6.1 Exploratory Data Analysis of X-Ray Images

Multiple datasets from different sources would be combined to form a generalized dataset, and then different machine learning algorithms would be applied to extract patterns and to obtain results with maximum accuracy.

3.6.2 Data Wrangling

In this section of the report will load in the data, check for cleanliness, and then trim and clean given dataset for analysis. Make sure that the document steps carefully and justify for cleaning decisions.

3.6.3 Data collection

The data set collected for predicting given data is split into Training set and Test set. Generally, we split the dataset into Training set and Test set. The Data Model which was created using machine learning algorithms are applied on the Training set and based on the test result accuracy, Test set prediction is done.

A total of 284 COVID-19 chest X-ray images can be accessed from GitHub. In addition, 234 normal chest X-ray images were obtained from GitHub as well. We combine both of the images into one data set. All the images are in different dimensions, so we need to resize them as 256×256.

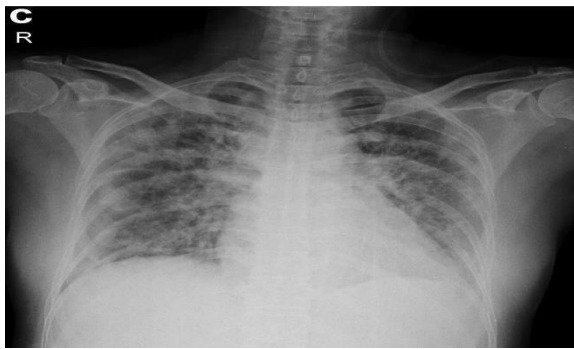
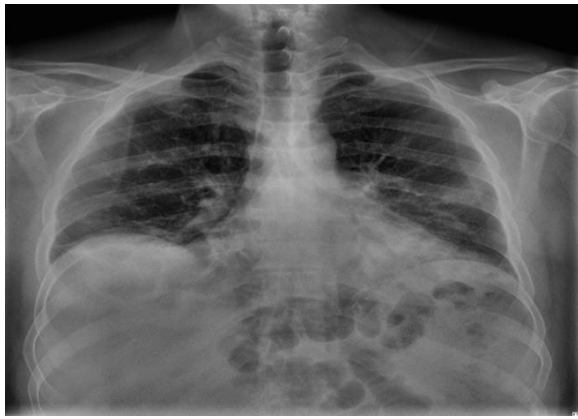


Fig 3.2 Covid Images



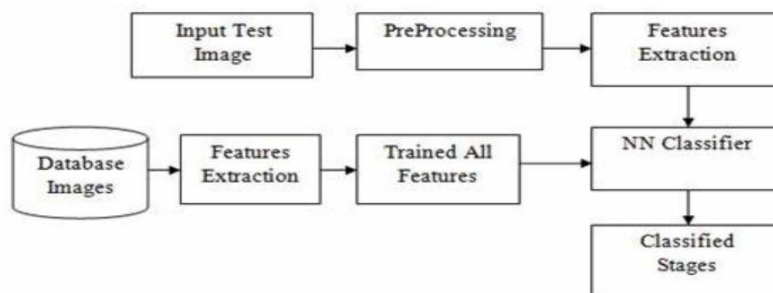
Fig 3.3 Normal Images

3.6.4 Building the classification model

For predicting the virus, ML algorithm prediction model is effective because of the following reasons: It provides better results in classification problem.

- It is strong in pre-processing outliers, irrelevant variables, and a mix of continuous, categorical and discrete variables.
- It produces out of bag estimate error which has proven to be unbiased in many tests and it is relatively easy to tune with.

Fig 3.4 Proposed system block diagram:



3.6.5 Advantages:

- Performance and accuracy of the algorithms can be calculated and compared.
- Requires less time to give results.

CHAPTER 4

EXPERIMENTAL METHODS AND ALGORITHMS USED

4.1 SYSTEM STUDY:

To develop this model, we use new modern technologies which are Machine Learning using Python for predicting and Gradio is for user interface.

4.1.1 System requirement specifications:

a) Hardware requirements:

◆	Processor :	Intel
◆	RAM :	2GB
◆	Hard Disk :	80GB

b) Software requirements:

◆	OS :	Windows
◆	Framework :	Flask
◆	Technology :	Machine Learning using Python
◆	Web Browser :	Chrome, Microsoft Edge
◆	Code editor :	Visual Studio

Code, Google Colab, Anaconda or Jupyter notebook.

4.2 SYSTEM SPECIFICATIONS:

4.2.1 Machine Learning Overview:

Machine learning is a field of study that looks at using computational algorithms to turn empirical data into usable models. The machine learning field grew out of traditional statistics and artificial intelligences communities. Through their business processes immense amounts of data have been and will be collected. This has provided an opportunity to re-invigorate the statistical and computational approaches to autogenerate useful models from data.

Machine learning algorithms can be used to (a) gather understanding of the cyber phenomenon that produced the data under study, (b) abstract the understanding of underlying phenomena in the form of a model, (c) predict future values of a phenomena using the above-generated model, and (d) detect anomalous behaviour exhibited by a phenomenon under observation.

4.2.2 Gradio Overview:

Gradio is an open-source python library which allows you to quickly create easy to use, customizable UI components for your ML model, any API, or any arbitrary function in just a few lines of code. You can integrate the GUI directly into your Python notebook, or you can share the link to anyone.

4.3 STEPS TO DOWNLOAD & INSTALL PYTHON:

Download the Latest version of the **Python** executable installer (<https://www.python.org/downloads/>). Watch the PIP list where pip is the package installer for python. Now upgrade the pip and setuptools using the command

Pip install --upgrade pip and Pip install --upgrade setuptools

4.3.1 IDE INSTALLATION FOR PYTHON

IDE stands for Integrated Development Environment. It is a GUI (Graphical User Interface) where programmers write their code and produce the final products. Best IDE is Pycharm. So download the pycharm new version and install the software

(<https://www.jetbrains.com/pycharm/download/>)

4.3.2 PYTHON FILE CREATION

GO To FILE MENU > CREATE > NEW > PYTHON FILE > (Name Your Python File as "HOUSE PRICE PPREDICTION" > SAVE

4.4 PYTHON LIBRARIES NEEDED

There are many libraries in python. In those we only use few main libraries needed

4.4.1 NUMPY LIBRARY

NumPy is an open-source numerical Python library. NumPy contains a multi- dimensional array and matrix data structures. It can be utilized to perform a number of mathematical operations on arrays such as trigonometric, statistical, and algebraic routines like mean, mode, standard deviation etc...,

Installation- (<https://numpy.org/install/>)

```
pip install NUMPY
```

Here we mainly use array, to find mean and standard deviation.

4.4.2 PANDAS LIBRARY

Pandas is a high-level data manipulation tool developed byWes McKinney. It is built on the Numpy package and its key data structure is called the Data Frame. Data Frames allow you to store and manipulate tabular data in rows of observations and columns of variables. There are several ways to create a Data Frame.

Installation- (https://pandas.pydata.org/getting_started.html)

```
pip install PANDAS
```

Here we use pandas for reading the csv files, for grouping the data, for cleaning the data using some operations.

4.4.3 MATPLOTLIB LIBRARY

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Use interactive figures that can zoom, pan, update, visualize etc.,

Installation- (<https://matplotlib.org/users/installing.html>)

```
pip install Matplotlib
```

Here we use pyplot mainly for plotting graphs. **matplotlib.pyplot** is a collection of functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.

4.4.4 TENSORFLOW LIBRARY

Pip install Tensorflow

TensorFlow is a free and open-source software library for machine learning and artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks.

Installation-(<https://www.tensorflow.org/install>)

4.4.5 Gradio Library

Pip install gradio

Gradio is an open-source python library which allows you to quickly create easy to use, customizable UI components for your ML model, any API, or any arbitrary function in just a few lines of code. You can integrate the GUI directly into your Python notebook, or you can share the link to anyone.

4.5 MODULES:

A modular design reduces complexity, facilitates change (a critical aspect of software maintainability), and results in easier implementation by encouraging parallel development of different part of system. Software with effective modularity is easier to develop because function may be compartmentalized and interfaces are simplified. Software architecture embodies modularity that is software is divided into separately named and addressable components called modules that are integrated to satisfy problem requirements.

Modularity is the single attribute of software that allows a program to be intellectually manageable. The five important criteria that enable us to evaluate a design method with respect to its ability to define an effective modular design are:

Modular decomposability, Modular Comps ability, Modular Understand ability, Modular continuity, Modular Protection.

4.6 MODULE DETAILS:

4.6.1 Data Pre-processing

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, to work with samples of data that may not be a true representative of the population of given dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters.

The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. The validation set is used to evaluate a given model, but this is for frequent evaluation. It as machine learning engineers use this data to fine-tune the model hyper parameters. Data collection, data analysis, and the process of addressing data content, quality, and structure can add up to a time-consuming to-do list. During the process of data identification, it helps to understand your data and its properties; this knowledge will help you choose which algorithm to use to build your model.

A number of different data cleaning tasks using Python Pandas library and specifically, it focuses on probably the biggest data cleaning task, missing values and it able to more quickly clean data. It wants to spend less time cleaning data, and more time exploring and modelling.

Some of these sources are just simple random mistakes. Other times, there can be a deeper reason why data is missing. It's important to understand these

different types of missing data from a statistics point of view. The type of missing data will influence how to deal with filling in the missing values and to detect missing values, and do some basic imputation and detailed statistical approach for dealing with missing data. Before, joint into code, it's important to understand the sources of missing data. Here are some typical reasons why data is missing:

- User forgot to fill in a field.
- Data was lost while transferring manually from a legacy database.
- There was a programming error.
- Users chose not to fill out a field tied to their beliefs about how the results would be used or interpreted.

Variable identification with Univariate, Bi-variate and multi-variate analysis:

- import libraries for access and functional purpose & read the given dataset
- General Properties of Analyzing the given dataset
- Display the given dataset in the form of data frame
- show columns
- shape of the data frame
- To describe the data frame
- Checking data type and information about dataset
- Checking for duplicate data
- Checking Missing values of data frame
- Checking unique values of data frame
- Checking count values of data frame
- Rename and drop the given data frame
- To specify the type of values
- To create extra columns

4.6.2 Data Validation/ Cleaning/Preparing Process

Importing the library packages with loading given dataset. To analyzing the variable identification by data shape, data type and evaluating the missing values, duplicate values. A validation dataset is a sample

of data held back from training your model that is used to give an estimate of model skill while tuning model's and procedures that you can use to make the best use of validation and test datasets when evaluating your models. Data cleaning / preparing by rename the given dataset and drop the column etc. to analyze the univariate, bi-variate and multi-variate process.

The steps and techniques for data cleaning will vary from dataset to dataset. The primary goal of data cleaning is to detect and remove errors and anomalies to increase the value of data in analytics and decision making.

MODULE DIAGRAM

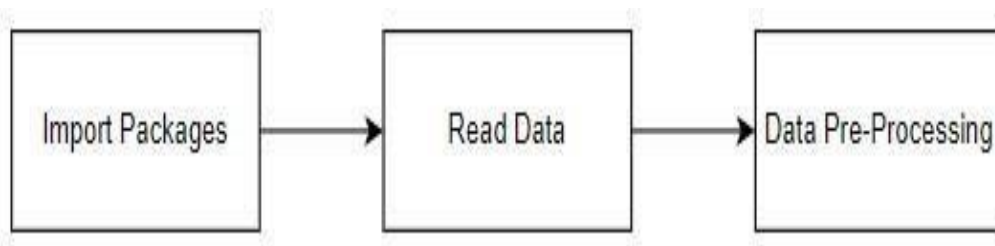


Fig 4.1 Module diagram 1

GIVEN INPUT EXPECTED OUTPUT

input: data output:

removing noisy data

4.6.3 Exploration data analysis of visualization

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data,

outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral and stakeholders than measures of association or significance. Data visualization and exploratory data analysis are whole fields themselves and it will recommend a deeper dive into some the books mentioned at the end.

Sometimes data does not make sense until it can look at in a visual form, such as with charts and plots. Being able to quickly visualize of data samples and others is an important skill both in applied statistics and in applied machine learning. It will discover the many types of plots that you will need to know when visualizing data in Python and how to use them to better understand your own data.

- How to chart time series data with line plots and categorical quantities with bar charts.
- How to summarize data distributions with histograms and box plots?

MODULE DIAGRAM

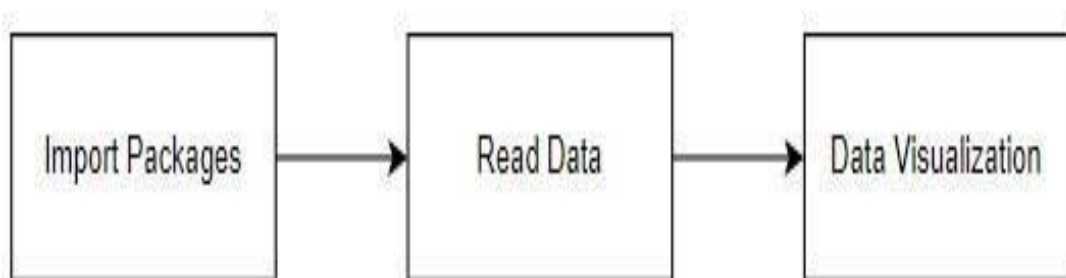


Fig 4.2 Module Diagram 2

GIVEN INPUT EXPECTED OUTPUT

input: data output:

visualized data

4.6.4 Comparing Algorithm with prediction in the form of best accuracy result

It is important to compare the performance of multiple different machine learning algorithms consistently and it will discover to create a test harness to compare multiple different machine learning algorithms in Python with scikit-learn. It can use this test harness as a template on your own machine learning problems and add more and different algorithms to compare. Each model will have different performance characteristics. Using resampling methods like cross validation, you can get an estimate for how accurate each model may be on unseen data. It needs to be able to use these estimates to choose one or two best models from the suite of models that you have created. When have a new dataset, it is a good idea to visualize the data using different techniques in order to look at the data from different perspectives. The same idea applies to model selection. You should use a number of different ways of looking at the estimated accuracy of your machine learning algorithms in order to choose the one or two to finalize. A way to do this is to use different visualization methods to show the average accuracy, variance and other properties of the distribution of model accuracies.

In the next section you will discover exactly how you can do that in Python with scikit-learn. The key to a fair comparison of machine learning algorithms is ensuring that each algorithm is evaluated in the same way on the same data and it can achieve this by forcing each algorithm to be evaluated on a consistent test harness.

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis. To achieving better results from the applied model in Machine

Learning method of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format, for example, Random Forest algorithm does not support null values. Therefore, to execute random forest algorithm null values have to be managed from the original raw data set. And another aspect is that data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in given dataset.

False Positives (FP): A person who will pay predicted as defaulter. When actual class is no and predicted class is yes. Ex. if actual class says this passenger did not survive but predicted class tells you that this passenger will survive.

False Negatives (FN): A person who default predicted as payer. When actual class is yes but predicted class in no. Ex. if actual class value indicates that this passenger survived and predicted class tells you that passenger will die.

True Positives (TP): A person who will not pay predicted as defaulter. These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes. Ex. if actual class value indicates that this passenger survived and predicted class tells you the same thing.

True Negatives (TN): A person who default predicted as payer. These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no. Ex. if actual class says this passenger did not survive and predicted class tells you the same thing.

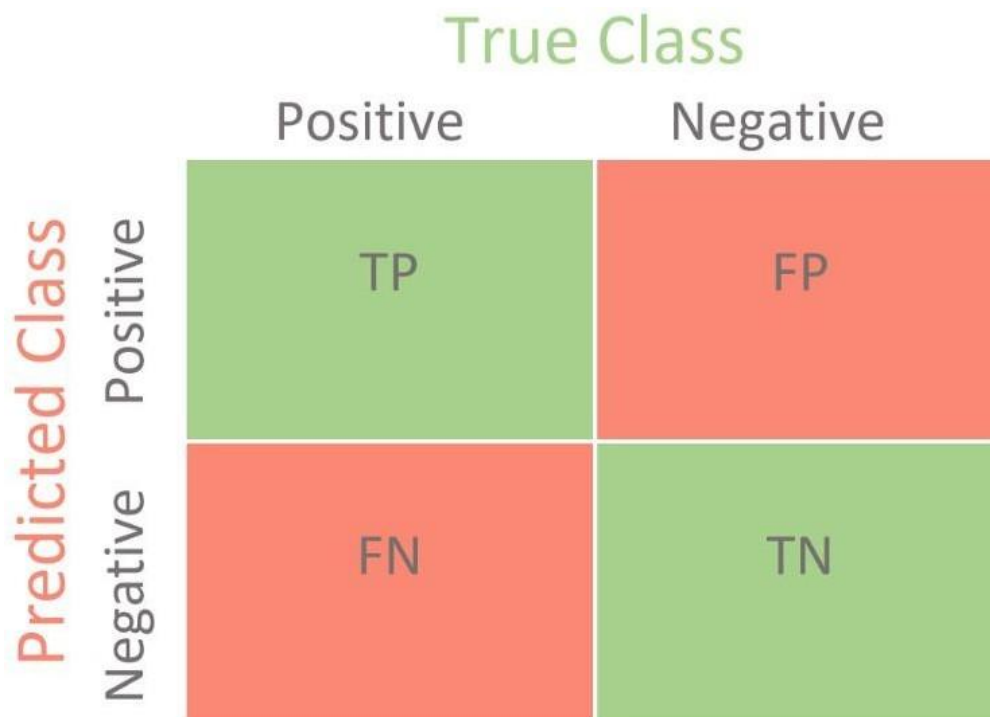


Fig 4.3 Class Diagram

Prediction result by accuracy:

Logistic regression algorithm also uses a linear equation with independent predictors to predict a value. The predicted value can be anywhere between negative infinity to positive infinity. It needs the output of the algorithm to be classified variable data. Higher accuracy predicting result is logistic regression model by comparing the best accuracy.

$$\text{True Positive Rate (TPR)} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{False Positive Rate (FPR)} = \text{FP} / (\text{FP} + \text{TN})$$

Accuracy: The Proportion of the total number of predictions that is correct otherwise overall how often the model predicts correctly defaulters and no defaulters.

Accuracy calculation:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same.

Precision: The proportion of positive predictions that are actually correct.

$$\text{Precision} = TP / (TP + FP)$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labelled as survived, how many actually survived? High precision relates to the low false positive rate. We have got 0.788 precision which is pretty good.

Recall: The proportion of positive observed values correctly predicted. (The proportion of actual defaulters that the model will correctly predict)

$$\text{Recall} = TP / (TP + FN)$$

Recall (Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution. Accuracy works best if false positives and false negatives have similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

General Formula:

$$F\text{- Measure} = 2TP / (2TP + FP + FN)$$

F1-Score Formula:

$$F1 \text{ Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

4.6.5 ALGORITHM AND TECHNIQUES**Algorithm Explanation**

In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation. This data set may simply be bi-class (like identifying whether the person is male or female or that the mail is spam or non-spam) or it may be multi-class too. Some examples of classification problems are: speech recognition, handwriting recognition, biometric identification, document classification etc. In Supervised Learning, algorithms learn from labeled data. After understanding the data, the algorithm determines which label should be given to new data based on pattern and associating the patterns to the unlabeled new data.

- In this project first we need to collect the dataset needed from different sources.
- Import the data using python libraries.
- Then we need to train the module to extract features and classifiers.
- Here we used CNNs to extract the features of CXR images.
- Check for accuracy if it not good means repeat the same process from start. (try with different datasets).

Used Python Packages:**Tensorflow:**

- **TensorFlow is a Python library** for fast numerical computing created and released by Google.
- **TensorFlow** is an open-source library developed by Google primarily for **deep learning applications**. It also supports traditional machine learning.

NumPy:

- It is a numeric python module which provides fast maths functions for calculations.
- It is used to read data in numpy arrays and for manipulation purpose.

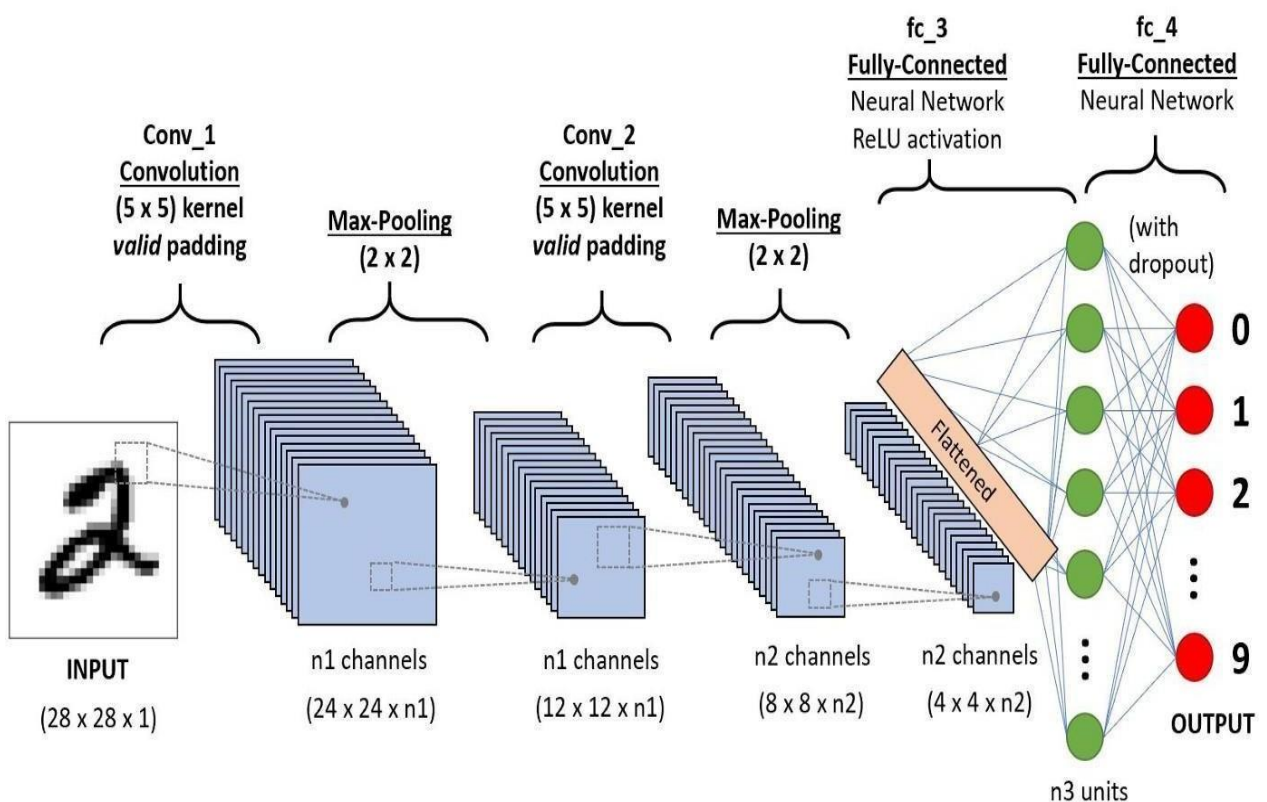
Pandas:

- Used to read and write different files.
- Data manipulation can be done easily with data frames.

Matplotlib:

- Data visualization is a useful way to help with identify the patterns from given dataset.
- Data manipulation can be done easily with data frames.

4.7 CNN – CONVOLUTION NEURAL NETWORK:



4.4 Convolution neural network systematic diagram

A **Convolutional Neural Network (CNN)** is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The preprocessing required in a CNN is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, CNN have the ability to learn these filters/characteristics.

The architecture of a CNN is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields' overlaps to cover the entire visual area.

A CNN is able to **successfully capture the Spatial and Temporal dependencies** in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better.

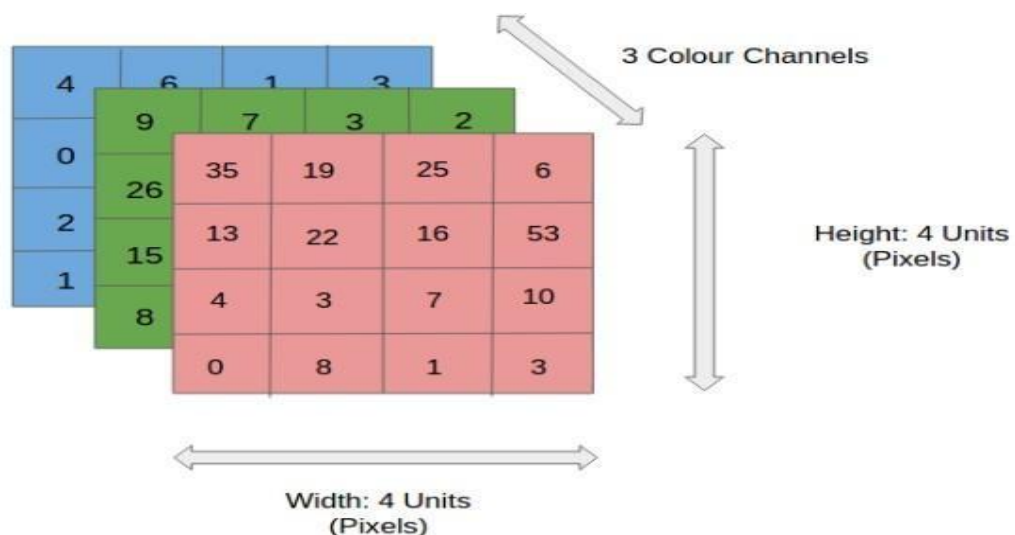
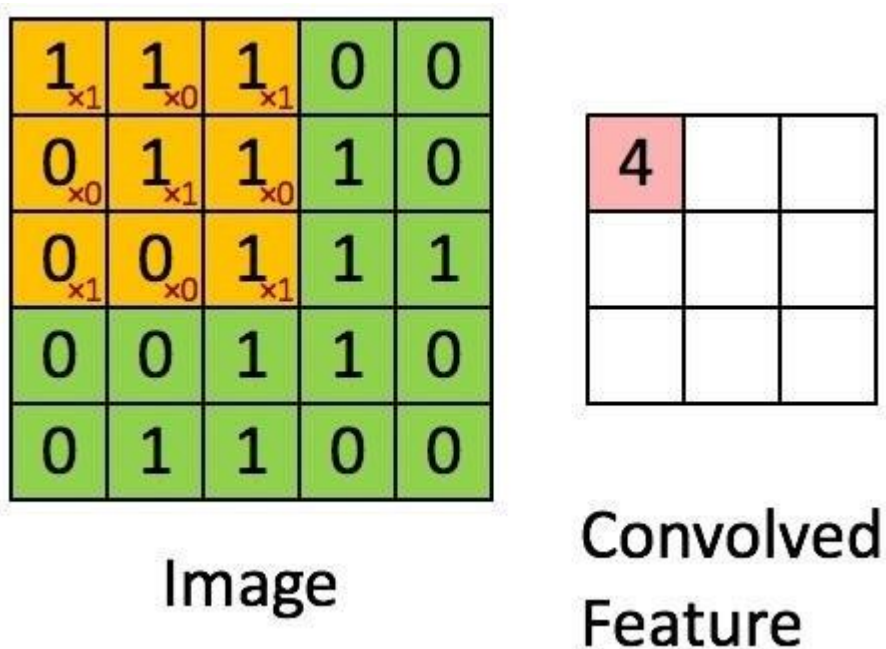


Fig 4.5 4x4x3 RGB image

In the figure, we have an RGB image which has been separated by its three color planes — Red, Green, and Blue. There are a number of such color spaces in which images exist — Grayscale, RGB, HSV, CMYK, etc.

You can imagine how computationally intensive things would get once the images reach dimensions, say 8K (7680×4320). The role of the CNN is to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction. This is important when we are to design an architecture which is not only good at learning features but also is scalable to massive datasets.

4.7.1 Convolution Layer — The Kernel



Convoluting a 5x5x1 image with a 3x3x1 kernel to get a 3x3x1 convolved feature
Fig 4.6 CNN Kernel

Image Dimensions = 5 (Height) x 5 (Breadth) x 1 (Number of channels, Ex. RGB)

In the above demonstration, the green section resembles our 5x5x1 input image, I.

The element involved in carrying out the convolution operation in the first part of a Convolutional Layer is called the Kernel/Filter, K, represented in the colour yellow.

We have selected K as a 3x3x1 matrix.

Kernel/Filter, K =

```
1 0 1  
0 1 0  
1 0 1
```

The Kernel shifts 9 times because of Stride Length = 1 (Non-Strided), every time performing a matrix multiplication operation between K and the portion P of the image over which the kernel is hovering.

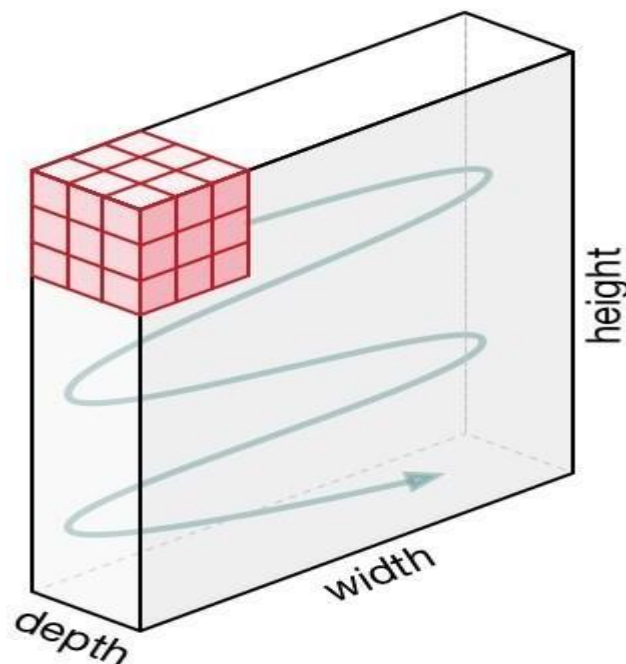


Fig 4.7 Movement of kernel

The filter moves to the right with a certain Stride Value till it parses the complete width. Moving on, it hops down to the beginning (left) of the image with the same Stride Value and repeats the process until the entire image is traversed.

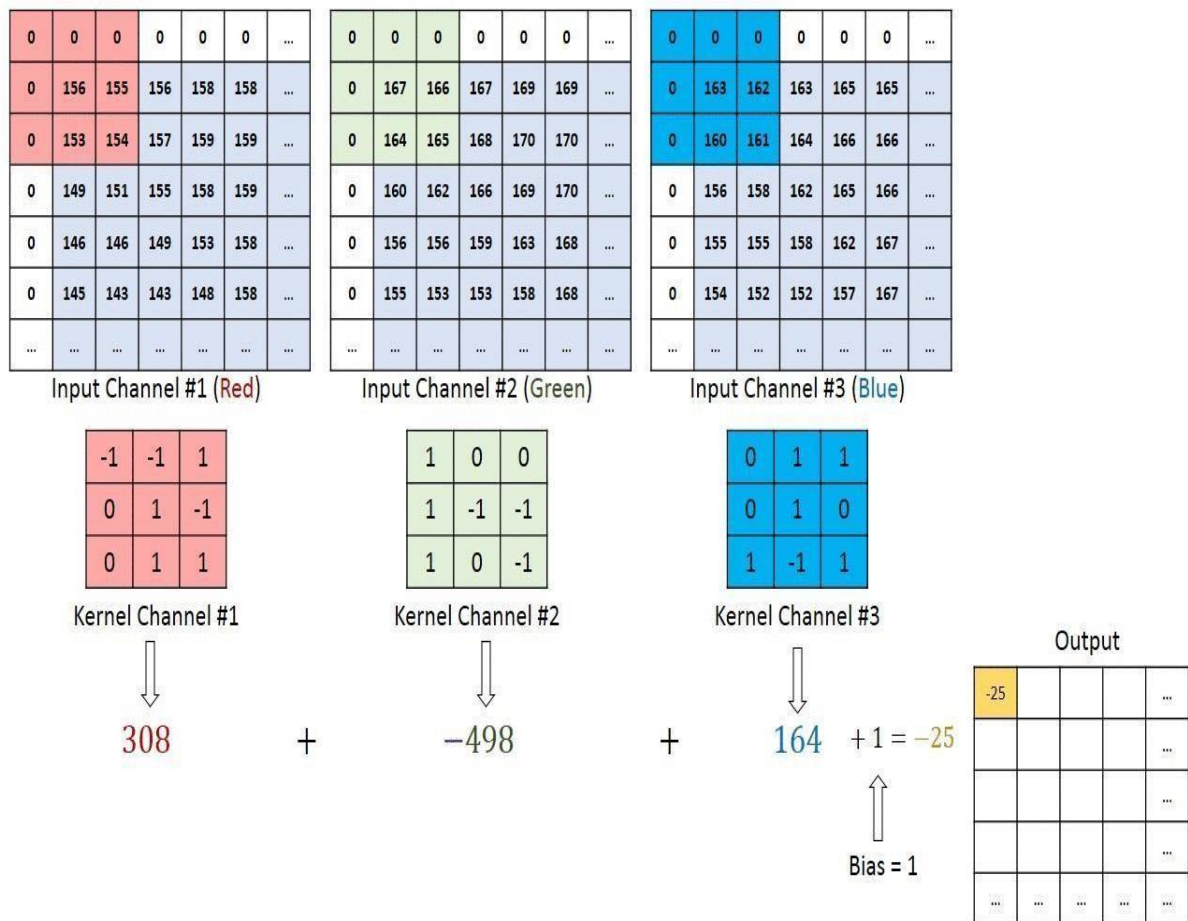


Fig 4.8 Convolution operation on a $M \times N \times 3$ image matrix with a $3 \times 3 \times 3$ Kernel

In the case of images with multiple channels (Ex. RGB), the Kernel has the same depth as that of the input image. Matrix Multiplication is performed between K_n and I_n stack ($[K_1, I_1]; [K_2, I_2]; [K_3, I_3]$) and all the results are summed with the bias to give us a squashed one-depth channel Convolved Feature Output.

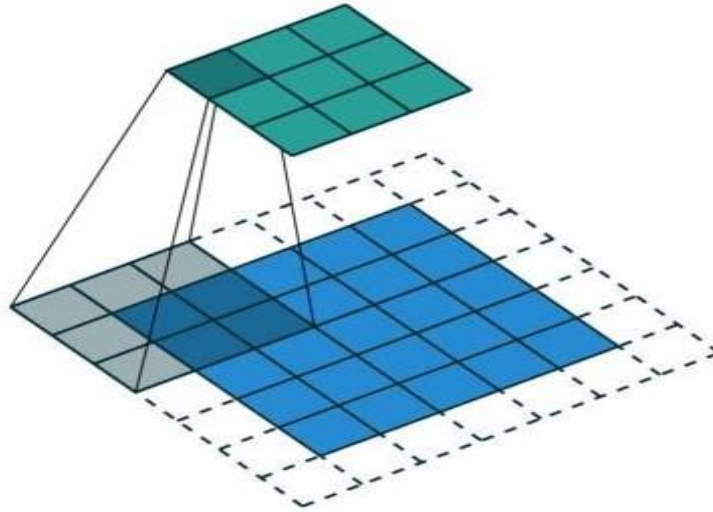


Fig 4.9 Convolution Operation with Stride Length = 2

The objective of the Convolution Operation is to extract the high-level features such as edges, from the input image. ConvNets need not be limited to only one Convolutional Layer. Conventionally, the first ConvLayer is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well, giving us a network which has the wholesome understanding of images in the dataset, similar to how we would. There are two types of results to the operation — one in which the convolved feature is reduced in dimensionality as compared to the input, and the other in which the dimensionality is either increased or remains the same. This is done by applying Valid Padding in case of the former, or Same Padding in the case of the latter.

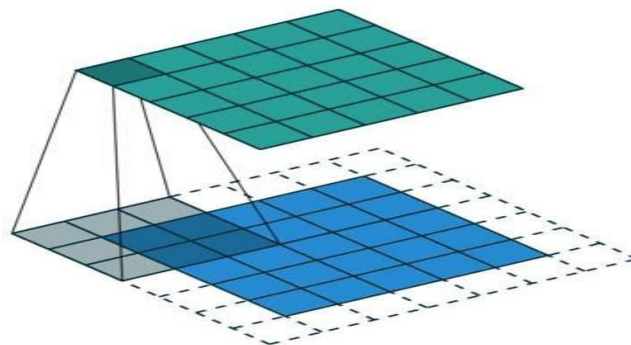


Fig 4.10 SAME padding: 5x5x1 image is padded with 0s to create a 6x6x1 image

When we augment the $5 \times 5 \times 1$ image into a $6 \times 6 \times 1$ image and then apply the $3 \times 3 \times 1$ kernel over it, we find that the convolved matrix turns out to be of dimensions $5 \times 5 \times 1$.

Hence the name — Same Padding.

On the other hand, if we perform the same operation without padding, we are presented with a matrix which has dimensions of the Kernel ($3 \times 3 \times 1$) itself — Valid Padding.

Similar to the Convolutional Layer, the Pooling layer is responsible for reducing the spatial size of the Convolved Feature. This is to decrease the computational power required to process the data through dimensionality reduction. Furthermore, it is useful for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training of the model.

There are two types of Pooling: Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel. On the other hand, Average Pooling returns the average of all the values from the portion of the image covered by the Kernel.

Max Pooling also performs as a Noise Suppressant. It discards the noisy activations altogether and also performs de-noising along with dimensionality reduction. On the other hand, Average Pooling simply performs dimensionality reduction as a noise suppressing mechanism. Hence, we can say that Max Pooling performs a lot better than Average Pooling.

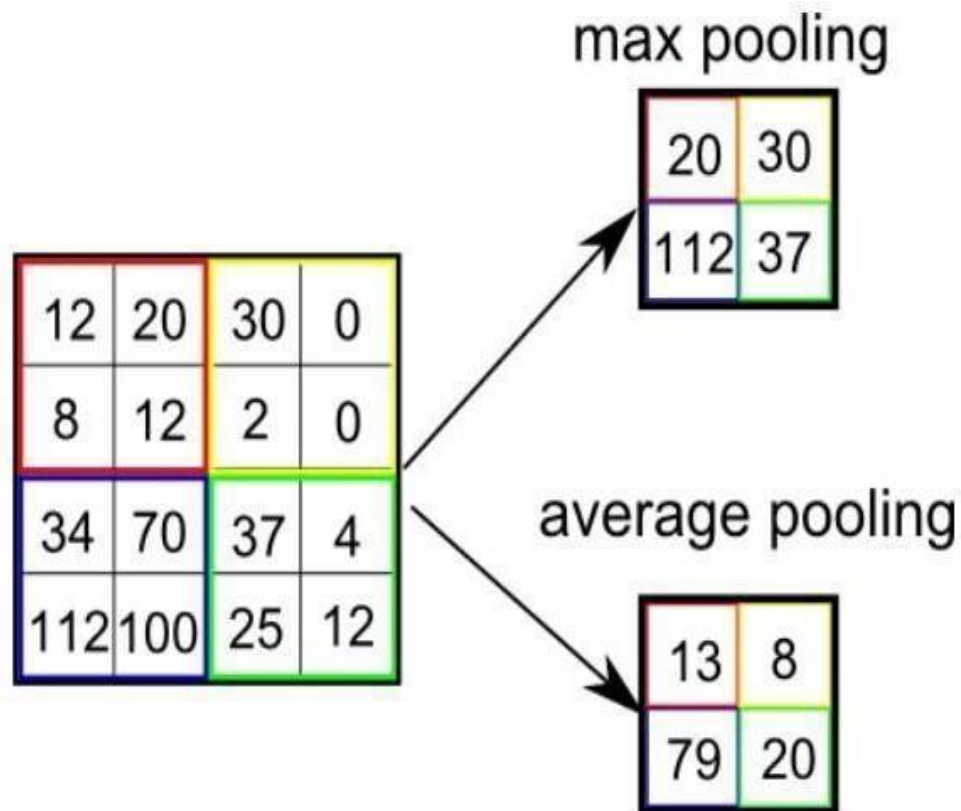


Fig 4.11 Types of Pooling

The Convolutional Layer and the Pooling Layer, together form the i-th layer of a Convolutional Neural Network. Depending on the complexities in the images, the number of such layers may be increased for capturing low-levels details even further, but at the cost of more computational power.

After going through the above process, we have successfully enabled the model to understand the features. Moving on, we are going to flatten the final output and feed it to a regular Neural Network for classification purposes.

4.7.2 Classification — Fully Connected Layer (FC Layer)

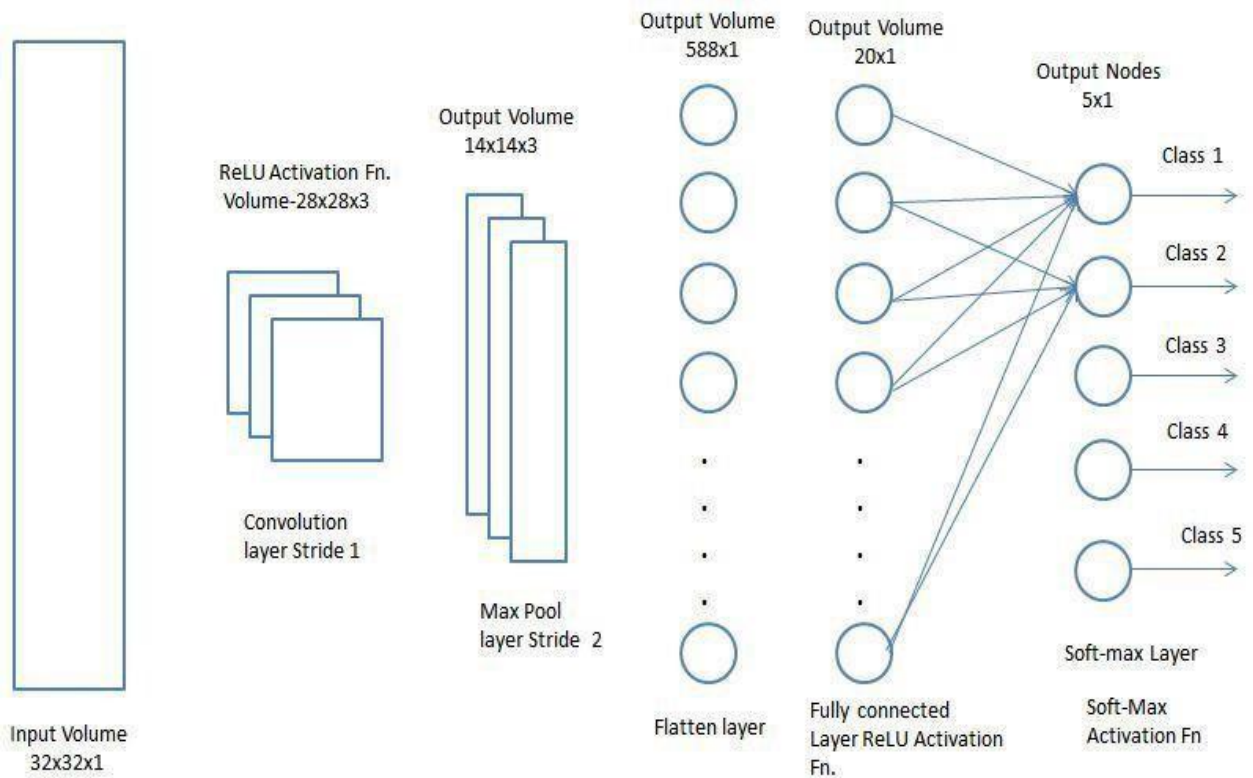


Fig 4.12 Fully connected layer in CNN

Adding a Fully-Connected layer is a (usually) cheap way of learning non-linear combinations of the high-level features as represented by the output of the convolutional layer. The Fully-Connected layer is learning a possibly non-linear function in that space.

Now that we have converted our input image into a suitable form for our Multi-Level Perceptron, we shall flatten the image into a column vector. The flattened output is fed to a feed-forward neural network and backpropagation applied to every iteration of training. Over a series of epochs, the model is able to distinguish between dominating and certain low-level features in images and classify them using the Softmax Classification technique.

There are various architectures of CNNs available which have been key in building algorithms which power and shall power AI as a whole in the foreseeable future.

Some of them have been listed below:

1. LeNet
2. AlexNet
3. VGGNet
4. GoogLeNet
5. ResNet
6. ZFNet
7. EfficientNet

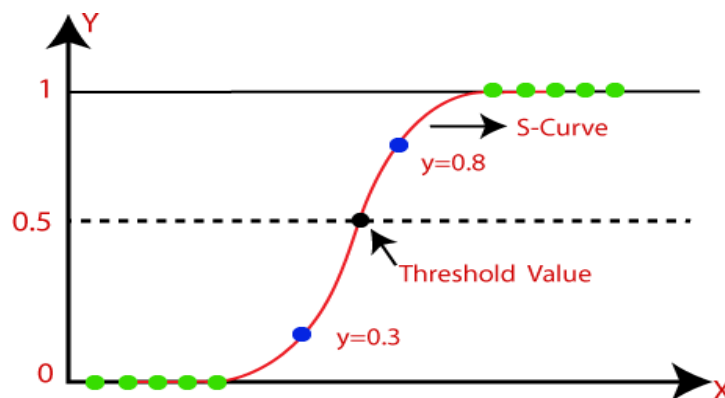


Fig 4.13 CNNs Threshold

GIVEN INPUT EXPECT OUTPUT

input: data output:

getting accuracy

- In this project we did experiments using Convolution neural networks (CNNs) on the images.
- These CNNs are capable of feature extraction and classification as well
- A traditional CNN consists of three layers: 1. convolution, 2. pooling, and 3. fully connected layers.

- Feature extraction will be performed in the convolutional layer by applying masks.
- Masking is the process of dividing images into a predefined dimension of segments and it also use filters to extract features from an image.
- By applying activation function, we will get a feature map.
- The function of feature map is projecting the features onto the 2d map.
- The activation function activates the most important and knowledgeable neurons in a non-linear way to remove the computational cost of the NNs.
- There are many activation functions available in CNNs, but the most common one is rectified linear unit (ReLU).
- The activation function (ReLU) doesn't activate all the neurons at the same time, so it provides a faster convergence when the weights find the optimum values to produce the trained response when training the model.
- Then we will perform pooling operation on the feature map to decrease the dimensions of the chest X-ray image.

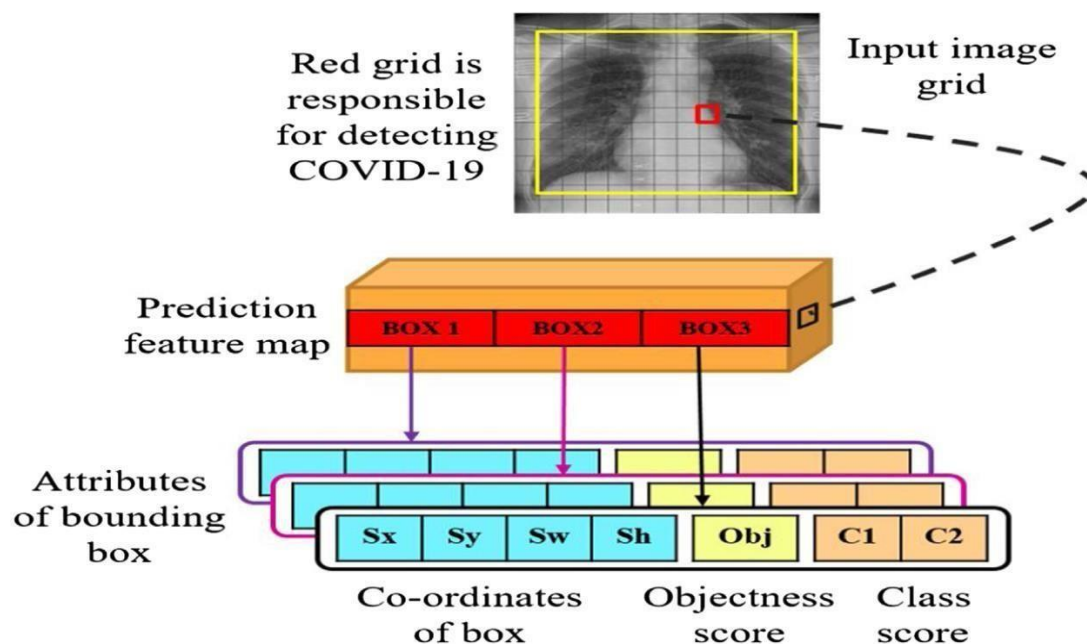


Fig 4.14 Detection mask through the trained model in an image.

- As we can see in above figure that red grid is showing us that the person with chest x ray image was infected be virus.
- And there is the prediction feature map to predict the features.
- Below are the screenshots gfor virtual neural networks for our machine learning model using CNNs

CHAPTER 5

RESULTS AND DISCUSSIONS, PERFORMANCE ANALYSIS

5.1 Results:

All training and testing of the deep learning models are done on Python language based framework in a system with Intel x86 processor equipped with graphics processing unit (GPU). Considering the memory limitations of the server, the batch size of the training model is taken as sixteen. Other factors like algorithm reaches the optimum accuracy, to regularize the training model are given in below table, with complete software and hardware specifications for training different CXR images. Training and validation sets were kept constant while test dataset was kept updating with new CXR images during the experiments. Apart from keeping images in the test set, the remaining dataset from all the sources is divided into training and validation set, which is further cross-validated. The deep learning model is trained with CXR images of different categories for several iterations until the loss gets saturated. Generated trained models are analyzed with multiple images in test dataset to get overall performance. Performance metrics used for the calculation of the experiment are:

$$\text{Sensitivity or Recall} = TP / (TP + FN)$$

$$\text{Specificity} = TN / (TN + FP)$$

$$\text{Precision} = TP / (TP + FP)$$

$$\text{F1 Score} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$$

$$\text{Accuracy} = (TP + TN) / (\text{Positivity} + \text{Negativity})$$

where true positive (TP) are those cases where the model correctly predicts the positive labelled image, false positive (FP) is the case where the model predicts as COVID-19 although the image is labelled as non-COVID-19. True negative (TN) is when the model correctly predicts negative image and false-negative (FN) is the case where the model incorrectly predicts a positive labelled negative image. The confusion matrix is used for measuring the performance of the machine learning classification problem. It is a combined representation of ground truth and predicted classes. First, image data is analyzed without augmentation and later the proposed methodology is tested with the application of augmentation techniques.

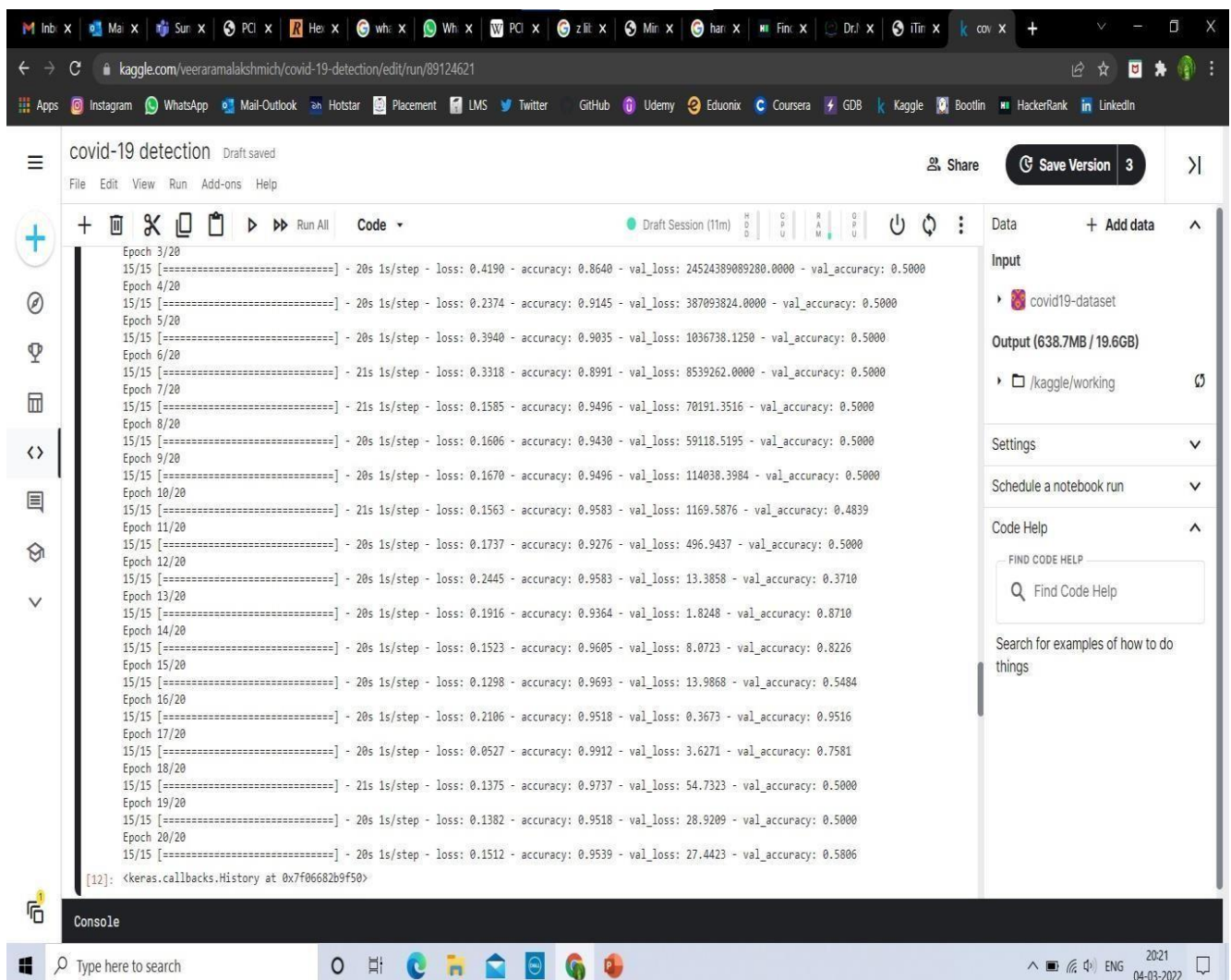


Fig 5.1 Snapshot for accuracy

As we seen in above screen shot for each epoch the value of accuracy will be increased. Since we are using CNNs for feature extraction, it will extract all possible features and classifiers for the covid image to get better results.

S.no	Loss	Accuracy	Val_loss	Val_accuracy
1	0.2448	0.6784	17651.643	0.2895
2	0.3752	0.7416	21980.125	0.4328
3	0.4190	0.8640	24537.000	0.5000
4	0.2374	0.9145	38715.000	0.5000
5	0.3940	0.9035	10370.125	0.5000
6	0.3318	0.8991	85392.000	0.5000
7	0.1585	0.9496	70200.352	0.5000
8	0.1606	0.9430	59200.519	0.5000
9	0.1670	0.9496	114040.398	0.5000
10	0.1563	0.9583	1169.587	0.4839
11	0.1737	0.9276	496.9437	0.5000
12	0.2445	0.9583	13.3858	0.3710
13	0.1916	0.9364	1.8248	0.8710
14	0.1532	0.9605	8.0723	0.8226
15	0.1298	0.9693	13.9868	0.5484
16	0.2106	0.9518	0.3673	0.9516
17	0.0527	0.9912	3.6271	0.7581
18	0.1375	0.9737	54.7323	0.5000
19	0.1382	0.9518	28.9209	0.5000
20	0.1512	0.9539	27.4423	0.5806

Table 5.1 Accuracy table

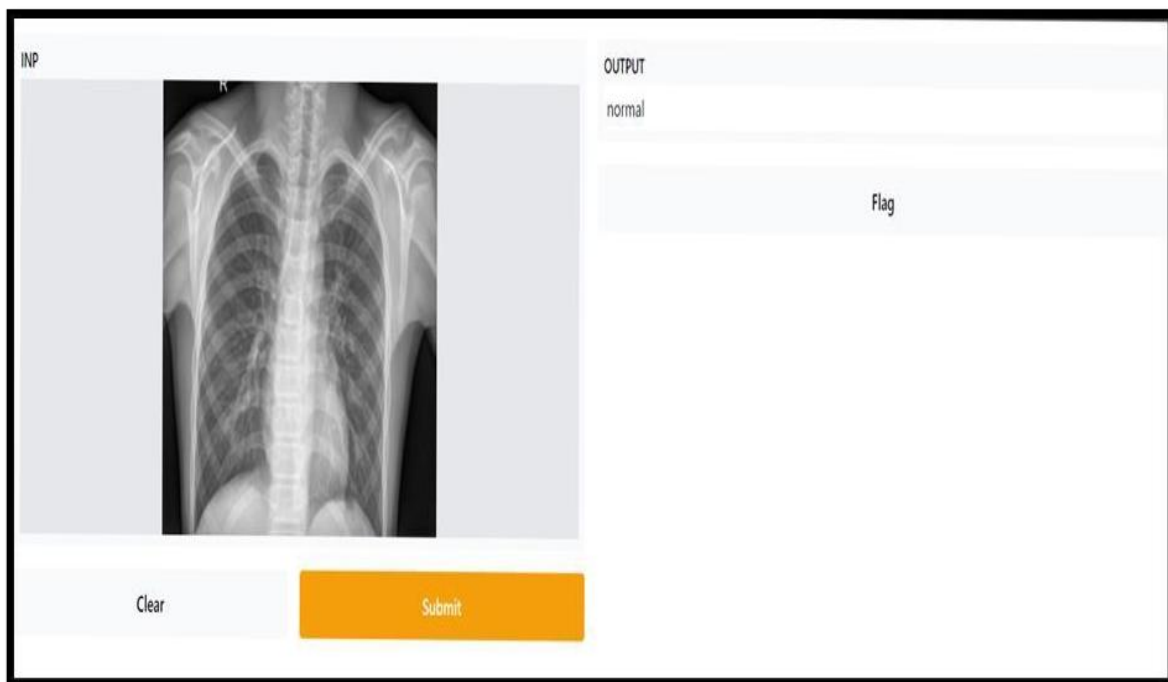
As we seen in above table for each epoch the value of accuracy will be increased. Since we are using CNNs for feature extraction, it will extract all possible features and classifiers for the covid image to get better results.

5.2 Discussions:

As we have seen in RTPCR Test it will take more than two days to produce the results. But in this experiment, we are using pre trained model called efficient net which means we are able to process the result within a minute. And if we search some other open-source data sets for pneumonia or flu and combine with the data set we have we will be able to detect both deceases at the same time.

5.3 Performance analysis and results:

In this project we obtained the accuracy as 96.4% which clearly specifying the success rate of this project. Since we are using Gradio library for getting the result/output which will take less than one minute to produce the result.



INP



OUTPUT

covid

Flag

Clear

Submit

Fig 5.2 Output snapshots

CHAPTER 6

CONCLUSION

6.1

Conclusion:

- Detection of COVID-19 from chest X-ray images will be useful for doctors as well as for patients to decrease the testing time and cost for the testing kit.
- AI and deep learning are capable of recognizing images for the tasks taught. In this project, we did many experiments to achieve the high-accuracy detection of COVID-19 which is nothing but to get much success rate in this project.
- This project results shows that the CNNs with minimized convolutional and fully connected layers are capable of detecting COVID-19 virus within the two- class, one in covid class and the other is normal class with accuracy rate of 96.4 respectively.
- Therefore, we came to know that the use of AI-based high-accuracy technologies will be useful to doctors in diagnosing COVID-19.

6.2 Future work:

Considering the rate of spread of *COVID-19*, automatic techniques are needed for the detection of this disease. In this report, a machine learning method is proposed for the detection of *COVID-19* disease. The evaluation demonstrated the effectiveness of the proposed method, by obtaining an average precision and recall equal to 0.965 in the discrimination between the *COVID-19* and *other* pulmonary diseases with similar symptoms. As future work, we plan to validate the proposed method considering also a set of healthy chest X-rays.

Moreover, we will investigate whether deep learning and formal verification techniques can obtain better performances in the *COVID-19* as demonstrated in similar contexts, from cancer detection to malware detection.

REFERENCES:

- Apostolopoulos, I. D., Mpesiana, T. Covid-19: Automatic Detection from XRay Images Utilizing Transfer Learning with Convolutional Neural Networks. Phys. Eng. Sci. Med. 2020.
- Cohen, J. P. COVID-19 Image Data Collection. ArXiv. 2020.
- Deng, X., Zhang, Y., Yang, S., et al. Joint Hand Detection and Rotation Estimation Using CNN. IEEE Trans. Image Proc. 2018.
- Haralick, R., Shapiro, L. Computer and Robot Vision, Vol. 1. Addison-Wesley: Reading (MA), 1992.
- Hartenstein, A., Lübke, F., Baur, A. D. J., et al. Prostate Cancer Nodal Staging: Using Deep Learning to Predict 68Ga-PSMA-Positivity from CT Imaging Alone. Sci. Rep. 2020.
- <https://github.com/ieee8023/covid-chestxray-dataset>.
- Huang, G., Liu, Z., van der Maaten, L., et al. Densely Connected Convolutional Networks. ArXiv. 2018.
- Kermany, D. S., Goldbaum, M., Cai, W., et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 2018.
- Melo, F. Area under the ROC Curve. In: Dubitzky, W., Wolkenhauer, O., Cho, K. H., Yokota, H. (eds.), Encyclopedia of Systems Biology. Springer: New York, 2013.
- Ozsahin, I., Sekeroglu, B., Mok, G. S. P. The Use of Back Propagation Neural Networks and 18F-Florbetapir PET for Early Detection of Alzheimer's Disease Using Alzheimer's Disease Neuroimaging Initiative Database. PLoS One. 2019.
- Roy, S. K., Krishna, G., Dubey, S. R., et al. HybridSN: Exploring 3-D–2-D CNN Feature Hierarchy for Hyperspectral Image Classification. IEEE Geosci. Remote. Sens. Lett. 2020.
- Rubin, G. D., Ryerson, C. J., Haramati, L. B.; et al. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society [published online ahead of print, 2020 Apr 7]. Chest. 2020.
- World Health Organization. WHO Coronavirus Disease (COVID-19) Dashboard. <https://covid19.who.int>
- Yilmaz, N., Sekeroglu, B. Student Performance Classification Using Artificial Intelligence Techniques. In: 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions (ICSCCW) [2019]. Adv. Intel. Sys. Comm. 2019.

