COVID-19 DETECTION USING CHEST CT IMAGES AND X-RAY IMAGES

**i**

**CONTENTS**

**Title page Pages**

ACKNOWLEDGEMENT ----------------------------------------------------------------------i

LIST OF TABLES ------------------------------------------------------------------------------ii

LIST OF FIGURES ----------------------------------------------------------------------------iii

**CHAPTER 1 - INTRODUCTION**

1.1. Title of the Project -----------------------------------------------------------------------1

1.2. Objective ----------------------------------------------------------------------------------1

1.3. Problem Statement ---------------------------------------------------------------------- 1

1.4. Motivation and Scope of the Project --------------------------------------------------1

**CHAPTER 2 – LITERATURE SURVEY**

2.1 Literature Survey----------------------------------------------------------------------- 2-3

2.2 Comparison --------------------------------------------------------------------------------3

**CHAPTER 3 – FEASIBILITY STUDY AND REQUIREMENT ANALYSIS**

3.1. System Requirements (Hardware and Software specifications)

3.1.1. Software Requirement -------------------------------------------------------4

3.1.2. Hardware Requirement ------------------------------------------------------4

3.2. Scheduled Feasibility

3.2.1. Work Breakdown Structure----------------------------------------------------5

3.2.2. Gantt Chart-----------------------------------------------------------------------6

3.3. Economic Feasibility

3.3.1. COCOMO Model--------------------------------------------------------------7-8

**CHAPTER 4 – PROPOSED PLAN**

4.1. Methodology------------------------------------------------------------------------------9

4.2. Dataset Description-------------------------------------------------------------------9-10

4.3. Performance evaluation metrics to be used--------------------------------------10-11

4.4. Flow chart------------------------------------------------------------------------------- 12

4.5 Block diagram----------------------------------------------------------------------------13

**CHAPTER 5 –IMPLEMENTATION**

5.1Importing Dataset -----------------------------------------------------------------------14

5.2 Importing Libraries ---------------------------------------------------------------------14

5.3 Implementing CNN -------------------------------------------------------------------- 14

5.4 Model Summary-------------------------------------------------------------------------15

5.5 Performing Operations -----------------------------------------------------------------15

5.6 Checking Accuracy ---------------------------------------------------------------------16

5.7 Loss Graph ------------------------------------------------------------------------------16

5.8 Accuracy Graph ------------------------------------------------------------------------17

**CHAPTER 6 – CONCLUSION**

6.1 Conclusion ------------------------------------------------------------------------------ 18

6.2 References ---------------------------------------------------------------------------18-19

6.3 Remark-----------------------------------------------------------------------------------20

**LIST OF TABLES**

**Table Title Page**

Table 3.3.1.1 COCOMO Model co-efficient tables ----------------------------- 7

ii

**LIST OF FIGURES**

**Figure Title Page**

3.1.1.1 Software Requirements --------------------------------------4

3.1.2.1 Hardware Requirements -------------------------------------4

3.2.1.1 Work Break down Structure----------------------------------5

3.2.2.1 Gantt chart -----------------------------------------------------6

4.4.1 Flow Chart----------------------------------------------------12

4.5.1 Block Diagram ---------------------------------------------- 13

5.1.1 Dataset--------------------------------------------------------- 14

5.2.1 Libraries ------------------------------------------------------ 14

5.3.1 CNN -----------------------------------------------------------14

5.4.1 Model Summary -------------------------------------------- 15

5.5.1 Performing Operation -------------------------------------- 15

5.6.1 Checking Accuracy -----------------------------------------16

5.7.1 Loss Graph -------------------------------------------------- 16

5.8.1 Accuracy Graph ---------------------------------------------17

**iii**

**CHAPTER 1**

**INTRODUCTION**

**1.1 TITLE OF THE PROJECT**

COVID-19 DETECTION USING CHEST CT IMAGES AND X-RAY IMAGES

**1.2 OBJECTIVE**

The primary objective of our project is to design a model using computer vision to detect COVID-19 by X-ray and Computed Tomography images to reduce the time taken to detect a COVID-19 patient and achieve higher accuracy than the rapid test which is present now a days. We aim to use thousands of readily available chest radiograph images and Computed Tomography images with clinical findings associated with COVID-19 as a training data set, mutually exclusive from the images with confirmed COVID-19 cases, which will be used as the testing data set.

**1.3 PROBLEM STATEMENT**

There has been a rapid increase in number of new and suspected COVID-19 cases in the world, and even the test takes more time to generate the result compared to the time for the virus to spread among people and sometimes it gives false results, therefore, to test the COVID-19 infection rapidly and in more efficient way, chest X-Ray or/and CT scan images of COVID-19 suspected individuals could be an answer.

**1.4 MOTIVATION AND SCOPE OF THE PROJECT**

The motivation behind this project is that several countries suffer from the shortage of test kits and high false negative rate in PCR test and covid is posed to be a deadly and infectious disease until recent times. Recognition of COVID-19 is a challenging task. Therefore, there is an urgent need for fast detection with clear visualization of infection which can be used on a suspected patient of COVID-19 and could be saved. Hence, we build a model to predict the COVID-19 cases at the earliest possible to control the spread of disease.

**CHAPTER 2**

**LITERATURE SURVEY**

1. **COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis** by Universiade Federal de OuroPreto (UFOP), MG, Brazil [1].

In this project they used techniques of pre-processing, Effective COVID net, training, data argumentation and the data set used are SARS-CoV-2 CT-scan dataset, COVID-CT dataset. The result they obtain is that the model could learn how to identify portions and patterns of one image that may indicate the presence (or absence) of COVID-19, although, those patterns may not appear in a different dataset.

1. **Identifying COVID19 from Chest CT Images: A Deep Convolutional Neural Networks Based Approach** by Arnab Kumar Mishra [2].

In this project they used techniques of various Deep CNN Architectures, Decision fusion approach and the datasets used here is COVID-CT dataset. The result they obtain is that Deep CNN based predictive models can be of very high value with regard to COVID19 identification from CT images. More specifically, the simple idea of decision fusion can improve the performance of Deep CNN models quite drastically, achieving above 86% results with respect to all the performance metrics under consideration.

1. **Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features** by Varalakshmi Perumal, Vasumathi Narayanan & Sakthi Jaya Sundar Rajasekar [3].

In this project they used techniques of Transfer Learning, Haralick Texture Feature Extraction and the data sets used are Chest X-Ray-14 dataset, COVID-CT dataset. The result they obtain is that the proposed model produces precision of 91%, recall of 90% and accuracy of 93% by VGG-16 using transfer learning, which outperforms other existing models for this pandemic period.

1. **Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks** by Dilbag Singh, Vijay Kumar, Manjit Kaur, Vaishali [4].

In this project they used techniques of Convolutional neural networks, Feature extraction, Classification and the datasets used are COVID-CT dataset. The result they obtain is that the proposed model outperforms competitive models, i.e., ANN, ANFIS, and CNN models in terms of accuracy, F-measure, sensitivity, specificity, and Kappa statistics by 1.9789%, 2.0928%, 1.8262%, 1.6827%, and 1.9276%, respectively.

1. **Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks** by Ali Narin1, Ceren Kaya, Ziynet Pamuk [5].

In this project they used techniques of Convolutional neural networks, Pre-trained models and datasets used are ChestX-ray8 datasets, Chest X-Ray Images (Pneumonia). The result they obtained that ResNet50 pre-trained model yielded the highest accuracy among five models for used three different datasets (Dataset-1: 96.1%, Dataset-2: 99.5% and Dataset-3: 99.7%).

**2.1 COMPARISIONS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Covid-19 detection using CT images & Deep Learning** | **Identifying Covid-19 using CT images** | **Detection of Covid-19 using Chest X-ray** | **Classification of Covid-1 using CT images** | | **Automatic Detection of Coronavirus Disease** |
| Major technique used here is pre-processing | Major technique used here is DEEP CNN architecture | Major technique used here is Haralick Feature Extraction | Major technique used here is CNN, feature extraction | | Major technique used here is CNN & pre trained models |
| Dataset used are SARS COV-2-CT Scan & COVID-CT | Dataset used is COVID-CT | Dataset used are Chest-X-ray & COVID-CT | | Dataset used is COVID-CT | Dataset used are Chest-X-ray |
| The model was able to identify patterns that may indicate the presence of Covid | CNN based model can be of very high value in detecting Covid | The model produced precision of 91% recall 90% which outperforms the existing models | | The models outperforms the competitive models | The pre trained model ResNet50 yielded the highest accuracy |

**CHAPTER 3**

**FEASIBILITY AND REQUIREMENT ANALYSIS**

A feasibility study is used to determine the viability of an idea, such as ensuring a project is legally and technically feasible as well as economically justifiable. It tells us whether a project is worth the investment. A feasibility study evaluates the project’s potential for success.

Our project is feasible as this method is present but most authorities are relying on rapid test which are not that accurate. The person who wants to test would just have to give us the CT scan photos or the X-Ray images and we will take it as input and compare the photo with our trained model and give him the results whether he has the disease or not and this process does not require any medical kit just a well-trained model.

The requirement analysis focuses on the tasks that determine the needs or conditions to meet the new or altered product or project, taking account of the possibly conflicting requirements of the various stakeholders, analyzing, documenting, validating and managing software or system requirements

**3.1. SYSTEM REQUIREMENTS**

|  |  |
| --- | --- |
| **SOFTWARE** | **SPECIFICATION** |
| **Operating System** | Windows OS (Windows 10,8.1) |
| **IDE** | Anaconda |
| **Programming language** | Python 3 |

**3.1.1. SOFTWARE REQUIREMENTS**

3.1.1.1 Software Requirements

**3.1.2. HARDWARE REQUIREMENTS**

|  |  |  |
| --- | --- | --- |
| **HARDWARE** | **MINIMUM** | **PREFERABLE** |
| **Processor** | 1 GHz | Modern processor |
| **Display** | 800x600px | 1280x720px |
| **Ram** | 2 GB for 64 bit | 16 GB for 64 bit |
| **Hard disk** | 500 GB for 32 bit OS | 1 TB for 64 bit OS |
| **GUI** | 2 GB | 16 GB |

3.1.2.1 Hardware Requirements

**3.2. SCHEDULED FEASIBILITY**

This assessment is the most important for project success, after all, a project will fail if not completed on time. In scheduling feasibility, it is estimated how much time the project will take to complete.

**3.2.1. WORK BREAK DOWN STRUCTURE**

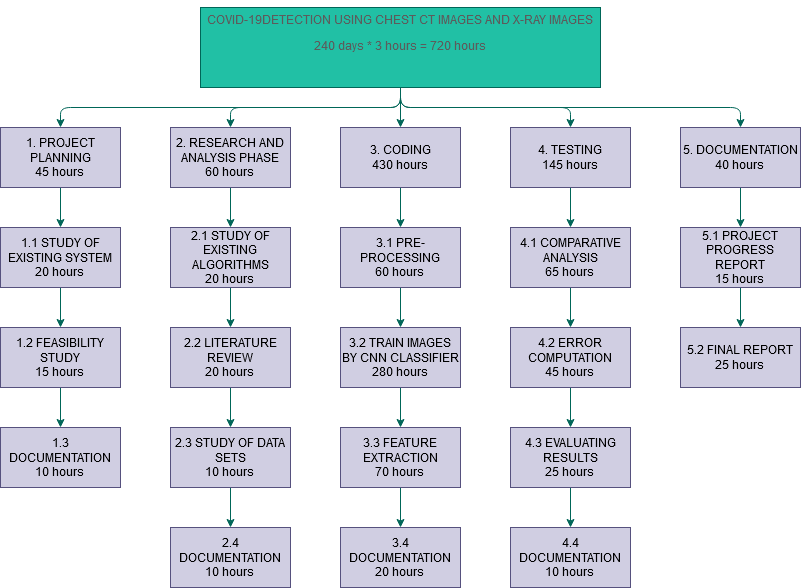


Fig 3.2.1.1 Work Break Down Structure

**3.2.2. GANTT CHART**

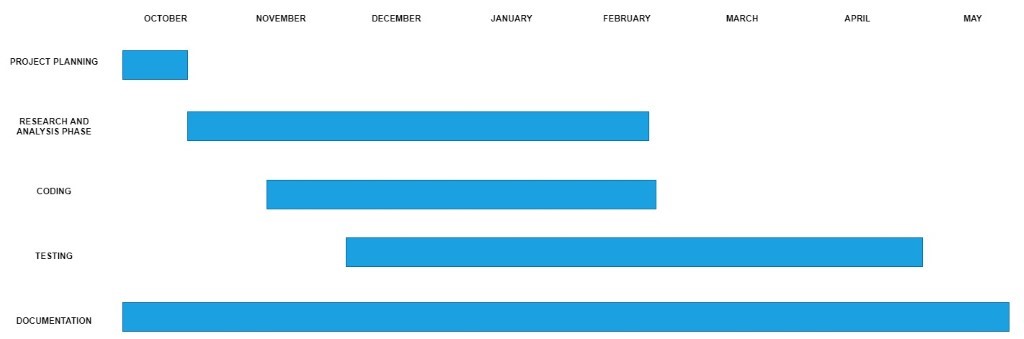


Fig 3.2.2.1 Gantt chart

**3.3. ECONOMIC FEASIBILITY AND OPERATIONAL FEASIBILITY**

This assessment typically involves a cost/ benefits analysis of the project, which helps to determine the viability, cost, and benefits associated with a project before financial resources are allocated. It also serves as an independent project assessment and enhances project credibility.

Our project is economically feasible as here for this we are using our own laptop and will make the model in the laptop itself and train it through present datasets and test it and finally we will be ready to do real time test.

Operational feasibility is the measure of how well a proposed system solves the problems, and takes advantage of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development.

Our system will solve the problem of detection of COVID-19 as our model will give more precious test results by analyzing the CT and X-Ray images provided by the patient, and our model will give very fast results.

**3.3.1. COCOMO MODEL**

The Constructive Cost Model (COCOMO) is an algorithm software cost estimation model developed by Barry W. Boehm. The model uses a basic regression formula with parameters that are derived from historical project data and current as well as future project characteristics. Boehm postulated that any software development project can be classified into one of the following three categories based on the development complexity: organic, semidetached, and embedded in order to classify a product into the identified categories.

COCOMO Model coefficient table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Software Project** | **a*b*** | **b*b*** | **c*b*** | **D*b*** |
| **Organic** | 2.4 | 1.05 | 2.5 | 0.38 |
| **Semi-detached** | 3.0 | 1.12 | 2.5 | 0.35 |
| **Embedded** | 3.6 | 1.20 | 2.5 | 0.32 |

Fig 3.3.1.1 COCOMO Model co-efficient table

The basic COCOMO estimation model is given by the following expression:

Effort=a1\*(KLOC)^a2 PM

Tdev=b1\*(Effort)^b2 Months

Where,

KLOC is the estimated size of the software product expressed in Kilo Lines of Code, a, b, c, d are constants for each category of software products.

Tdev is the estimated time to develop the software, expressed in months. Effort is the total effort required to develop the software product, expressed in person months (PMs).We assume that our project will be an Organic project.

Calculation of effort and development time:

Effort=a1\*(KLOC)^a2

PM=2.4\*(7.5)^1.05

=19.9PM  
Tdev=b1\*(Effort)^b2 Months= 2.5\*(19.9)^0.38=7.7 Months (approx.)

Number of persons= Effort/Tdev= 19.9/7.7= 2.58= 3 person

**CHAPTER 4**

**PROPOSED PLAN**

**4.1. METHODOLOGY**

We present a system to identify covid-19 infection using deep learning techniques. It exhibits higher classification accuracy for CT scan and X-Ray images respectively so the effectiveness of our proposed system will be demonstrated on two systematic datasets of chest radiographs and CT scans. The evaluation will be done on publicly available databases containing both chest X-Ray and CT scan images. The proposed system will extract simple features and then learn the pattern of COVID-19 cases obtained from the patients’ X-Ray and CT-Scan images and compare them to give the results.

**4.2. DATASET DESCRIPTION**

1. SARS-COV-2 Ct-Scan Dataset

The SARS-CoV-2 CT-scan dataset [6] consists of 2482 CT scans from 120 patients, with 1252 CT scans of 60 patients infected by SARS-CoV-2 from males (32) and females (28), and 1230 CT scan images of 60 non-infected patients by SARS-CoV-2 from males (30) and females (30), but presenting other pulmonary diseases. Data was collected from hospitals of São Paulo, Brazil.

2. COVID-CT dataset

To assemble the COVID-CT dataset [7], CT images of patients infected with COVID-19 were collected from scientific articles (pre-prints) deposited in the medRxiv and biRxiv repositories, from January 19 to March 25 and also some images were donated by hospitals. The PyMuPDF software was used to extract images from the manuscripts, in order to maintain high quality. Meta data were manually extracted and associated with each image: patient age, gender, location, medical history, scan time, severity of COVID-19, and medical report. A total of 349 images were collected, from 216 patients.

3. COVID-19 Lung CT Scans

CT scans are promising in providing accurate, fast, and cheap screening and testing of COVID-19 [8]. COVID-CT dataset, containing 275 CT scans that are positive for COVID-19, to foster the research and development of deep learning methods which predict whether a person is affected with COVID-19 by analyzing his/her CTs. We train a deep convolutional neural network on this dataset and achieve an F1 of 0.85 which is a promising performance but yet to be further improved.

4. Chest X-Ray Images (Pneumonia)

The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal) [9]. There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal). Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children’s Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients’ routine clinical care.

**4.3. PERFORMANCE EVALUTION MATRICES**

1. CLASSIFICATION ACCURACY

Classification Accuracy is what we usually mean, when we use the term accuracy. It is the ratio of number of correct predictions to the total number of input samples.

Image for post

It works well only if there are equal number of samples belonging to each class.

1. CLASSIFICATION REPORT

This report consists of the scores of Precisions, Recall, F1 and Support. They are explained as follows –

Precision

Precision, used in document retrievals, may be defined as the number of correct documents returned by our ML model. We can easily calculate it by confusion matrix with the help of following formula −

Precision=TP/TP+FP

Recall or Sensitivity

Recall may be defined as the number of positives returned by our ML model. We can easily calculate it by confusion matrix with the help of following formula −

Recall=TP/TP+FN

Specificity

Specificity, in contrast to recall, may be defined as the number of negatives returned by our ML model. We can easily calculate it by confusion matrix with the help of following formula −

Specificity=TN/TN+FP

Support

Support may be defined as the number of samples of the true response that lies in each class of target values.

F1 Score

This score will give us the harmonic mean of precision and recall. Mathematically, F1 score is the weighted average of the precision and recall. The best value of F1 would be 1 and worst would be 0. We can calculate F1 score with the help of following formula −

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

F1 score is having equal relative contribution of precision and recall.

We can use classification report function of sklearn.metrics to get the classification report of our classification model.

**4.4. FLOW CHART**

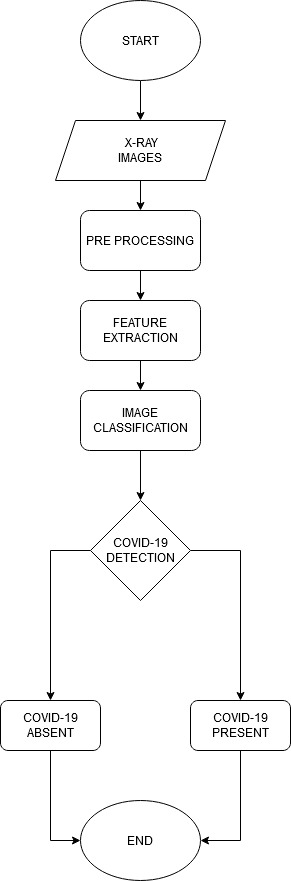
****

Fig 4.4.1 Flow Chart

**4.5 BLOCK DIAGRAM**

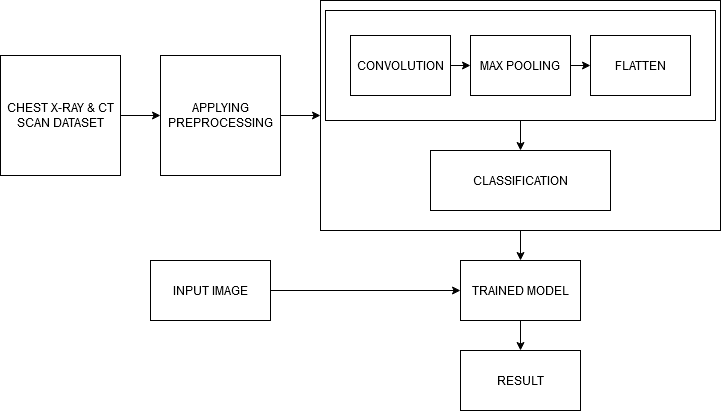
****

Fig 4.5.1 Block diagram

**CHAPTER 5**

**IMPLEMENTATION**

**5.1 IMPORTING DATASET**

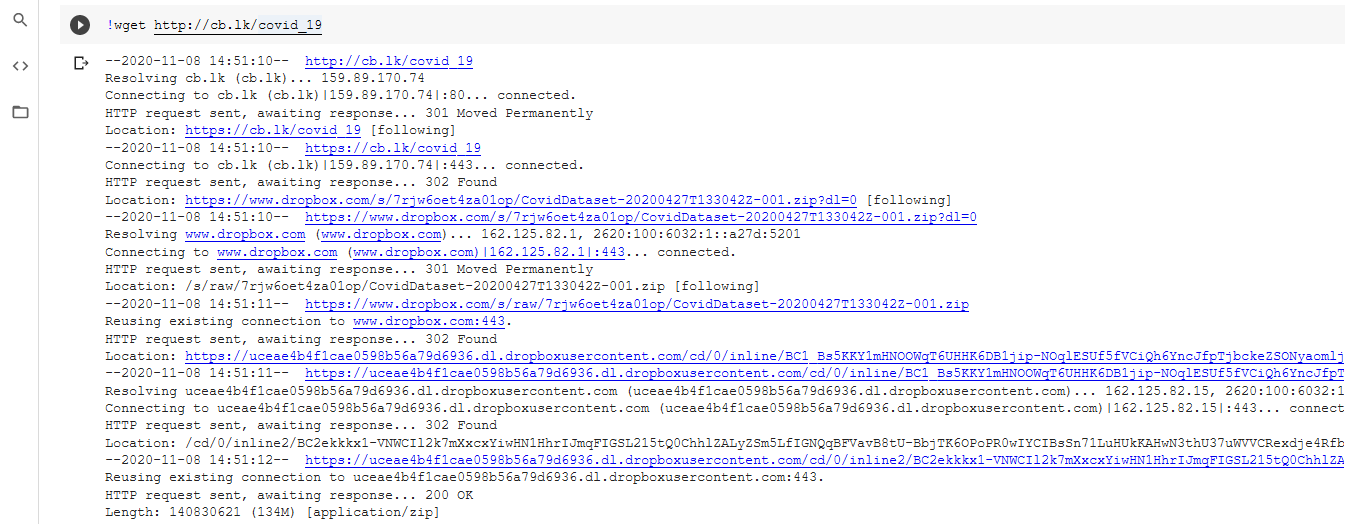
****

Fig 5.1.1 dataset

**5.2 IMPORTING LIBRARIES**

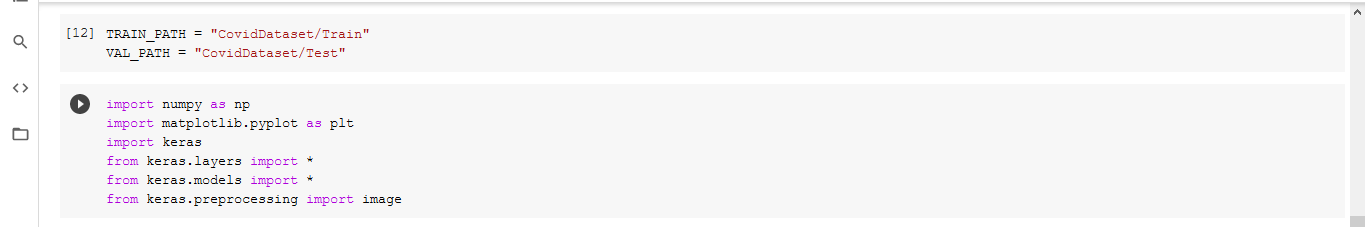
****

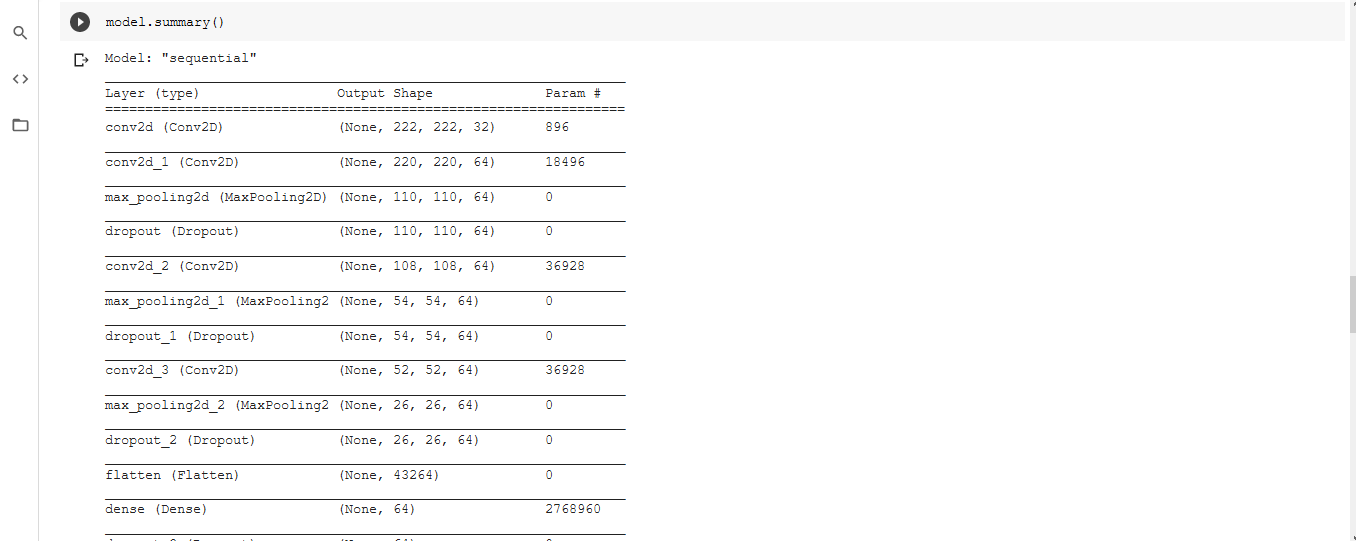
Fig 5.2.1 Libraries

**5.3 IMPLEMENTING CNN**



Fig 5.3.1 CNN

**5.4 MODEL SUMMARY**



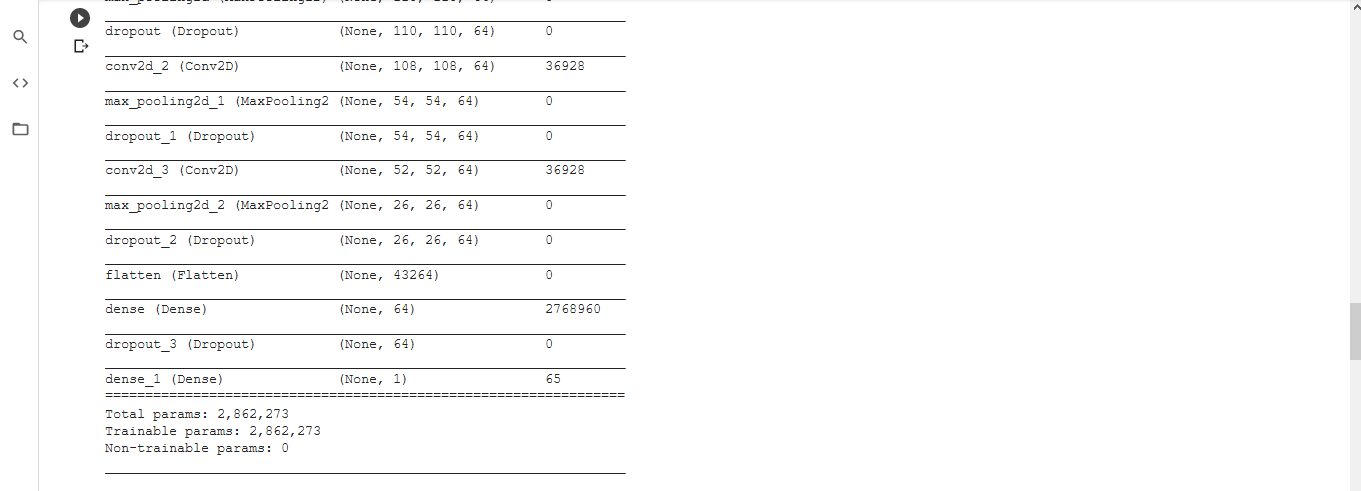
****

Fig 5.4.1 Model Summary

**5.5 PERFORMING OPERATIONS**



Fig 5.5.1 Performing Operation

**5.6 CHECKING ACCURACY**

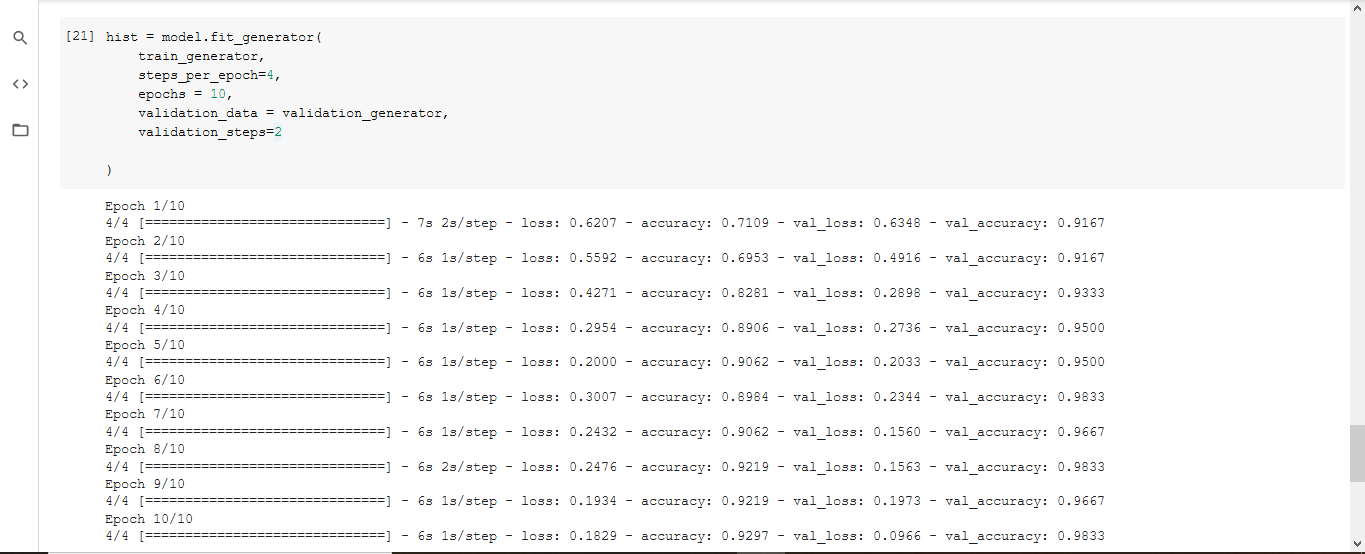


Fig 5.6.1 Checking Accuracy

**5.7 LOSS GRAPH**

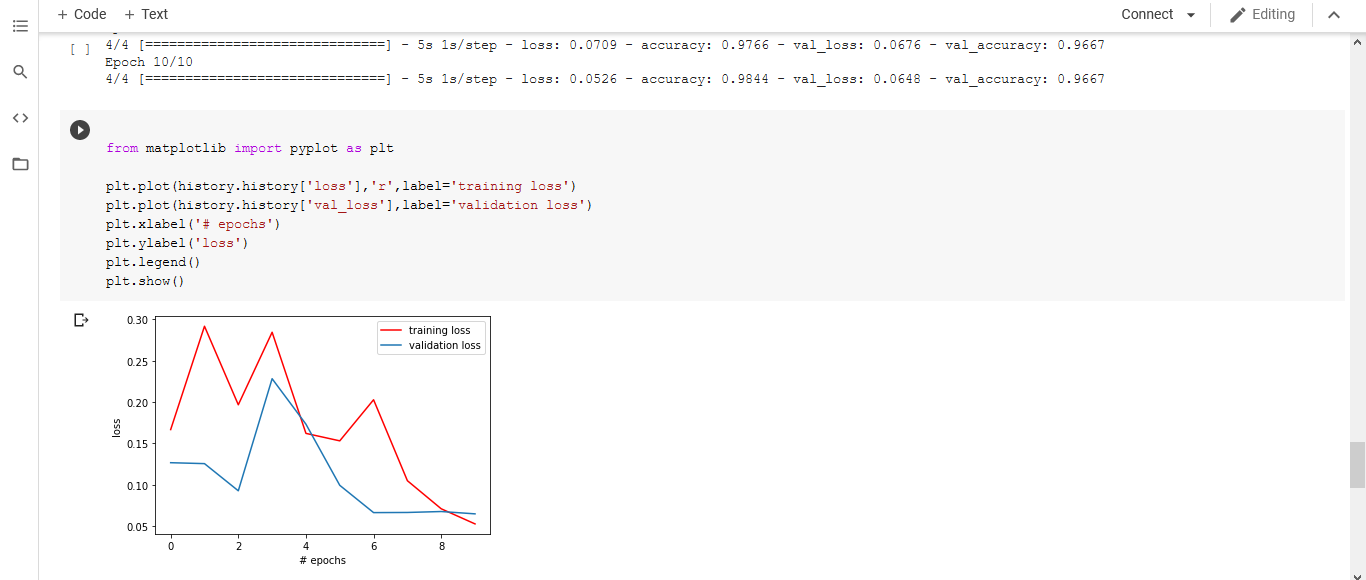
****

Fig 5.7.1 Loss Graph

**5.8 ACCURACY GRAPH**

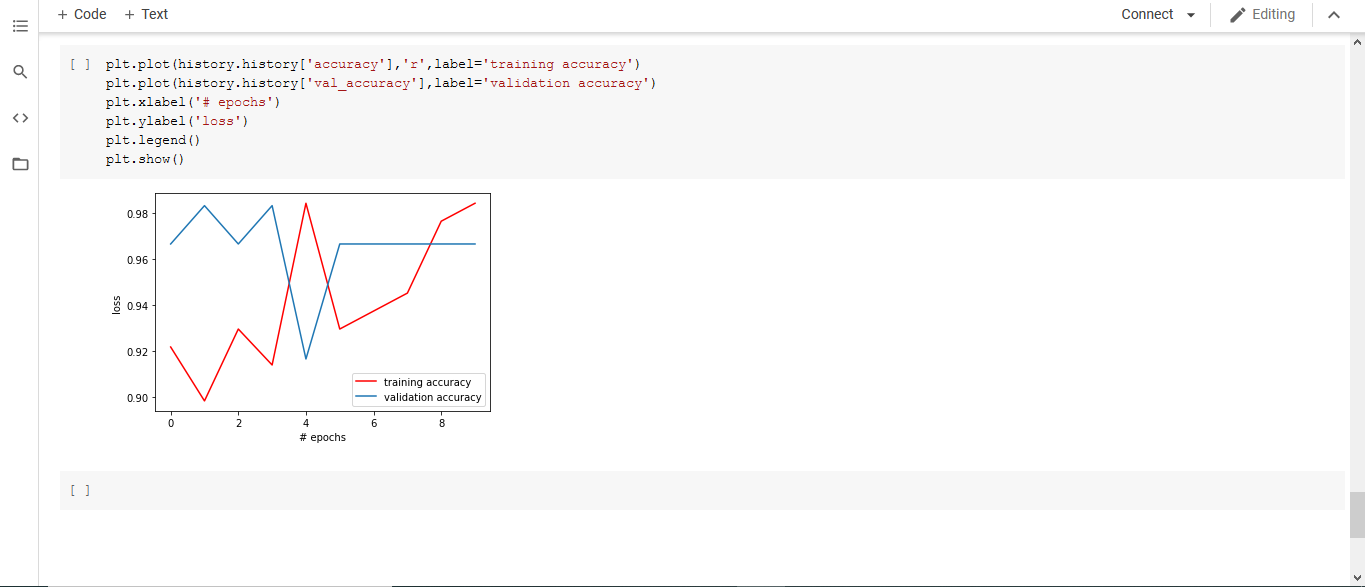
****

Fig 5.8.1 Accuracy Graph

**CHAPTER 6**

**CONCLUSION**

**6.1 CONCLUSION**

Through this project we will try to detect COVID-19 by analyzing the chest CT and X-Ray images, our project will solve the problem of ineffective results and also try to decrease the time taken to detect a COVID-19 patient and hence successfully stop some careless spread of COVID-19 among the population.

**6.2 REFERENCES**

[1] Silva, P., Luz, E., Silva, G., Moreira, G., Silva, R., Lucio, D., & Menotti, D. (2020). COVID-19 detection in CT images with deep learning: A voting-based scheme and cross-datasets analysis. Informatics in Medicine Unlocked, 100427.

[2] Mishra, A. K., Das, S. K., Roy, P., & Bandyopadhyay, S. (2020). Identifying COVID19 from Chest CT Images: A Deep Convolutional Neural Networks Based Approach. Journal of Healthcare Engineering, 2020.

[3] Perumal, V., Narayanan, V., & Rajasekar, S. J. S. (2020). Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features. Applied Intelligence, 1-18.

[4] Singh, D., Kumar, V., & Kaur, M. (2020). Classification of COVID-19 patients from chest CT images using multi-objective differential evolution–based convolutional neural networks. European Journal of Clinical Microbiology & Infectious Diseases, 1-11.

[5] Narin, A., Kaya, C., & Pamuk, Z. (2020). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. arXiv preprint arXiv:2003.10849.

[6] E. Soares, P. Angelov, S. Biaso, M.H. Froes, D.K. AbeSars-cov-2 ct-scan dataset: a large dataset of real patients CT scans for sars-cov-2 identification medRxiv (2020)

[7] J. Zhao, Y. Zhang, X. He, P. Xie Covid-ct-dataset: a CT scan dataset about covid-19(2020) arXiv preprint arXiv:2003.13865

[8] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie, “COVID-19 Lung CT Scans.” Kaggle, 2020, doi: 10.34740/KAGGLE/DS/584020.

[9] Paul MooneyChest X-Ray Images (Pneumonia) Scans Kaggle 2020

**6.3 REMARKS**