A Software Requirement Specification on

**CoVid Detection using Chest Radiography Scans (v1.0.0)**

Submitted to Manipal University, Jaipur

Towards the partial fulfillment for the Award of the Degree of

**BACHELORS OF TECHNOLOGY**

In Computers Science and Engineering

2018-2022

By

Ashray Gupta - 189301068

Niyati Jain - 189301051

Vishwas Tanwar - 189301125

Abhishek Singh - 189301112

Under the guidance of - Ms. Anubha Parashar



**Department of Computer Science and Engineering**

**School of Computing and Information Technology**

**Manipal University Jaipur**

**Jaipur, Rajasthan**

# **1. Introduction**

## **1.1 Purpose**

The purpose of this documentation is to provide a debriefed view of requirements and specifications of a project of Detection of Covid-19 using Chest Radiography scans.

The goal of this project is to input the Chest Radiography scan of the user through our full-fledged functioning Application and predict as an output with the help of our Deep learning model, whether the said patient is suffering from Corona, Pneumonia or none.

In this section(Section 1), a review of the entire document is provided. The reader would get familiarized with the contents before the further details are described.

## **1.2 Documentation**

* All special terms are stated in italics style.
* Main features or important terms are in bold.
* TBD means “To be Decided”. These are the components that are not yet decided.
* For more references, see terminology.

## **1.3 Intended Audience and Reading suggestions**

Anyone with some programming experience, with familiarity in Python and C, can understand this document. The document is intended for developers, software architects, testers, project managers and document writers.

This Software Requirement Specification also includes:

* Overall description of the model
* External interface requirements
* System features
* Other non-functional requirements

## **1.4 Product scope**

COVID-19 (coronavirus disease 2019) is a highly infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first cases were reported in Wuhan; China, in December 2019 before the rapid global spread. The outbreak was subsequently declared as a pandemic on 11th March 2020. Ghana has confirmed 834 COVID cases with 9 deaths (as at April 18, 2020) according to the Ghana health service.

The chest radiographs of patients infected by the novel coronavirus demonstrate characteristic pneumonia-like patterns that can help in the diagnosis, according to a case report by the Chinese Center for Disease Control and Prevention published in the New England Journal of Medicine.

The chest imaging findings are nonspecific and most commonly show atypical or organising pneumonia, often with a bilateral, peripheral and bi-basal predominant distribution.

## **1.5 References**

* IEEE Standard 830-1998 Recommended Practice for Software Requirements Specifications.
* This document is written in GitHub flavored Markdown.

## **1.6 Terminology**

|  |  |
| --- | --- |
| **Term** | **Terminology** |
| User | Any living being who is interacting with the software is a user. |
| System | The package of all the components which takes input and gives output to demonstrate the features of the software is called a system. |
| Database | Collection of information on different topics related to each other. |
| Library | The collection of tracks inside a directory or across multiple directories forms up a library. |
| Store | This is the persistence layer of the whole system. |
| Classifier | An algorithm that implements classification, especially in a concrete implementation. |
| Tag | A label attached to the track which gives extra information about it. |
| GCP | Google Cloud Platform for running the machine learning algorithm. |

# **2. Overall Description**

## **2.1 Product Perspective**

This system consists of the following components:

* Chest-Radiography Scans
* Application Part
* Image Pre-Processing Model
* Feature Extraction Model
* Machine Learning Model

**Chest Radiography Scans** - We need the Radiography scans of the chest of a person who needs to predict whether the given scan is COVID +ve, Normal or Pneumonia infected.

**Application Part -** Application made on Flask framework using Rest API.

A cloud Service Provider for deployment - GCP

**Image Pre-Processing Model -** Pre-processing of images has been done by resizing the images to size 512 x 512, rescaled using min-max normalization technique, enhanced using adaptive histogram equalization and finally saved the images by converting it to an unsigned byte format. The final count of images in our dataset after preprocessing is 2,046 out of which 1,855 belong to the training dataset and the rest belong to the validation dataset.

**Feature Extraction Model -** For feature extraction from an image we have considered both spatial and frequency domain methods. Using spatial domain methods Texture [22], GLCM [23] and GLDM [26] 14, 56, 56 features are extracted and using frequency domain methods i.e., Wavelet [25] and FFT [24] 112, 14 image features are obtained respectively. GLCM and GLDM feature extraction methods are implemented in four directions (different) and whereas eight sub bands are used for wavelet transform.

Global standard features are extracted such as Uniformity, Mean, Std Gradient, Skewness, Max, Entropy, Deviation, Min, Range, RMS, Median, Mean Gradient, Kurtosis, Energy using spatial and frequency domain methods. At last Total 252 features are extracted from each X-Ray image using above methods and by concatenating these feature arrays a Feature pool is derived.

**Machine Learning Model -** After extracting the feature, we constructed a feature pool by concatenating all the feature arrays. Further, we have applied two supervised classifiers XG boost, SVM for multi-classification of data into three classes COVID (+), Normal, Pneumonia and their performance results are compared.

**XGBoost** is a tree boosting system and it is end to end scalable. XGBoost supports regression, ranking, classification and various objective functions as well [12]. The value of parameters of XGBoost in the proposed method were as follows:

*base\_score=0.5, eval\_metric='auc', learning\_rate=0.1, max\_depth=3,n\_estimators=100, n\_jobs=1, nthread=None, num\_class=3, objective='multi:softprob'*

The classes for the classification are 0,1 and 2, 0 if COVID(+), 1 if Normal and 2 if Pneumonia.

**SVM** is a supervised machine learning algorithm and a powerful classifier highly used for classification purposes. SVM also works well for multi-class classifications, for our proposed model we have constructed 3 classifiers and train the data from all the three target classes i.e., COVID (+), Normal and Pneumonia.

The values of the parameters of SVM in the proposed method were as follows:

*C=8, cache size=200, class weight='balanced’, degree=5, kernel='poly'*

The classes for the classification are 0,1 and 2, 0 if COVID (+), 1 if Normal and 2 if Pneumonia.

## **2.2 Product Functions**

* CoVid Prediction
* Pneumonia Virus Prediction
* Complete Report of Person

## **2.3 User Classes and Characteristics**

1. People who have symptoms of COVID-19.
2. People who have had close contact (within 6 feet of an infected person for at least 15 minutes) with someone with confirmed COVID-19.
3. People who have been asked or referred to get testing by their healthcare provider, local external icon or state ​health department.

## **2.4 Operating Environment**

A system with the following minimum specifications:

* Operating System: Windows distribution
* Processor: 2.90Ghz 64 bit
* Memory: 1GB or more

Additional Requirements: A cloud server able to host a TensorFlow application and storage platform for storing the datasets.

We have used Google Cloud Platform as our cloud platform to run our TensorFlow application and Google Drive Platform as our storage platform to save our Training and Validation Datasets.

## **2.5 Design and Implementation Constraints**

* Heavy processing power for running neural networks for classification, thus necessitating the use of a cloud platform.
* Stable and fast internet connection.
* Huge data set so that probability of obtaining the right result is close to 1.

## **2.6 Assumption and Dependencies**

### **2.6.1 Assumptions**

* The basic assumption is that everyone who needs to get tested will have a chest radiography scan saved with himself in a jpg/jpeg format..

### **2.6.2 Dependencies**

* Tensorflow
* Cloud ML
* Google Drive
* REST API

# **3. External Interface Requirements**

## **3.1 User Interface**

The user interface for this project is going to be an app which we have developed with the help of Flask using rest API.

Flask is a commonly used lightweight framework for developing web services in Python.

In my opinion , it is probably the easiest way to implement a web service.

## 

## **3.2 Hardware Requirements**

Following are the minimal hardware requirements to run the application -

1. Disk space - 1GB or more

1. Processor - 2.90Ghz 64 bit
2. Memory - 8 GB
3. Display - (1366 × 768) Capable video adapter and monitor

## **3.3 Software Interface**

Following are the softwares that have been used in developing our Covid Detection application -

Software Used Description

Operating System We have chosen Windows Operating System for it’s best support and user friendliness.

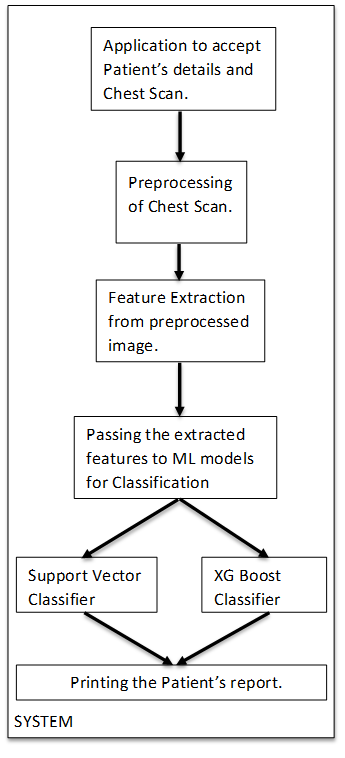
Framework To develop our application , we have chosen xfdvsvsdvsdvsvvsdvvsvsvvsvsvdvvsFlask as our framework due it’s ease of use.

Language Most of the code for this application has vvsdvdsvsvsvsvdssvdvsvsvvsvvsdvsdvs been written in Python language .

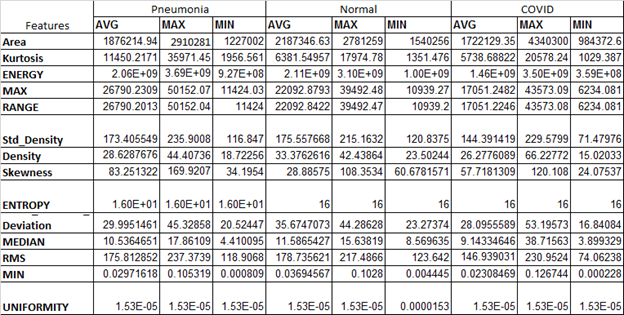
Cloud for Hosting Google cloud has been chosen as our hosting dsvddvsdvsvvsvsdvsvdsvdsvsdvsdvsvsv platform.

**4. System Features**

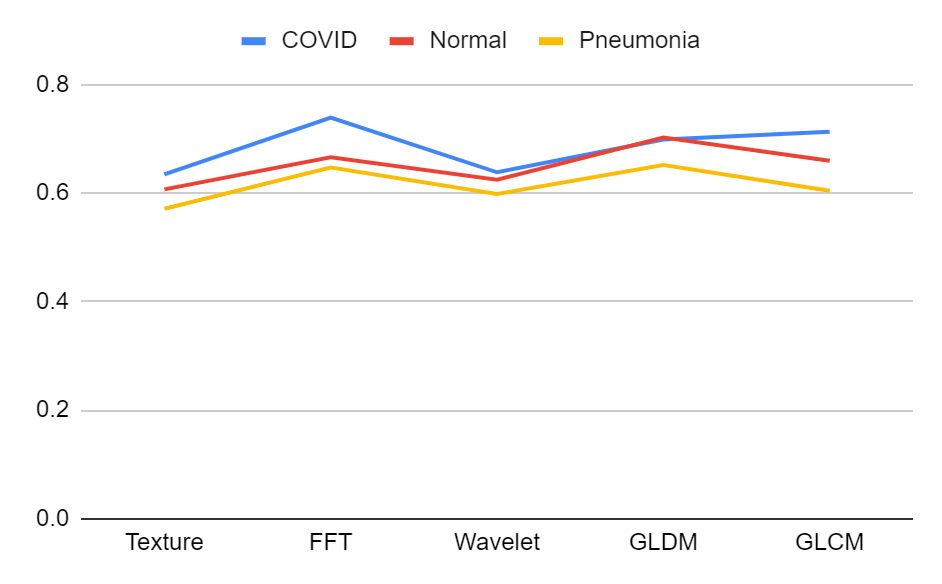
## **4.1 Product Diagram**



## **4.2 Average, Maximum and Minimum value of target classes COVID (+), Normal and Pneumonia for all features extracted by FFT method.**



## **4.3 Average AUC value for COVID (+), Normal and Pneumonia for all feature extraction methods.**



**5. Non-Functional Requirements**

## **5.1 Performance**

The Results obtained from the proposed model have high accuracy . It performs much better than the existing methods, thus, it can be effectively used for detecting the novel coronavirus.

xgboost === 1.1.1

numpy === 1.18.5

Pandas === 1.0.4

Scikit-learn === 0.23.2

Tensorflow === 2.2.0

Flask === 1.1.2

## **5.2 Quality Assurance**

The following are some of the aspects of a Machine Learning model that needs to be tested/quality assured :

* Quality of data
* Quality of features
* Quality of ML Algorithms

Quality to be maintained by :

1. Retraining all of the models and evaluating performance.
2. Tracking the performance of all the models with new data set at regular intervals.
3. Raising the defect if another model starts giving greater accuracy or performing better than the existing model.

## **5.3 Conclusion**

In our software, a frontend application has been designed using the FLASK framework for accepting patient’s details, novel ML is designed by using the global extraction features used for X-ray images for prediction of COVID (+), Normal and Pneumonia cases. Firstly, global image features are used for feature extraction instead of feature extraction from the lesion segmented. This reduces the errors in finding or segmentation of wrong lesions and to find ROI for lesion segmentation. Secondly, multi-classification is being performed instead of binary classification using SVM and XG Boost. Even after finding encouraging results there will be always a scope to extend our work i.e., we can increase the dataset and selection of some other feature extraction methods and feature reduction methods such as LPP [20], recursive feature elimination [21], or combining them with our feature reduction approach. We can even try to make this application more accurate and go on large scale for wider use.