

Title of the project
“Deep Learning based Detection of
Depression”

Under the guidance of
Prof. Ashritha R Murthy

Project Batch No: A1

Project Team:

Sl No	USN	Name	Section	Roll No
1	01JST17CS014	Akshay Suryanarayan Hegde	A	07
2	01JST17CS054	Ganesh S	A	22
3	01JST17CS018	Aniket Kharad	A	08
4	01JST17CS084	Manzoor Ahmed	C	23

Signature of Guide

(Prof Ashritha R Murthy)

Signature of HOD

(Dr. M P Pushpalatha)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

2020-21

Contents

1 Problem Statement	4
2 Aim	4
3 Objectives	4
4 Review of Literature	5
5 Feasibility Study	8
5.1 Technical Feasibility	8
5.2 Financial Feasibility	8
6. Requirement Collection and Categorisation	9
6.1 Stakeholders	9
6.2 Questionnaires	9
6.3 Functional requirements	10
6.4 Non-functional requirements	10
7. Software Requirement Specification	11
7.1 Introduction	12
7.1.1 Purpose	12
7.1.2 Document Conventions	12
7.1.3 Intended Audience and Reading Suggestions	12
7.1.4 Product Scope	12
7.1.5 References	12
7.2 Overall Description	14
7.2.1 Product Perspective	14
7.2.2 Product Functions	14
7.2.3 User classes and characteristics	15
7.2.4 Operating Environment	16
7.2.5 Design and Implementation Constraints	16
7.2.6 User Documentation	17
7.2.7 Assumptions and Dependencies	17
7.3 External Interface Requirements	18
7.3.1 User interface	18
7.3.2 Hardware interfaces	18
7.3.3 Software interfaces	18
7.3.4 Communication Interfaces	18
7.4 System Features	19
7.4.1 System Feature 1	19
7.4.2 System Feature 2	19
7.5 Other Non-functional Requirements	20

7.5.1 Performance requirements	20
7.5.2 Safety requirements	20
7.5.3 Security requirements	20
7.5.4 Software quality attributes	20
7.5.5 Business rules	20
7.5.6 User requirements	21
8 PROJECT TEAM DETAILS	22

1 Problem Statement

A widespread psychiatric illness is depression. It has a direct and indirect effect on economic development. Not only for those affected, but also for their families and their social and work related environments, depression also has significant consequences. It can be the psychological basis for symptoms of panic and anxiety. Increasingly, panic disorder has been focused on health care and the media, impacting young people aged 20-40. Social media is increasingly used by different levels of age group. Patients of psychiatric disorders also turn to online social media and web forums for specific conditions and emotional support information. Although social media can be used to transform the life of an individual as a really beneficial tool, it can create certain conflicts that can have a negative effect. The practices turn out to be extremely difficult with the growing number of users and their content. This motivates for detecting depression from the user posts. The main idea is to detect such psychological problems from user's posts as early as possible.

2 Aim

Deep Learning based detection of Depression.

3 Objectives

- The first step is to pre-process the data brought from text mining into form that is predictable and analyzable for depression detection.
- It is imperative for extracting features from the structured textual data for natural language processing.
- With large pool of features we have to generate models using different machine learning and deep learning techniques.
- To analyze and optimize the models for assessment for depressive symptoms of an individual.

4 Review of Literature

According to DEY, SHARMISTHA, et. al., [1] more than 264 million people across globe suffer from mental health conditions and depression, between 76% to 85% of lower and middle income groups don't receive any treatment for depression. The author also mentions that since the dawn of the beginning of the internet people have begun to dump their thoughts on the Internet which may prove to be a very powerful tool for the diagnosis of a person's mental health problems.

They [1],[2],[3] have given a survey on various algorithms applied on depressions data which includes Machine Learning, Deep Learning, Data Mining etc. also mentioned about various trends in the usage of Random Forest, Hidden Markov Model, Naïve Bayes models for the better analysis of the model.

Mandar Deshpande, Vignesh Rao [2] have proposed a concrete idea on data extraction and data preprocessing using Natural Language Processing. The model proposed incorporates SVM, Multinomial Naïve Bayes as the primary algorithms. Author has taken into consideration F1-Score, Precision and Recall as the accuracy measures.

Md. Rafqul Islam et. al., [3] proposed a model which uses various Machine Learning Algorithms like KNN, SVM, Decision Tree and variations of it, the model has made use of Ensemble learning techniques to increase the performance and Accuracy of the model. There was also a special mention about the tool called LIWC which is used to extract relevant data from Social Media Sites. In the feature extraction step, there is a thorough analysis of characteristics of the data.

Hao Guoa et. al., [4] proposed a model on the basis of "Resting state functional brain networks" which has been widely studied in brain disease, Resting state functional brain networks were constructed for 38 major depressive disorders. The model has an average accuracy of 79.27% and 78.22% for SVM and Neural Network with RBF kernel respectively, with 28 features.

DEY, SHARMISTHA, et. al., [1] has proposed a cumulative survey of various models and approaches which they have taken and summarized in the form of a table and it is shown in the below table which gives us clear insights on how different models give different accuracies.

Author	Year	Approach	Platform Used	Sample Size	Performance and Future Scope
Eichstaedt C. J.[1]	2017	Logistics Regression with 10-fold cross validation	Facebook	683 patients, 114 depressed	
Aldarwish M.M., Hafiz F. A[4]	2017	SVM and Naïve Bayes Classifier).	Twitter, Facebook	2073 post indicating depression and 2073 posts indicating non-Depression	The authors have calculated accuracy, precision and recall.
Reece G. A. and et. al.,[6]	2017	Random forest, Hidden Markov Model	Twitter	204 persons (105 having depression, 99 healthy), 74,990 daily observations	85% accuracy they have achieved
Islam R, Kabir A, Wang H and Ulhaq[2]	2019	Decision Tree, KNN, Support Vector Machine	Facebook	7145 comments, comments that shows depression: 4149, Non depression indicative comments: 2996	Decision tree Recall 98% Precision 58%
Jonathon C and et. al.,[3]	2018	probability calculation, using 10- fold cross validation	Facebook post	683, 114 of whom is detected with depression	AUC is 0.69, give same performance as manual survey

Jana M. Havigerova et.al proposed an unique feature from the text based analysis, the author found out that people suffering from depression tend to use some peculiar semantic linguistics that is they use second person singular to address a person and use quantifiers of extreme poles like “never”, “everything” etc. Some of the second person singular is “I”, “We” etc. Making use of this peculiar feature of Text the author has proposed various models.

Thirteen linguistic variables (6 single morpho-syntactic characteristics, 7 indexes combining more morphosyntactic characteristics) were included into the predictive models. Eight predictive models (for 4 different texts and 2 genders) were created and compared with each other. Author has considered gender depression for the purposes of analysing the models and specifies there are different symptoms based on the gender.

5 Feasibility Study

Depression being the suffered by more than 300 million people across the globe demands for a quick and efficient solution making use of the existing data and the stream of incoming data from various sources, From the dawn of this decade technology has proven to play a significant role in the research and development for vaccines of various diseases, its causes and symptoms. After the analysis we found out the significance, vitality of building a model around the topic of Depression was needed for the hour.

5.1 Technical Feasibility

Since the sole aim of the project is to detect depression from text, it will help most of the people suffering from depression with medical diagnosis. Our model uses deep learning, NLP, feature extractions which are highly CPU intensive tasks. So minimum specifications of the Laptop is essential and our solution is compatible across all the Operating systems installed with the Functional Requirements specified.

But building a complex Deep Learning Model with a lot of hidden layers may prove to be a challenge for the computation. So to accommodate these changes or complex features we have to extend the RAM capacity and also GPU for smooth performance and also a CPU with high computational power, currently these high end systems seems to be infeasible but we can incorporate these once we get our hands on the required resources.

But overall for building a model with good accuracy the current systems and Hardware and Software seems to be sufficient which leads to the technical feasibility of the Project.

5.2 Financial Feasibility

For the system to be highly scalable and responsive, we require the use of good scraping tools, computation, storage. The dataset maybe huge which needs to be present in the RAM all the time, using Google Colab with additional virtual memory solves the problem

The dataset could be fetched and stored in Cloud Services so that it handles the infrastructure and dynamic scaling when a lot of data is fetched and stored.

The most feasible solution for the project considering the Technical, Financial Feasibility with optimal tradeoffs to achieve better performance is chosen and is subjected to implementation.

6. Requirement Collection and Categorisation

6.1 Stakeholders

The major stakeholders of this applications would be:

- Research Professionals
- Medical Professionals
- Health workers and NGO's uplifting the mental health
- Higher authority
 1. Ministry of Health and Family Welfare | GOI
 2. National mental health programme

6.2 Questionnaires

To get a clear picture of the current status of the depression following questions are asked to the respective individuals.

- Research Professionals
 1. What is the current study going on under the field of depression?
 2. What are the various models proposed to solve the problem?
 3. What are the existing solutions for detection of depression?
 4. How is the dataset collected using different tools?
 5. What are the different data preprocessing steps have you applied?
- Medical Professionals
 1. What are the ways in which they diagnose depression?
 2. What are the different symptoms of depression?
 3. How is it that they cure depression by medical diagnosis?
 4. What are the inputs that could be considered for detecting depression?
 5. What extent technology is useful in detection of depression?
- Higher Authority
 1. How could the technology for detection of depression reach every individual?
 2. Which is the most feasible way of distribution of this software across medical institutions?

6.3 Functional requirements

The following are to be considered while developing this project:

1. Provide separate profile for authorized users and research analyst
2. Scan and filter the individual who are at high risk for developing depression
3. Provide a detailed visualization of the current and past cases and their effect in uplifting mental illness through many NGOs and government bodies
4. To verify the timely recovery and regular checkups of individuals suffering from depressive symptoms
5. A platform that provides a session window for interacting between patients and health care workers

6.4 Non-functional requirements

Although not necessary the following are to have utmost importance for smooth and easy usage and maintenance of the application:

1. Every screen of the application should load under 3 seconds.
2. The app should be compatible with previous Android versions.
3. Processing of messages and notifications should be quick when number of users are high.
4. Appropriate measures should be taken to prevent unauthorized access or spamming of the application

7. Software Requirement Specification

Software Requirements for Specification

Deep Learning Based Detection of Depression

Version 1.0

Prepared by:

Akshay Suryanarayan Hegde

Ganesh S

Aniket Kharad

Manzoor Ahmed

JSS Science and Technology University Mysuru

30th Dec 2020

7.1 Introduction

7.1.1 Purpose

The main objective of this document is to illustrate the requirements of the project “Deep learning based detection of depression”. Our project helps in detecting the depression of a person based on their social media activity especially based on their text feedback or comments. The document gives the detailed description of the both functional and non-functional requirements proposed by each and every stakeholder. It also specifies the hardware and software interface requirements.

7.1.2 Document Conventions

Font type: Times New Roman

Font size

- Titles : Bold with 18 size
- Headings : Bold with 14 size with Verdana Font Type
- Body : Normal text with 12 size

Links to ppt are used within the document.

Pictures are shown and flow diagrams are defined within the document.

7.1.3 Intended Audience and Reading Suggestions

This document is meant for the developers of this project, the project guide, and panel of evaluators, government agencies and whoever wants to get an overview of our project. Developers can go through user classes, design and implementation, interfaces and functional and non-functional requirements to get technical details of our project.

7.1.4 Product Scope

Due to Covid-19, the number of people using social media has increased and we can see all the reasons for depression. By using this data we can get more accurate results than previous models, and with the use of several Machine Learning Techniques such as Naïve Bayes, SVM, KNN we can improve the results. The hybrid Deep Learning model has also proven to be of considerable significance in terms of accuracy. The scope of the product is to detect depression using various Text mining techniques and also it can be used for medical diagnosis consulting the experts in the field of medical sciences. The scope of this product can also be extended in getting to know the state of mind of an individual.

7.1.5 References

- [1] Eichstaedt C. J. Facebook language predicts depression in medical records, Psychological And Cognitive Science, vol. 115, No. 44, pp: 11203– 11208, October 2018.

- [2] Islam R, Kabir A, Wang H and Ulhaq A, Depression detection from social network data using machine learning Techniques, Islam et al. Health Inf Sci Syst, vol. 6, No. 8, pp:1-12, 2019.
- [3] Jonathon C and et. al, Facebook language predicts depression in medical records, PNAS ,vol.115 ,No.44, pp: 11203-11208; October, 2018.
- [4] Aldarwish M.M., Hafiz F. A. “Predicting Depression Levels Using Social Media”, King Saud University for Health Science National Guard, 2017 IEEE 13th International Symposium on Autonomous Decentralized Systems.
- [5] Sadeque F, Xu D. and Bethard S.”Measuring the Latency of Depression Detection in Social Media”, WSDM '18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp.: 495-503, October 2018.
- [6] Reece G. A. and et. al., Forecasting the onset and course of mental illness with Twitter data, Scientific Reports | 7: 13006.

7.2 Overall Description

7.2.1 Product Perspective

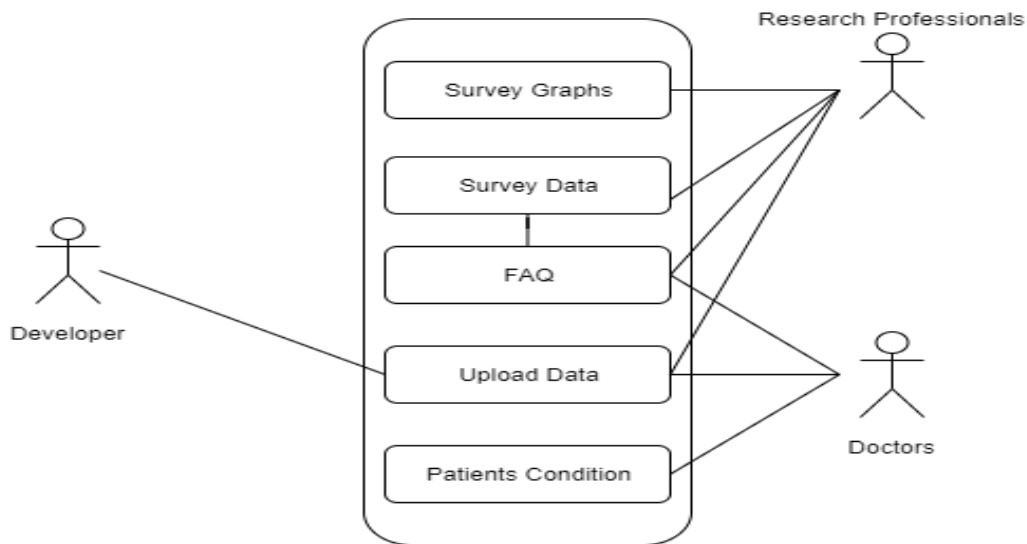


Figure: Use case diagram

7.2.2 Product Functions

The application collects data from twitter,facebook,reddit,linkedIn etc... and applies preprocessing techniques mostly includes Natural Language Processing(NLP) and extracts features from preprocessed data by applying lexicons, the extracted features are passed through machine learning and deep learning models and creates evaluation metrics which includes accuracy, precision, recall etc...

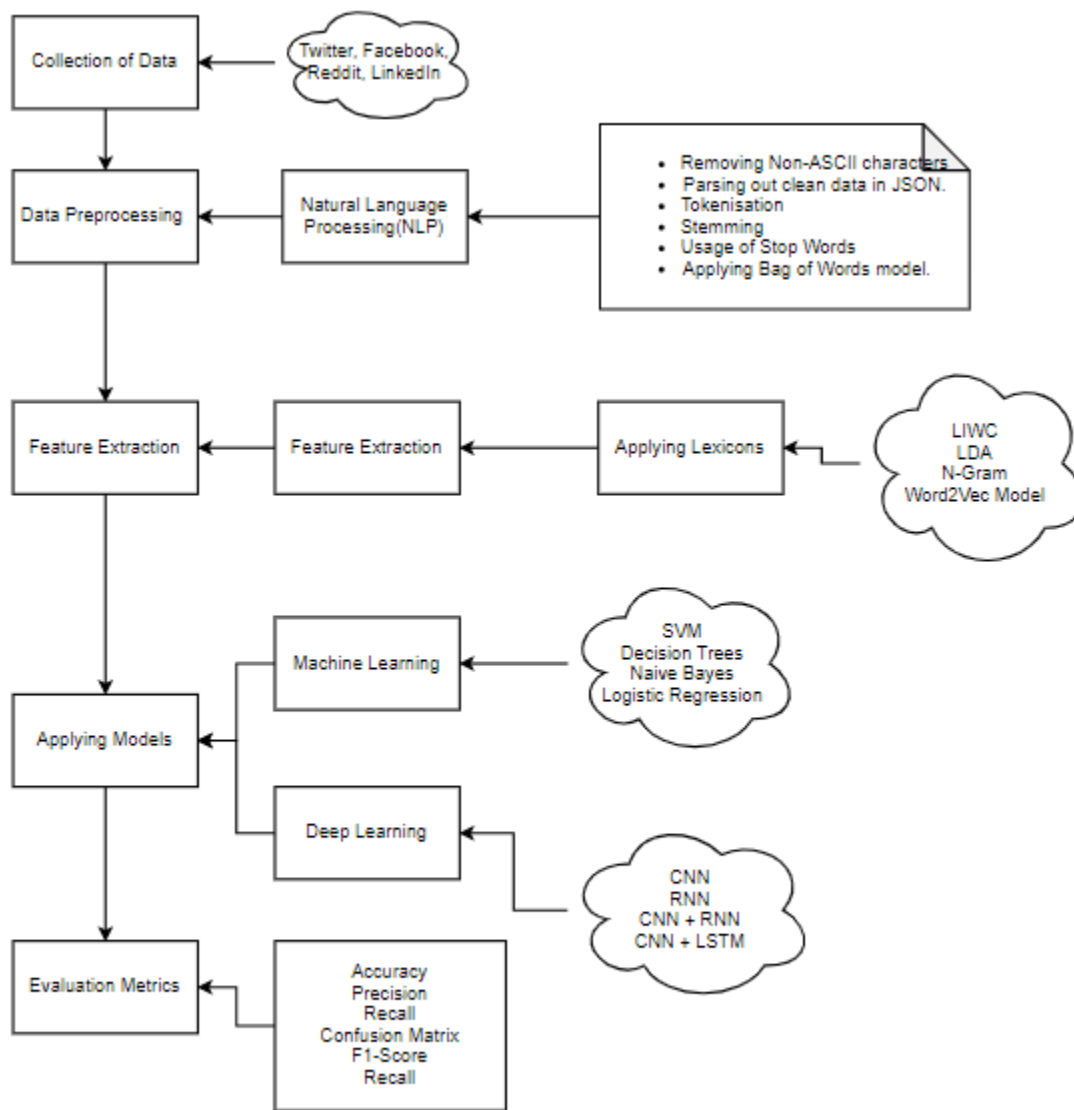


Figure: Architecture

7.2.3 User classes and characteristics

- Doctors/Medical professionals: People who can treat the patients in depression
- Research professionals/Government authorities: These are the people who can stop people from getting depression by providing facilities.

Features available for doctors/medical authorities:

- Can upload data sets and get results.
- Can predict whether a patient is depressed or not.

- Can access FAQs

Features available for Research professionals/Government authorities:

- Can access FAQs
- Can Upload Data sets
- Can get graphs of surveys
- Can get survey data

7.2.4 Operating Environment

- **Software Configuration**

This application will be developed by using python for creating and training of models, for UI web technologies will be used.

-Operating System: Mac Os, Windows 8 9 10, Linux

-Language: Python, Node JS

- **Hardware Configuration**

- 8 GB RAM

- 150 GB of available disk space

- Modern web browser

7.2.5 Design and Implementation Constraints

This software has a Deep Learning model at its core. With thorough learning and understanding of large datasets it is still a challenging task to know and predict the general mental state of an individual at a given particular geographic location at a given period. Although learning from a huge plethora of textual data would make a model strong predictor for a scenario without the knowledge of all socio-demographic variables of a person hinders the ability to gauge the extent of their mental health.

When the person's behaviour is not monitored, there is a chance of misinterpreting actions taken. Now for example, when the variables such as people interacting with the individual and their interaction are not taken into consideration and their chronological order for the study of an individual's behaviour, there is a chance of predicting false positive or false negative for the short

temporary change in the subjects mental health. Other major factors are the cyclic mood swing that a normal person would endure. Now considering many such factors we consider understanding mental health by text is vital yet insufficient as interactions amongst individuals not pertain only to text but the time consumed in creating and publishing the same as well as the frequency in which such textual contents are created and published.

7.2.6 User Documentation

User documentation will be available on the website to guide through our application as soon as it gets released. As the intended use case is for medical practitioners and authorised personals and educated tour of the platform and its intended use case, its security features and confidentiality of the patients is considered and conveyed the same.

7.2.7 Assumptions and Dependencies

The assumptions taken while design the project are as follows:

- The users should have access to the internet and know how to use it.
- The application is hosted on cloud and has low level of latency and moderate traffic.
- The security is class A as the sensitive information such as the textual property of an individual is subject to study and prediction by our model.
- There is potential for data loss. The application should have only authorized access and regulatory policies are to be set.

The dependencies required for the development and deploying the applications are as follows:

The intended use case is to know the general mental state of an individual in particular for depression and predict if such symptoms are possible to develop in the individual in near future and to deem a person to be mentall ill and suffering from depression. This later definition is to be determined by the authorized medical practitioners such as physiatrist.

7.3 External Interface Requirements

7.3.1 User interface

This describes the characteristics of each interface that the required audience has to interact with the software that is to be designed.

As the evaluation of the individual with depressive symptoms is to be analysed the interface that the medical professionals and healthcare workers use would be of simple text based web page which would consist of the various social media accounts of the patient and different kinds of texts as well as different kinds of inferences that the patient concludes to various situations.

At the root of this project we intend to deliver a model precise enough to accurately describe the mental state of the patient given specific information such as their digital behaviour and their interaction and reaction to specific scenarios.

The backend would be a simple ML-engine/DL engine when deployed will render the results seamlessly.

As this is a text based model the time should be minimum against the input in a few tens of milliseconds. Secondly the scalability of this product is for the designated professionals and need to be designed for moderate traffic.

7.3.2 Hardware interfaces

The well trained Deep Learning-model is to be deployed in the cloud and as a web based interface this is available amongst many devices.

7.3.3 Software interfaces

Our applications use Interfaces like :-

1. TWINT a scraping tool written in python to scrape the comments of the user from Twitter website. There are many other scraping tools for scraping various other websites like LIWC for Facebook etc.

7.3.4 Communication Interfaces

The required authentication with the use of product is restrained to medical professionals and authorized health care workers.

7.4 System Features

7.4.1 System Feature 1

Scrapping of various social media websites such as LinkedIn, Twitter, Reddit to get the access to the comments of the users

- **Description**

Scrapping the entire social media using the scraping tools for example we can use TWINT to scrap the Twitter.

- **Stimulus/Response Sequences**

Scrapping the social media we get the comments from the users in an unstructured fashion including hashtags, unicode characters etc.

- **Functional Requirements**

The standard Dataset available for the analysis of depression.

The scraping tools like TWINT to get the extracted data of the users from the social media sites.

7.4.2 System Feature 2

A feature for the medical professionals to upload their own dataset or their conversations with the patients as a part of the dataset.

- **Description**

We provide a facility for the medical professionals to upload their own dataset in the format specified to run it against our model and perform exploratory data analysis.

- **Stimulus/Response Sequences**

Performing so we get the output for every dataset and we store the accuracy and the features of every dataset in the form of logs.

- **Functional Requirements**

The new dataset in the specified format.

Tools for Exploratory Data analysis.

7.5 Other Non-functional Requirements

7.5.1 Performance requirements

This request from various clients should be responded to in real time. The development of this project is provide feasibility to the analyst to choose between the choice of having to understand the patient by knowing few sentences that he/she writes or to dive deep into their digital world and excavate their trail through social engineering

7.5.2 Safety requirements

This product is designed to understand and try to tell before if a person will go into depression. The intended use is to suggest and cannot be confirmed without a medical practitioner.

7.5.3 Security requirements

Privacy is the most precious property of an individual in this world and we keep it safe. The user data is not used for any study or research work without their consent. The models are trained from the data that is publicly accessible, and the user data is discarded as soon as possible. Appropriate measures should be taken to prevent unauthorized access or spamming of the application.

7.5.4 Software quality attributes

As a web based solution for the medical field this is available through the world is most flexible as can be accessed by any kind of device. Maintaining the product is necessary as a cloud based solution it has to be robust, reliable and accessible all the time.

Even though the web based solution is correct most of the time, it has its limitations as the data that is trained may not be sufficient to encompass all behaviour patterns in this gargantuan population to meet all the social and demographic variables of an individual.

7.5.5 Business rules

The set of operations that can be carried out as strictly accessible through medical professionals and authorized health care workers. The patient's details are to be filled by them.

7.5.6 User requirements

Every screen of the application should load under 3 seconds. The app should be compatible with previous Android versions. Processing of messages and notifications should be quick when number of users are high so that the latency is set to minimum. The webapp should be responsive and the control and navigation should be simple to ensure ease of use. The proper user interface, user manual, online help and the guide must be sufficient to educate the users on how to use the app without any problems.

8 PROJECT TEAM DETAILS



Akshay Suryanarayan Hegde
01JST17CS014



Ganesh S
01JST17CS054



Aniket Kharad
01JST17CS018



Manzoor Ahmed
01JST17CS084