



**TRIBHUVAN UNIVERSITY**  
**INSTITUTE OF ENGINEERING**  
**HIMALAYA COLLEGE OF ENGINEERING**

[CODE: CT 755]

**A**

**FINAL YEAR PROJRCT REPORT**

**ON**

**MEDICAL REPORT ANALYSIS SYSTEM**

**BY**

**Prabesh Kumar Shrestha (070/BCT/32)**

**Praveen Dhungel (070/BCT/33)**

**Sahit Baral (070/BCT/41)**

**Season Maharjan (070/BCT/48)**

**A REPORT SUBMITTED TO DEPARTMENT OF ELECTRONICS AND  
COMPUTER ENGINEERING IN PARTIAL FULFILLMENT OF THE  
REQUIREMENT FOR BACHELOR'S DEGREE IN  
COMPUTER ENGINEERING**

**DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING  
LALITPUR, NEPAL**

August 2017

# **MEDICAL REPORT ANALYSIS SYSTEM**

**BY**

Prabesh Kumar Shrestha (070/BCT/32)

Praveen Dhungel (070/BCT/33)

Sahit Baral (070/BCT/41)

Season Maharjan (070/BCT/48)

**PROJECT SUPERVISOR**

**Er. Pratap Sapkota**

A Report Submitted In Partial Fulfillment of the Requirements for the  
Degree of Bachelors of Computer Engineering

Department Of Electronics and Computer Engineering

**HIMALAYA COLLEGE OF ENGINEERING**

Tribhuvan University

Lalitpur, Nepal

*August, 2017*

## **ACKNOWLEDGEMENT**

We the entire project team express our gratitude to Institute of Engineering, Pulchowk for including provision of major project in the syllabus of BCT fourth year.

Also, we are thankful to the management of Himalaya College of Engineering for providing us the chance to build software during the major project & managing the resources and specialists to assist our project.

We are grateful to our Head Project Coordinator Er. Alok Kafle who will be the part of our project team with their assistance, advices and suggestions in whatever way they could.

We are indebted to our respected HOD Er. Ashok GM and Deputy HOD Er. Devendra Kathayat, who are the actual heroes of this action for they have managed everything we students need.

### **Group Members**

## ABSTRACT

This project titled “Medical Report Analysis System” is supposed to be the guide for those persons from non-medical fields as well for those who are from medical fields. This project ensures that the medical diagnosis could be much easier to understand to every person once the project is completed. This project holds different features that enhances the medical field and its diagnostic reports.. As every people may not be familiar with the health reports that they receive from the Laboratory and the research lab, this application when developed helps the people to understand the report. This application tend to guide through the complex result generated by the research lab into the simpler ones by providing necessary data on the report.

This application is targeted for all common people and to those who are related to medical field. It contains different features so as to enhance the technology in terms of medical field. The application imports digital image and by the use of CVIP we can process the image and generate terms we need and the application helps to crawl through the internet to generate the results. This application will use different algorithms like C4.5, PageRank and AdaBoost. The application will find the best result comparing different sites. Initially the user will be asked to import the report and the processing of the image is done afterwards. The processed image has some terms that the application will extract from different websites and the user can print the result. This applications contains BMI calculator, symptoms checker, online doctor support, chat services, forum for users etc. This application refers to the systems and processes as the interaction between patient, doctors and different users who are in connection via internet. Those who don't have sufficient time to spend on their checkup or visit the doctor can now easily diagnose their report in their home/office. The main aim of this application is to enhance the technology in terms of medical sector and make an easy environment for the users to be familiar with their health status without visiting any hospital.

**Keywords:** *image processing, medical, diagnose, research lab, C4.5, PageRank*

# TABLE OF CONTENTS

ACKNOWLEDGEMENT .....	ii
ABSTRACT.....	iii
LIST OF FIGURES .....	vi
ABBREVIATIONS .....	vii
1. INTRODUCTION .....	1
1.1 Background.....	1
1.2 Problem statement.....	2
1.3 Objectives .....	2
1.4 Features .....	2
1.5 Scope and application .....	3
2. LITERATURE REVIEW .....	4
3. REQUIREMENT ANALYSIS AND FEASIBILITY STUDY .....	6
3.1 Feasibility Analysis of the project: .....	6
3.1.1 Economic feasibility:.....	6
3.1.2 Operational feasibility: .....	7
3.1.3 Schedule Feasibility: .....	7
3.1.4 Technical feasibility: .....	7
3.2 Project Requirement.....	8
3.2.1 Functional Requirement .....	8
3.2.2 Non Functional Requirement .....	8
3.2.3 Model and Software Process .....	9
4. SYSTEM DESIGN AND ARCHITECTURE .....	10
4.1 System Flow.....	10
4.2 Use Case Diagram.....	11
4.3 DFD level 0.....	12
4.4 DFD level 1 .....	13
4.5 Sequence Diagram .....	14
4.6 State Chart.....	15

5. METHODOLOGY .....	16
5.1 Optical character recognition .....	16
5.1.1 Techniques for OCR .....	17
5.2 C4.5 Algorithm:.....	20
5.3 PageRank.....	22
5.4 Agile Methodology.....	23
5.4.1 Why Agile.....	24
5.5 Testing .....	25
5.5.1 Unit Testing .....	25
5.5.2 System Testing.....	25
5.5.3 Performance Testing .....	26
5.5.4 Verification and Validation.....	26
6.RESULT ANALYSIS AND FUTURE ENHANCEMENT .....	27
6.1 Analysis and Requirement Collection .....	27
6.2 Backend Work .....	27
6.3 Frontend Work.....	28
6.4 Future Enhancement .....	28
7. CONCLUSION.....	29
REFERENCES .....	30
SCREENSHOTS: .....	31

## **LIST OF FIGURES**

Fig 1 System Diagram of MRAS	10
Fig 2 Use Case Diagram	11
Fig 3 DFD_0	12
Fig 4 DFD_1	13
Fig 5 Sequence Diagram	14
Fig 6 State Chart	15
Fig 7 C4.5 Algorithm	20

## **ABBREVIATIONS**

App	: Application
BCT	: Bachelor in Computer Engineering
BMI	: Body Mass Index
ES	: Expert System
FIG	: Figure
GUI	: Graphical User Interface
HCOE	: Himalaya College of Engineering
HOD	: Head of Department
IDE	: Integrated Development Environment
JSON	: JavaScript Object Notation
MRAS	: Medical Report Analysis System
MY SQL	: MY Structured Query Language
OCR	: Optical Character Recognition
UI	: User Interface



# **1. INTRODUCTION**

## **1.1 Background**

This document is an introduction of the “Medical Report Analysis System” an android application. After the analysis, we prepared this document giving description for overview of this application and its development criteria. This application is supposed to simplify the medical reports with necessary and related details on the particular sector. Our application is supposed to be developed as an expert system. The strength of an ES derives from its knowledge base – an organized collection of facts and heuristics about the system’s domain. It even uses the inferences engine to combine the facts of a specific case with the knowledge contained in the knowledge base to come up with some recommendations. MRAS is simply for the use of ordinary people for the reading of medical reports using OCR and suggestion if simple facts in respect to the obtained values in the reports. This application also helps people to know their BMI by filling the necessary information on the application by the user after this application. MRAS is simply a package application for the normal people who can have benefits as report scanning and optical character recognition with suggestions for the medical report values and the BMI calculation helps people for their body mass index values defining their fitness level i.e. health related awareness, the emergency call feature here deals with calling the blood bank, ambulance service, police control and many hospitals. As everyone visits a medical personnel or the doctor for the review of medical report, this application takes that extra work out as to instantly analyze the medical report just in a click of the image of medical report, we think the application is helpful and finds its true scope and improved features in the near future. The application MRAS can be very helpful for everyone and provides a basic scope of technology towards the medical field. MRAS being on any android device as world is contracting with the growth of the smartphone technology, the number of users is increasing day by day which helps the application reach in more hands.

## **1.2 Problem statement**

With the technological advancement and urbanization more people are captivated by their regular works. People lack enough time to visit doctors and get concerned about their health status. Due to this reason people are unaware of their health and are suffering from different sorts of diseases. Hence we came up with an idea so that the people may find their health care easier than before. The lab test are easily done anywhere else like in research labs, medical stores, pharmacies, clinics. Here our application afterwards scans the report and provides efficient results on certain topics. This application requires OCR which includes scanner or from the drive to upload the data. There are no specific devices to scan the report. Also this application needs internet connection in order to get accurate results. This application can be operated via any Android smart phones.

## **1.3 Objectives**

- To use the smartphone technology in the field of health
- To be able to read the medical report
- To provide necessary information to the user

## **1.4 Features**

- Use the OCR technology to read the medical report image and provide the result to the user
- Use BMI calculator of an individual and provide suggestions
- Use the emergency call feature within the app

## **1.5 Scope and application**

This system is primarily targeted for people from any sector. This application can be handy for people those who are concerned with their health activities. There is no specific target for the people using it because this application can be used by anyone from anywhere else having an internet connection. The main application of this system is to provide suggestions to people in instant without visiting a doctor. Likewise in case of BMI and other simple additional features like calling blood bank and ambulance just in a click within the app. The main scope of the application is to use the medical report of the people and suggesting the people about the related values respectively.

## **2. LITERATURE REVIEW**

In context of Nepal, there are very limited services on medical field using latest technologies. Government of Nepal has announced Patan hospital as the Central hub to connect the telemedicine center. But due to lack of resources and proper manpower the work is conducted much slowly. This service is provided to more than 40 districts of Nepal and far from Kathmandu, people who are in rural areas are using this services. We realized that the system can manage the information of the patient efficiently. It would be much great if the people from different rural location could have an application which could use different features combined in one form to connect to different hospitals, chat with the doctors, have recommendations and find the medicine.

One of the largest area of applications of artificial intelligence is expert system. While referencing the history we can find the expert system [1] which were built in health sector but there are very small attempts in case of Nepal. Since the health depends upon the geographic area and people's life style it was felt that there is lack of such system.so the idea of our project emerged to provide the platform to discovery of the disease and recommend the patients.

In the beginning of the research we decided to monitor the hospital reports related to specific diagnosis and find out there were only the numerical values in respect to each type of test, the main theme of the project will be reading the values from the reports and process those data using optical character recognition (OCR) and following the process using some algorithm related to AI and conclude the reports by giving an exact conclusion on that specific tests.

During our research phase we also get a system i.e. mayo clinic[2] which is an nonprofit medical practice and medical research group based in rochester, minnesota. It is the first and largest integrated nonprofit medical group practice in the world, employing approx. 4000 of physicians and scientists etc. this system dealed with a web based software that needed some input symptoms.

In the research we find the recognition of the characters were done in various fields using various algorithms of AI. Sandhya Arora [3]in, used four feature extraction techniques namely, intersection, shadow feature, chain code histogram and straight line fitting features. Shadow features are computed globally for character image while intersection features, chain code histogram features and line fitting features are computed by dividing the character image into different segments. On experimentation with a dataset of 4900 samples the overall recognition rate observed was 92.80% for Devanagari characters.

M. Hanmandlu, O.V. Ramana Murthy have presented in their study the recognition of handwritten Hindi and English numerals by representing them in the form of exponential membership functions which serve as a fuzzy model.[4] The recognition is carried out by modifying the exponential membership functions fitted to the fuzzy sets. These fuzzy sets are derived from features consisting of normalized distances obtained using the Box approach. The membership function is modified by two structural parameters that are estimated by optimizing the entropy subject to the attainment of membership function to unity. The overall recognition rate is found to be 95% for Hindi numerals and 98.4% for English numerals.

### **3. REQUIREMENT ANALYSIS AND FEASIBILITY STUDY**

#### **3.1 Feasibility Analysis of the project:**

A feasibility study is a preliminary study which investigates the information needs of prospective users and determines the resources requirements, costs, benefits, and feasibility of a proposed system. A feasibility study takes an account of various constraints within which the systems should be implemented and operated. The main objective of the feasibility study is to determine whether the project would be feasible in terms of economic, technical and operational feasibility. The results of this analysis are used in making the decision whether to proceed with project or not. Feasibility study includes:

##### **3.1.1 Economic feasibility:**

Economic feasibility attempts to deal with the costs of developing and implementing a new system, against the present day condition. This feasibility study gives the top management economic justification for the new system. As the system requires android devices which also comes in affordable price, all people can use the system in low cost. A simple economic analysis which gives the actual comparison of costs and benefits are much more meaningful in this case. This app can replace many other apps to use for medical purposes. The creation of the application is also not costly.

### **3.1.2 Operational feasibility:**

Operational feasibility is a measure of how well a proposed system solves the problems, and take advantage of the opportunities identified during scope definition. A system with easy interface will always help the user to easily use the system. Nowadays more people are using android devices, so the system can be easily operated by the people without having high experience. Thus the system is said to be operationally feasible.

### **3.1.3 Schedule Feasibility:**

A project will fail if it takes too long to be completed before it is useful. Typically this means estimating how long the system will take to develop, and if it can be completed in a given time period using the methods like payback period. Schedule feasibility is a measure of how reasonable the project timetable is. Given our technical expertise, are the project deadlines reasonable? Some project is initiated with specific deadlines. It is necessary to determine whether the deadlines are mandatory or desirable.

### **3.1.4 Technical feasibility:**

Evaluating the technical feasibility is the trickiest part of a feasibility study. This is because, at this point of time, not too many detailed design of the system, making it difficult to access issues like performance, costs etc. Before commencing the project, we have to be very clear about what are the technologies that are to be required for development of the new system. The system uses more software component rather than hardware. All the tools necessary for this system is easily available. It uses Android Studio for the application development. Nowadays the network is almost available in all parts of country so the application can be accessed easily. As using android device over a 3G network also makes it feasible to access from different location. Thus the system is said to be technically feasible. Though all the tools seems to be easily available, there will be other challenges too.

## **3.2 Project Requirement**

### **3.2.1 Functional Requirement**

Functional requirements are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this, the functional requirements that the system must require are as follows:

- System must perform optical character recognition
- System must be able to show the results after analysis of medical report
- System should process the BMI calculation
- System must be able to perform the emergency call function

### **3.2.2 Non Functional Requirement**

Non-functional requirement is a description of features, characteristics and attribute of the system as well as any constraints that may limit the boundaries of the proposed system. The non-functional requirements are essentially based on the performance, information, economy, control and security efficiency and services. Based on these, the non-functional requirements are as follows:

- User friendly
- To perform with efficient throughput and response time
- System should provide better accuracy



### **3.2.3 Model and Software Process**

For the development of the software starting from beginning to the completion of the final product, specific models have to be used. In software development lifecycle, it is possible to use any software models like waterfall, incremental, prototype, agile etc. Based on project requirements, agile model has been used. It is a type of incremental model in which software is developed in rapid cycles. Whenever new changes have to be made, agile model allows easy implementation at very little cost. It minimizes the risk by developing software in small iteration. Planning, developing and testing phase has been iteratively used to implement new changes easily that reduces the risk of project failure.

## 4. SYSTEM DESIGN AND ARCHITECTURE

### 4.1 System Flow

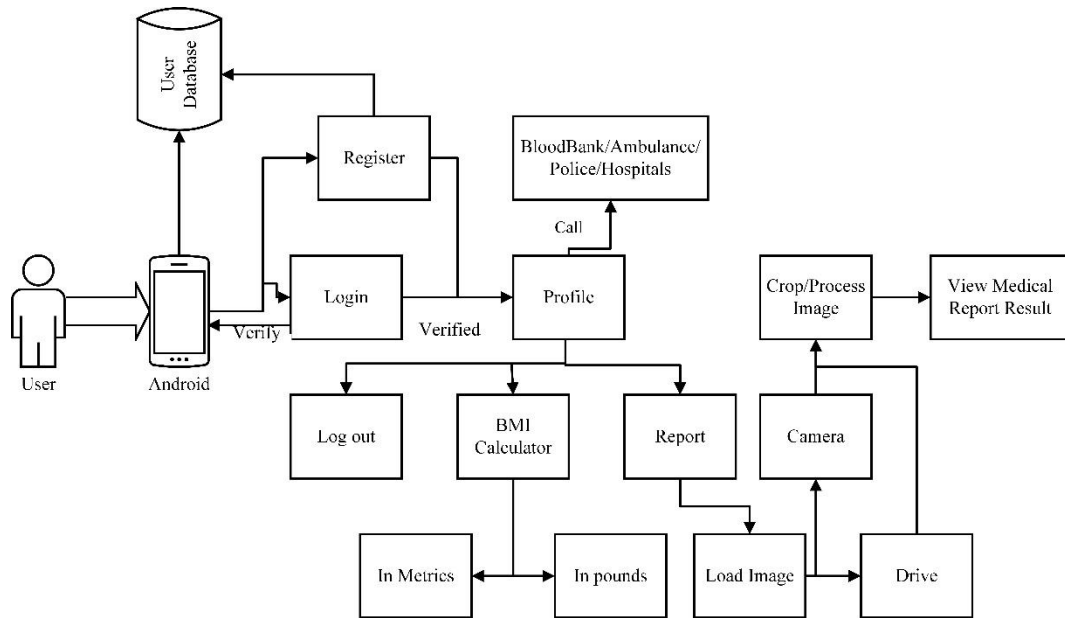


Fig.1: System Diagram of the application

The system flow above shows the MRAS application flow in basic level as we can observe above that the user's scope and the applications features and availability of the basic notations what user can do or will be doing. simply describing the user can register his/her email id and then login to his/her profile, then simply he can use the features like Medical report analysis, BMI calculation, Emergency call feature, suggestions based on analysis of the reports etc.

## 4.2 Use Case Diagram

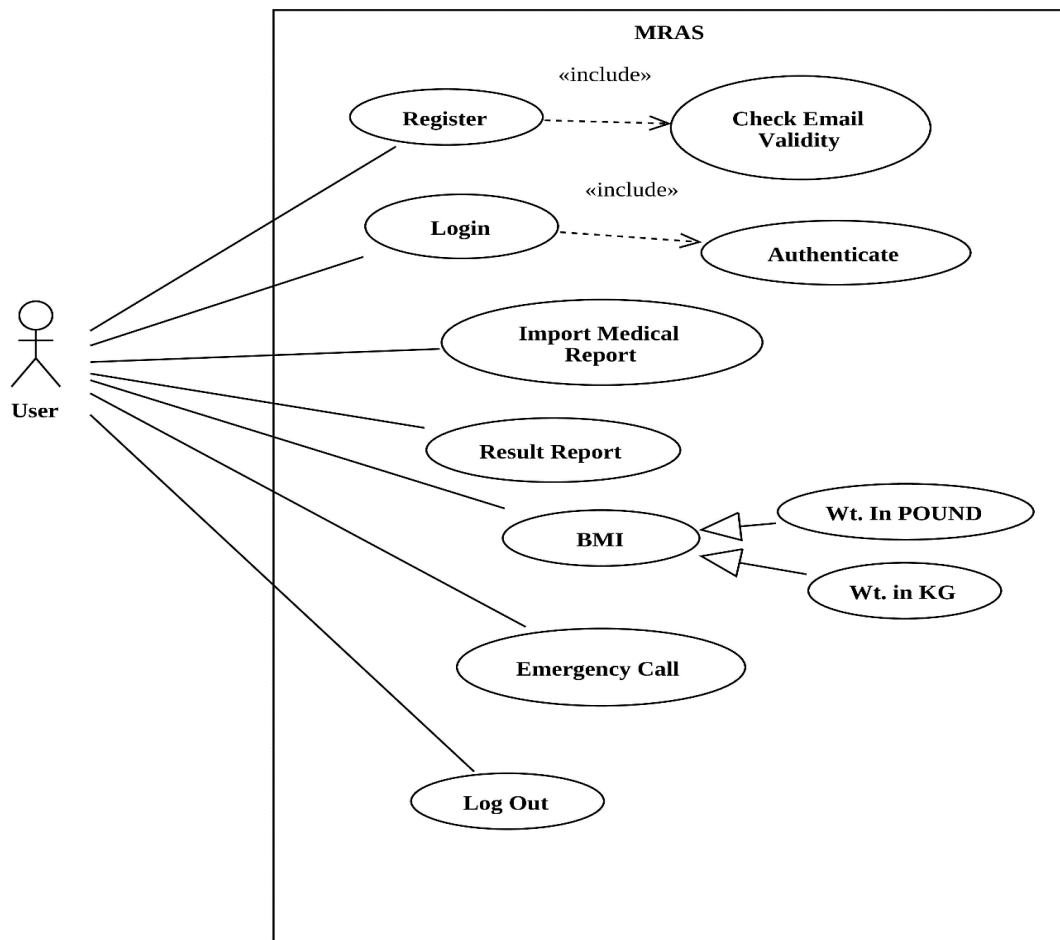


Fig 2: Use Case Diagram

The Use case diagrams are usually referred to as behavior diagrams used to describe a set of actions i.e. use cases that some system or systems should or can perform in collaboration with one or more external users of the system i.e. actors. Above diagram consist of the user as an actor and its cases inside the box representing the system.

### 4.3 DFD level 0

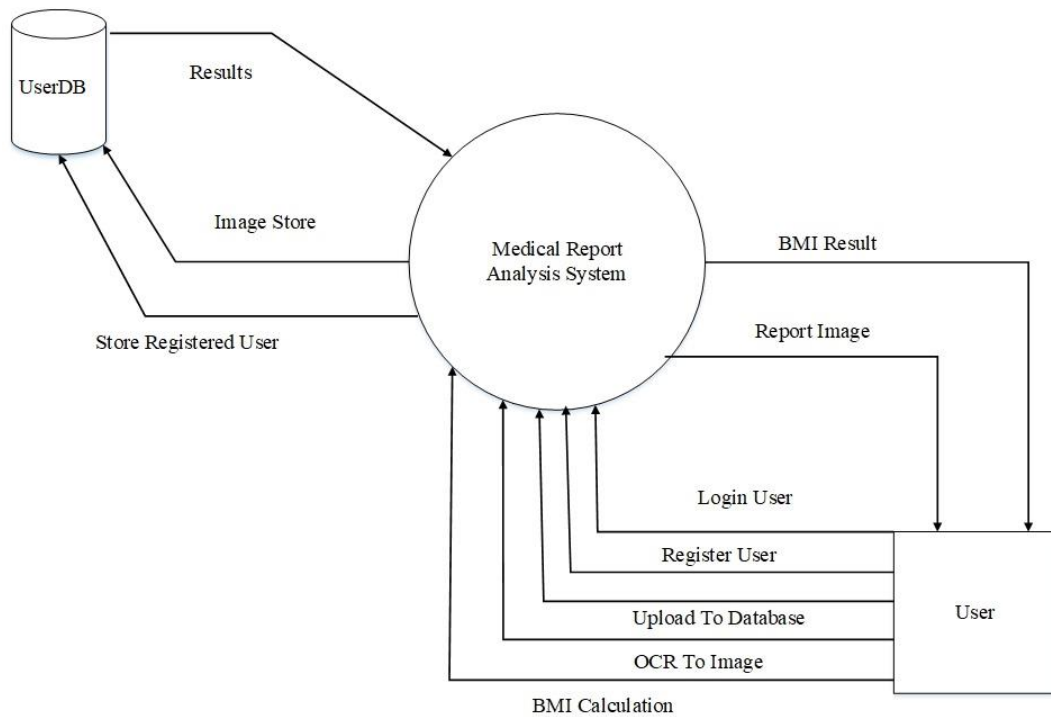


Fig 3: DFD Level 0

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modelling its process aspects. Above DFD level 0 shows the flow through the user to the application and application to database and viseversa respectively.

#### 4.4 DFD level 1

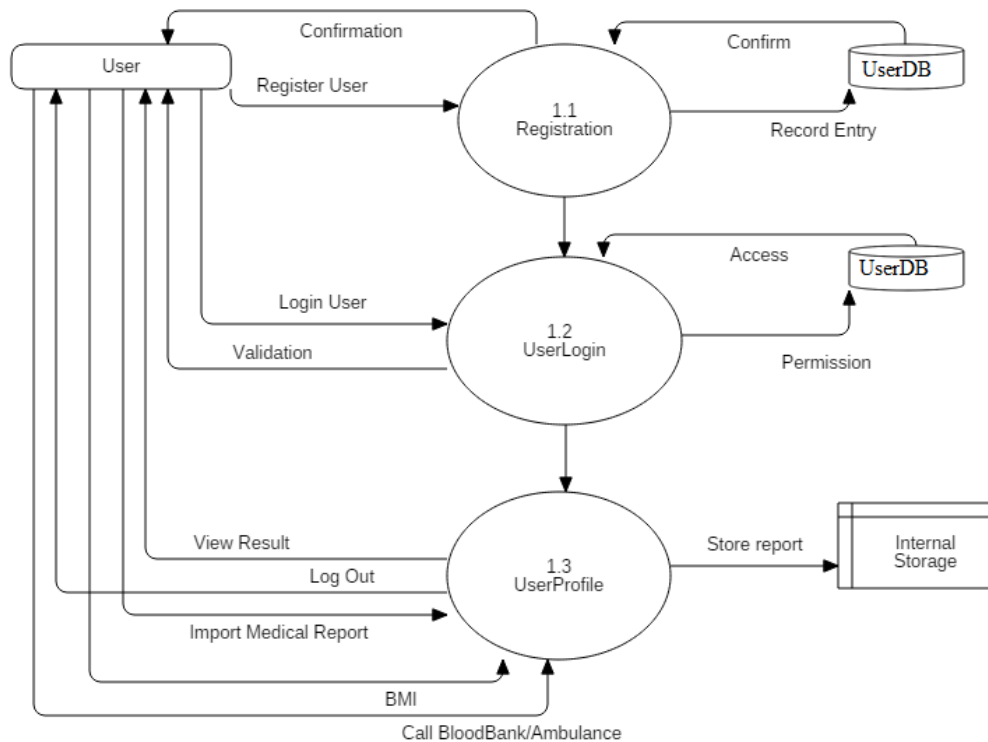


Fig 4: DFD Level 1

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modelling its process aspects. The DFD level 1 shows the flow in more detail than the DFD level 0 and shows the interactions too.

## 4.5 Sequence Diagram

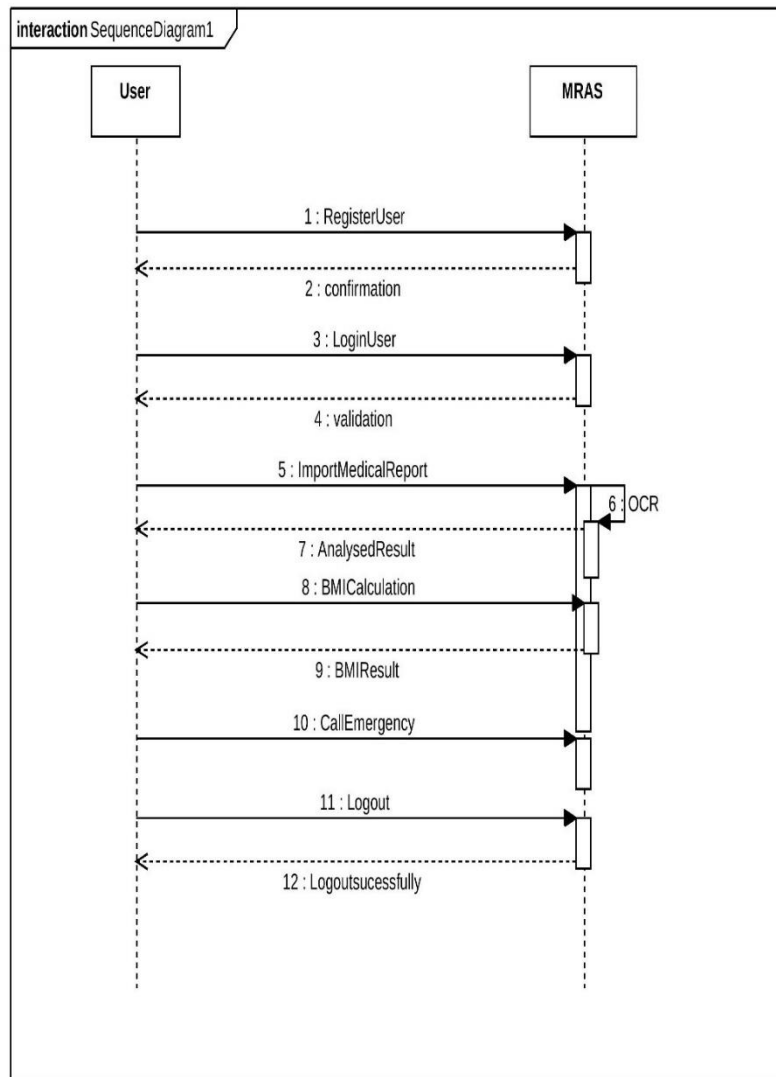


Fig 5: Sequence Diagram

## 4.6 State Chart

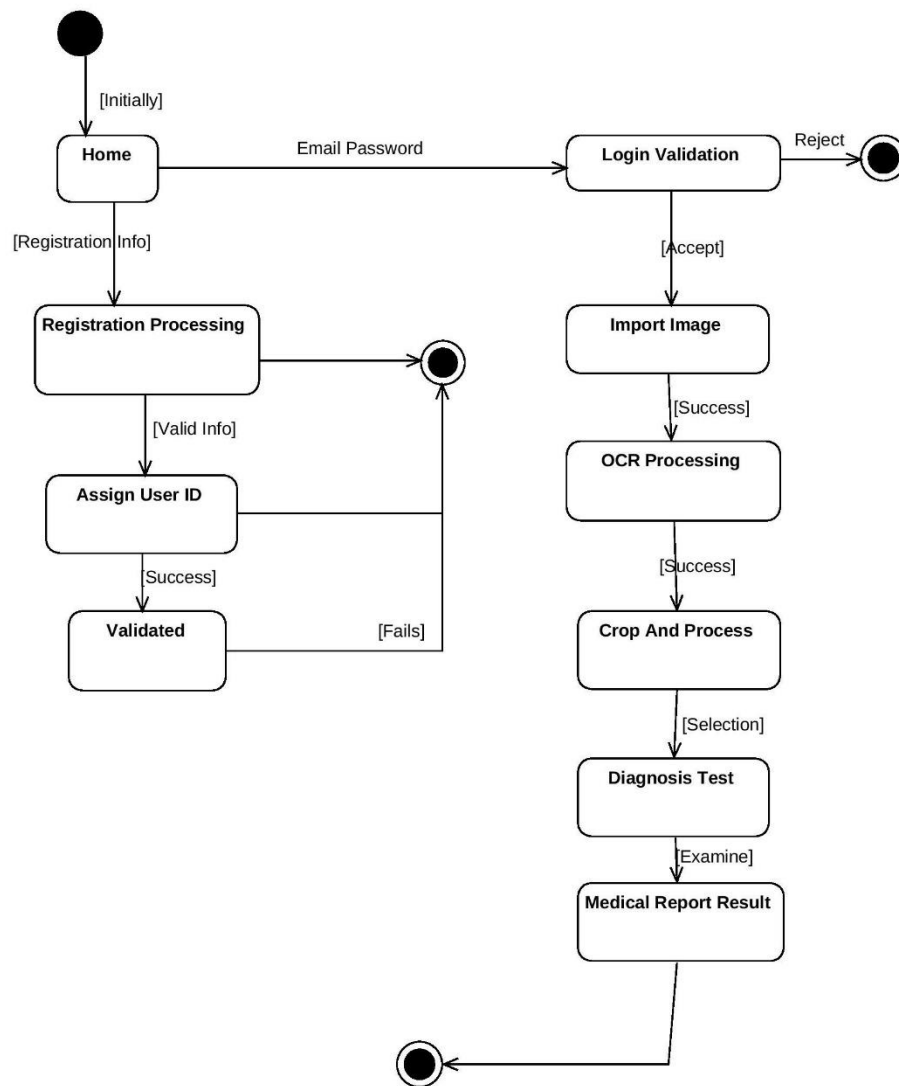


Fig 6: State Chart

## **5. METHODOLOGY**

The main aim of this project is to create the expert system to diagnose the medical report uploaded by the user into the application. Human health is a complex and sensitive topic where it needs a lot of research. Thus a large datasets are required for data mining. The internet is a vast and insanely large collection of different information. And thus extracting correct and reliable information from the internet becomes tough when the topic is of medicals sector. This application has different sorts of level for its works. The main methods of this application process includes image processing and data mining. There are a lot of algorithms to process the image and extract certain characters which by means we will be using OCR. The OCR will extract the specific texts that are needed into the application and using various algorithms, we will be crawling through different websites to fetch the necessary results we will be needing.

### **5.1 Optical character recognition**

Optical character recognition (also optical character reader, OCR) is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image (for example from a television broadcast). It is widely used as a form of information entry from printed paper data records, whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static-data, or any suitable documentation. It is a common method of digitizing printed texts so that they can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes such as cognitive computing, machine translation, (extracted) text-to-speech, key data and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.



### **5.1.1 Techniques for OCR**

- Pre-processing
- Character recognition
- Post processing

#### **5.1.1.1 Pre-Processing**

Character Recognition software often "pre-processes" images to improve the chances of successful recognition. Techniques include:

- De-skew – If the document was not aligned properly when scanned, it may need to be tilted a few degrees clockwise or counterclockwise in order to make lines of text perfectly horizontal or vertical.
- DE speckle – remove positive and negative spots, smoothing edges
- Binarisation – Convert an image from color or greyscale to black-and-white. The task of binarisation is performed as a simple way of separating the text from the background. The task of binarisation itself is necessary since most commercial recognition algorithms work only on binary images since it proves to be simpler to do so. In addition, the effectiveness of the binarisation step influences to a significant extent the quality of the character recognition stage and the careful decisions are made in the choice of the binarisation employed for a given input image type; since the quality of the binarisation method employed to obtain the binary result depends on the type of the input image (scanned document, scene text image etc.).
- Line removal – Cleans up non-glyph (an alphabetic or numeric font) boxes and lines
- Layout analysis – Identifies columns, paragraphs, captions, etc. as distinct blocks. Especially important in multi-column layouts and tables.

- Line and word detection – Establishes baseline for word and character shapes, separates words if necessary.
- Script recognition – In multilingual documents, the script may change at the level of the words and hence, identification of the script is necessary, before the right OCR can be invoked to handle the specific script.
- Character isolation or "segmentation" – For per-character OCR, multiple characters that are connected due to image artifacts must be separated single characters that are broken into multiple pieces due to artifacts must be connected.
- Normalize aspect ratio and scale

Segmentation of fixed-pitch fonts is accomplished relatively simply by aligning the image to a uniform grid based on where vertical grid lines will least often intersect black areas. For proportional fonts, more sophisticated techniques are needed because whitespace between letters can sometimes be greater than that between words, and vertical lines can intersect more than one character.

#### **5.1.1.2 Character recognition**

There are two basic types of core OCR algorithm, which may produce a ranked list of candidate characters.

Matrix matching involves comparing an image to a stored glyph on a pixel-by-pixel basis; it is also known as "pattern matching", "pattern recognition", or "image correlation". This relies on the input glyph being correctly isolated from the rest of the image, and on the stored glyph being in a similar font and at the same scale. This technique works best with typewritten text and does not work well when new fonts are encountered. This is the technique the early physical photocell-based OCR implemented, rather directly.

Feature extraction decomposes glyphs into "features" like lines, closed loops, line direction, and line intersections. The extraction features reduces the dimensionality of the representation and makes the recognition process computationally efficient. These features are compared with an abstract vector-like representation of a character, which might reduce to one or more glyph prototypes. General techniques of feature detection in computer vision are applicable to this type of OCR, which is commonly seen in "intelligent" handwriting recognition and indeed most modern OCR software. Nearest neighbor classifiers such as the k-nearest neighbor's algorithm are used to compare image features with stored glyph features and choose the nearest match.

Software such as Cuneiform and Tesseract use a two-pass approach to character recognition. The second pass is known as "adaptive recognition" and uses the letter shapes recognized with high confidence on the first pass to recognize better the remaining letters on the second pass. This is advantageous for unusual fonts or low-quality scans where the font is distorted (e.g. blurred or faded).

### **5.1.1.3 Post Processing:**

OCR accuracy can be increased if the output is constrained by a lexicon – a list of words that are allowed to occur in a document. This might be, for example, all the words in the English language, or a more technical lexicon for a specific field. This technique can be problematic if the document contains words not in the lexicon, like proper nouns. Tesseract uses its dictionary to influence the character segmentation step, for improved accuracy.

The output stream may be a plain text stream or file of characters, but more sophisticated OCR systems can preserve the original layout of the page and produce, for example, an annotated PDF that includes both the original image of the page and a searchable textual representation.

The Levenshtein Distance algorithm has also been used in OCR post-processing to further optimize results from an OCR API.

## **5.2 C4.5 Algorithm**

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

C4.5 builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. The training data is a set of already classified samples. Each sample consists of a p-dimensional vector, where they represent attribute values or features of the sample, as well as the class in which falls.

At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting

criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision.

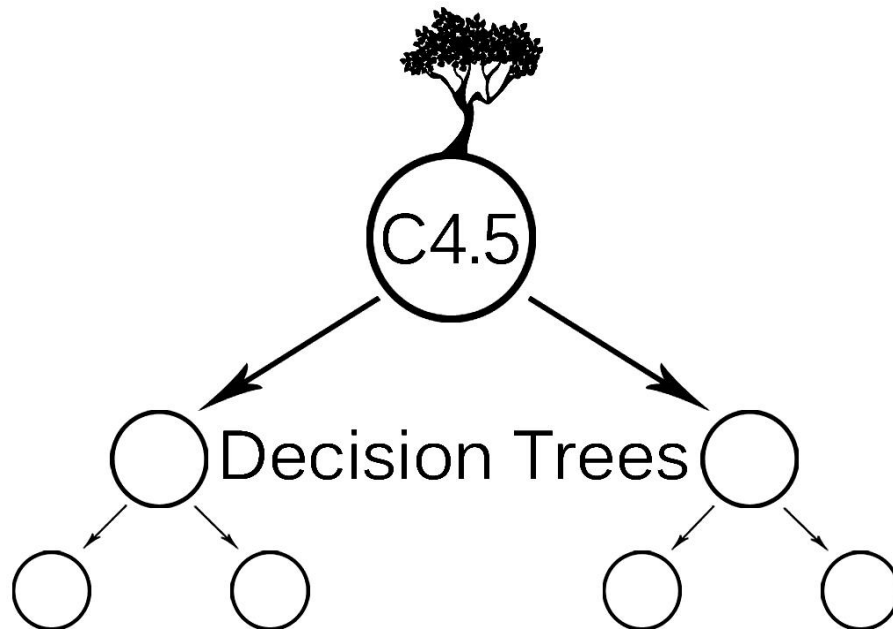


Fig 7: C4.5 Algorithm

This algorithm has a few base cases.

- All the samples in the list belong to the same class. When this happens, it simply creates a leaf node for the decision tree saying to choose that class.
- None of the features provide any information gain. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class.
- Instance of previously-unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected value

## 5.3 PageRank

PageRank is an algorithm used by Google Search to rank websites in their search engine results. PageRank was named after Larry Page. It is one of the founders of Google. PageRank is a way of measuring the importance of website pages. PageRank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set. The algorithm may be applied to any collection of entities with reciprocal quotations and references. The numerical weight that it assigns to any given element  $E$  is referred to as the PageRank of  $E$  and denoted by other factors like Author Rank can contribute to the importance of an entity.

A PageRank results from a mathematical algorithm based on the web graph, created by all World Wide Web pages as nodes and hyperlinks as edges, taking into consideration authority hubs such as [cnn.com](http://cnn.com) or [usa.gov](http://usa.gov). The rank value indicates an importance of a particular page. A hyperlink to a page counts as a vote of support. The PageRank of a page is defined recursively and depends on the number and PageRank metric of all pages that link to it ("incoming links"). A page that is linked to by many pages with high PageRank receives a high rank itself.

The PageRank algorithm outputs a probability\_distribution used to represent the likelihood that a person randomly clicking on links will arrive at any particular page. PageRank can be calculated for collections of documents of any size. It is assumed in several research papers that the distribution is evenly divided among all documents in the collection at the beginning of the computational process. The PageRank computations require several passes, called "iterations", through the collection to adjust approximate PageRank values to more closely reflect the theoretical true value.

A probability is expressed as a numeric value between 0 and 1. A 0.5 probability is commonly expressed as a "50% chance" of something happening. Hence, a

PageRank of 0.5 means there is a 50% chance that a person clicking on a random link will be directed to the document with the 0.5 PageRank.

$$PR(A) = (1-d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$$

Where

- $PR(A)$  is the PageRank of page A,
- $PR(Ti)$  is the PageRank of pages  $Ti$  which link to page A,
- $C(Ti)$  is the number of outbound links on page  $Ti$  and
- $d$  is a damping factor which can be set between 0 and 1.

## 5.4 Agile Methodology

- In traditional software development methodologies like Waterfall model, a project can take several months or years to complete and the customer may not get to see the end product until the completion of the project.
- At a high level, non-Agile projects allocate extensive periods of time for Requirements gathering, design, development, testing and User Acceptance Testing, before finally deploying the project.
- In contrast to this, agile projects have Sprints or iterations which are shorter in duration (Sprints/iterations can vary from 2 weeks to 2 months) during which pre-determined features are developed and delivered.
- Agile projects can have one or more iterations and deliver the complete product at the end of the final iteration.

### 5.4.1 Why Agile

Agile development methodology provides opportunities to assess the direction of a project throughout the development lifecycle. This is achieved through regular cadences of work, known as sprints or iterations, at the end of which teams must present a potentially shippable product increment. By focusing on the repetition of abbreviated work cycles as well as the functional product they yield, agile methodology is described as “iterative” and “incremental.” In waterfall, development teams only have one chance to get each aspect of a project right. In an agile paradigm, every aspect of development requirements, design, etc. is continually revisited throughout the lifecycle. When a team stops and re-evaluates the direction of a project every two weeks, there’s always time to steer it in another direction.

The results of this “inspect-and-adapt” approach to development greatly reduce both development costs and time to market. Because teams can develop software at the same time they’re gathering requirements, the phenomenon known as “analysis paralysis” is less likely to impede a team from making progress. And because a team’s work cycle is limited to two weeks, it gives stakeholders recurring opportunities to calibrate releases for success in the real world. Agile development methodology helps companies build the right product. Instead of committing to market a piece of software that hasn’t even been written yet, agile empowers teams to continuously replan their release to optimize its value throughout development, allowing them to be as competitive as possible in the marketplace. Development using an agile methodology preserves a product’s critical market relevance and ensures a team’s work doesn’t wind up on a shelf, never released.



## **5.5 Testing**

### **5.5.1 Unit Testing**

Unit testing refers to the process of testing modules against the detailed design. The inputs to the unit testing are the successfully compiled modules from the coding process. These are assembled during unit testing to make the largest units, i.e. the components of architectural design.

Testing has been performed in each phase of the project design and coding. The module interface is tested to ensure that information properly flows into and out of the program unit under testing. The local data structure is examined to ensure that data stored temporarily maintains an integrity during all steps in an algorithm's execution. And finally all error-handling paths are tested.

### **5.5.2 System Testing**

System testing process is concerned with finding errors that results from unanticipated interaction between sub-systems and system components. Once source code has been generated, software must be tested to uncover as many errors as possible before delivery to the customers. Our goal is to design a series of test cases that have a high likelihood of finding errors. To uncover the errors, software techniques are used. These techniques provide systematic guidance for designing test that

- Exercise the internal logic of software components,
- Exercise the input and output domains of the programs to uncover errors

Software is tested from two different perspectives:

- Internal program logic is exercised using “white box” test case design techniques.
- Software requirements are exercised using “black box” test case design techniques.

### **5.5.3 Performance Testing**

Performance testing is designed to test the run-time performance of software within the context of an integral system. Performance testing occurs throughout all the steps in the testing process. Even at the unit level, the performance of an individual module may be assessed as white-box tests are conducted.

### **5.5.4 Verification and Validation**

Software testing is one element of broader topic than is often referred to as verification and validation. The purpose of verification and validation is to confirm system specification and to meet the requirements of the system customers. Verification represents the set of activities that are carried out to confirm that the software correctly implements the specific functionality. Validation represents the set of activities that ensure that the software has been built is satisfying the customer requirements. Boehm stated this another way:

Verification: Are we building the product right?

Validation: are we building the right product?

## **6. RESULT ANALYSIS AND FUTURE ENHANCEMENT**

After a long and continuous effort from the beginning of the project time we have come to some of the result which is really enough to make us to forget all such stress we faced during the whole of the project period. We have come up with the progress of our project are:

### **6.1 Analysis and Requirement Collection**

- Research at different pathology labs
- Analysis on many collection of medical reports
- Problem Analysis
- Requirement Collection
- System Design

### **6.2 Backend Work**

- Database designed
- API developed
  - ✓ User Registration
  - ✓ User Login
  - ✓ Image Upload to Database
  - ✓ Camera Feature
  - ✓ UI development
  - ✓ BMI Calculation
  - ✓ Emergency Call Feature
  - ✓ OCR (optical character recognition)

## **6.3 Frontend Work**

- Android Application
  - ✓ User Login
  - ✓ User Registration
  - ✓ Image upload to database
  - ✓ OCR
  - ✓ Simple suggestion to user
  - ✓ Camera Feature
  - ✓ BMI Calculation
  - ✓ Emergency Call Feature

## **6.4 Future Enhancement**

- Better Suggestion to the user
- Medical Dictionary
- More User Friendly
- Recommendation system
- Coverage of all medical fields

## **7. CONCLUSION**

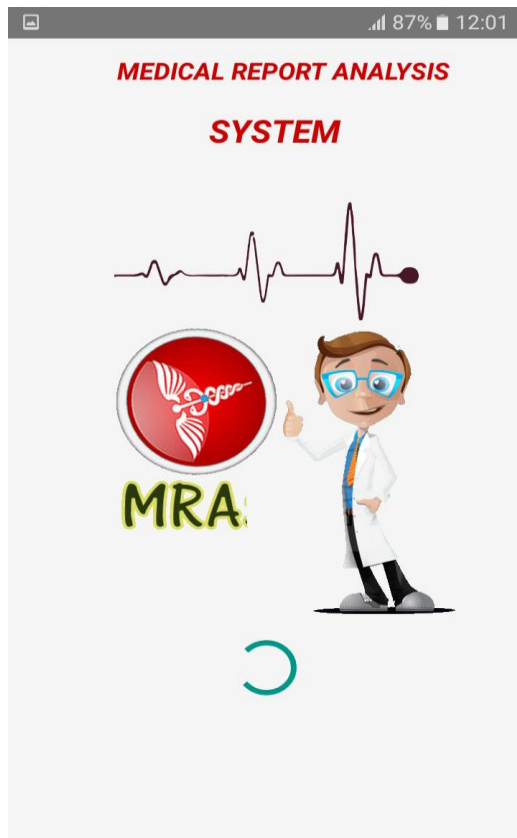
Finally, it is concluded that the application can be used over internet. This application is categorized as android based and web based. This application can run on the android mobile and this application also works on mobile network so it is user friendly in case of our country like Nepal.

The main goal of the project titled “Medical Report Analysis System” is to provide the results on basis of analysis of the medical report. Hence, this project has been a great experience for us. As no application can be perfect and this line also plays a role in this project and it contains few limitations but such limitations can be solved with great dedication, devotion and continuous effort we come to this point and it is really a wonderful feeling for all of us.

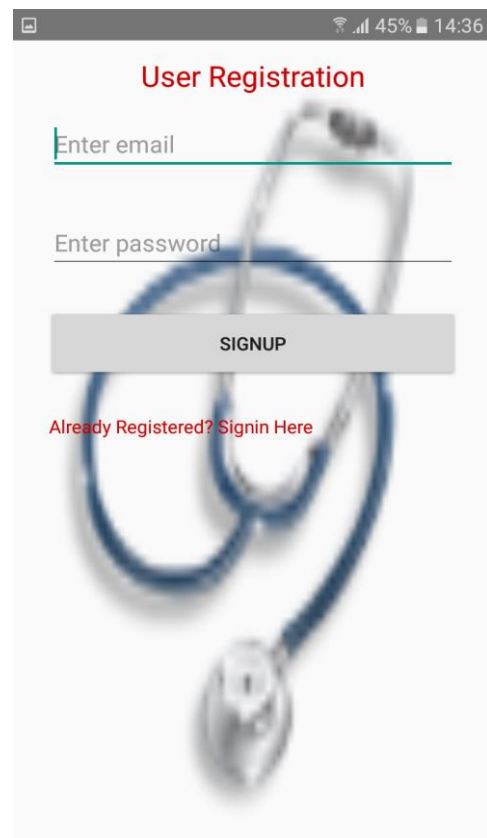
## REFERENCES

- [1] Z. Huang, X. Lu, H. Duan, and C. Zhao, “Collaboration-based medical knowledge recommendation,” *Artif. Intell. Med.*, vol. 55, no. 1, pp. 13–24, May 2012.
- [2] “Mayo Clinic.” [Online]. Available: <http://www.mayoclinic.org/>. [Accessed: 11-Dec-2016].
- [3] S. Arora, D. Bhattacharjee, M. Nasipuri, D. K. Basu, and M. Kundu, “Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition,” *ArXiv10054032 Cs*, May 2010.
- [4] A. Purohit and S. S. Chauhan, “A Literature Survey on Handwritten Character Recognition.”
- [5] Edward P Hoffer, Mitchell J Feldman and Richard J Kim “Explains patterns of use of a mature expert system” AMIA Annual Symposium Proceedings Archive.”
- [6] Available: <http://www.cs.cmu.edu/~javabayes/Home/node2.html> (Referred on 14 Dec, 2016)
- [7] Herk, Egbert Van. A diagnostic decision support system. Amsterdam: Drukwerk, 19 January 1994.

## SCREENSHOTS:



SPLASH



REGISTRSTION

86% 12:02

### User Login

snmaharjan56@gmail.com

.....

SIGN IN

Not have an account? [Signup Here](#)


86% 12:02

### User Login

snmaharjan56@gmail.com

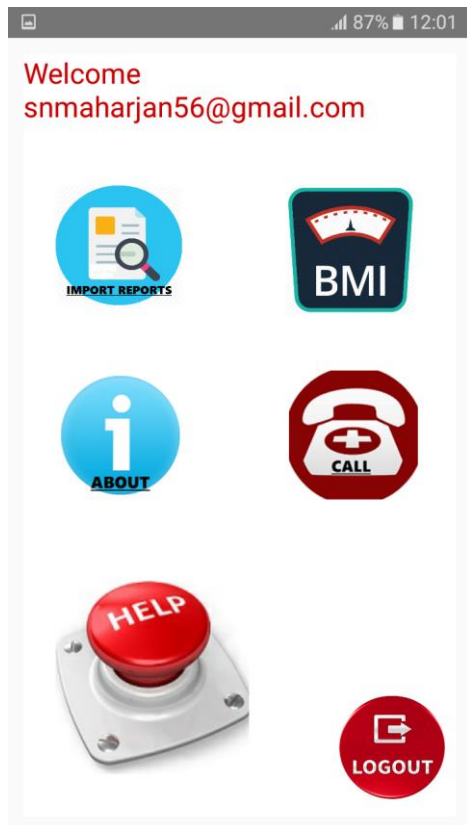
.....

SIGN IN

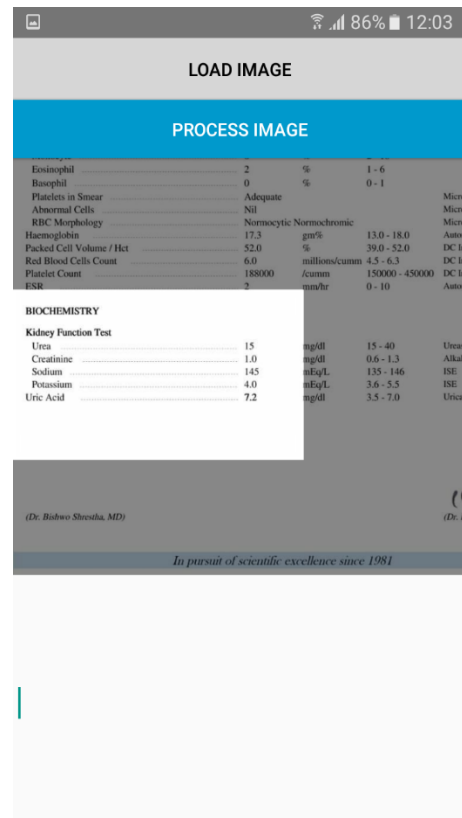
 Logging Please Wait...

LOGIN FORM

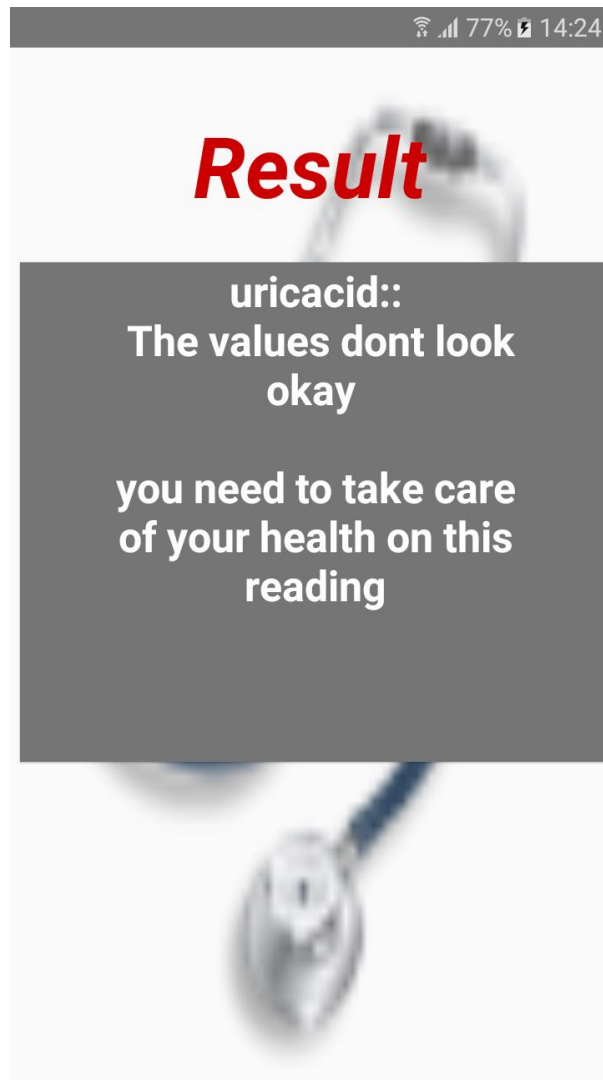




USER PROFILE



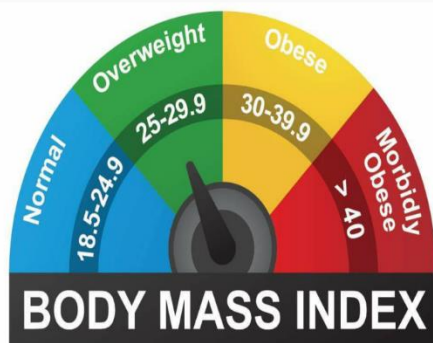
IMPORT MEDICAL REPORT



RESULT

Weight in Pounds :  
 \_\_\_\_\_  
 HEIGHT :  
 Feet :  
 \_\_\_\_\_  
 Inches :  
 \_\_\_\_\_

METRIC FORM    CALCULATE



Weight in KG :  
 65  
 HEIGHT :  
 CM :  
 170

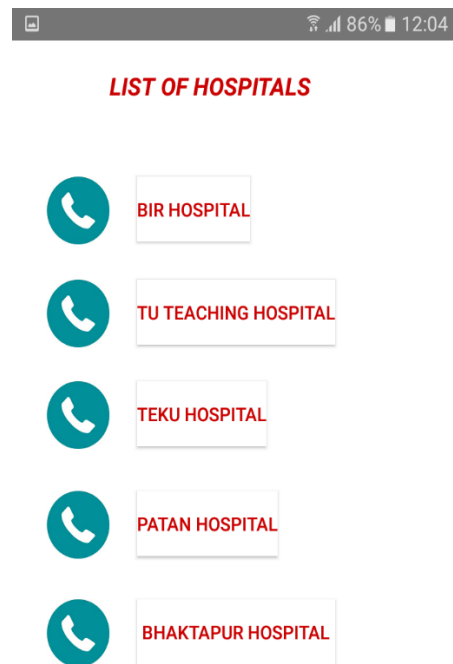
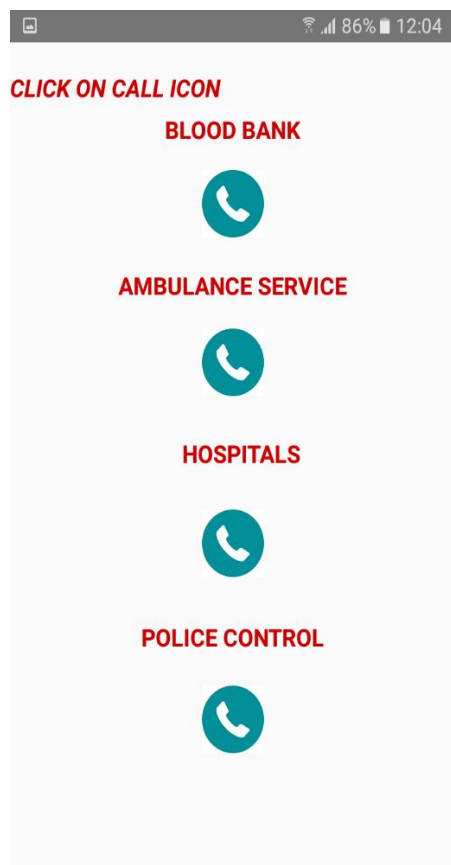
BACK    CALCULATE

Your BMI is : 22.491348  
 OK

 <b>HEALTHY WEIGHT</b> 14.50-20.00 (FEMALE) 17.50-22.50 (MALE)	 <b>OBSE</b> 25.00-HIGHER (FEMALE) 27.50-HIGHER (MALE)
<b>UNDERWEIGHT</b> 0 - 14.50 ( FEMALE) 0 - 17.50 (MALE)	<b>OVERWEIGHT</b> 20.00-25.00 (FEMALE) 22.50-27.50 (MALE)

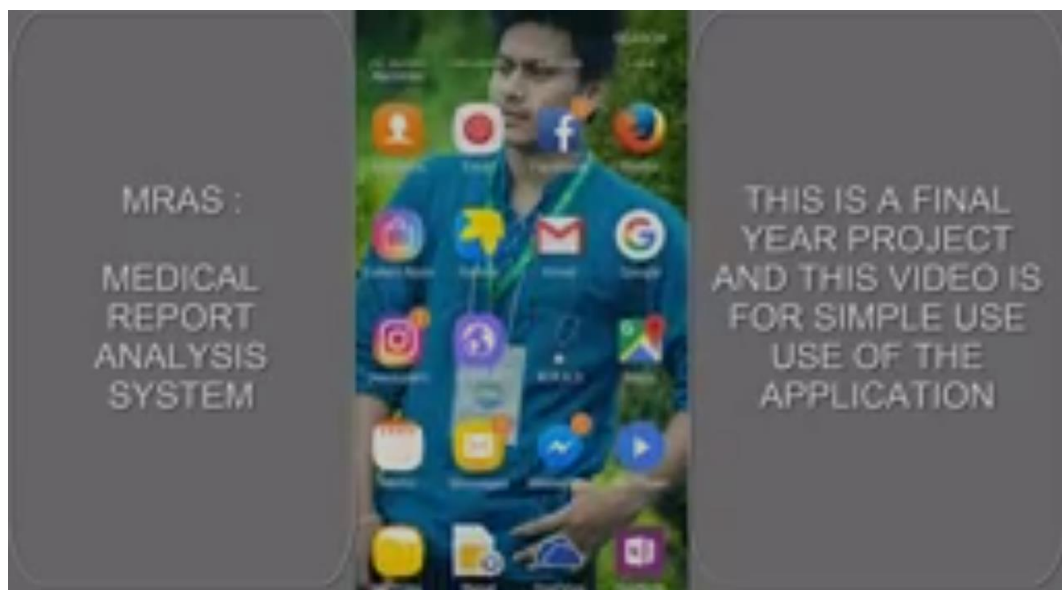
©MRAS

BMI CALCULATION



## EMERGENCY CALL FEATURE

|



HELP



ABOUT