Early Prediction of Pre-Diabetes Using Machine Learning

Submitted in partial fulfillment of the requirements of the degree of
Bachelors of Engineering in Information Technology By

Neha Joisher

Malav Shah

Pujan Sheth

Supervisor

Prof. Radhika Kotecha

DEPARTMENT OF INFORMATION TECHNOLOGY

K.J. SOMAIYA INSTITUTE OF ENGINEERING AND INFORMATION TECHNOLOGY, AYURVIHAR, SION, MUMBAI-400022

2020

CERTIFICATE

This is to certify that the project entitled “Early Prediction of Pre-Diabetes Using Machine Learning” is a bonafide work of students Pujan Sheth, Malav Shah and Neha Joisher submitted to University of Mumbai in partial fulfillment of the requirement for the award of the degree of Bachelor of Engineering in Information Technology.

Prof. Radhika Kotecha
Project Guide
Department of Information Technology

Dr. Radhika Kotecha
Head of Department

Dr. Suresh Ukarande
Principal

Place: Sion, Mumbai-400022

Date:

Project Report Approval for B.E.

This project report entitled "**Early Prediction of Pre-Diabetes Using Machine Learning**" by

Neha Joisher (44)

Malav Shah (62)

Pujan Sheth (65)

is approved for the degree of Bachelor of Engineering in Information Technology.

Examiners: 1.____________________

2.____________________

Date:

Place: Sion, Mumbai-400022.

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Neha Joisher

Malav Shah

Pujan Sheth

Date:

Place: Sion, Mumbai-400022

ACKNOWLEDGEMENT

We would like to convey our sincere thanks to everyone who guided us throughout this project endeavour, before we present our project entitled “Early Prediction of Pre-Diabetes Using Machine Learning”.

We wish to express our sincere thanks to our Principal Dr. Suresh Ukarande and Vice-Principal Dr. Sunita Patil, for providing us with all the necessary facilities for the research.

We place on record, our sincere thank you to Prof. Radhika Kotecha, Head of Department and our Project Guide, for the continuous encouragement and providing her valuable time by providing us with the information which is needed that helped us to complete this project in a better manner.

We are also grateful to Ms. Reena Lokare, Project Supervisors. We are extremely thankful and indebted to them for sharing expertise, and sincere and valuable guidance and encouragement extended to us.

We take this opportunity to express gratitude to all the Department faculty members for their help and support. We also thank our parents for unceasing encouragement, support and attention.

Neha Joisher

Malav Shah

Pujan Sheth

Date:

Abstract

With the advancement of computer and mobile technologies, mobile health (mHealth) can be leveraged for patient self-management, patient diagnosis, and determining the possibility of being affected by some disease. Diabetes mellitus is a chronic and lifestyle disease and millions of people from all over the world fall victim to it. Although there are some mobile apps keeping track of calories, sugar taken, medicine doses, lifestyle, blood glucose, blood pressure, weight of individuals and giving suggestions about food, exercises to prevent or control diabetes, no application has been found that was explicitly developed to analyse the risk of being a diabetic patient. Therefore, the objective of this project is to develop an intelligent mHealth application to assess his/her possibility of being diabetic, prediabetic or nondiabetic. The application uses novel machine learning techniques to predict diabetes levels for the users. At the same time, the system also provides knowledge about diabetes and some suggestions on the disease. A comparative analysis of machine learning (ML) algorithms were performed. The Decision Tree (DT) classifier performs amongst the ML algorithms. Hence, DT classifier is used to design the machinery for the mobile application for diabetes prediction.

TABLE OF CONTENT

List of Abbreviations

DM - Diabetes Mellitus

IDF - International Diabetes Federation

SVR - Support Vector Regression

GP - Gaussian Processes

PIDD - Pima Indian Diabetes Dataset

OSA - Obstructive Sleep Apnea

RBF - Radial Basis Function

KNN - k- Nearest Neighbor

DT - Decision Tree

ML - Machine Learning

XGB - eXtreme Gradient Boosting

SVM - Support Vector Machine

NB - Naïve Bayes

SDLC - Software Development Life Cycle.

List of Figures

List of Tables

CHAPTER 1
INTRODUCTION

1.1 Overview

Diabetes Mellitus (DM) simply named diabetes is a disorder caused when the pancreas does not produce Insulin or body cells no longer respond to Insulin. Our body cells are fuelled by Insulin which is one type of hormone which acts as a key that allows glucose from the blood to enter into our cells. If in the pancreas, the insulin producing beta cells are put down, then the glucose in the blood is not adequately regulated, as a consequence, the glucose level in the blood increases abruptly this causes an individual to be diabetic. On that point are primarily four types of diabetes. Prediabetes is categorized by the glucose level higher than a normal but not yet high enough to be characterized as diabetes. Type1 diabetes is an autoimmune disease that causes the insulin producing beta cells in the pancreas to be destroyed, inhibiting the body from being able to yield enough insulin to effectively regulate the blood glucose levels. Type2 diabetes is a metabolic disorder that results from insulin resistance, the body cells no longer react to insulin hormone, a situation in which cells fail to utilize insulin properly. Gestational diabetes refers to higher than normal blood glucose level occurring during gestation in adult females who were not diabetic before pregnancy. It is usually developed between twenty-four and twenty-eight week of gestation. Diabetes is becoming a possible epidemic in India with more than 63 million individuals found to be diabetic. In a survey conducted in 2000, it is found that India topped the world with 31.7 million of diabetic individuals, second position occupied by China with 20.8 million of diabetics, US comes next to China with 17.7 million are found to be diagnosed with diabetes. This statistic became double in the year 2014 with 65.1 million of diabetics found in India and also it is foreshadowed that in 2030 this statistic crosses 101 million of people. According to IDF (International Diabetes Federation), 4.4 million Indians in their most productive years aren't aware that they have diabetes, due to this lack of knowledge 1 million Indians were passing in the year 2011. The concerning factor is that the predicted prevalence rate of diabetes in India in the future is very high, therefore in regional and as well as at national level the campaign started to spread knowledge about diabetes. The danger of diabetes can be minimized by supplying knowledge to the masses, so that the masses can do their regular body processes to minimize the issue of diabetes.

1.2 Aims and Objectives

- Main purpose of this system is people nowadays are very health conscious and every certain interval of time they checkup their body to see that they are suffering from any diseases or not.
- Diabetes is most common diseases is seen in people of India nowadays so our system will predict whether people will suffer from diabetes or not
- We use various data mining and machine learning algorithm to predict

CHAPTER 2
LITERATURE REVIEW

2.1 Review of Literature

1. Explains the current state of diabetes mellitus in India by Seema Abhijeet Kaveeshwar and Jon Cornwall. Eleni I. Georga et al.
2. Employed two kernel based algorithms SVR (Support Vector Regression) and GP (Gaussian Processes) for short-term prediction of glucose concentration for using a patient's detailed knowledge of hyperglycemia on a regular basis.Yang Guo et al.
3. Uses Naïve Bayes classifier to predict patients developing Type2 diabetes, WEKA tool is utilized for making predictions on PIDD (Pima Indian Diabetes Dataset) and obtained good accuracy. Altikardes Z.A et al.
4. Designs an expert system to predict the Non-Dipping/Dipping Patterns by using data collected from various sources. Decision Tree and Naïve- Bayes classifiers are used. Data sets are tested using J48 decision tree algorithm.Nirmala Devi M et al.
5. Developed a Hybrid model for classifying PIDD, this model combines *k-means* and *KNN* (k- Nearest Neighbor) to enhance the accuracy. *k-means* clustering is used to improve the effectiveness of data set by eliminating the outliers after this, result given to *KNN* which classifies the data with improved accuracy.Thangarasu G and Petronas B S I
6. Proposes a new methodology for prediction of diabetes as well as its type and complications associated with diabetes in an effectively, efficiently and economically faster manner. Ramachandran A and Pai V.V.S
7. Proposes the importance of mobile apps for proper care management of chronic diseases like diabetes. Ming-Hseng Tseng et al.
8. Develops an Android app that can provide an easy and efficient way to pre-screen, high-risk of OSA (Obstructive Sleep Apnea: highly associated with diabetes) potential patients, aid in condition management to achieve initial diagnosis and treatment determinations, prevent the existence of complications, and reaches the goal of preventive medicine.
9. Proposes Classification of lung cancer subtypes by data mining technique by DassM.V, Rasheed M.A and Ali M.M.

10. Othman M.F.B, Abdullah N.B and Kamal N.F.B, explains MRI brain classification using a support vector machine algorithm.
11. Muhammed L.A.N explains data mining techniques to diagnose heart disease.
12. Explains the application of the error function in analyzing the learning dynamics near singularities of the multilayer perceptrons by Guo Weili, Wei Haikun, Zhao Junsheng, Li Weiling and Zhang Kanjian.
13. In these authors defined the predictive algorithm's accuracy varying before and after pre-processing. After pre-processing the dataset, Decision tree technique demonstrated good accuracy compared to J48.
14. In these, authors have proposed a new diabetes risk prediction model. Firstly, the median of each group was utilized to replace missing values and outliers to optimize the original diabetes dataset. Next to that, the updated data set was selected to extract the optimal feature subset by the weighted feature selection algorithm of Random Forest. After this step, they constructed the XGB classifier for diabetes risk prediction.
15. In these, authors have evaluated the performance of Tree based classifiers in WEKA. In order to improve the performance, remove noise from the dataset.
16. In these, authors have discussed WEKA as it is a powerful tool as it contains both supervised and unsupervised learning techniques.
17. In these authors have performed the classification of diabetes patients using three classification techniques including Radial Basis Function (RBF), Multi-Layer perceptron and multilevel counter propagation network and performing the simulation tests by WEKA software tool.
18. In these, authors have used ID3, C4.5, & CART classification algorithms on diabetes dataset.
19. In these authors used the Naïve Bayes classifier along with WEKA tool to forecast patients with Type2 diabetes with good accuracy.
20. In these, authors developed the Non-Dipping/Dipping Patterns prediction method by using data obtained from different sources. Here, Decision Tree along with Naïve- Bayes classifiers are used and data sets are tested using J48 decision tree machine learning

algorithm.

21. In these. authors have developed a Hybrid classification model, which combines k-means and k- Nearest Neighbour (KNN) to enhance the accuracy.
22. In these, authors have created a smart phone-based application that provides a quick and simple approach to pre-screen, high-danger of OSA.

CHAPTER 3

METHODOLOGIES AND IMPLEMENTATION

3.1 Detailed Statement of Problem

The primary goal of this study is to develop an android-based healthcare application, which can assist the users to monitor their health-related conditions for improving their health. Methods: The application is developed using the android operating system environment. A Java programming language, namely Android Studio is used to develop the system. The modification is presented as: (1) integration of different modules and their offline usage, (2) history facility, (3) user friendly. The qualitative method is used to study the objective. Findings: The research paper depicts a brief study of existing systems and the new development that has been made in the application and also it is better in the manner that it works as a guide to control risk factors. The descriptive analysis points out that the application is effective to deal with health-related issues. Applications/Improvement: Integration of modules is performed on the android platform of different applications that are located on different websites, the storage facility is added by using Tiny DB, guidance in the form of charts and text is provided to the users. Such features are not provided in the previous work. Diabetes mellitus (DM) is reaching possibly epidemic proportions in India. The degree of disease and destruction due to diabetes and its potential complications are enormous, and originated a significant health care burden on both households and society. The concerning factor is that diabetes is now being proven to be linked with a number of complications and to be occurring at a comparatively younger age in the country. In India, the migration of people from rural to urban areas and corresponding modification in lifestyle are all moving the degree of diabetes. Deficiency of knowledge about diabetes causes untimely death among the population at large. Therefore, acquiring a proficiency that should spread awareness about diabetes may affect the people in India. In this work, a mobile/android application-based solution to overcome the deficiency of awareness about diabetes has been shown. The application uses novel machine learning techniques to predict diabetes levels for the users. At the same time, the system also provides knowledge about diabetes and some suggestions on the disease. A comparative analysis of four machine learning (ML) algorithms were performed. The Decision Tree (DT) classifier outperforms amongst the 4 ML algorithms. Hence, DT classifier is to design the machinery for the mobile application for diabetes prediction using real world dataset.

3.2 Scope

In Our system we will make a hardware device which has insulin and through that insulin blood will go to our hardware device and check the glucose level and send to android app wireless or wired and android app will perform data mining and machine learning algorithm in background and check the user glucose or sugar level with the dataset present in database and say whether user when user will get diabetes or not and if user will get diabetes than how many years later user will suffer from diabetes and we also suggest how to maintain diabetes and how to prevent or delay the diabetes we will use blood to measure diabetes in future through retina or laser diabetes can be measure and perform this` task and also suggest the daily routine to user which help user to control the diseases and help people to stay away from diabetes.

3.3 Implementation

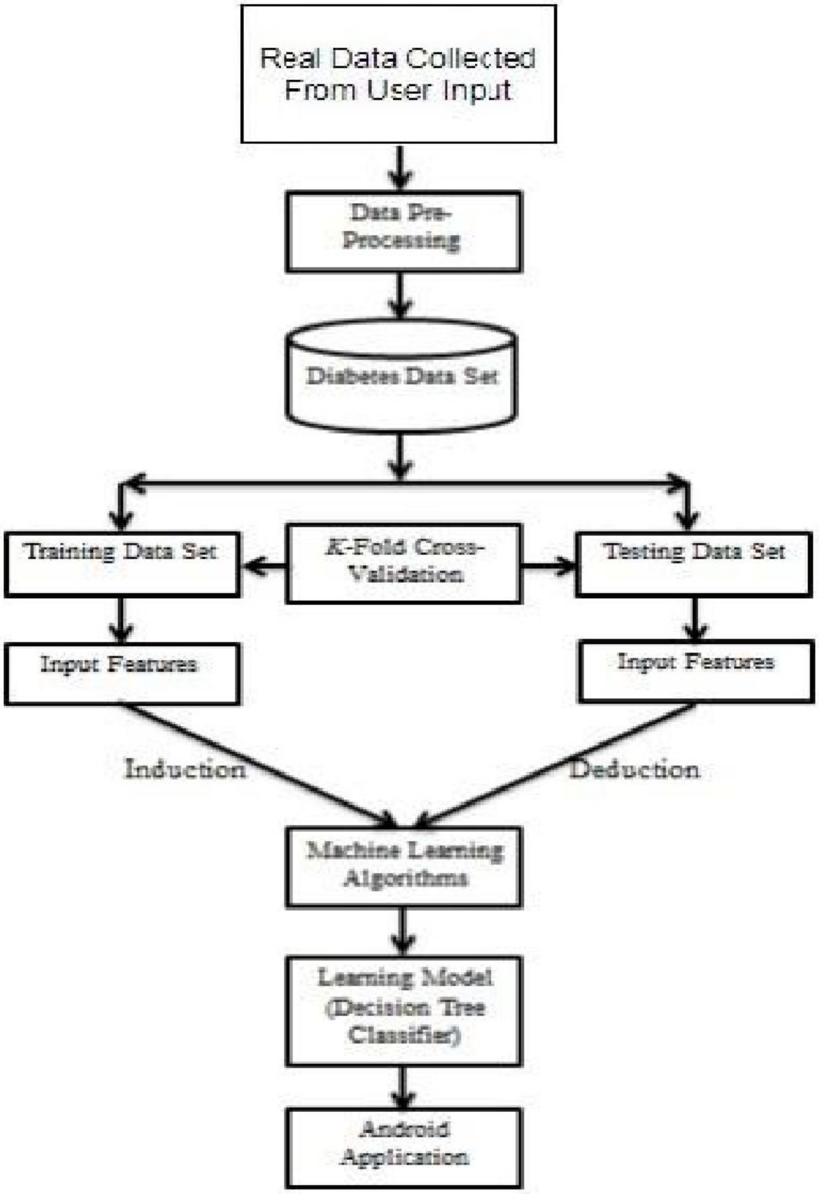


Fig 3.1: Flow Diagram of Proposed System

The system takes height and weight as input from the user through an android application and it will calculate BMI of the user using predefined formula. After getting input from the user, it will process data and apply machine learning algorithms on it. The system will predict the user's glucose or sugar level with the data present in the dataset and it will predict whether the user will suffer from diabetes or not. The system will also predict as to how many years down the line the user will suffer from diabetes. We also suggest daily routine and food habits to users which will help them to control the sugar level and help people to stay away from diabetes.

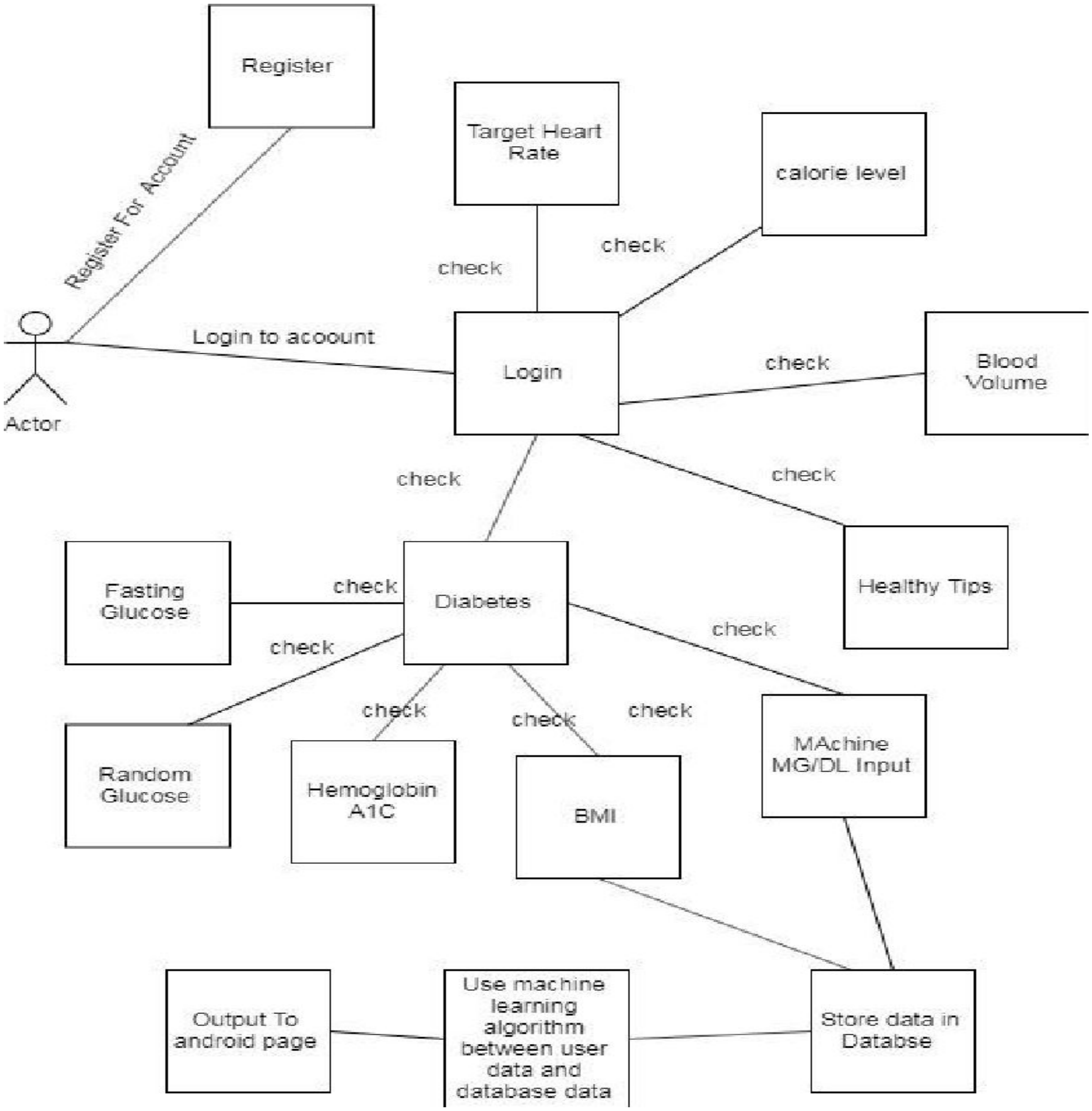


Fig 3.2: Communication UML of The System

Communication diagram (called collaboration diagram in UML 1.x) which shows interactions between objects and/or parts using sequenced messages in a free-form arrangement. It Models message passing between objects or roles that deliver the functionalities of use cases and operations

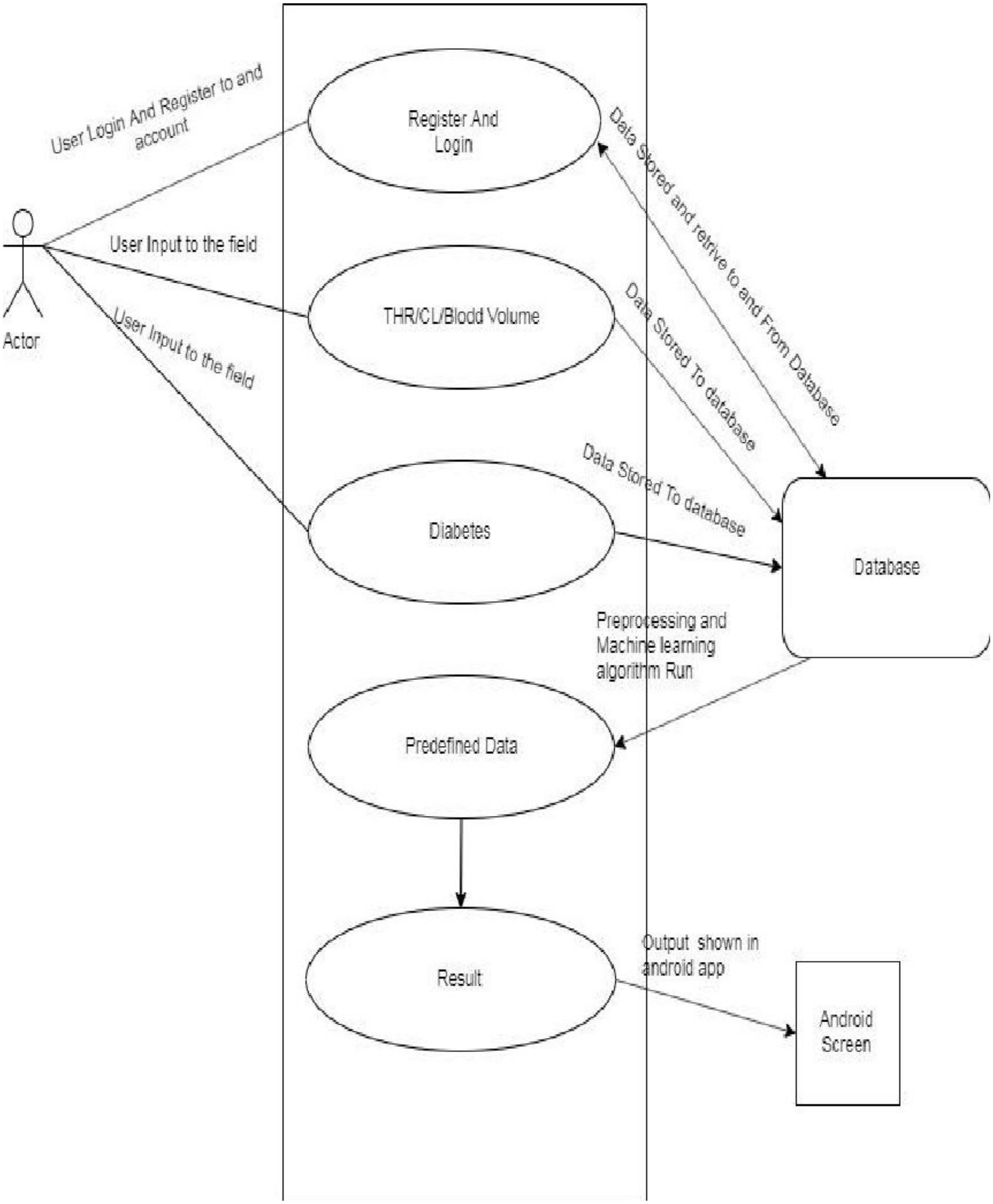


Fig 3.3: UML Diagram of The System

Within the circular containers, we express the actions that the actors perform. In our system actors are the users and the actions performed are: Register/Login, Check target heart rate, blood volume, calorie level, diabetes.

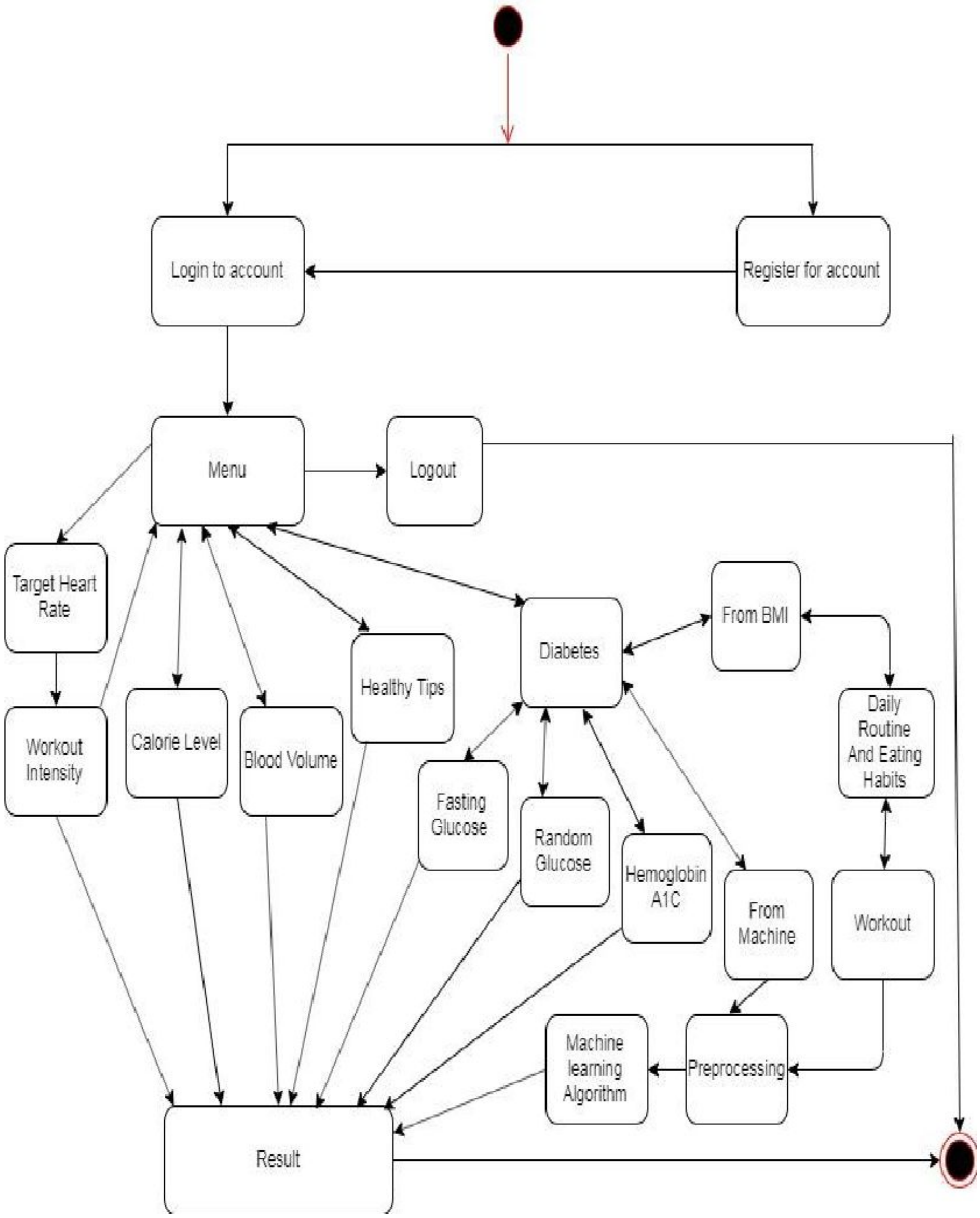


Fig 3.4: Activity Diagram of The System

This diagram describes the flow of control in a system. It consists of activities and links. Activities are nothing but the functions of a system. This is used to visualize the flow of controls in a system.

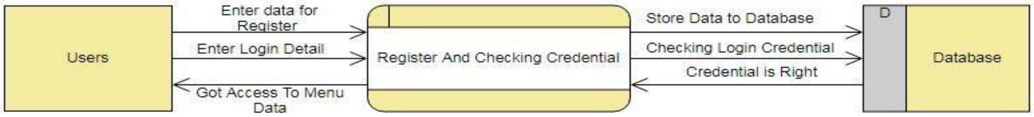


Fig 3.5: Data Flow Diagram Level 0

This diagram gives an abstraction view, showing the system as a single process with its relationship to external entities. It represents the entire system as a single bubble with input and output data indicated by incoming/outgoing arrows.

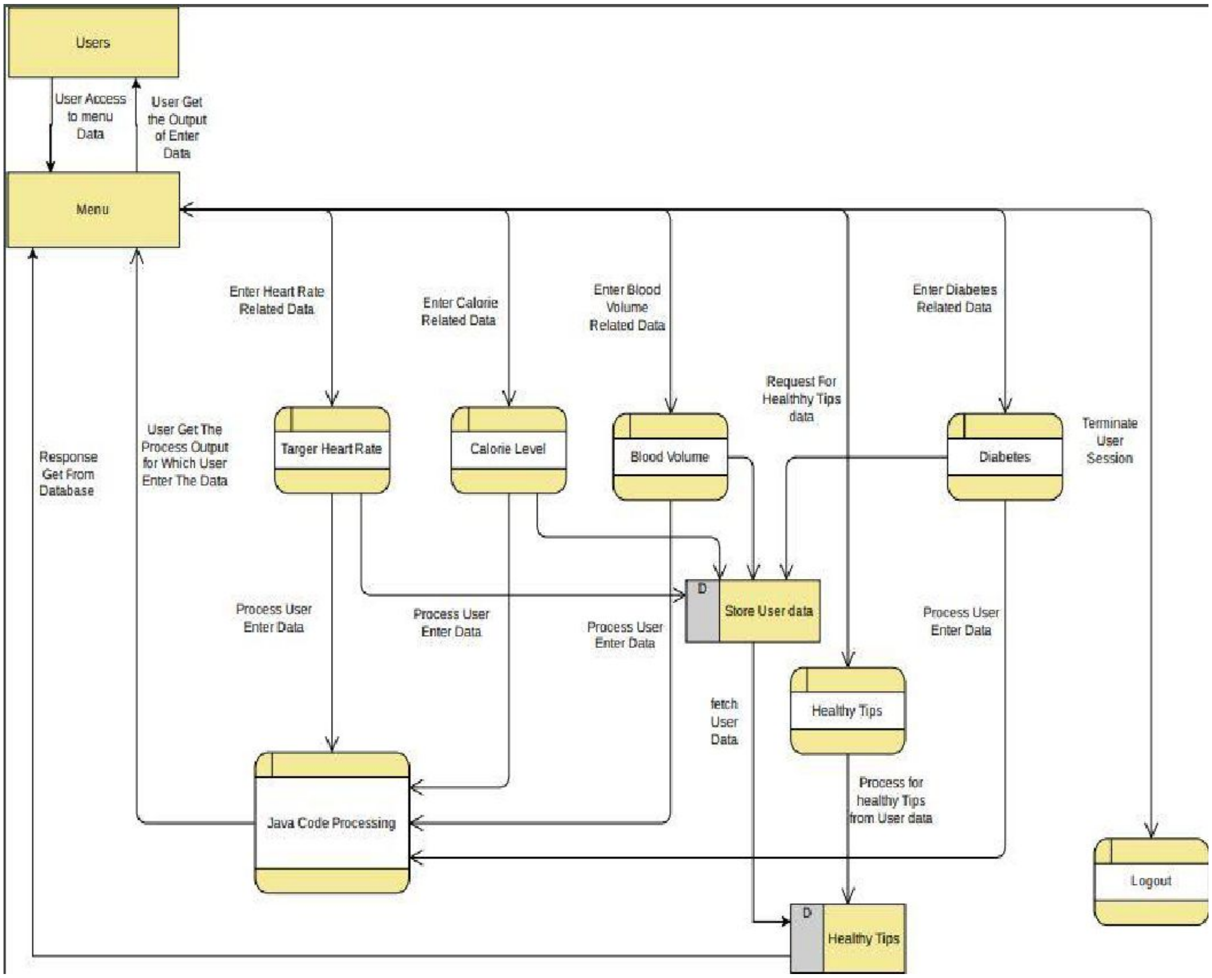

Users
User Access to menu Data
User Get the Output of Enter Data
Menu
Enter Heart Rate Related Data
Enter Calorie Related Data
Enter Blood Volume Related Data
Enter Diabetes Related Data
Request For Healthhy Tips data
Terminate User Session
Response Get From Database
User Get The Process Output for Which User Enter The Data
Targer Heart Rate
Calorie Level
Blood Volume
Diabetes
D
Store User data
Process User Enter Data
Process User Enter Data
Process User Enter Data
Process User Enter Data
fetch User Data
Healthy Tips
Process for healthy Tips from User data
Java Code Processing
Logout
D
Healthy Tips

Fig 3.6: Data Flow Diagram Level 1

In this diagram DFD-0 is decomposed into multiple bubbles/processes. In this level we highlight the main functions of the system and break down the high level process of 0-level DFD into sub processes.

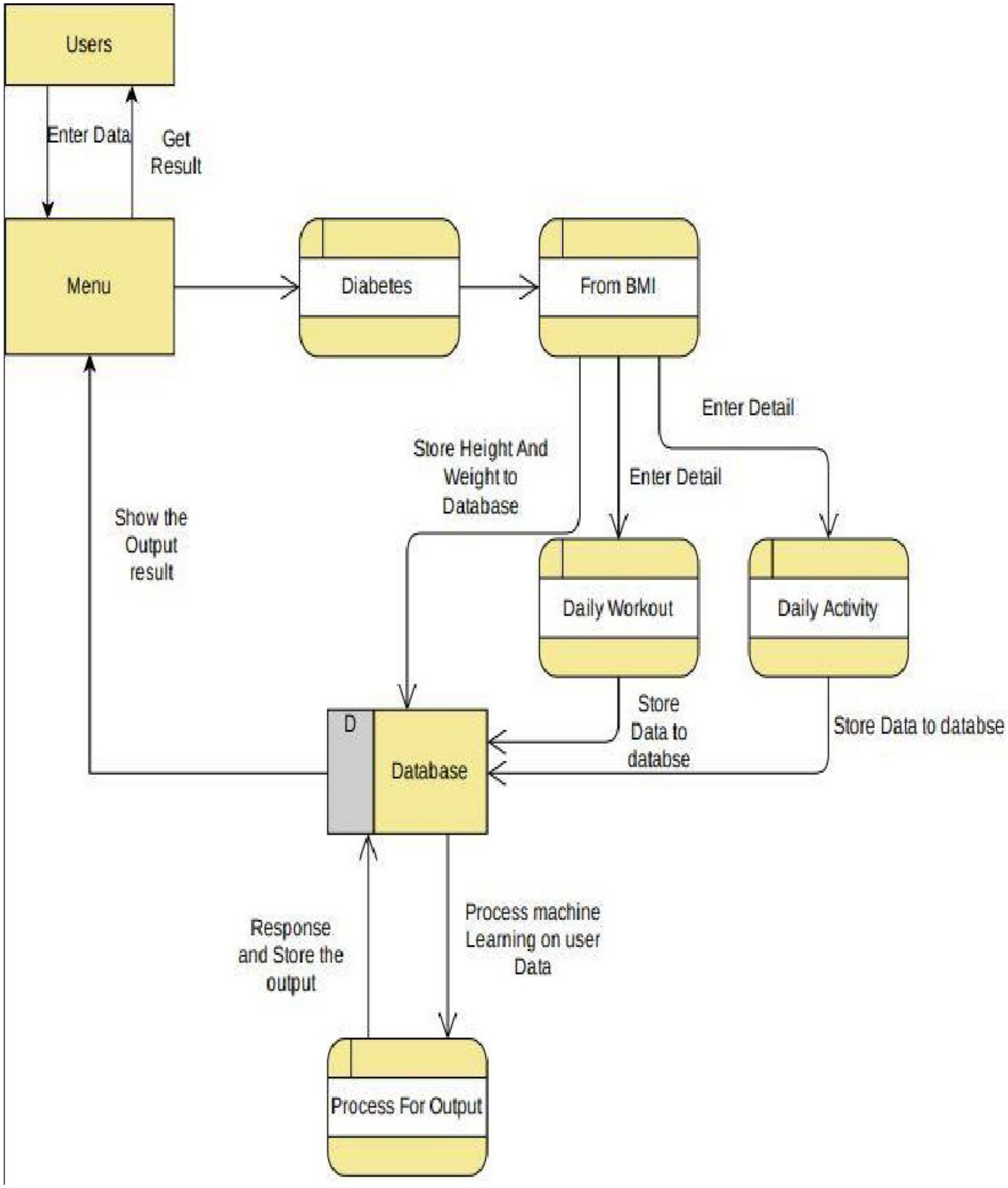


Fig 3.7: Data Flow Diagram Level 2

This diagram goes one step deeper into parts of 1-level DFD. It is used to plan and record the specific/necessary detail about the system’s functioning.

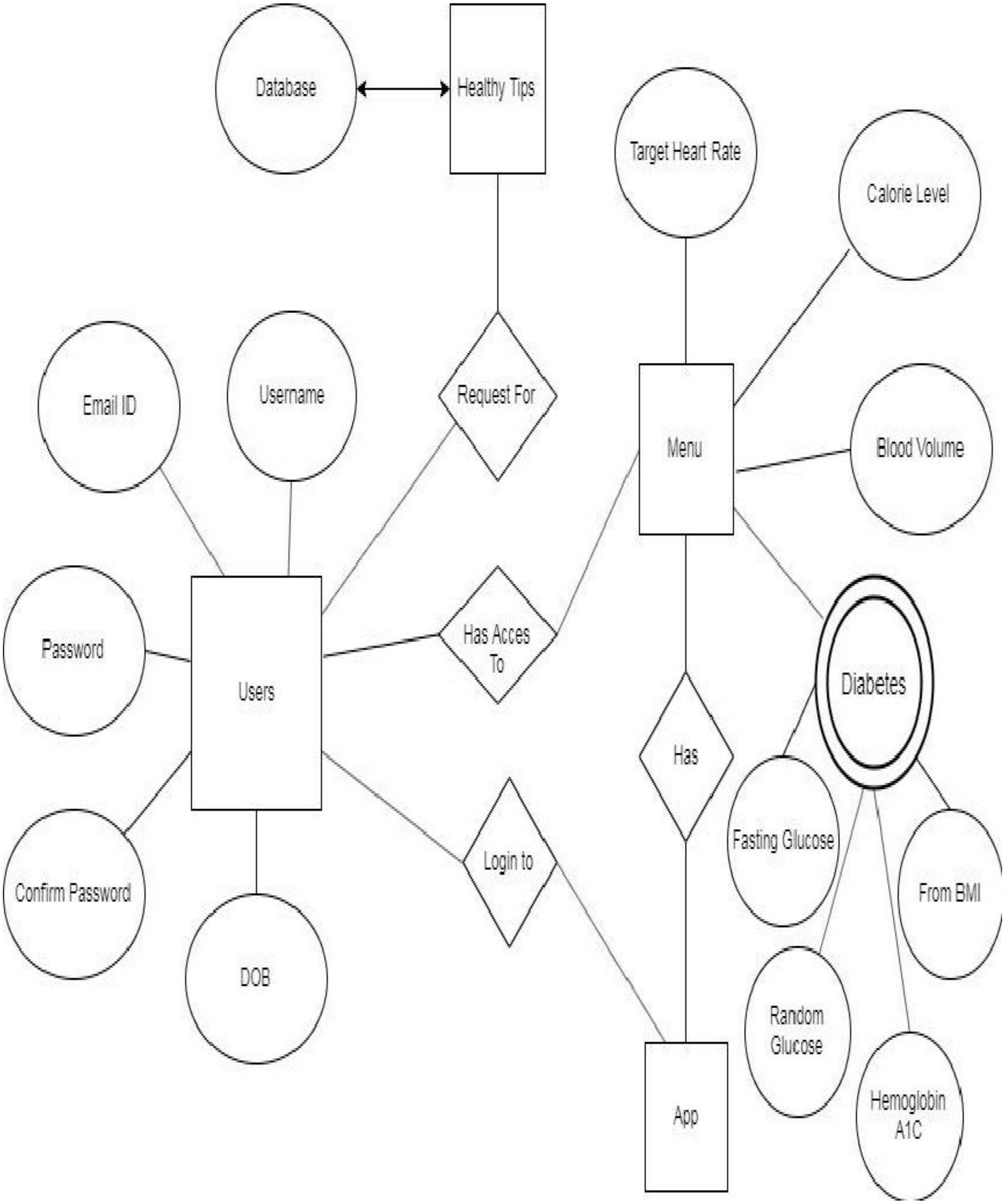


Fig 3.8: Entity Relationship Diagram

This diagram is a type of flowchart that illustrates how “entities” such as people, objects or concepts relate to each other within a system. It shows the interrelation between users, application, Healthy tips.

3.4 Design Details

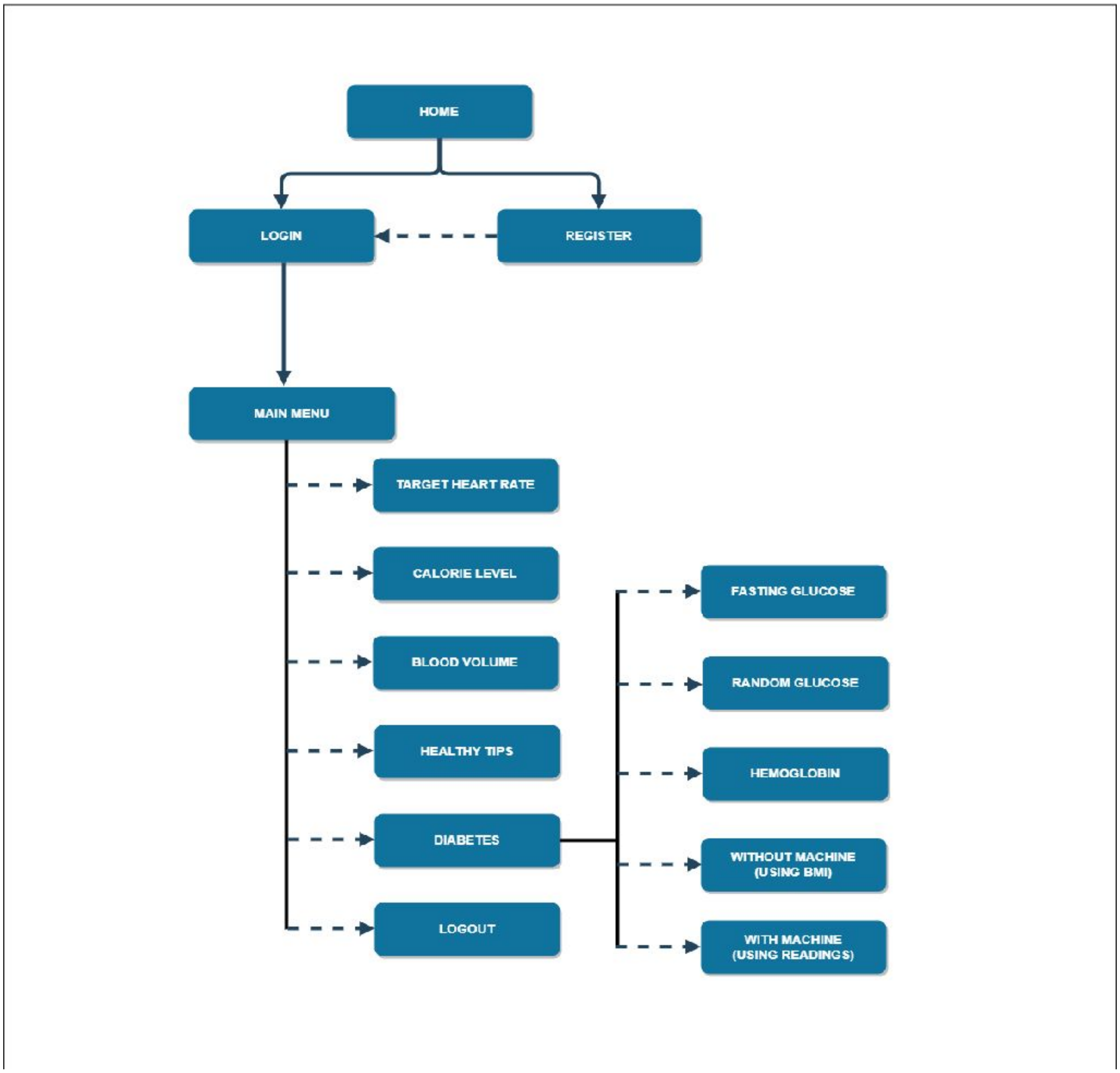


Fig 3.9: Flow Diagram of Android System

Basically, an user will first register himself with the application. After he registers, he can login into the application and he will be redirected to the home page. Here he will be displayed five options for checking/predicting diabetes. These options are target heart rate, calorie level, blood volume, healthy tips, diabetes. In the first option i.e. target heart rate. The target heart rate module allows the user to get information about the different readings related to heart-beat levels, which assists in keeping the heartbeat level at desired level. Firstly, the user has to give certain inputs like resting heart rate, age and activity level. These inputs are then calculated to get the desired output, such as it computes the extreme pulse rate and the higher and lower pulse rate limits. The second module determines the caloric demands of the user based on his/her age, weight, height and activity level, and gives recommendations accordingly. Age, weight, height and activity level are the inputs required from the user. These inputs are used in the calculation of the final result, reflecting how much calories is in the human body. The blood volume component aims at determining the quantity of blood in a human body subject to height and weight. It requires inputs including age, height and weight from the users required for calculating the blood volume. This module records the reading of blood sugar to assist the users for tracking their diet. The user first has to choose from one of the three test types, namely (1) Fasting blood glucose level, (2) Random blood glucose level, and (3) haemoglobin A1C. When the user selects the test type of Fasting Blood Glucose, then he chooses the blood glucose value from required ranges given to the user. The user then gets informed about his blood glucose value that either it is in the normal range, pre- diabetes or diabetes. Similarly, when the user selects the test type of Random Blood Glucose or haemoglobin A1C, then it intimates the user about their blood glucose value, i.e. whether it is in the normal range, pre diabetes or diabetes.

3.5 Techniques and Algorithms

Data Collection:

Data collected through a hardware device and sent to an android application.

Data Pre-processing:

Missing values are replaced with most frequent occurred values (i.e. Median) and also a unit of measurement is standardized e.g. height is obtained in inches or cm, so all values are standardized into cm.

Data Partitioning:

The whole dataset is partitioned into two parts: for example, say, 75% of the dataset is used to train the model and 25% of data is set aside for testing the model.

Machine Learning Algorithms:

eXtreme Gradient Boosting (XGB): XGB is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework.

Random Forest: are an ensemble learning method for classification regression and other tasks that operate by constructing a mess of decision trees at training time and outputting the class that is mode of the classes or mean prediction of the individual trees.

Support Vector Machine (SVM): SVM are supervised learning models with associated learning algorithms that analyze data used for classification and multivariate analysis.

Naïve Bayes (NB): Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a one algorithm but a family of algorithms where all of them share a standard principle

Decision Tree (DT): Is a support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is a technique to display an algorithm that only contains conditional control statements.

3.6 Implementation

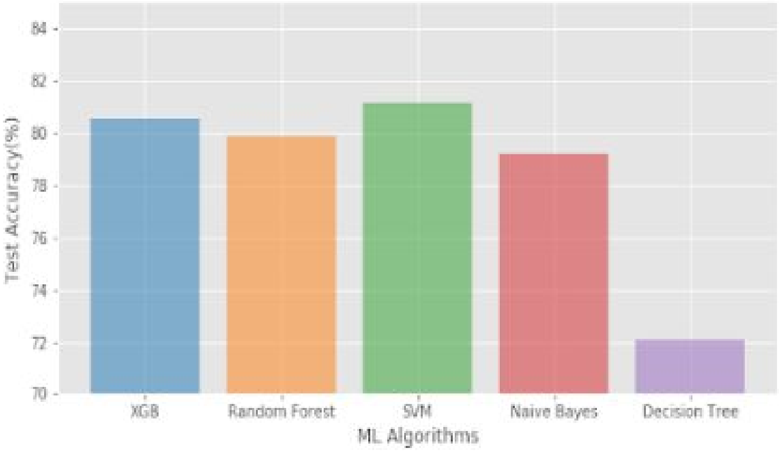


Fig 3.10: Testing Accuracy

To demonstrate the effectiveness of the proposed algorithm, we compare all the algorithms as shown in Figure 3.10. Decision tree pruning may neglect some key values in training data, which can lead the accuracy for a toss. In the case of the Naïve Bayes algorithm, we cannot rely on it for larger datasets as the recall and precision will be very low in such cases. XGB works better when the data size is large. But since the data is sparse in our case, the performance observed in XGB was inefficient. Whereas in the case of Random Forest, which takes the advantage of grid search-like features to find the best tree structure for classification, it has been observed that it works well even with sparse data. But due to excellent scaling and memory-efficient capabilities of the SVM algorithm, it proves to be a better algorithm than the other considered algorithms. And thus, after comparing and analyzing various machine learning algorithms based on testing accuracy, the SVM algorithm gives the highest accuracy and hence, we use the SVM algorithm to build this model.

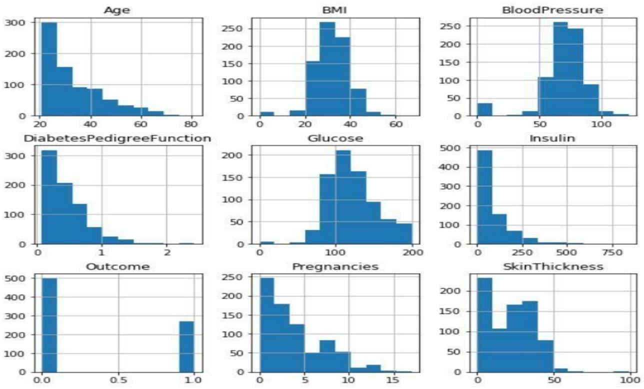


Fig 3.11: Data Distribution

Figure 3.11 Describes how the data is distributed across all attributes

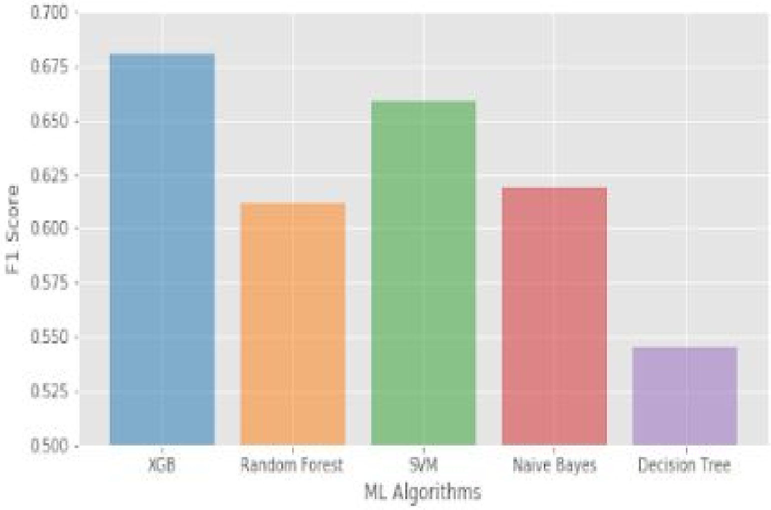


Fig 3.12: F1 Score

In Figure 3.12, we have compared various machine learning algorithms. F1 score a harmonic mean of precision and recall. As depicted in Eq. (1), this score can help to solve any contradiction that may appear between Precision and Recall scores.

For α ∈ R, α > 0.

$$F_\alpha = \frac{(1+\alpha)(\text{Prec} \times \text{Recall})}{(\alpha \times \text{Prec}) + \text{Recall}}$$

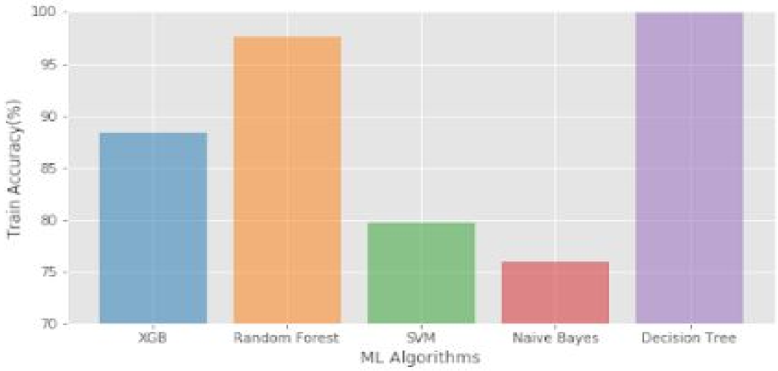


Fig 3.13: Training Accuracy

In Figure 3.13, we have compared various machine learning algorithms. Training accuracy is used to check whether the model is overfitting or under-fitting on the data.

ML Algorithm	Testing Accuracy (%)	Training Accuracy (%)	F1 Score
XGB	80.51	88.82	0.677
RF	79.87	97.63	0.615
SVM	81.16	79.26	0.662
NB	79.22	77.39	0.623
DT	72.07	100	0.538

Table 3.1: Comparative analysis of model performance

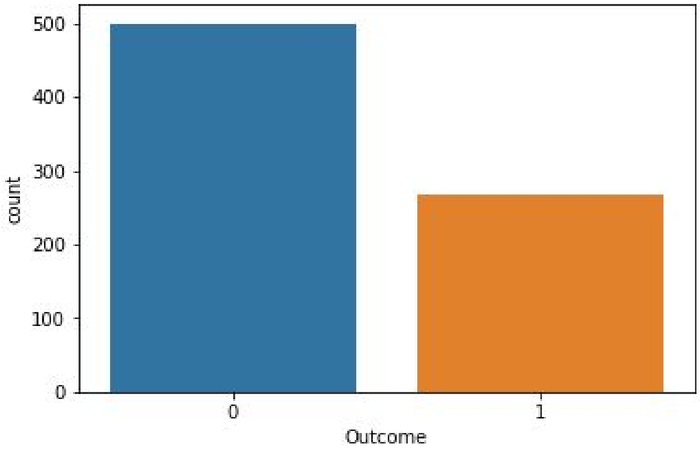


Fig 3.14: Relation of Data

Figure 3.14 describes how we can observe that the data set contains 768 rows and 9 columns. '*Outcome*' is the column which we are going to predict, which says if the patient is diabetic or not. 1 means the person is diabetic and 0 means a person is not. We can identify that out of the 768 persons, 500 are labelled as 0 (non-diabetic) and 268 as 1 (diabetic).

CHAPTER 4
PROJECT ANALYSIS

4.1 Project TimeLine

4.1.1 Gantt Chart:

A Gantt chart is a graphical depiction of a project schedule. A Gantt chart is a type of bar chart that shows the start and finish dates of several elements of a project that include resources, milestones, tasks and dependencies. Henry Gantt, an American mechanical engineer, designed the Gantt chart.

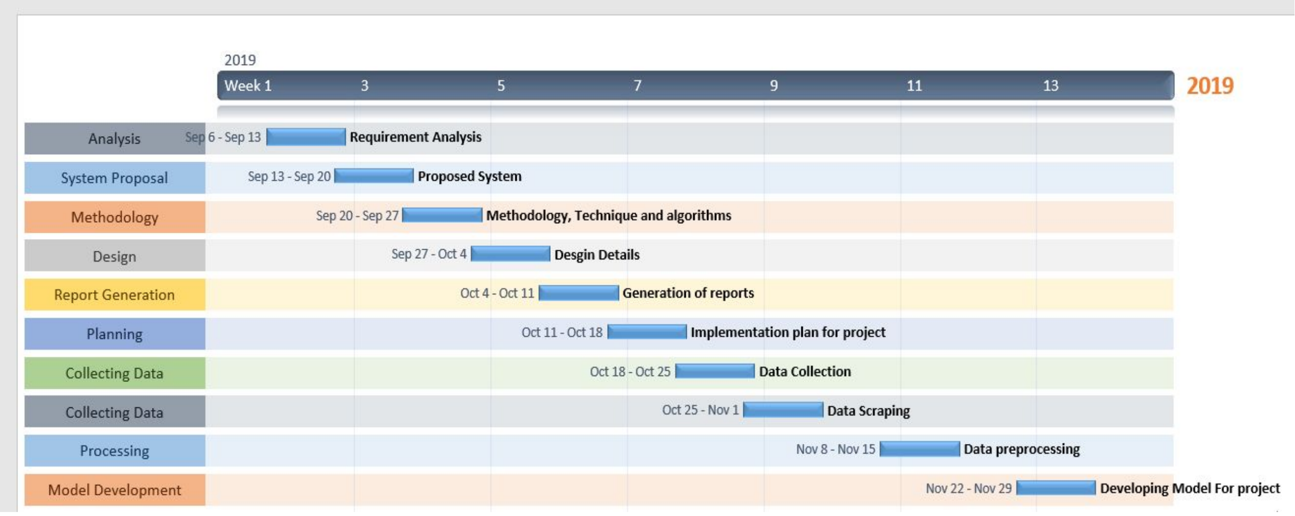


Fig 4.1: Timeline 1

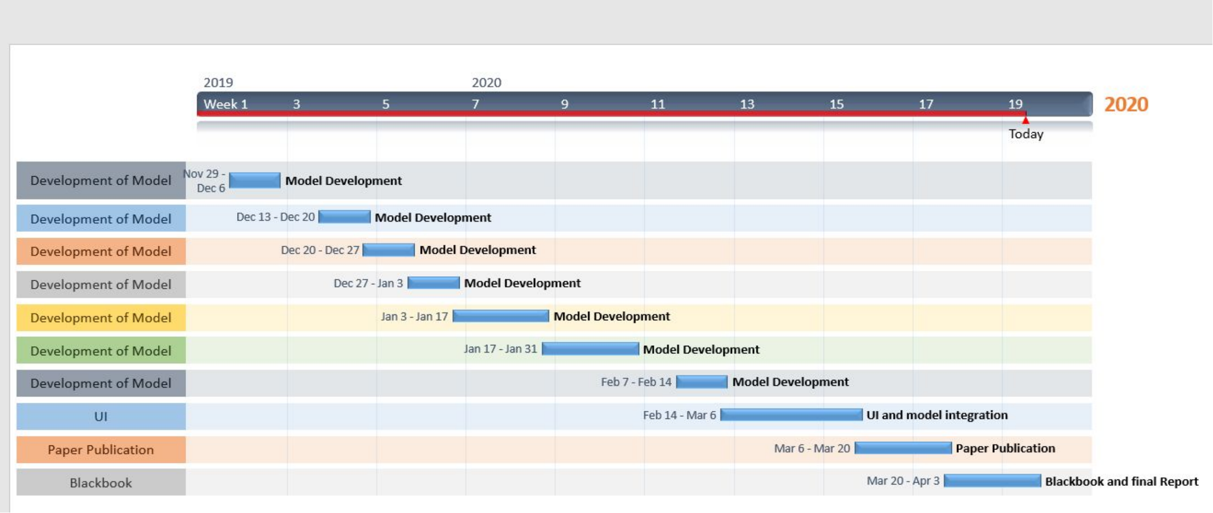


Fig 4.2: Timeline 2

4.2 Task Distribution:

TASK LIST	ASSIGNED TO	STATUS
Defining Project	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	
Literature Review	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	
Survey Paper	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	
Project Plan	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	
Project Analysis	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	
Input Page Design	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	
Documentation of Synopsis	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	
Database Design	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	
Implementation	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	
Testing	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	
Final Report	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	

Final Presentation	Neha Joisher	Complete
	Malav Shah	
	Pujan Sheth	

Table 4.1: Task Distribution

4.3 Development Methodology

This section describes the project as per the various stages of the Software Development life cycle. The model of software development life cycle used in this project is the waterfall method. The Waterfall Method comprises a series of very definite phases, as shown below in figure 4.7, each one run intended to be started sequentially only after the last has been completed, with one or more tangible deliverables produced at the end of each phase of the waterfall method of SDLC. Essentially, it starts with a heavy, documented, requirements planning phase that outlines all the requirements for the project, followed by sequential phases of design, coding, test-casing, optional documentation, verification (alpha-testing), validation (beta-testing), and finally deployment/release.

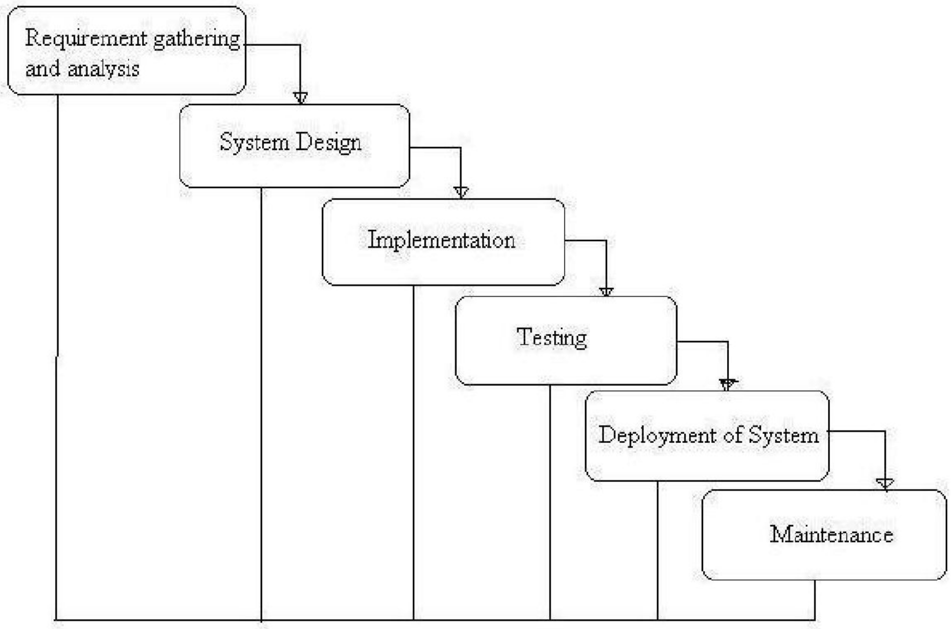


Fig 4.3: Waterfall Model

CHAPTER 5
SYSTEM REQUIREMENTS

5.1 Hardware Requirements

Processor: Intel ® Core (TM) i7-8550U
Main Memory (RAM): 8 GB or More
Cache Memory: 8 MB
Monitor: 13.3” Colour Monitor
Keyboard: 108 keys
Mouse: Optical Mouse
Hard Disk: 320GB or more
System Requirements: 64-bit OS, x64-based processor

5.1.1 Mobile Device

Android Version: 4.0 or More
Ram: 1 GB or More
Screen Resolution: 5.0 or More

5.2 Software Requirements

Front End/Language: Python 3, Java
Back End/Database: Java, Firebase
Platform: Jupyter Notebook, WEKA, Android Studio, Google Firebase
Operating System: Windows 7/Windows 8/ Windows 10

CHAPTER 6
TESTING

6.1 Test Approach

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software testing also provides an objective, independent view of the software to allow the business to appreciate and understand the risks of software implementation.

6.1.1 Black box testing

In black box testing we test the system at random for some random functionalities and depending on the output that we get we come to the conclusion that whether the system we have built is right or wrong. Internal system design is not considered in this type of testing. Tests are based on requirements and functionality. The number of modules and number of java files required for each module is checked.

6.1.2 White box testing

This testing is based on knowledge of the internal logic of an application code. Also known as Glass box Testing. Internal software and code working should be known for this type of testing. Tests are based on coverage of code statements, branches, paths, conditions. All the modules are tested for their logic whether it functions properly or not. Code is checked by inserting different inputs to check its functionality.

6.1.3 Unit testing

Testing of individual software components or modules. Each module was run separately to check the output. Unit testing focuses first on the modules, independently of one another, to locate errors. This enables the tester to detect errors in coding and logical errors that are contained within that module alone. Those resulting from the interaction between modules are initially avoided. Here we test each module individually and integrate the overall system. Unit testing focuses verification efforts even in the smallest unit of software design in each module.

6.1.4 Integration testing

Integration testing is the testing process in software testing to verify that when two or more modules interact and produce a result that satisfies its original functional requirement or not. Integrated testing will start after completion of unit testing.

6.1.5 User Acceptance Testing

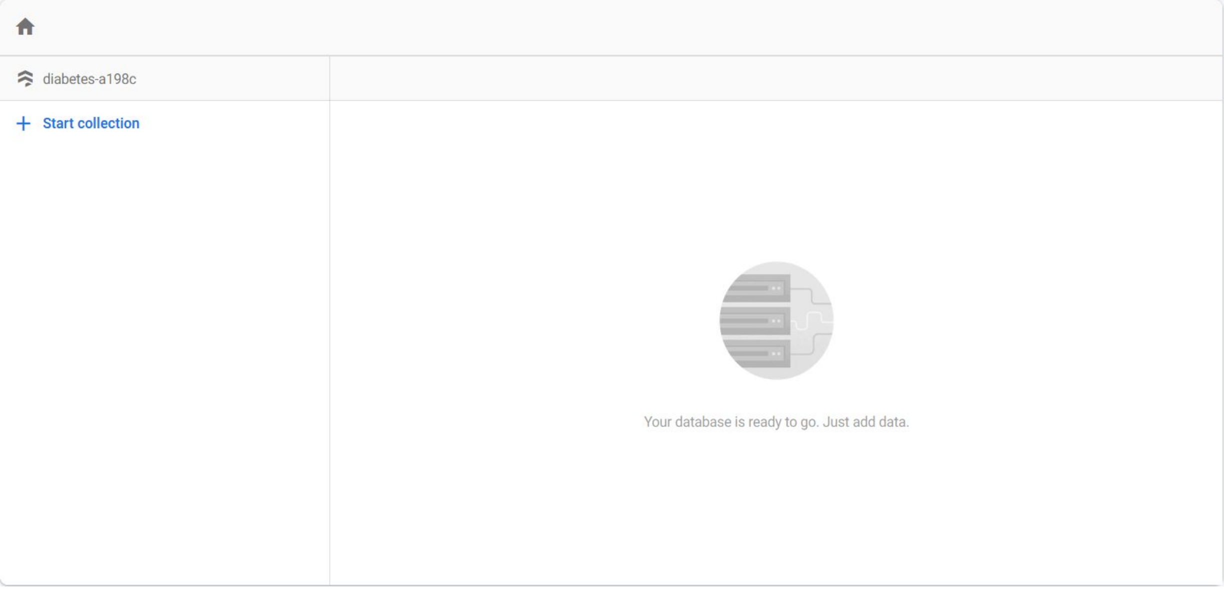
User acceptance testing of the system is the key factor for the success of any system. A system under consideration is tested for user acceptance by constantly keeping in touch with the prospective system at the time of development and making changes whenever required. This is done with regard to the input screen design and output screen design. Here we will test whether the proposed system is having well defined UI so that the citizens can interface the application more easily.

6.1.6 Functional Testing

Functional testing is a technique in which all the functionalities of the program are tested to check whether all the functions that were proposed during the planning phase are fulfilled. This is also to check that if all the functions proposed are working properly. This is further done in two phases One before the integration to see if all the unit components work properly. Second to see if they still work properly after they have been integrated to check if some functional compatibility issues arise.

CHAPTER 7
RESULT ANALYSIS

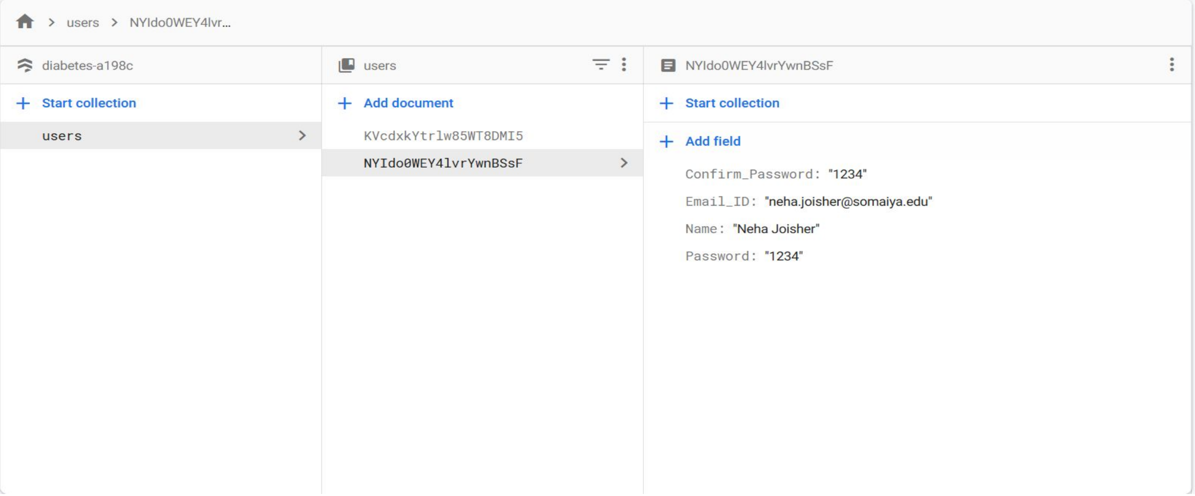
7.1 Result



Your database is ready to go. Just add data.

Fig 7.1: Database Configuration

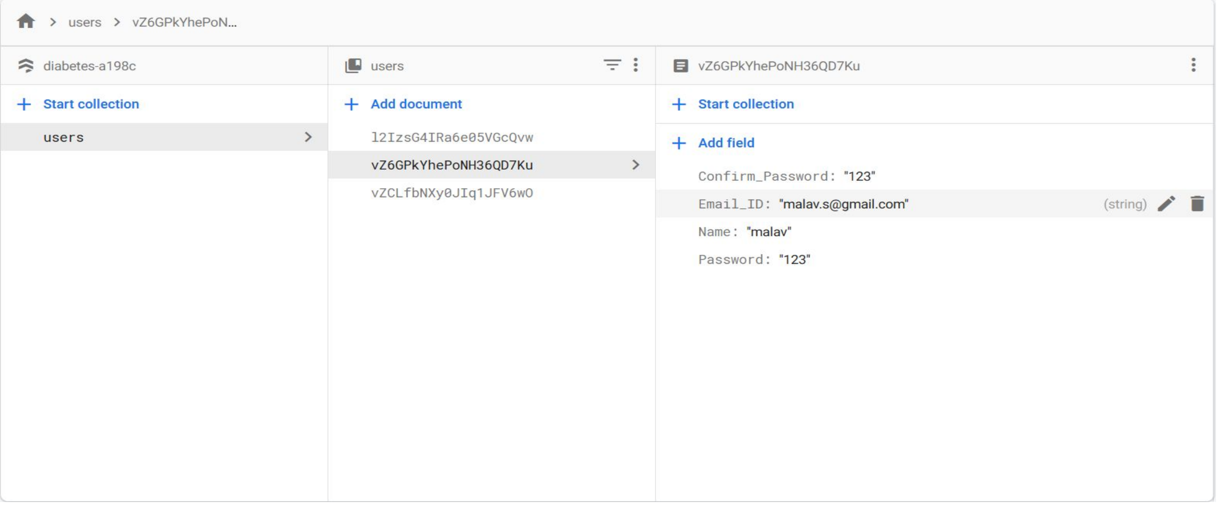
The database selected for backend processing is Google Firebase, which we configured and integrated with the application.



Confirm_Password: "1234"
Email_ID: "neha.joisher@somaiya.edu"
Name: "Neha Joisher"
Password: "1234"

Fig 7.2: Data Insert Into Database After Registration

This picture depicts the data taken from the user during the registration process. This includes the email-id, name, password.



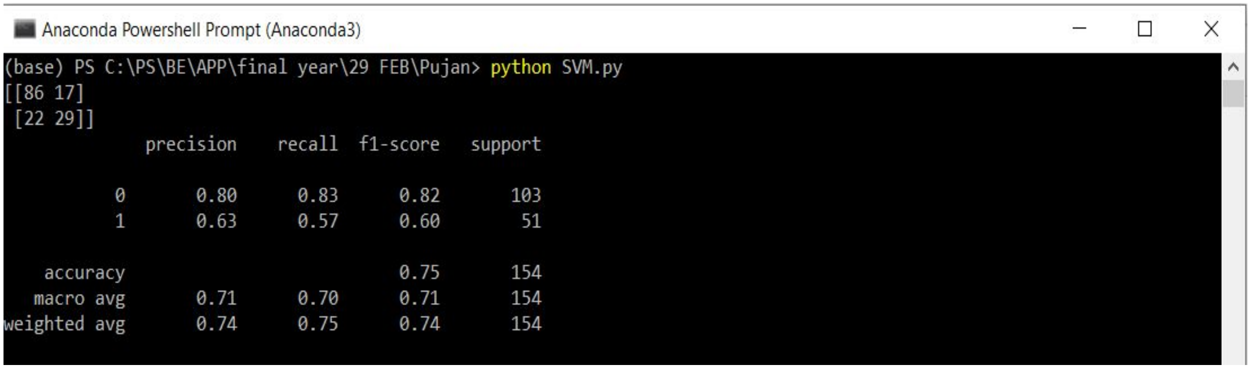
users

12IzsG4IRa6e05VGcQvw
vZ6GPkYhePoNH36QD7Ku
vZCLfbNXy0JIq1JFV6wO

```
Confirm_Password: "123"
Email_ID: "malav.s@gmail.com"   (string)
Name: "malav"
Password: "123"
```

Fig 7.3: Data in Database

This picture shows how the entered data is verified using backend data during logging into the application**.**



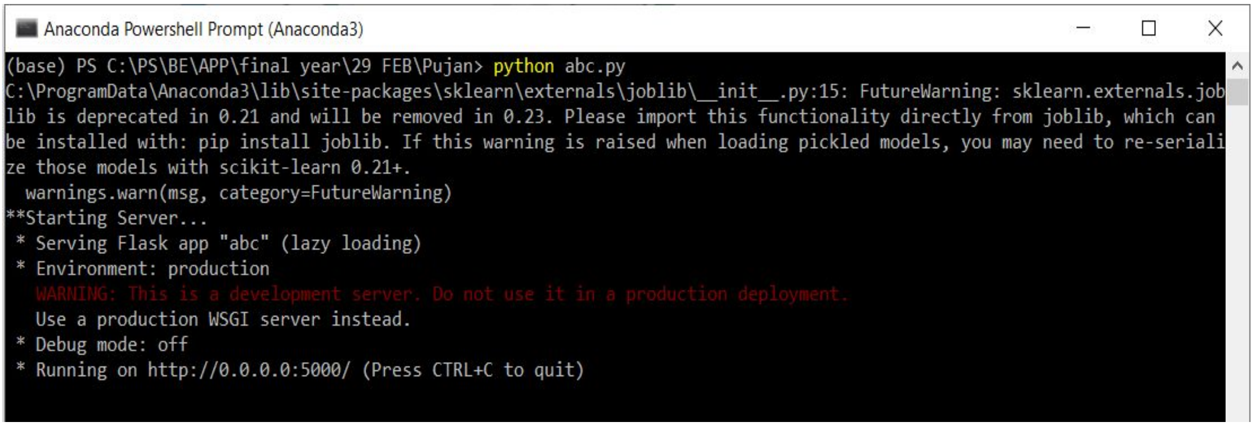
```
(base) PS C:\PS\BE\APP\final year\29 FEB\Pujan> python SVM.py
[[86 17]
 [22 29]]
              precision    recall  f1-score   support

           0       0.80      0.83      0.82       103
           1       0.63      0.57      0.60        51

    accuracy                           0.75       154
   macro avg       0.71      0.70      0.71       154
weighted avg       0.74      0.75      0.74       154
```

Fig 7.4: SVM Model Code Output

This picture shows the accuracy of the used algorithm (SVM algorithm) that is 75%.



```
(base) PS C:\PS\BE\APP\final year\29 FEB\Pujan> python abc.py
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\externals\joblib\__init__.py:15: FutureWarning: sklearn.externals.job
lib is deprecated in 0.21 and will be removed in 0.23. Please import this functionality directly from joblib, which can
be installed with: pip install joblib. If this warning is raised when loading pickled models, you may need to re-seriali
ze those models with scikit-learn 0.21+.
  warnings.warn(msg, category=FutureWarning)
**Starting Server...
 * Serving Flask app "abc" (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: off
 * Running on http://0.0.0.0:5000/ (Press CTRL+C to quit)
```

Fig 7.5: Server Code Output

This picture shows the connection between the server and the application. The application is hosted on 0.0.0.0:5000 on the web.

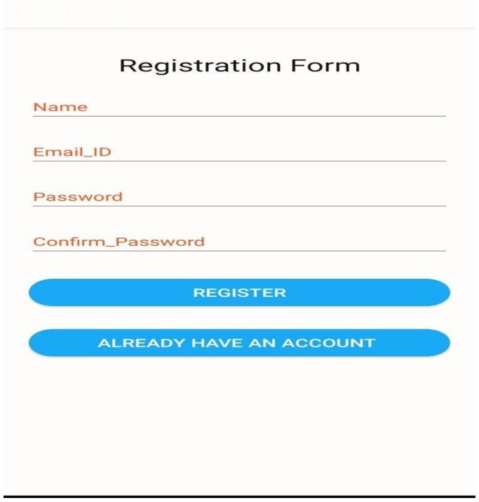


Fig 7.6: Registration Page

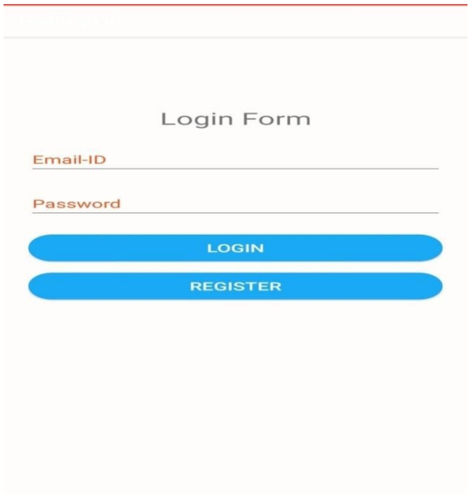


Fig 7.7: Login Page

Fig 7.6 and Fig 7.7 shows the registration and login page of the application respectively.

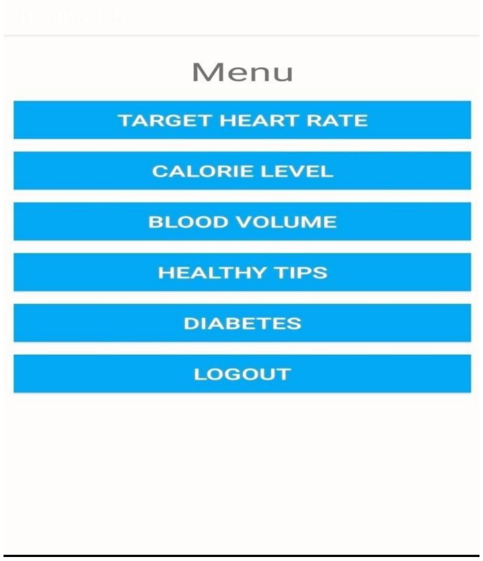
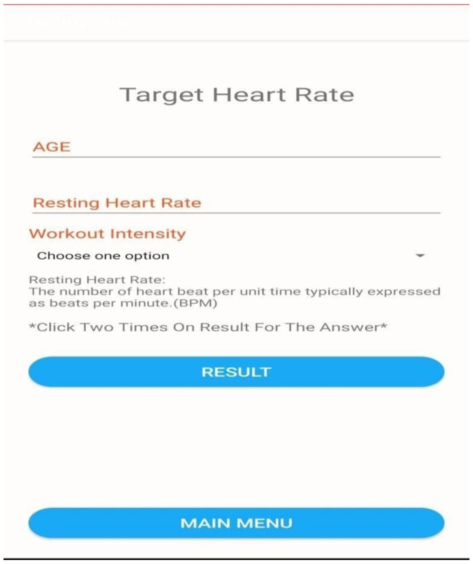


Fig 7.8: Main Menu Page

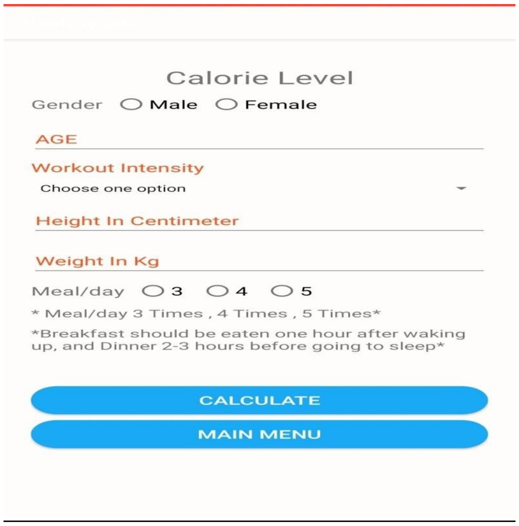


Resting Heart Rate:
The number of heart beat per unit time typically expressed as beats per minute.(BPM)

Click Two Times On Result For The Answer

Fig 7.9: Target Heart Rate (THR) Page

Fig 7.8 and Fig 7.9 shows the main menu page and target heart rate page of the application. Menu page contains options to navigate within the app. Target heart rate page is used to calculate the maximum and minimum heart rate using age and resting heart rate and workout intensity.



Calorie Level

Gender ○ Male ○ Female

AGE

Workout Intensity

Choose one option

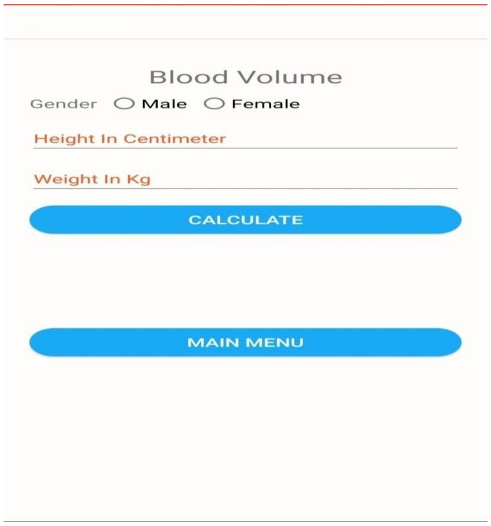
Height In Centimeter

Weight In Kg

Meal/day ○ 3 ○ 4 ○ 5

* Meal/day 3 Times , 4 Times , 5 Times*

Breakfast should be eaten one hour after waking up, and Dinner 2-3 hours before going to sleep



Blood Volume

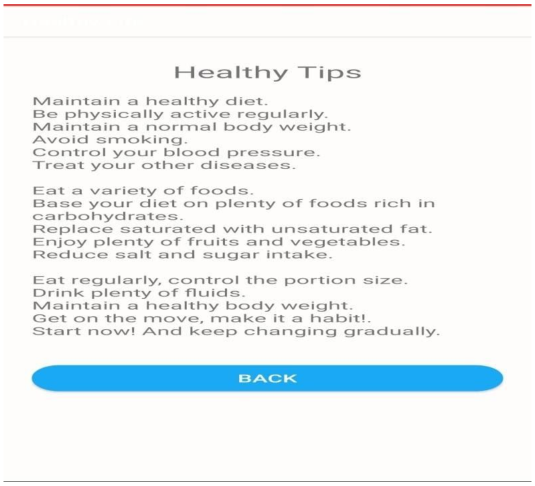
Gender ○ Male ○ Female

Height In Centimeter

Weight In Kg

Fig 7.10: Calorie Level (CL) Page **Fig 7.11: Blood Volume (BV)**
Page

Fig 7.10 and Fig 7.11 shows the Calorie Level page and Blood Volume Page respectively.We calculate the calorie level using age, workout intensity, height, weight, number of meals. Blood Volume is calculated using the height and weight and gender.



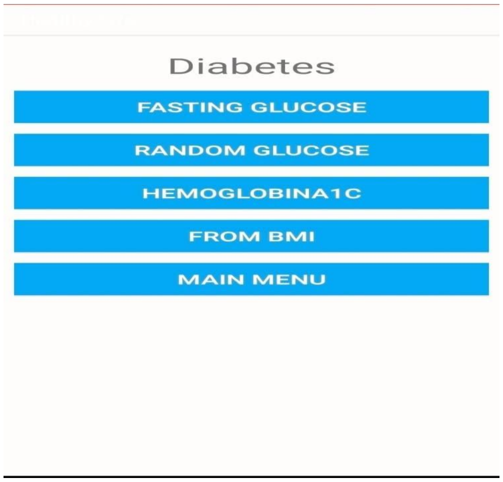
Healthy Tips

Maintain a healthy diet.
Be physically active regularly.
Maintain a normal body weight.
Avoid smoking.
Control your blood pressure.
Treat your other diseases.

Eat a variety of foods.
Base your diet on plenty of foods rich in carbohydrates.
Replace saturated with unsaturated fat.
Enjoy plenty of fruits and vegetables.
Reduce salt and sugar intake.

Eat regularly, control the portion size.
Drink plenty of fluids.
Maintain a healthy body weight.
Get on the move, make it a habit!.
Start now! And keep changing gradually.

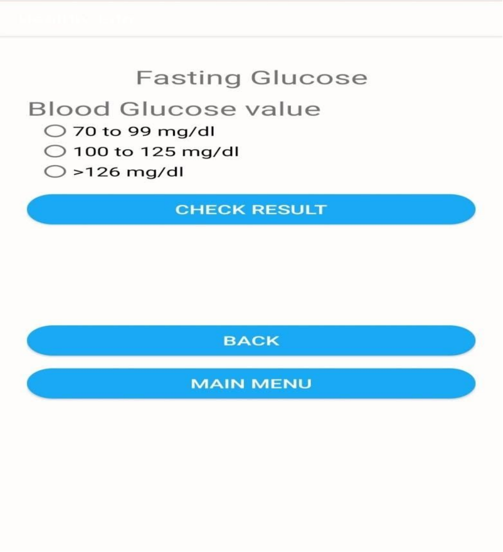
Fig 7.12: Healthy Tips Page



Diabetes

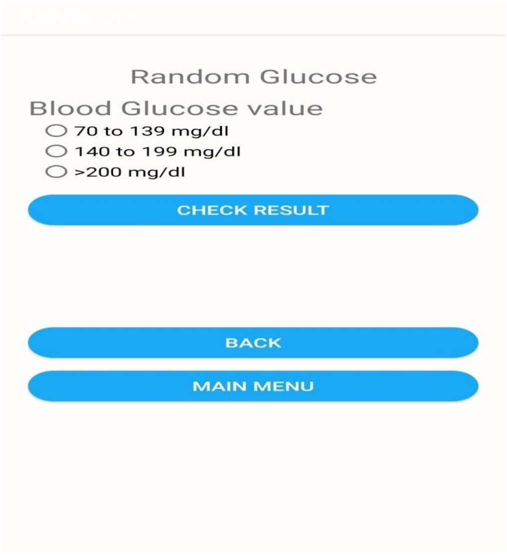
Fig 7.13: Diabetes Menu Page

Fig 7.12 and Fig 7.13 shows the Healthy tips page and Diabetes Menu Page Respectively. Healthy tips page shows the tips to prevent diabetes. Under the diabetes we have different option to cater to the requirements of user such as fasting glucose, random glucose, hemoglobinA1C and BMI


Fasting Glucose

Blood Glucose value

○ 70 to 99 mg/dl
○ 100 to 125 mg/dl
○ >126 mg/dl


Random Glucose

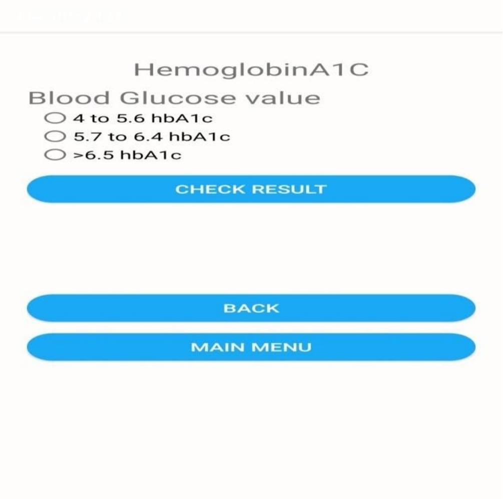
Blood Glucose value

○ 70 to 139 mg/dl
○ 140 to 199 mg/dl
○ >200 mg/dl

Fig 7.14: Fasting Glucose Page

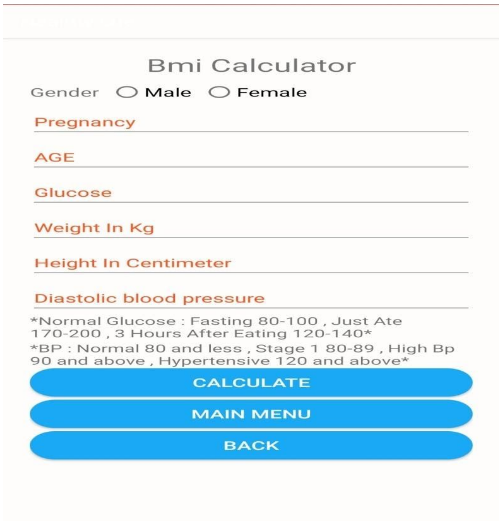
Fig 7.15: Random Glucose Page

Fig 7.14 and Fig 7.15 shows Fasting glucose page and Random glucose page of the application. These pages are provided with different options to calculate diabetes.


HemoglobinA1C

Blood Glucose value

○ 4 to 5.6 hbA1c
○ 5.7 to 6.4 hbA1c
○ >6.5 hbA1c


Bmi Calculator

Gender ○ Male ○ Female

Pregnancy

AGE

Glucose

Weight In Kg

Height In Centimeter

Diastolic blood pressure

Normal Glucose : Fasting 80-100 , Just Ate 170-200 , 3 Hours After Eating 120-140
BP : Normal 80 and less , Stage 1 80-89 , High Bp 90 and above , Hypertensive 120 and above

Fig 7.16: HemoglobinA1C Page

Fig 7.17: BMI Calculator Page

Fig 7.16 and Fig 7.17 shows the Hemoglobin A1C page and BMI calculator page.The Hemoglobin A1C page will show the result that whether or not the user is suffering from diabetes while the BMI calculator page will apply machine learning algorithm for the prediction.

CHAPTER 8
CONCLUSION

8.1 Conclusion

The main purpose and focus of developing the healthcare application are to help people to maintain their health. This healthcare application includes the four modules, namely (1) Target heart rate, (2) calorie level, (3) blood volume, and (4) diabetes app. The first module describes the pulse rate (in beats per minute) that allows the user to exercise safely while getting the maximum benefits from your workout. It includes THR zones which range from low to vigorous i. e (50 to 85) % of MaxHR. The second module is the calorie level, all essential processes of our body, use this measurement unit of energy. In order to encounter the energy needs of our body the speed at which the calorie is used alters continually. Throughout different phases of life, it changes from individual to individual. It is used to determine the caloric needs based on the age, weight, and height and activity level. The third module is the blood volume, which reflects the amount of the blood in the human body. This app assists in answering about how much blood is in the human body, more precisely in your own body depending on the height and weight. The fourth module is the diabetes app that tells about when the body does not properly use or store glucose. Its records, the blood sugar readings, and assists users to track their Diabetes properly.

8.2 Future Scope

In future, we’ll also integrate more apps to our main application to make it a more sophisticated auto-help tool and to provide a wide range of facilities to the end user. These apps will include: (1) Measuring blood pressure and Measuring Weight of the body, (2) Provide reminders to users about their medications which help them to take medicine on time. Therefore, through these reminders, the user can take care of their health, and (3) Graphs of the output obtained will help the user to keep track of the changes in diabetes-related readings and to manage their diet and health in a more effective way.

BIBLIOGRAPHY

Journal Paper

1. Seema Abhijeet Kaveeshwar and Jon Cornwall, "The current state of diabetes mellitus in India," Australas Med J;. PMCID: PMC3920109, pp: 45–48, January 2014.
2. Eleni I. Georgia, Vasilios C. Protopappas, Stavroula G. Mougiakakou and Dimitrios I. Fotiadis, "Short-term vs. Long-term Analysis of Diabetes Data: Application of Machine Learning and Data Mining Techniques," IEEE: 13th International Conference on Bioinformatics and Bioengineering (BIBE), 2013.
3. Yang Guo, Karlskrona, S Guohua Bai and Yan Hu, "Using Bayes Network for Prediction of Type-2 diabetes," IEEE: International Conference on Internet Technology And Secured Transactions, pp: 471 - 472, Dec. 2012.
4. Altikardes Z.A, Erdal H, Baba A.F, Tezcan H, Fak A.S and Korkmaz, H, "A study to classify Non-Dipper/Dipper blood pressure pattern of type 2 diabetes mellitus patients without Holter device," IEEE: Computer Applications and Information Systems (WCCAIS), World Congress, pp: 1-5, 2014.
5. NirmalaDevi M, Appavu S and Swathi U.V, "An amalgam KNN to predict diabetes mellitus," IEEE: International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), pp: 691 - 695, March 2013.
6. Thangarasu G and Petronas Bandar Seri Iskandar, "Prediction of hidden knowledge from Clinical Database using data mining techniques," IEEE: International Conference on Computer and Information Sciences (ICCOINS), pp: 1-5, June 2014.
7. Ramachandran A and Pai V.V.S, "Patient-centered mobile apps for chronic disease management," IEEE: International Conference on Computing for Sustainable Global Development (INDIACom), pp: 948 - 952, March 2014.
8. Ming-Hseng Tseng, Chung-Shan, Hsueh-Chen Hsu, Che-Chia Chang, Hua Ting, Hui-Ching Wu and Ping-Hung Tang, "Development of an Intelligent App for Obstructive Sleep Apnea Prediction on Android Smartphone Using Data Mining Approach," IEEE: 9th International Conference on Autonomic & Trusted Computing (UIC/ATC), pp:774 - 779, Sept. 2012.
9. DassM.V , Rasheed M.A and Ali M.M, "Classification of lung cancer subtypes by data mining technique," IEEE: International Conference on Control, Instrumentation, Energy and Communication (CIEC), pp: 558 - 562, Feb 2014.
10. Othman M.F.B, Abdullah N.B and Kamal N.F.B, "MRI brain classification using support vector machine," IEEE: 4th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO), pp: 1-4, April 2011.
11. Muhammed L.A.N, "Using data mining techniques to diagnose heart disease," IEEE: International Conference on Statistics in Science, Business, and Engineering (ICSSBE), pp: 1-3, Sept. 2012.
12. Guo Weili, Wei Haikun, Zhao Junsheng, Li Weiling and Zhang Kanjian, "Application of

the error function in analyzing the learning dynamics near singularities of the multilayer perceptrons," IEEE: 31st Chinese Control Conference (CCC), pp: 3240 - 3243, July.2012.

13. K.R. Pradeep and N C. Naveen, in the second International conference on contemporary computing and informatics (IC3I), IEEE, pp. 347-352, December 2016.
14. W. Xu, J. Zhang, Q. Zhang and X. Wei, type 2 diabetes on the random forest model. In the International conference on advances in electronics, information, communication.
15. A.S , "Performance analysing of classification algorithms under various datasets, "Computing for sustainable global development, IEEE, pp. 1584-1589, March 2016.
16. C M .Velu and K R. Kashwan, "Classification of diabetic patients using visual data mining methods," In the 3rd IEEE International Advance Computing Conference (IACC), IEEE, pp. 1070-1076, February 2013.
17. V. V. Vijayan and C. Anjali, "Prediction of diabetes mellitus using decision support systems," In 2015 Global conference on Communication Technologies (GCCT), December 2015.
18. S. bashir, U. qamar, F H. khan and M Javed, “Detection and classification of diabetes using ID3, C4.5, & CART ensembles," In 2014 12th International conference on information technology, IEEE, pp. 226-231, June 2015.
19. Yang , "Bayes network for prediction of Type-2 diabetes," In the International conference of 2012 for internet technology and secured transactions, IEEE, pp. 471-472, December 2012.
20. Altikardes, H. Erdal, A. Baba, "Conducted study to classify non dipper or dipper blood pressure patterns of type 2 diabetes mellitus patients without using the Holter device, "The World Congress Computer Applications and Information Systems (WCCAIS), IEEE, pp. 1-5, January 2014.
21. NirmalaDevi and Balamurugan, "Prediction of diabetes with the help of an amalgam KMN," The International conference in 2013 on Emerging trends in computing, communication and nanotechnology (ICECCN) IEEE, pp. 691-695, March 2013.
22. M. Tseng, "Intelligent application for obstructive sleep apnea prediction on android smartphone using data mining approach," The 9th International Conference on Ubiquitous Intelligence and Computing in and 9th International Conference on Autonomic and Trusted Computing, IEEE, pp. 774-779, September 2012.

PUBLICATIONS & CERTIFICATES

“Early Prediction of Pre-Diabetes Using Machine Learning”, 3rd International Conference on Advances in Science & Technology (ICAST 2020) SSRN, Elsevier - Abstract id -.

Early Prediction of Pre-Diabetes Using Machine Learning

Pujan Sheth
K J Somaiya Institute of Engineering & Information
Technology, Sion, Mumbai
University of Mumbai, India
pujan.sheth@somaiya.edu

Malav Shah
K J Somaiya Institute of Engineering & Information
Technology, Sion, Mumbai
University of Mumbai, India
malav.s@somaiya.edu

Neha Joisher
K J Somaiya Institute of Engineering & Information
Technology, Sion, Mumbai
University of Mumbai, India
neha.joisher@somaiya.edu

Radhika Kotecha
K J Somaiya Institute of Engineering & Information
Technology, Sion, Mumbai
University of Mumbai, India
radhika.kotecha@somaiya.edu

Abstract: With the advancement of computer and mobile technologies, mobile heath(mHealth) can be leveraged for patient self-management, patient diagnosis, and determining the possibility of being affected by some disease. Diabetes is a chronic and major disease wherein the body's sugar content is very high over a prolonged period of time and is possibly reaching epidemic proportions in the world. Although there are some mobile applications keeping track of calories, sugar intake, medicine doses, blood glucose, blood pressure, there is no application exceptionally developed to analyze the risk of being diabetic or to identify if a patient is pre-diabetic. Therefore, the objective of this work is to develop an intelligent mHealth application to assess a person's possibility of being pre-diabetic, diabetic & non-diabetic. The proposed approach uses machine learning for prediction and demonstrates promising results.

Keywords: Pre-Diabetes, Machine Learning, mHealth, Random Forest, Support Vector Machine (SVM), Decision Tree

I. INTRODUCTION

Diabetes mellitus, generally referred to as diabetes, is a gathering of metabolic disorders characterized by a high blood glucose level over a constant period of your time. Diabetes is caused either when the pancreas of the body is not producing enough insulin hormones, or the cells of the body not responding properly to the insulin produced. Our body cells are supplemented by Insulin, a type of hormone which serves as a key that allows blood glucose to reach our cells [1].

Diabetes can be classified either as Type1 or Type2. Type1 diabetes is a condition where the cells producing insulin are destroyed and thus repressing the body from being able to yield enough insulin to adequately manage the blood glucose levels. Type 2 diabetes is a metabolic condition caused by insulin resistance in which the body cells never again react to insulin hormone, a circumstance wherein insulin is not used by our body. The global incidence of diabetes was estimated at 422 million within the year 2014, and its prevalence among the adult population has seen in increase from 4.780 to 8.5 % in 2014 [2]. In 2015, an estimated 1.6 million deaths worldwide were linked directly to diabetes.

Early prediction of such disease can be controlled over and can be used to save human life. Machine learning approaches offer powerful ways to derive knowledge by creating predicting models from diagnostic medical data sets gathered from the diabetic patients which might be further accustomed predict the possibility of a person of being pre-diabetic, diabetic and non-diabetic. We have compared five common machine learning algorithms, i.e. Decision tree, Extreme Gradient Boosting, Support Vector Machine, Naive Bayes, Random Forest and decided to use Support Vector Machine algorithm for further predictions. We have combined various modules into one android application like calorie level, Target heart rate, blood volume, diabetes. We use Google Firebase for the database [3].

Following are the contributions of this Paper:

While awareness raising about diabetes is a crucial worry, there are, as far as of our knowledge, no real-world tool or resources to help build this understanding for the people on an individual level. Referable to the lack of awareness about diabetes, many people also don't know about their probability of suffering, which will lead to a severe problem, as length of the untreated condition often raises the hazards associated with it. A mobile application 'Healthy Life' is being presented as a solution for this problem in this report. The smartphone application will serve as an effective tool to help predict the likelihood of becoming pre-diabetic, diabetic and non- diabetic.

II. RELATED WORK

In [4] the authors defined the predictive algorithm’s accuracy varying before and after pre-processing. After pre-processing the dataset, Decision tree technique demonstrated good accuracy compared to J48.

In [5] the authors have proposed a new diabetes risk prediction model. Firstly, the median of each group was utilized to replace missing values and outliers to optimize the original diabetes dataset. Next to that, the updated data set was selected to extract the optimal feature subset by the weighted feature selection algorithm of Random Forest. After this step, they constructed the XGB classifier for diabetes risk prediction.

In [6] the authors have evaluated the performance of Tree based classifiers in WEKA. In order to improve the performance, remove noise from the dataset.

In [7] the authors have discussed on WEKA as it is a powerful tool as it contains both supervised and unsupervised learning techniques.

In [8] the authors have performed the classification of diabetes patients using three classification techniques including Radial Basis Function (RBF), Multi-Layer perceptron and multilevel counter propagation network and performing the simulation tests by WEKA software tool. In [10] the authors have used ID3, C4.5, & CART classification algorithm on diabetes dataset.

In [12] the authors used the Naïve Bayes classifier along with WEKA tool to forecast patients with Type2 diabetes with a good accuracy.

In [13] the authors developed the Non-Dipping/Dipping Patterns prediction method by using data obtained from different sources. Here, Decision Tree along with Naïve- Bayes classifiers are used and data sets are tested using J48 decision tree machine learning algorithm.

In [14] the authors have developed a Hybrid classification model, which combines k-means and k-Nearest Neighbour (KNN) to enhance the accuracy.

In [15] the authors have created a smart phone-based application that provides with a quick and simple approach to pre-screen, high-danger of OSA (Obstructive Sleep Apnea)

III. PROPOSEDAPPROACH

The proposed approach is indicated in the Figure 1.

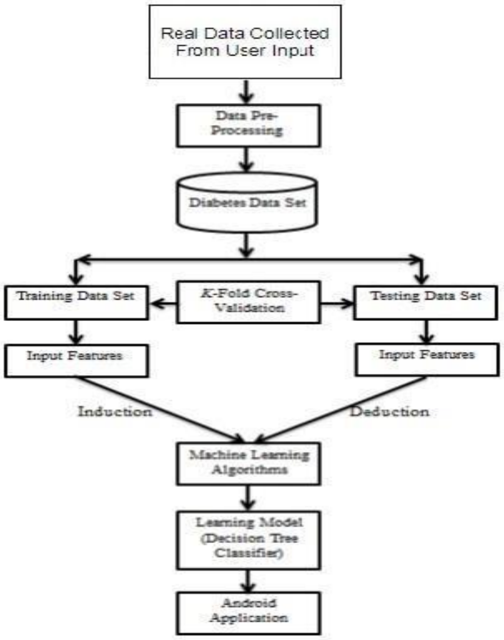


Figure 1: Working of proposed system

The system takes height and weight as input from user through android application and it will calculate BMI of user using predefined formula. After getting input from the user, it will process data and apply machine learning algorithm on it. The system will predict the user’s glucose or sugar level with the data present in dataset and it will predict whether the user will suffer from diabetes or not. The system will also predict as to how many years down the line the user will suffer from diabetes. We also suggest daily routine and food habits to users which will help them to control the sugar level and help people to stay away from diabetes.

IV. EXPERIMENTAL SETUP

We have collected diagnostic data set with 9 diabetic attributes of 768 patients for the study. These attributes are Body Mass Index, plasma glucose concentration, blood- pressure, insulin, etc. Based on these attributes, we compared five common algorithms, namely Decision tree (DT), Extreme Gradient Boosting (XGB), Support Vector Machine (SVM), Random Forest, Naive Bayes (NB):

eXtreme Gradient Boosting (XGB): XGB is a decision-tree-based group Machine Learning algorithm that utilizes a gradient boosting framework.

Random Forest: Random Forests are an ensemble learning approach for classification, regression and other tasks that operate by building a wreck of decision trees at training time and outputting the class that is the mode of the classes or imply mean prediction of the individual trees.

Support Vector Machine (SVM): SVM are supervised learning models with associated learning algorithms that examine data utilized for classification and multivariate analysis.

Naïve Bayes (NB): Naive Bayes classifiers are a set of classification algorithms based on Bayes' Theorem. It is not just one algorithm however a circle of relatives of algorithms where they all share a common principle, i.e. each pair of features being categorized is independent of every other.

Decision Tree (DT): A decision tree is a decision support tool that makes use of tree-like model of decisions and their viable consequences, including chance event outcomes, aid costs, and utility. It is a method to show set of rules that best includes conditional control statements.

To demonstrate the effectiveness of the proposed algorithm, we compare all the algorithms as shown in Figure 2. In case of Decision tree, pruning may neglect some key values in training data, which could lead the accuracy for a toss. In the case of the Naïve Bayes algorithm, we cannot rely on it for larger datasets as the recall and precision will be very low in such cases. XGB works better when the data size is large. But since the data is sparse in our case, the performance observed in XGB was inefficient. Whereas in the case of Random Forest, which takes the advantage of grid search like feature to find the best tree structure for classification, it has been observed that it works well even with sparse data. But due to excellent scaling and memory-efficient capabilities of the SVM algorithm, it proves to be a better algorithm than the other considered algorithms. And thus, after comparing and analyzing various machine learning algorithms based on testing accuracy, the SVM algorithm gives the highest accuracy and hence, we use the SVM algorithm to build this model.

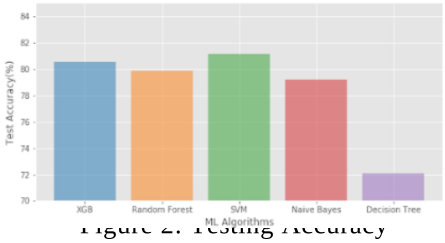


Figure 2: Testing Accuracy

V. RESULTS

In this section, we present analysis of the data and following are the results obtained:

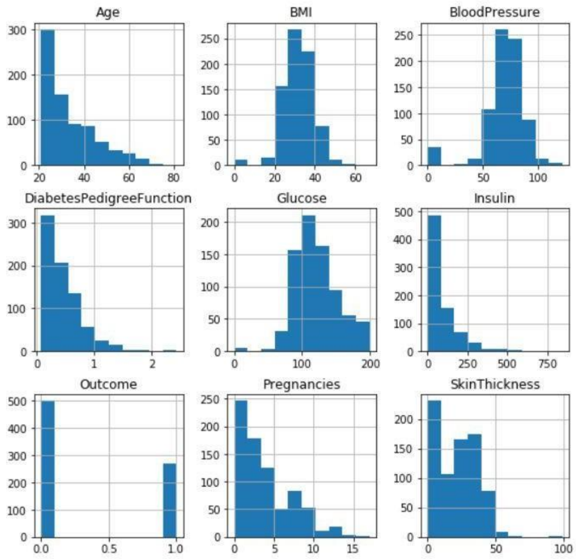


Figure3: Data Distribution

Figure 3 describes how the data is distributed across all attributes.

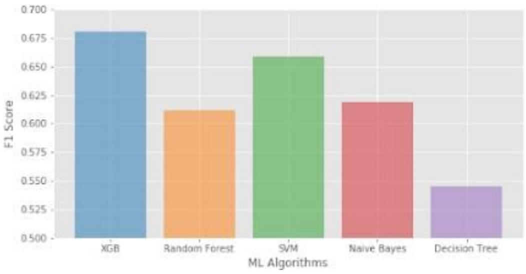


Figure 4: F1 Score

In Figure 4, we have compared various machine learning algorithms. F1 score a harmonic mean of precision and recall. As depicted in Eq. (1), this score can help to solve any contradiction that may appear between Precision and Recall scores.

For α ∈ R, α > 0.

$$F_\alpha = \frac{(1+\alpha)(\text{Prec} \times \text{Recall})}{(\alpha \times \text{Prec}) + \text{Recall}}$$

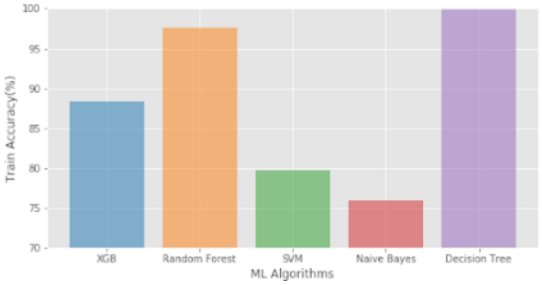


Figure 5: Training Accuracy

In Figure 5, we have compared various machine learning algorithms. Training accuracy is used to check whether the model is over-fitting or under-fitting on the data.

ML Algorithm	**Testing Accuracy (%)**	**Training Accuracy (%)**	**F1 Score**
XGB	80.51	88.82	0.677
RF	79.87	97.63	0.615
SVM	81.16	79.26	0.662
NB	79.22	77.39	0.623
DT	72.07	100	0.538

Table 1: Comparative analysis of model performance

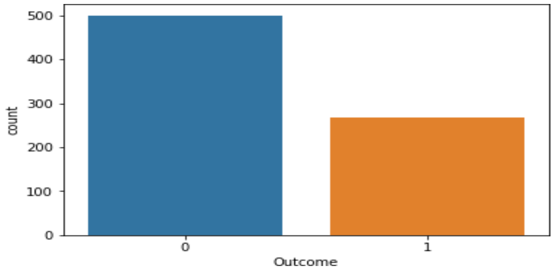


Figure 6: Relation of Data

Figure 6 describes we can observe that the data set contain 768 rows and 9 columns. ‘'Outcome' is the segment which we will foresee, which says if the patient is diabetic or not. 1 signifies that an individual is diabetic and 0 signifies that an individual isn't. Of the 768 people, 500 are listed as non-diabetic and 268 as diabetic.

VI. CONCLUSION & FUTURE WORK

The main purpose and focus of developing this health care application is to assist folks to take care of their health. This application ‘Healthy Life’ includes the four modules, specifically (1) Target heart rate, (2) calorie level, (3) blood volume, and (4) diabetes disease application. This research lays the ground work for building a system that can predict the possibility of a person suffering from pre- diabetes or diabetes. We collected diagnostic data set with 9 diabetic attributes of 768 patients for the study. Based on these attributes, we compared five common machine learning algorithms and after comparing and analyzing various machine learning algorithms based on testing accuracy, the SVM algorithm gives the highest accuracy and hence, we use this algorithm to build the model.

In future,we’ll provide reminders to users concerning their medications that facilitate them to require medication on time.Graphs of the data collected can allow the patient to more accurately monitor improvements in diabetes-related readings and to control their diet and health.

VII. REFERENCES

[1] A. Singhal , K. Sowjanya and C. choudhary , In 2015 International Advance Computing Conference (IACC) IEEE, pp. 397-402, 2015

[2] H. Abbas, L. Alic, M. Rios, M. Abdul-Ghani and K Learning," Symposium on Computer-Based Medical Systems (CBMS), IEEE, pp. 567- 570, June 2019.

[3] M. F. Faruque, Asaduzzaman and I. H. Sarker, on electrical, computer and communication engineering (ECCE), Cox'sBazar, IEEE, pp. 1-4, February 2019.

[4] K R. Pradeep and N C. Naveen, in the second International conference on contemporary computing and informatics (IC3I), IEEE, pp. 347-352, December 2016.

[5] W. Xu, J. Zhang, Q. Zhang and X. Wei, type 2 diebetes on the random forest model. In the International conference in on advances in electronics, information, communication.

[6] A.S , "Performance analysing of classification algorithms under various datasets, "Computing for sustainable global development, IEEE, pp. 1584-1589, March 2016.

[7] C M .Velu and K R. Kashwan, "Classification of diabetic patients using visual data mining methods," In the 3rd IEEE International Advance Computing Conference (IACC), IEEE, pp. 1070-1076, February 2013.

[8] V. V. Vijayan and C. anjali, "Prediction of diabetes mellitus using decision support systems," In 2015 Global conference on Communication Technologies (GCCT), December 2015.

[9] Velu and K. R. Kashwan, "Counter Propagation of multi-level Network for diabetes classification," In 2013 International Conference on Signal Processing, Image Processing & Pattern Recognition, IEEE, pp. 190-194, April 2013.

[10] S. bashir, U. qamar, F H. khan and M Javed, “Detection and classification of diabetes using ID3, C4.5, & CART ensembles," In 2014 12th International conference on information technology, IEEE, pp. 226-231, June 2015.

[11] B. Lee, B. Ku and J Kim, " Using anthropometric measures prediction of fasting plasma glucose for the detection of diabetes," In IEEE Journal of Biomedical and Health Informatics, IEEE, vol. 18, no. 2, pp. 555-561, March 2014.

[12] Yang , "Bayes network for prediction of Type-2 diabetes," In the International conference of 2012 for internet technology and secured transactions, IEEE, pp. 471-472, December 2012.

[13] Altikardes, H. Erdal, A. Baba, "Conducted study to classify non dipper or dipper blood pressure pattern of type 2 diabetes mellitus patients without using the Holter device, "The World Congress in 2014 on Computer Applications and Information Systems (WCCAIS), IEEE, pp. 1-5, January 2014.

[14] NirmalaDevi and Balamurugan, "Prediction of diabetes with the help of an amalgam KMN," The International conference in 2013 on Emerging trends in computing, communication and nanotechnology (ICECCN) IEEE, pp. 691-695, March 2013.

[15] M. Tseng, "Intelligent application for obstructive sleep apnea prediction on android smartphone using data mining approach," The 9th International Conference on Ubiquitous Intelligence and Computing in and 9th International Conference on Autonomic and Trusted Computing, IEEE, pp. 774-779, September 2012.[