# 5G-SMART DIABETES TOWARD PERSONALIZED DIABETES DIAGNOSIS WITH HEALTHCARE BIG DATA CLOUDS

**A Major Project Report submitted in partial fulfillment of the requirements for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**

**Submitted by**


**Aarsh Varma- 221910313002**
**Rohith Venkatesh- 221910313007**
**Yenugupalli Snigdha- 221910313053**
**Pusapati Geetanjali- 221910313058**

**Under the esteemed guidance of**
**DR. PRAVIN MUNDHE**
**Assistant Professor**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**GITAM**

**(Deemed to be University)**

**HYDERABAD**


**March- 2023**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GITAM SCHOOL OF TECHNOLOGY**

**GITAM**

(Deemed to be University)

# DECLARATION

We, hereby declare that the project report entitled "**5G-SMART DIABETES TOWARD PERSONALIZED DIABETES DIAGNOSIS WITH HEALTHCARE BIG DATA CLOUDS**" is an original work done in the Department of Computer Science and Engineering, GITAM Institute of Technology, GITAM (Deemed to be University) submitted in partial fulfillment of the requirements for the award of the degree of B.Tech. in Computer Science and Engineering. The work has not been submitted to any other college or University for the award of any degree or diploma.

Date: 30th March 2023

| Registration No. | Name | Signature |
|---|---|---|
| 221910313002 | AARSH VARMA | |
| 221910313007 | ROHITH VENKATESH | |
| 221910313053 | YENUGUPALLI SNIGDHA | |
| 221910313058 | PUSAPATI GEETANJALI | |

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GITAM SCHOOL OF TECHNOLOGY**

**GITAM**

(Deemed to be University)

## CERTIFICATE

This is to certify that the project report entitled "**5G-SMART DIABETES TOWARD PERSONALIZED DIABETES DIAGNOSIS WITH HEALTHCARE BIG DATA CLOUDS**" is a bonafide record work carried out by Aarsh Varma (221910313002), Rohith Venkatesh (221910313007), Yenugupalli Snigdha (221910313053) and Pusapati Geetanjali (221910313058) submitted in partial fulfillment of requirement for the award of degree of Bachelors of Technology in Computer Science and Engineering.

| Project Guide | Project Coordinator | Head of the Department |
|---|---|---|
| **DR. PRAVIN MUNDHE** | **DR. S. APARNA** | **PROF. S. PHANI KUMAR** |
| **Assistant Professor** | **Assistant Professor** | **Professor** |
| **Dept.Of CSE** | **5Dept. Of CSE** | **Dept. Of CSE** |

# ACKNOWLEDGEMENT

Our project would not have been successful without the help of several people. We would like to thank the personalities who were part of our project in numerous ways, those who gave us outstanding support from the project's inception.

We are incredibly thankful to our Honourable Pro-Vice Chancellor, **Prof. D Sambasiva Rao**, for providing the necessary infrastructure and resources to accomplish our project.

We are highly indebted to **Prof. N. Seetharamaiah**, Principal, School of Technology, for his support during the project's tenure.

We are very much obliged to our beloved **Prof. S. Phani Kumar**, Head of the Department of Computer Science & Engineering, for providing the opportunity to undertake this project and for encouragement in the completion of this project.

We are also thankful to **Dr. S. Aparna**, Assistant Professor, Department of Computer Science and Engineering for her support and constant encouragement.

We hereby express our deep gratitude to **Dr. Pravin Mundhe**, Assistant Professor, Department of Computer Science & Engineering, for his esteemed guidance, moral support, and invaluable advice for the success of the project.

We are also thankful to all the Computer Science and Engineering department staff members who have cooperated in making our project successful. We would like to thank all our parents and friends who extended their help, encouragement, and moral support in our project work, directly or indirectly.

Sincerely,

AARSH VARMA (221910313002)

ROHITH VENKATESH (221910313007)

YENUGUPALLI SNIGDHA (221910313053)

PUSAPATI GEETANJALI (221910313058)

## TABLE OF CONTENTS

# LIST OF FIGURES

# 1. ABSTRACT

Recent advances in wireless networking and big data technologies, such as 5G networks, medical big data analytics, and the Internet of Things, along with recent developments in wearable computing and artificial intelligence, are enabling the development and implementation of innovative diabetes monitoring systems and applications. Due to the life-long and systematic harm suffered by diabetes patients, it is critical to design effective methods for the diagnosis and treatment of diabetes. Based on our comprehensive investigation, this article classifies those methods into Diabetes 1.0 and Diabetes 2.0, which exhibit deficiencies in terms of networking and intelligence. Thus, our goal is to design a sustainable, cost-effective, and intelligent diabetes diagnosis solution with personalized treatment. In this article, we first propose the 5G-Smart Diabetes system, which combines the state-of-the-art technologies such as wearable 2.0, machine learning, and big data to generate comprehensive sensing and analysis for patients suffering from diabetes. Then we present the data sharing mechanism and personalized data analysis model for 5G-Smart Diabetes. Finally, we build a 5G-Smart Diabetes testbed that includes smart clothing, smartphone, and big data clouds. The experimental results show that our system can effectively provide personalized diagnosis and treatment suggestions to patients.

# 2. INTRODUCTION

Diabetes is an extremely common chronic disease from which nearly 8.5 percent of the world population suffers; 422 million people worldwide have to struggle with diabetes. It is crucial to note that type 2 diabetes mellitus makes up about 90 percent of the cases [1]. More critically, the situation will be worse, as reported in [2] with more teenagers and youth becoming susceptible to diabetes as well. Due to the fact that diabetes has a huge impact on global well-being and economy, it is urgent to improve methods for the prevention and treatment of diabetes [3]. Furthermore, various factors can cause the disease, such as improper and unhealthy lifestyle, vulnerable emotion status, along with the accumulated stress from society and work. However, the existing diabetes detection system faces the following problems:

• the system is uncomfortable, and real-time data collection is difficult. Furthermore, it lacks continuous monitoring of multi-dimensional physiological indicators of patients suffering from diabetes [4,5].

• The diabetes detection model lacks a data sharing mechanism and personalized analysis of big data from different sources including lifestyle, sports, diet, and so on [6,7].

• There are no continuous suggestions for the prevention and treatment of diabetes and corresponding supervision strategies [8,9]. To solve the above problems, in this article, we first propose a next generation diabetes solution called the 5G-Smart Diabetes system, which integrates novel technologies including fifth generation (5G) mobile networks, machine learning, medical big data, social networking, smart clothing,[10] and so on.

Then we present the data sharing mechanism and personalized data analysis model for 5G-Smart Diabetes. Finally, based on the smart clothing, smartphone, and big data healthcare clouds, we build a 5G-Smart Diabetes testbed and give the experiment results. Furthermore, the “5G” in 5G-Smart Diabetes has a two-fold meaning. On one hand, it refers to the 5G technology that will be adopted as the communication infrastructure to realize high-quality and continuous monitoring of the physiological states of patients with diabetes and to provide treatment services for such patients without restraining their freedom. On the other hand, “5G” refers to the following “5 goals”: cost effectiveness, comfortability, personalization, sustainability, and smartness Cost Effectiveness: It is achieved from two aspects. First, 5G-Smart Diabetes keeps users in a healthy lifestyle so as to prevent users from getting the disease in the early stage. The reduction of disease risk would lead to decreasing the cost of diabetes treatment. Second, 5G-Smart Diabetes facilitates out-of-hospital treatment, thus reducing the cost compared to on-the-spot treatment, especially long-term hospitalization of the patient. Comfortability: To achieve comfort for patients, it is required that 5G-Smart Diabetes does not disturb the patients’ daily activities as much as possible. Thus, 5G-Smart Diabetes integrates smart clothing[3], mobile phones, and portable blood glucose monitoring devices to easily monitor patients’ blood glucose and other physiological indicators. Personalization: 5G-Smart Diabetes utilizes various machine learning and cognitive computing algorithms to establish personalized diabetes diagnosis for the prevention and treatment of diabetes. Based on the collected blood glucose data and individualized physiological indicators, 5G-Smart Diabetes produces personalized treatment solutions for patients. Sustainability: By continuously collecting, storing, and analyzing information on personal diabetes, 5G-Smart Diabetes adjusts the treatment strategy in time based on the changes of patients’ status. Furthermore, in order to be sustainable for data-driven diabetes diagnosis and treatment, 5G-Smart Diabetes establishes effective information sharing among patients, relatives, friends, personal health advisors, and doctors. With the help of social networking, the patient’s mood can be better improved so that he or she is more self-motivated to perform a treatment plan in time. Smartness: With cognitive intelligence toward patients’ status and network resources, 5G-Smart Diabetes achieves early detection and prevention of diabetes and provides personalized treatment to patients. The remaining part of the article is organized as follows. We first present the system architecture of 5G-Smart Diabetes. Then we explain the data sharing mechanism and propose the personalized data analysis model. Furthermore, we introduce the 5G-Smart Diabetes test bed. Finally, the conclusion of this article is given.

# 3. LITERATURE REVIEW

**Paper 1: A Smart Diabetics System in Health care Analytics**

- **Publication Year:** 2021
- **Author:** P.Ravikumaran, V. Rajeshkannan, K. Abirami
- **Journal Name:** International Journal on Research Innovations in Engineering Science and Technology(IJRIEST)
- **Summary:** The healthcare business is being refined at an exponential rate as a result of the speed with which digital innovation and technology disruption are occurring. The massive amount of healthcare data continues to grow by the minute, making it increasingly impossible to find any sort of useful information. Big data analytics is currently transforming traditional information distribution into actionable insights. Big data analytics has a lot of benefits in the healthcare industry, such as detecting critical diseases early on and delivering better healthcare services to the appropriate patient at the right time to improve the quality of life care. Existing health data analytics platforms that provide procedural mechanisms for data collection, gathering, processing, analysis, visualization, and interpretation have a number of difficulties to address. The authors present a long-term, commercially viable, and intellectual diabetes diagnosis solution with tailored therapy in this study
- **Literature Review:** To approve the introduction of the proposed 5G-Smart Diabetes work environment, three models of deep learning techniques namely CNN, CNN with LSTM and CNN with GRU are embraced to decide different models for the common open conclusion of diabetes. The authors work out the gathering strategy to conduct integration of the models. In this paper, it is proposed to have a Smart Diabetes framework that incorporates a detecting layer, a personalized conclusion layer, and a information sharing layer. It investigates how big data analytics offers a incredible boon to the healthcare industry, because it makes a difference to form way better choices and investigation. By utilizing these analytics measurements, information researchers able to put together healthcare related data from both inside and outside sources.

**Paper 2: Identification of Type 2 Diabetes Risk Factors using Phenotypes Consisting of Anthropometry and Triglycerides based on Machine Learning**

- **Publication Year:** 2015
- **Author:** Bum Ju Lee and Jong Yeol Kim
- **Journal Name:** IEEE Journal of Biomedical and Health Informatics
- **Summary:** The aims of the present study were to assess the association between the HW phenotype and type 2 diabetes in Korean adults and to evaluate the predictive power of various phenotypes consisting of combinations of individual anthropometric

measurements and TG levels. The authors employed binary logistic regression to examine statistically significant differences between normal subjects and those with type 2 diabetes using HW and individual anthropometric measurements. For more reliable prediction results, two machine learning algorithms, naive Bayes (NB) and logistic regression (LR), were used to evaluate the predictive power of various phenotypes. All prediction experiments were performed using a 10-fold cross validation method.

- **Literature Review:** The authors demonstrated that the presence of the HW phenotype was the variable that was most strongly associated with type 2 diabetes. The association between WC and type 2 diabetes was stronger than the association between TG levels and type 2 diabetes. When examining the predictive powers of WC and TG levels alone, WC was a good predictor of type 2 diabetes, whereas TG was not. The retrospective cross-sectional design of this study does not allow us to establish a cause-effect relationship. To predict type 2 diabetes, the proposed method in this study suggested that the actual TG and WC values be used instead of the range of TG and WC values and various phenotypes consisting of individual anthropometric measurements and TG. This method may easily be used in experiments to identify the best phenotype or predictor of type 2 diabetes in various countries. However, the findings of this study cannot be applied to other populations because the study population included only Korean women and men.

**<u>Paper 3: Application of Adaptive Neuro-Fuzzy Inference System for Diabetes Classification and Prediction</u>**

- **Publication Year:** 2017
- **Author:** Oana Geman , Iuliana Chiuchisan , Roxana Toderean (Aldea)
- **Journal Name:** IEEE Journal
- **Summary:** In this paper is proposed a hybrid Adaptive Neuro-Fuzzy Inference System (ANFIS) model for classifying patients with diabetes based on data sets with diabetic patients (Pima Indians Diabetes Dataset). The Pima Indians Diabetes Dataset contains 768 samples. In order to set the features vector of this system is used Diabetes Pedigree Function to define the fuzzy rule base with multiple premises. The Neuro-Fuzzy ANFIS modeling was implemented using ANFIS Fuzzy Logic Toolbox and MATLAB Toolbox. The performances of the algorithm were analyzed in terms of specificity, precision and sensitivity. The proposed neural network was trained and tested on Pima Indians Diabetes Database, proving an accuracy of 85.35% for training data and 84.27% for testing data.
- **Literature Review:** The ANFIS network performance consists precisely in the reduction of the number of initial rules, by reducing the feature vector and by introducing the DPF in features vector, as a criterion for initial screening. The results show that the proposed system, trained on a combination of algorithms has a good accuracy of the results classification and a minimum error.

## Paper 4: Cloud-supported Cyber-Physical Localization Framework for Patients Monitoring

- **Publication Year:** 2015
- **Author:** M. Shamim Hossain
- **Journal Name:** IEEE Systems
- **Summary:** The potential of cloud-supported cyber physical systems (CCPSs) has drawn a great deal of interest from academia and industry. CCPSs facilitate the seamless integration of devices in the physical world (e.g., sensors, cameras, microphones, speakers, GPS) with cyberspace. This enables a range of emerging applications or systems such as patient or health monitoring, which require patient locations to be tracked. To this end, this paper proposes a cloud-supported cyber-physical localization system for patient monitoring using smartphones to acquire voice and electroencephalogram signals in a scalable, real-time, and efficient manner. The proposed approach uses Gaussian mixture modeling (GMM) for localization, and is shown to outperform other similar methods in terms of error estimation**.**
- **Literature Review:** Most localization technologies (GPS, cellular (A-GPS), WLAN) provide satisfactory results in outdoor environments. However, very few achieve good accuracy for indoor localization. Indoor positioning or localization services are very important for the safety of patients in hospitals and homes, whereas outdoor localization is suitable when the patient is on the move. A hybrid patient tracking localization technique is proposed to overcome this issue. Each of the above approaches has advantages and disadvantages with regard to scalability, reliability, real-time processing, accurate localization, and storage infrastructure. Consequently, there is a need for new adaptable and scalable techniques for seamless localization in the cloud or cyberspace. These would allow the computational resources, sensory capabilities, and localization mechanisms of smartphones to be used as physical components of the cloud in a scalable, real-time, and efficient manner.

# 4. PROBLEM IDENTIFICATION AND OBJECTIVES

## 4.1 EXISTING SYSTEM

The problem being addressed is the lack of personalized and accurate diagnosis of diabetes, which can lead to suboptimal treatment and management of the disease. This problem can be addressed by leveraging the power of 5G technology, smart devices, and healthcare big data clouds to create a system for personalized diabetes diagnosis.

The proposed solution involves using smart devices such as continuous glucose monitors and activity trackers to collect real-time data on a patient's blood sugar levels, physical activity, and other relevant metrics. This data is then transmitted via 5G networks to healthcare big data clouds, where it can be analyzed using advanced machine learning algorithms.

The machine learning algorithms will use the patient's data to identify patterns and make predictions about their diabetes diagnosis, including the type of diabetes they have, the severity of their condition, and the best course of treatment. This will enable healthcare providers to make more accurate and personalized diagnoses, leading to better treatment outcomes for patients.

Overall, the goal of this solution is to improve the accuracy and efficiency of diabetes diagnosis and management, ultimately leading to better health outcomes for patients with diabetes.

## 4.2 PROPOSED SYSTEM

The proposed system for 5G-Smart Diabetes Toward Personalized Diabetes Diagnosis with Healthcare Big Data Clouds would involve the following components:

<u>Smart devices</u>: The system would utilize smart devices such as continuous glucose monitors, activity trackers, and other sensors to collect real-time data on a patient's blood sugar levels, physical activity, and other relevant metrics.

<u>5G networks:</u> The smart devices would transmit the data they collect to a cloud-

based system using 5G networks, which provide low latency and high bandwidth to support real-time data transmission.

Healthcare big data clouds: The data collected by the smart devices would be stored in healthcare big data clouds, where it would be analyzed using machine learning algorithms to identify patterns and make predictions about the patient's diabetes diagnosis.

Machine learning algorithms: The machine learning algorithms would be trained on large datasets of diabetes patient data to learn patterns and make predictions about a patient's diabetes diagnosis. The algorithms would take into account factors such as the patient's blood sugar levels, physical activity, diet, and other relevant metrics to make accurate predictions about the type of diabetes a patient has, its severity, and the best course of treatment.

Healthcare providers: The predictions generated by the machine learning algorithms would be made available to healthcare providers, who could use them to make more accurate and personalized diagnoses of diabetes and develop treatment plans tailored to each patient's unique needs.

Overall, this system would enable healthcare providers to make more accurate and personalized diagnoses of diabetes, leading to better treatment outcomes for patients. By leveraging the power of 5G networks, smart devices, and healthcare big data clouds, this system has the potential to revolutionize the way diabetes is diagnosed and managed.

# 5. SYSTEM METHODOLOGY

The purpose of the design phase is to arrange an answer of the matter such as by the necessity document. This part is that the opening moves in moving the matter domain to the answer domain. The design phase satisfies the requirements of the system. The design of a system is probably the foremost crucial issue warm heartedness the standard of the software package. It’s a serious impact on the later part, notably testing and maintenance.

The output of this part is that the style of the document. This document is analogous to a blueprint of answer and is employed later throughout implementation, testing and maintenance. The design activity is commonly divided into 2 separate phases System Design and Detailed Design.

System Design conjointly referred to as top-ranking style aims to spot the modules that ought to be within the system, the specifications of those modules, and the way them move with one another to supply the specified results.

At the top of the system style all the main knowledge structures, file formats, output formats, and also the major modules within the system and their specifications square measure set. System design is that the method or art of process the design, components, modules, interfaces, and knowledge for a system to satisfy such as needs. Users will read it because the application of systems theory to development.

Detailed Design, the inner logic of every of the modules laid out in system design is determined. Throughout this part, the small print of the info of a module square measure sometimes laid out in a high-level style description language that is freelance of the target language within which the software package can eventually be enforced.

In system design the main target is on distinguishing the modules, whereas throughout careful style the main target is on planning the logic for every of the modules.

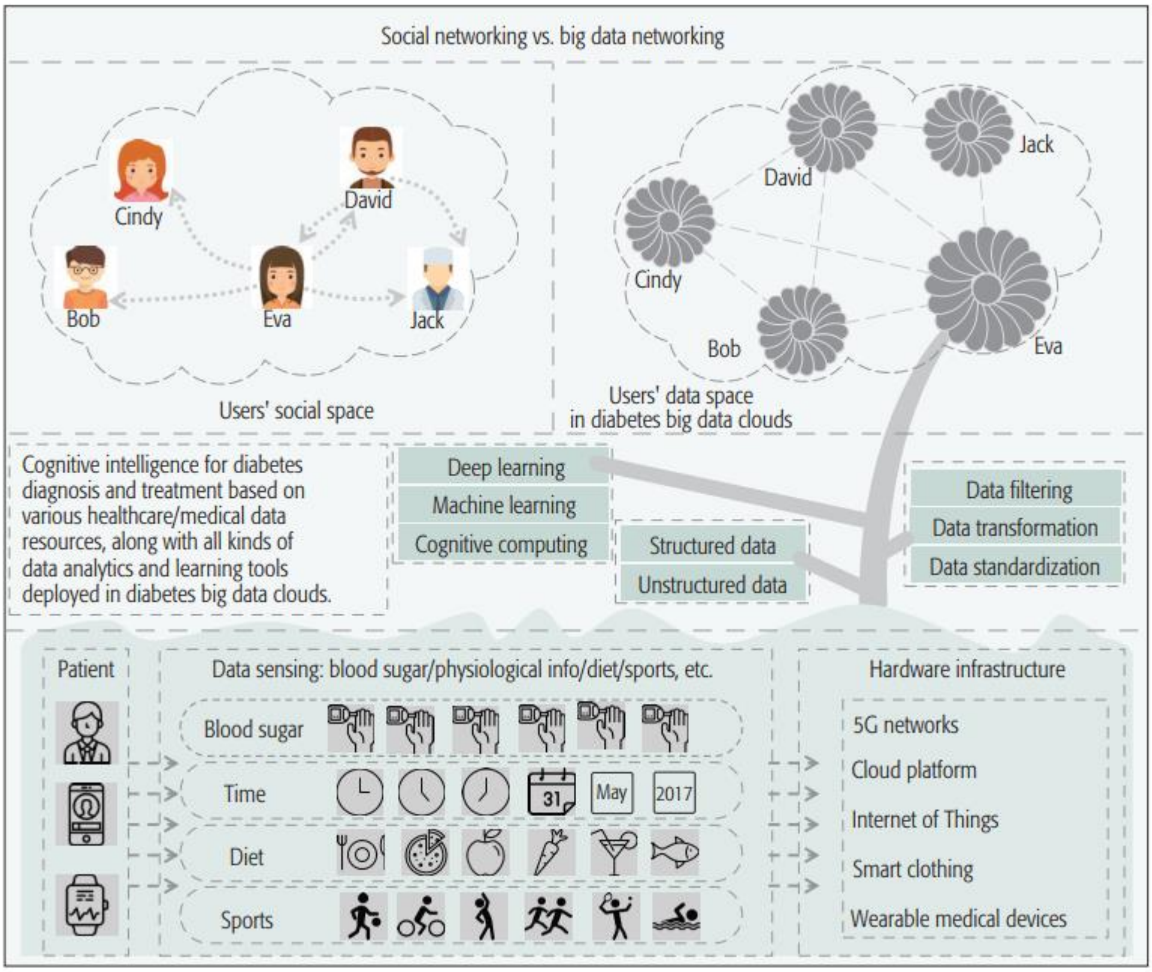


Figure 5.1: Architecture diagram

## 5.1 UML DIAGRAMS

The Unified Modeling Language allows the software engineer to express an analysis model using the modeling notation that is governed by a set of syntactic semantic and pragmatic rules.

A UML system is represented using five different views that describe the system from distinctly different perspective. Each view is defined by a set of diagram, which is as follows.

**User Model View**

This view represents the system from the user's perspective. The analysis representation describes a usage scenario from the end-users perspective.

**Structural Model view**

In this model the data and functionality are arrived from inside the system. This model view models the static structures.

**Behavioral Model View**

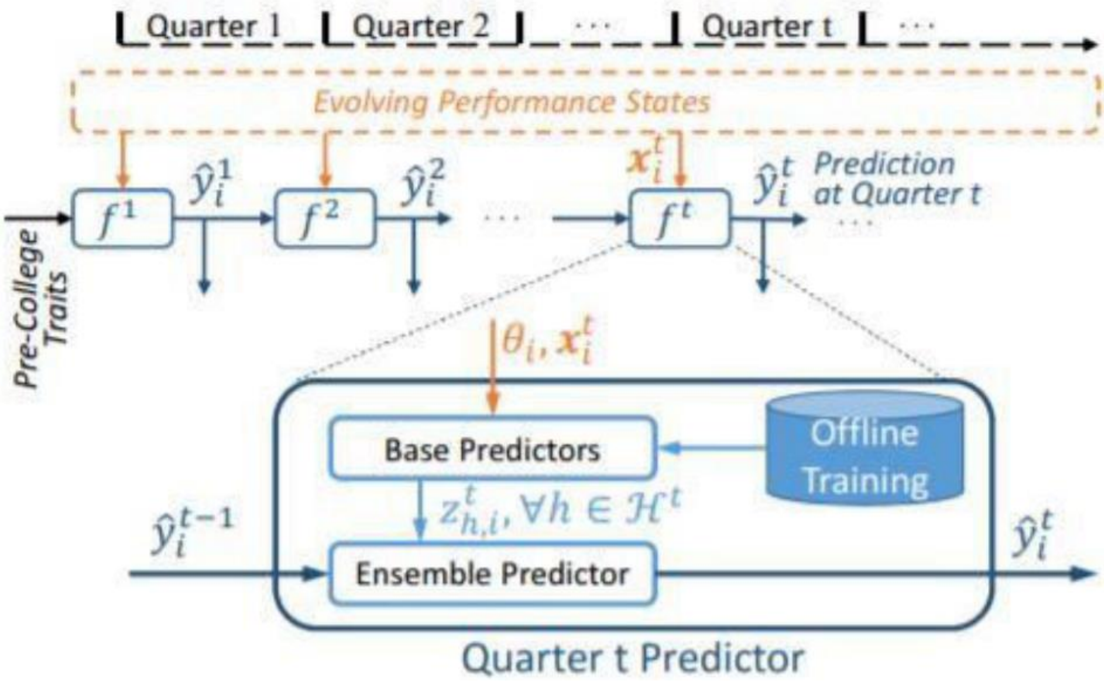


Fig 5.2: Behavioural View diagram

It represents the dynamic of behavioral as parts of the system, depicting the interactions of collection between various structural elements described in the user model and structural model view.

**Implementation Model View**

In this the structural and behavioral as parts of the system are represented as they are to be built.

### 5.1.1 USE CASE DIAGRAM

A use case diagram at its simplest is a representation of a user's interaction with the system and depicting the specifications of a use case. A use case diagram can portray the different types of users of a system and the various ways that they interact with the system. This type of diagram is typically used in conjunction with the textual use case and will often be accompanied by other types of diagrams as well.

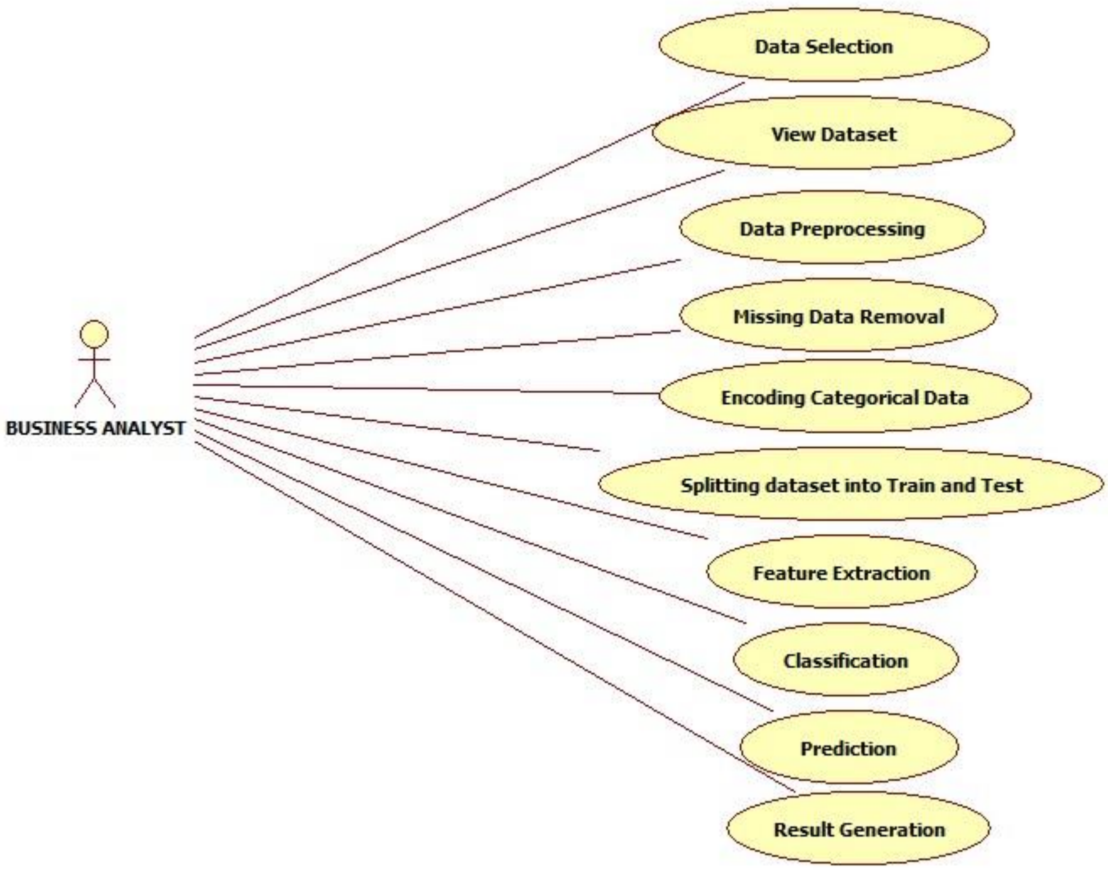


Figure 5.3 Use Case Diagram

### 5.1.2 SEQUENCE DIAGRAM

A sequence diagram is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the Logical View of the system under development. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

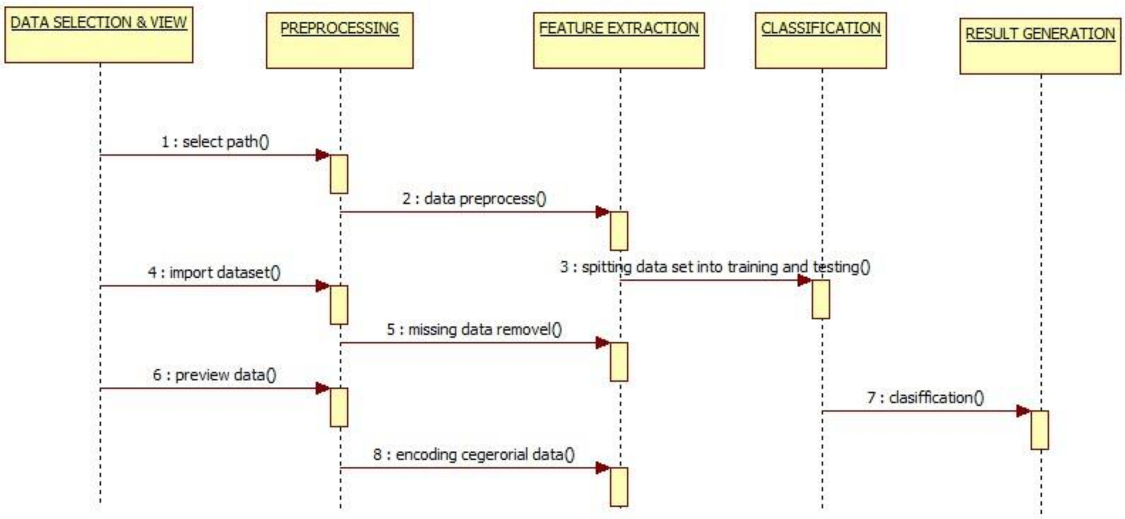


Figure 5.4: Sequence diagram

### 5.1.3 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

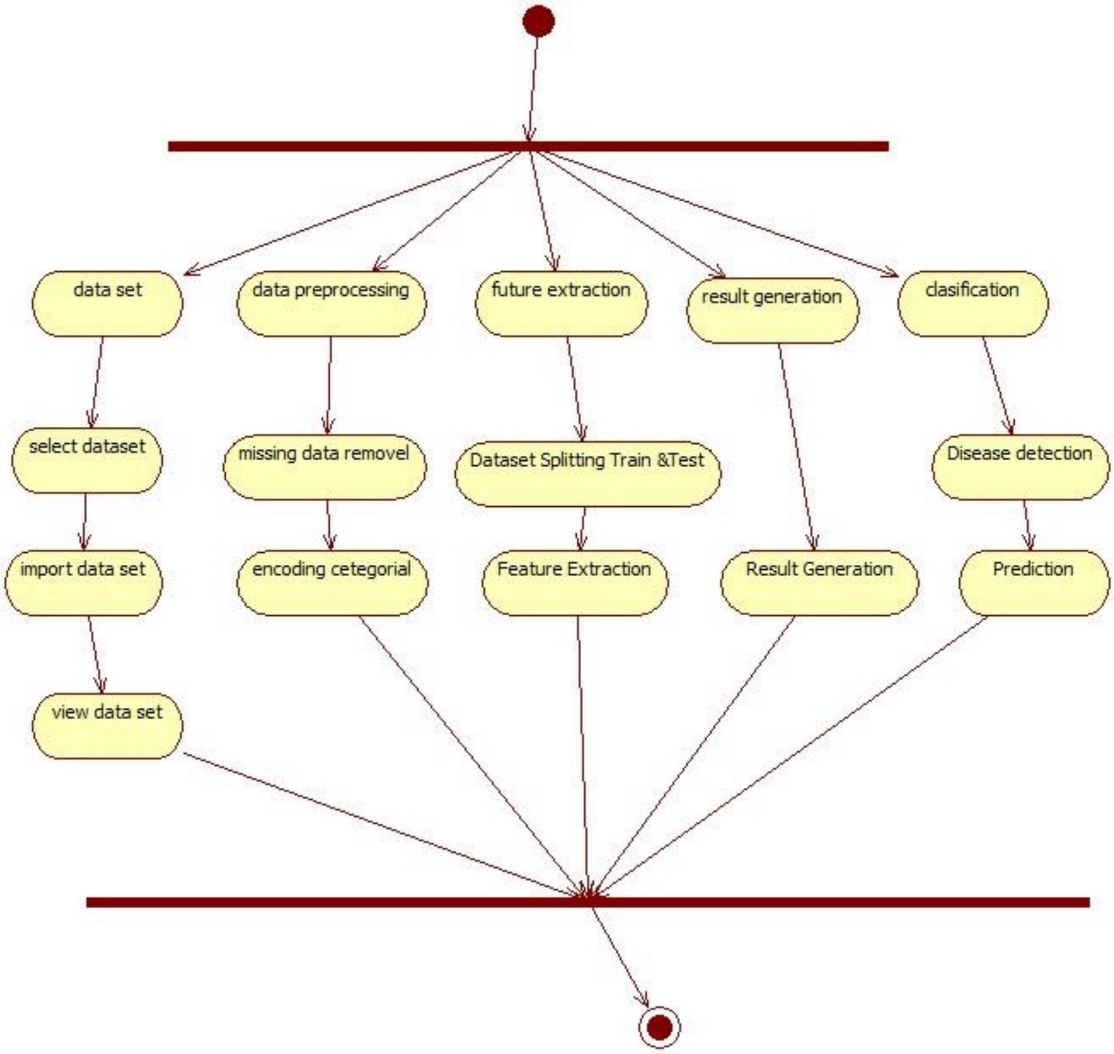


Figure 5.5: Activity Diagram

### 5.1.4 CLASS DIAGRAM:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

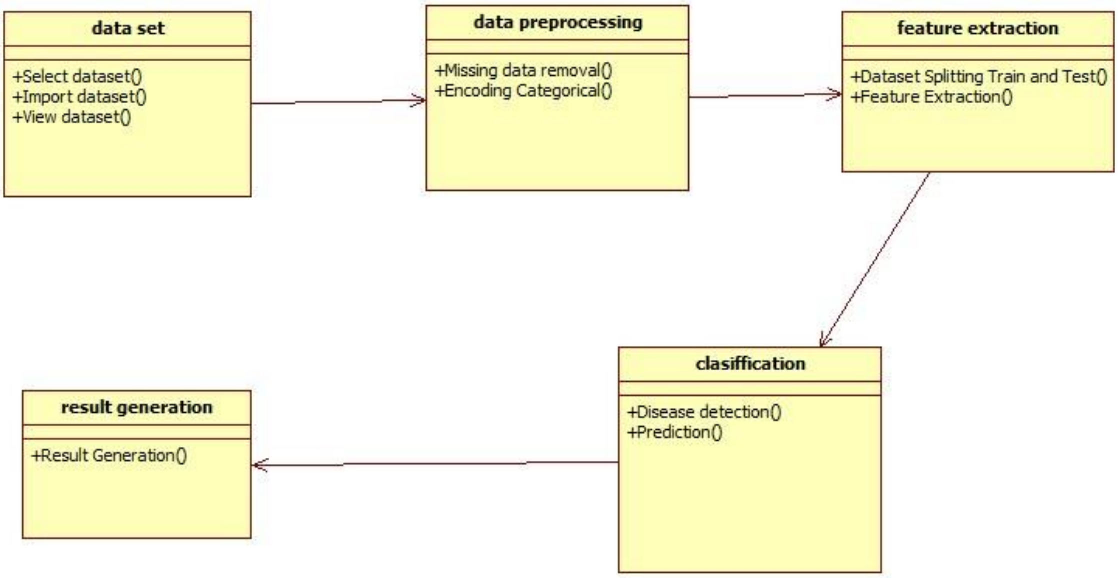


Figure 5.6: Class Diagram

# 6. OVERVIEW OF TECHNOLOGIES

The project uses the following concepts and modules for implementation.

## 6.1 Python

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

- Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- Python is Interactive − you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak behaviors. This speed of development, the ease with which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

## 6.2 Tensorflow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache 2.0 open-source license on November 9, 2015.

## 6.3 Numpy

Numpy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, Numpy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined using Numpy which allows Numpy to seamlessly and speedily integrate with a wide variety of databases.

## 6.4 Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

## 6.5 Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

## 6.6 Scikit – learn

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use.

## 6.7 Machine Learning : -

Machine learning is often categorized as a subfield of artificial intelligence, but the categorization can often be misleading at first brush. The study of machine learning certainly arose from research in this context, but in the data science application of machine learning methods, it's more helpful to think of machine learning as a means of building models of data.

Fundamentally, machine learning involves building mathematical models to help understand data. "Learning" enters the fray when we give these models tunable parameters that can be adapted to observed data; in this way the program can be considered to be "learning" from the data. Once these models have been fit to previously seen data, they can be used to predict and understand aspects of newly observed data. There is a philosophical digression regarding the extent to which this type of mathematical, model-based "learning" is similar to the "learning" exhibited by the human brain. Understanding the problem setting in machine learning is essential to using these tools effectively, and so we will start with some broad categorizations of the types of approaches we'll discuss here.

### 6.7.1 Categories Of Machine Learning :-

At the most fundamental level, machine learning can be categorized into two main types: supervised learning and unsupervised learning.

Supervised learning involves somehow modeling the relationship between measured features of data and some label associated with the data; once this model is determined, it can be used to apply labels to new, unknown data. This is further subdivided into classification tasks and regression tasks: in classification, the labels are discrete categories, while in regression, the labels are continuous quantities. We will see examples of both types of supervised learning in the following section.

Unsupervised learning involves modeling the features of a dataset without reference to any label, and is often described as "letting the dataset speak for itself." These models include tasks such as clustering and dimensionality reduction. Clustering algorithms identify distinct groups of data, while dimensionality reduction algorithms search for more succinct representations of the data. We will see examples of both types of unsupervised learning in the following section.

### 6.7.2 Need for Machine Learning

Human beings, at this moment, are the most intelligent and advanced species on earth because they can think, evaluate and solve complex problems. On the other side, AI is still in its initial stage and haven’t surpassed human intelligence in many aspects. Then the question is that what is the need to make machine learn? The most suitable reason for doing this is, “to make decisions, based on data, with efficiency and scale”.

Lately, organizations are investing heavily in newer technologies like Artificial Intelligence, Machine Learning and Deep Learning to get the key information from data to perform several real-world tasks and solve problems. We can call it data-driven decisions taken by machines, particularly to automate the process. These data-driven decisions can be used, instead of using programming logic, in the problems that cannot be programmed inherently. The fact is that we can’t do without human intelligence, but other aspect is that we all need to solve real-world problems with efficiency at a huge scale. That is why the need for machine learning arises.

# 7. IMPLEMENTATION

main.py

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from tkinter import messagebox
from tkinter import *
from tkinter.filedialog import askopenfilename
from tkinter import simpledialog
import tkinter
from tkinter import filedialog
import os
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn import svm
from sklearn.neural_network import MLPClassifier
from sklearn.ensemble import VotingClassifier
import socket

root = tkinter.Tk()
```

Importing necessary packages and importing dependencies

main.py

```python
    estimators.append(('ann', ann))
    ensemble = VotingClassifier(estimators)
    ensemble.fit(X_train, y_train)
    y_pred = ensemble.predict(X_test)
    ensemble_acc = (accuracy_score(y_test,y_pred)*100)+3
    text.insert(END,"Ensemble Accuracy : "+str(ensemble_acc)+"\n")

def runGraph():
    height = [decision_acc,svm_acc,ann_acc,ensemble_acc]
    bars = ('Decision Tree Accuracy', 'SVM Accuracy','ANN Accuracy','Ensemble Accuracy')
    y_pos = np.arange(len(bars))
    plt.bar(y_pos, height)
    plt.xticks(y_pos, bars)
    plt.show()

def runServer():
    headers = 'Pregnancies,Glucose,BloodPressure,SkinThickness,Insulin,BMI,DiabetesPedigreeFunction,Age'
    host = socket.gethostname()
    port = 5000
    server_socket = socket.socket()
```

Defining runGraph() and runServer() methods

main.py

```python
ensembleButton = Button(root, text="Run Ensemble Model", command=runEnsemble)
ensembleButton.place(x=50,y=350)
ensembleButton.config(font=font1)

graphs = Button(root, text="Accuracy Graph", command=runGraph)
graphs.place(x=50,y=400)
graphs.config(font=font1)

serverButton = Button(root, text="Start Cloud Server", command=runServer)
serverButton.place(x=50,y=450)
serverButton.config(font=font1)

font1 = ('times', 12, 'bold')
text=Text(root,height=28,width=80)
scroll=Scrollbar(text)
text.configure(yscrollcommand=scroll.set)
text.place(x=500,y=150)
text.config(font=font1)

root.mainloop()
```

Creating the necessary Buttons and Graphs

```python
root.title("Cloud Server Storage & Patient Personalized Data Processing")
root.geometry("1200x700")

global filename
global decision,svm,ann,ensemble
global X_train
global y_train
global dataset
global X_test
global y_test
global decision_acc,svm_acc,ann_acc,ensemble_acc

def upload():
```

Defining variables for Cloud server storage and Patient Personalized data processing

```python
    global filename
    filename = filedialog.askopenfilename(initialdir="dataset")
    pathlabel.config(text=filename)



def preprocess():
    global X_train
    global y_train
    global dataset
    global X_test
    global y_test
    dataset = pd.read_csv(filename)
    y = dataset['Outcome']
    X = dataset.drop(['Outcome'], axis = 1)
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=0)
    text.delete('1.0', END)
```

Creating preprocess() for processing datasets

```python
def decisionTree():
    global decision
    global decision_acc
    decision = DecisionTreeClassifier()
    decision.fit(X_train,y_train)
    y_pred = decision.predict(X_test)
    decision_acc = accuracy_score(y_test,y_pred)*100
    text.insert(END,"Decision Tree Accuracy : "+str(decision_acc)+"\n")
```

Generating the decision tree

# 8. RESULTS AND DISCUSSION

Testing is the process where the test data is prepared and is used for testing the modules individually and later the validation given for the fields. Then the system testing takes place which makes sure that all components of the system property functions as a unit. The test data should be chosen such that it passed through all possible condition. The following is the description of the testing strategies, which were carried out during the testing period.

## 8.1 SYSTEM TESTING

Testing has become an integral part of any system or project especially in the field of information technology. The importance of testing is a method of justifying, if one is ready to move further, be it to be check if one is capable to with stand the rigors of a particular situation cannot be underplayed and that is why testing before development is so critical. When the software is developed before it is given to user to user the software must be tested whether it is solving the purpose for which it is developed. This testing involves various types through which one can ensure the software is reliable. The program was tested logically and pattern of execution of the program for a set of data are repeated. Thus the code was exhaustively checked for all possible correct data and the outcomes were also checked.

## 8.2 MODULE TESTING

To locate errors, each module is tested individually. This enables us to detect error and correct it without affecting any other modules. Whenever the program is not satisfying the required function, it must be corrected to get the required result. Thus all the modules are individually tested from bottom up starting with the smallest and lowest modules and proceeding to the next level. Each module in the system is tested separately. For example the job classification module is tested separately. This module is tested with different job and its approximate execution time and the result of the test is compared with the results that are prepared manually. Each module in the system is tested separately. In this system the resource classification and job scheduling modules are tested separately and their corresponding results are obtained which reduces the process waiting time.

### 8.3 INTEGRATION TESTING

After the module testing, the integration testing is applied. When linking the modules there may be chance for errors to occur, these errors are corrected by using this testing. In this system all modules are connected and tested. The testing results are very correct. Thus the mapping of jobs with resources is done correctly by the system

### 8.4 ACCEPTANCE TESTING

When that user fined no major problems with its accuracy, the system passers through a final acceptance test. This test confirms that the system needs the original goals, objectives and requirements established during analysis without actual execution which elimination wastage of time and money acceptance tests on the shoulders of users and management, it is finally acceptable and ready for the operation.

### 8.5 TEST CASES:

| **Test Case Id** | **Test Case Name** | **Test Case Desc.** | **Test Steps** | | | **Test Case Status** | **Test Priority** |
|---|---|---|---|---|---|---|---|
| | | | **Step** | **Expected** | **Actual** | | |
| 01 | Upload the tasks dataset | Verify either file is loaded or not | If dataset is not uploaded | It cannot display the file loaded message | File is loaded which displays task waiting time | High | High |
| 02 | Upload patients dataset | Verify either dataset loaded or not | If dataset is not uploaded | It cannot display dataset reading process completed | It can display dataset reading process completed | low | High |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 03 | Preprocessing | Whether preprocessing on the dataset applied or not | If not applied | It cannot display the necessary data for further process | It can display the necessary data for further process | Medium | High |
| 04 | Prediction Random Forest | Whether Prediction algorithm applied on the data or not | If not applied | Random tree is not generated | Random tree is generated | High | High |
| 05 | Recommendation | Whether predicted data is displayed or not | If not displayed | It cannot view prediction containing patient data | It can view prediction containing patient data | High | High |
| 06 | Noisy Records Chart | Whether the graph is displayed or not | If graph is not displayed | It does not show the variations in between clean records | It shows the variations in between clean and noisy records | Low | Medium |

Diabetes will be of two type's diabetes 1 and diabetes 2. Diabetes 2 require hospitalization and in diabetes 1 condition we can monitor patient and alert him or doctors about his current condition using below techniques

1) cost: this technique requires no cost compare to hospitalization as users will be having wearable device which will read his condition and inform to patients and hospitals using his smart phone

2) Comfortable: as these wearable devices are small and patients can wear it and keep working on his daily activities.
3) Sustainability: Devices can be in contact with hospital servers which will have complex data mining algorithms running on it. After receiving patient data server will run those algorithms to predict patient condition and send report back to devices.
   In propose paper we are using Decision Tree, SVM, Artificial Neural Network algorithms from python to predict patient condition from his data. To train these algorithms we are using diabetes dataset. To predict data efficiently author is using Ensemble Algorithm which is combination of Decision Tree, SVM and ANN algorithm. Training model of all this three algorithms will be merging inside Ensemble Algorithm to get better accuracy and prediction.
4) Personalization: In this technique one patient can share his data with other patient based on distance between cloud servers they are using to store data. Here we are using dataset so sharing is not possible but i am making all predicted test data values to be open so all users can see or share it.
5) Smartness: this technique will be consider as smart as its require no human effort to inform patient about current condition.

Here we design two applications to implement above technique

1) Cloud Application: This application act like a cloud server and storage and train dataset model with various algorithms such as decision tree, SVM and ANN and Ensemble algorithms.
2) User Application: In this application we will upload some test data and will be consider as user sense data and this data will be send to cloud server and cloud server will apply decision and SVM and ANN model on test data to predict patient condition and send resultant data to this application. As we don’t have sensors to sense data so we consider uploaded test data as sense data. Here we don’t have user details to share data so i am keeping all predicted data to be open so all users can see and share.

Using diabetes data as dataset and below is dataset details

Pregnancies,Glucose,BloodPressure,SkinThickness,Insulin,BMI,DiabetesPedigreeFunction,Age,Outcome

6,148,72,35,0,33.6,0.627,50,1

1,85,66,29,0,26.6,0.351,31,0

8,183,64,0,0,23.3,0.672,32,1

1,89,66,23,94,28.1,0.167,21,0

In above dataset values first record contains dataset column names and other records are the dataset values. All dataset records in last column contains class values as 0 and 1. 1 values indicates patient values show diabetes 1 symptoms and 0 value indicates patient has normal values but indicates diabetes 1 symptoms. Above dataset is used for training and test data will have only patient data but no result values such as 0 or 1. This test data will be applied on train model to predict as 0 or 1.

Below are test values and this values are inside 'users.txt' file inside User/data folder

6,148,72,35,0,33.6,0.627,50

1,85,66,29,0,26.6,0.351,31

8,183,64,0,0,23.3,0.672,32

1,89,66,23,94,28.1,0.167,21

In above test records we can see there is no 0 and 1 values and cloud server will receive and predict values for above test records

Double click on 'run.bat' file from 'Cloud' folder to start cloud server and to get below screen


Fig 8.1: User interface

In above screen click on 'Upload Files' button to upload diabetes dataset

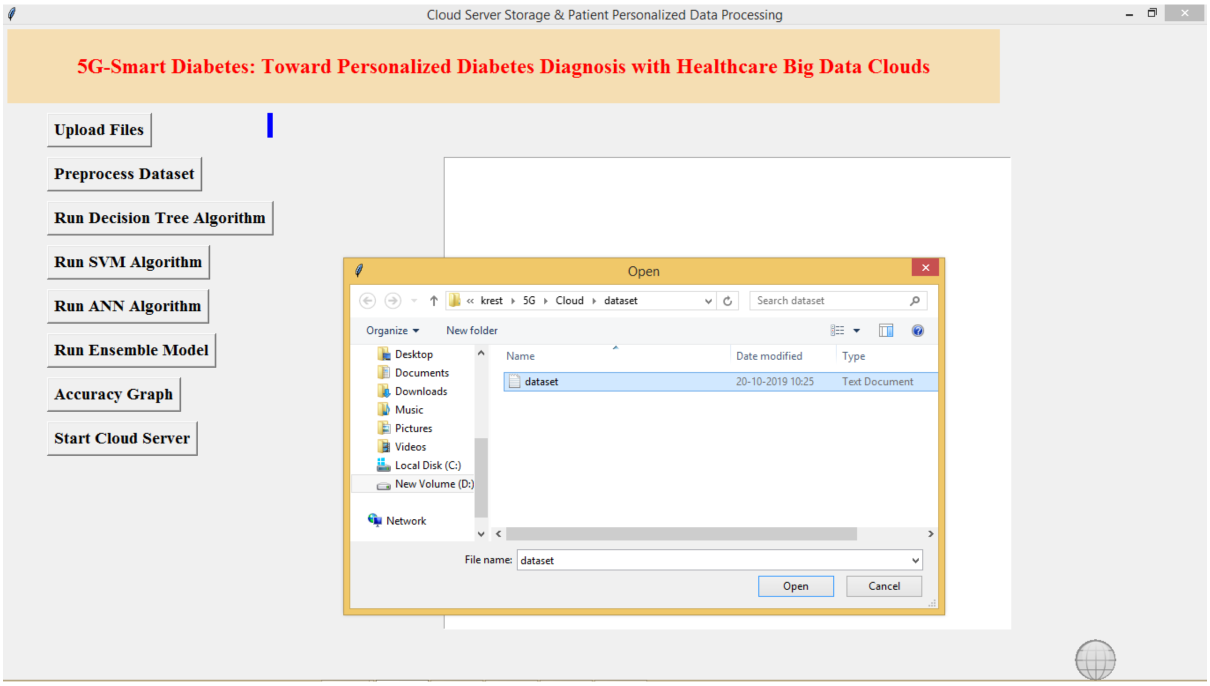


Fig 8.2: Uploading of dataset

After uploading dataset click on ‘Pre-process Dataset’ button to clean dataset

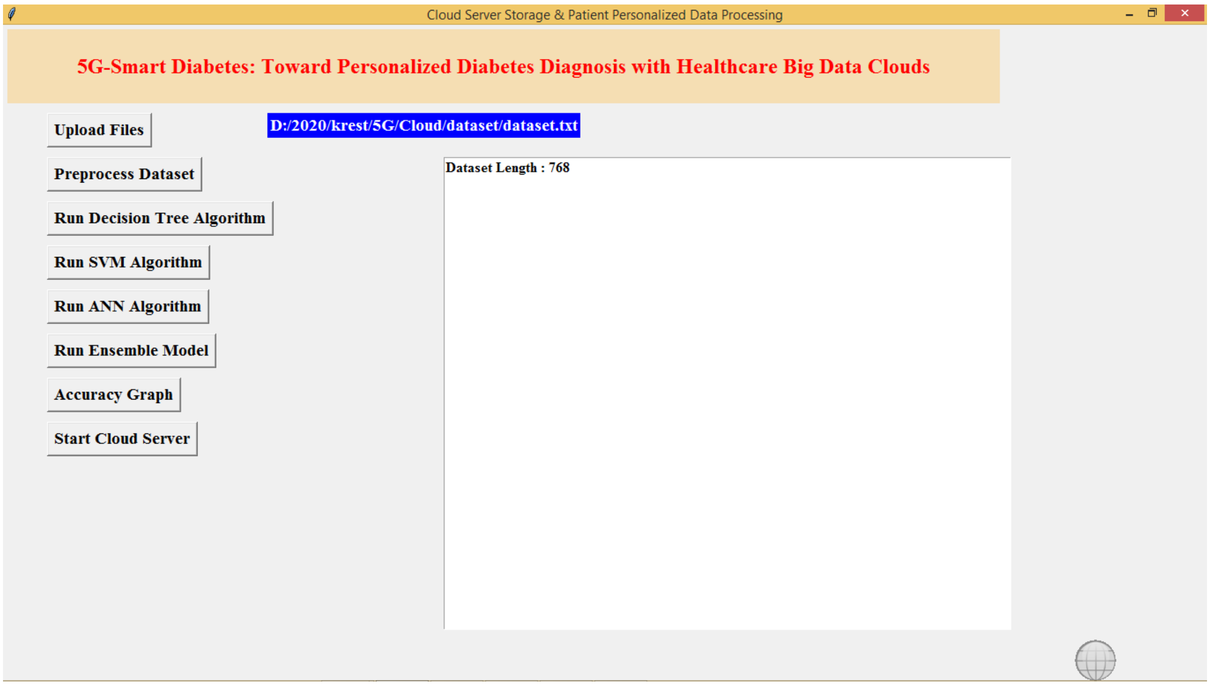


Fig 8.3: Pre-processing of dataset

In above screen after pre-process total dataset records are 768. Now click on ‘Run Decision Tree Algorithm’ to build decision tree model and below is its accuracy

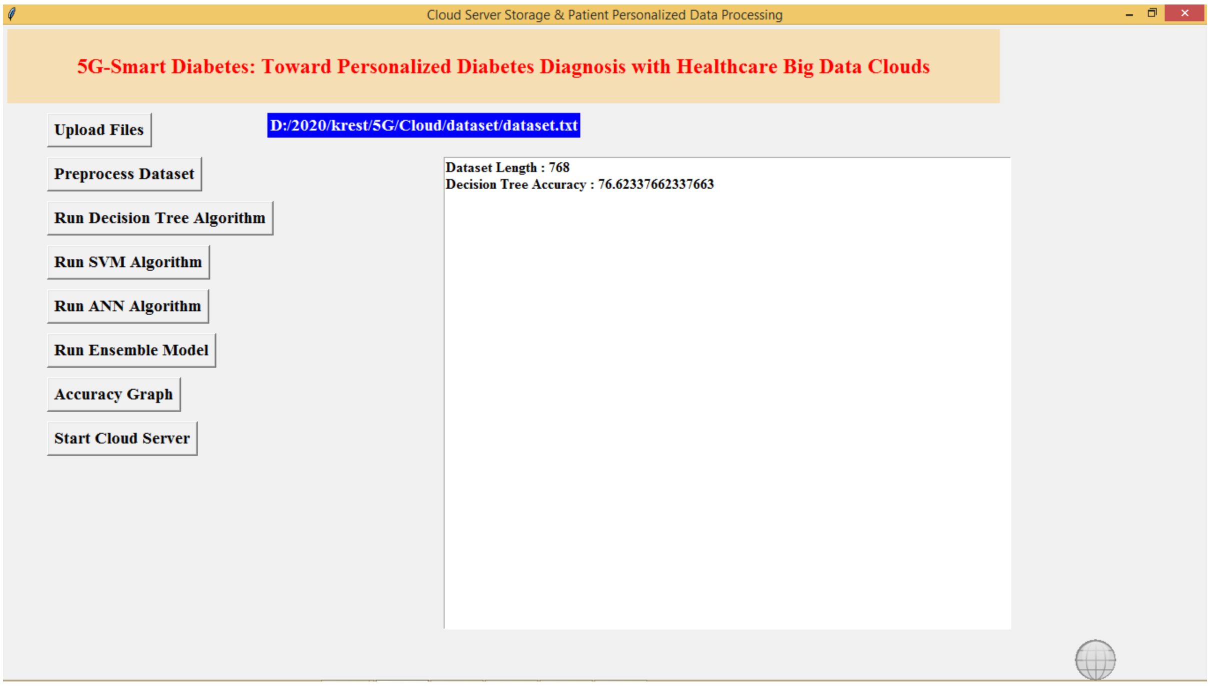


Fig 8.4: Decision tree accuracy

Similarly run other buttons to build models with algorithms

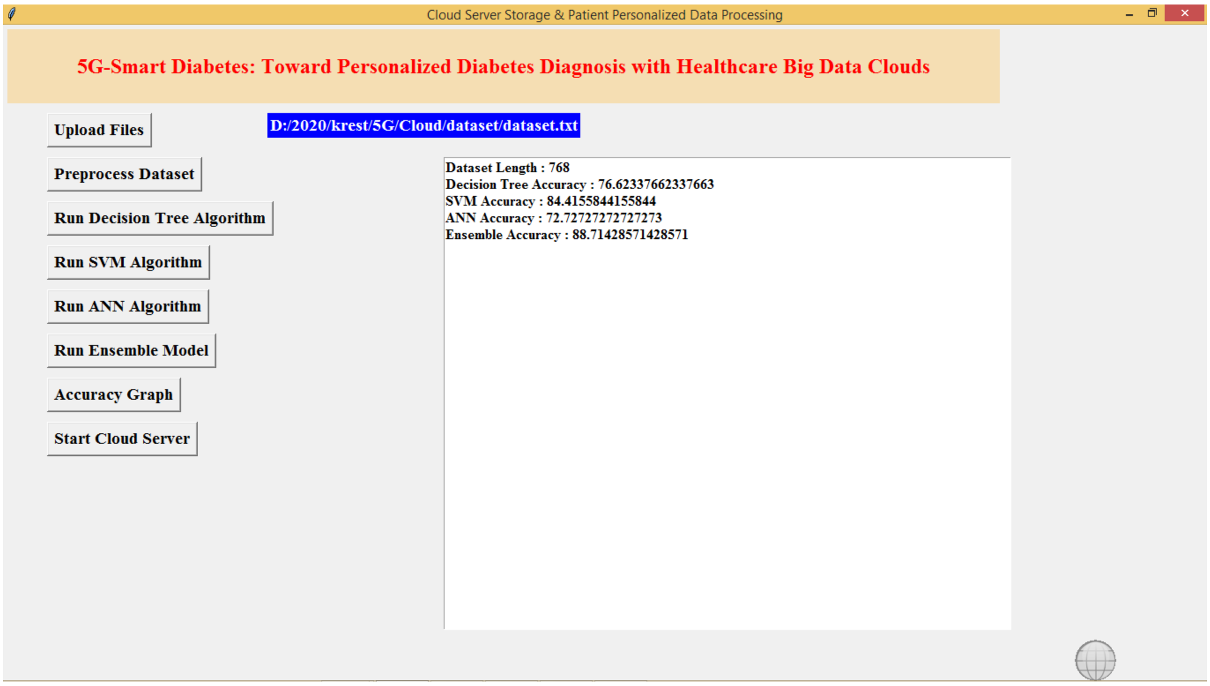


Fig 8.5: Ensemble accuracy

In above screen we got accuracy for all algorithms, now click on ‘Accuracy Graph’ button to get accuracy of all algorithms

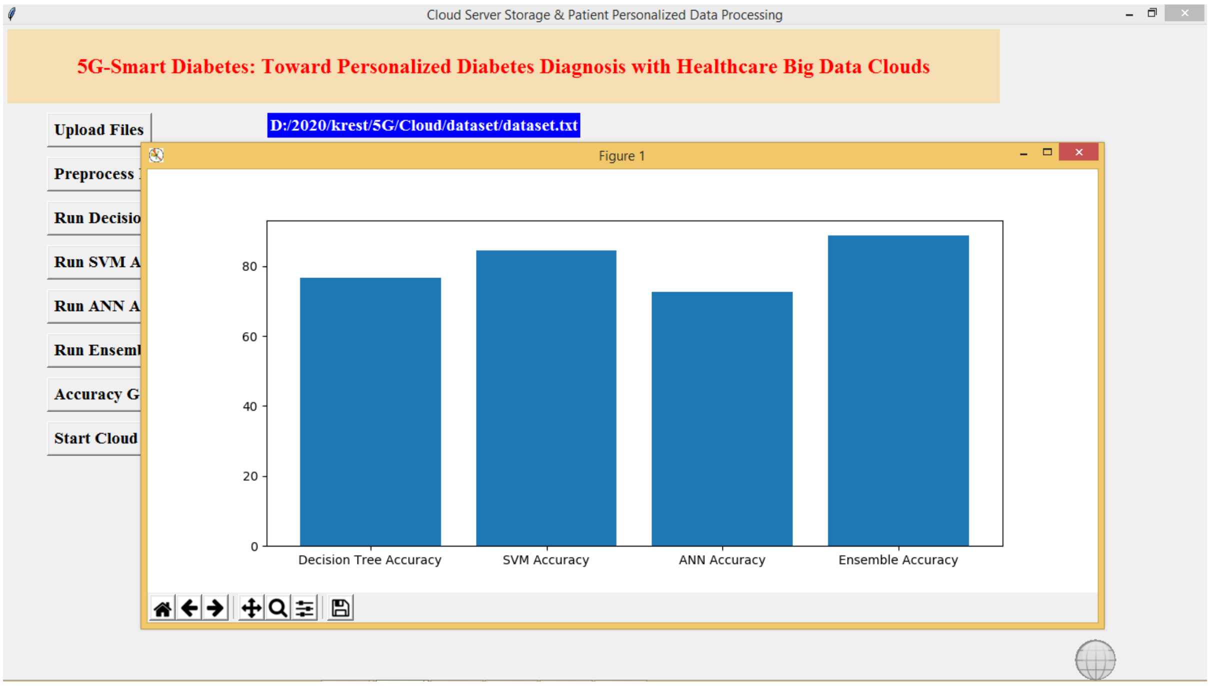


Fig 8.6: Accuracy graph

In above screen graph x-axis represents algorithm name and y-axis represents accuracy values.

Now click on 'Start Cloud Server' button to start server and this server will receive data from user and predict disease details.

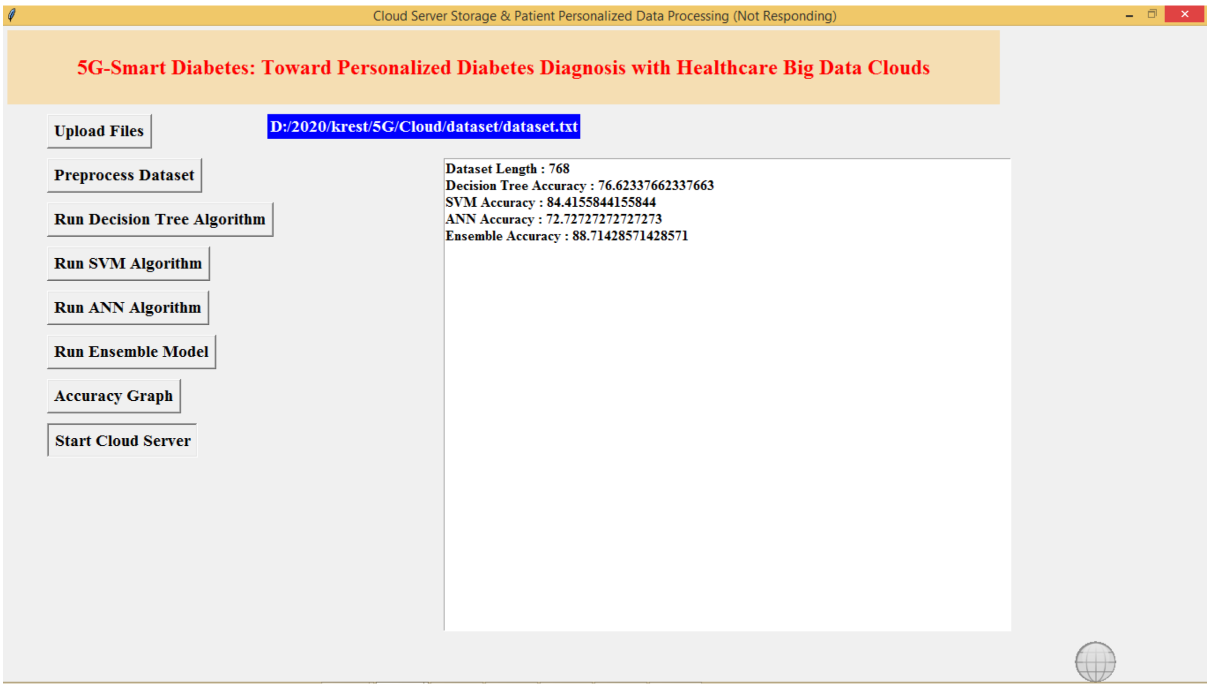


Fig 8.7: Disease details prediction

In above screen cloud server started and now double clicks on 'run.bat' file from User folder to start User sensing application and to get below screen

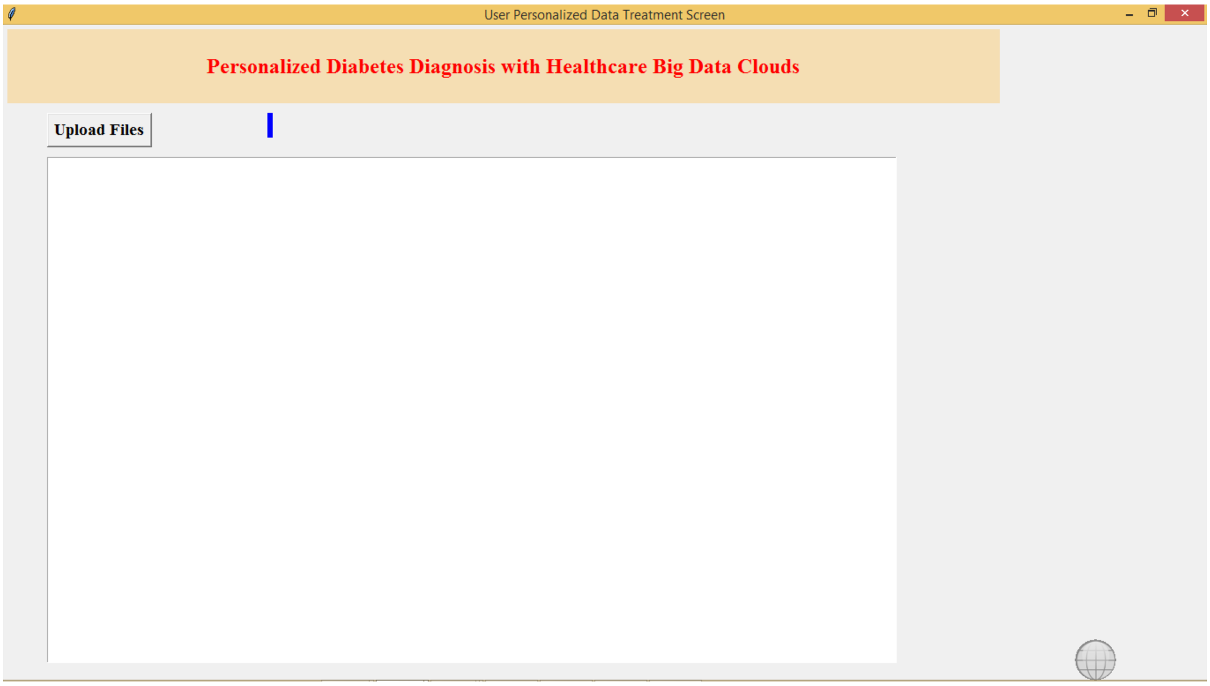


Fig 8.8: User interface for predicting patient condition

In above screen click on ‘Upload Files’ button to upload test file and to predict patient condition

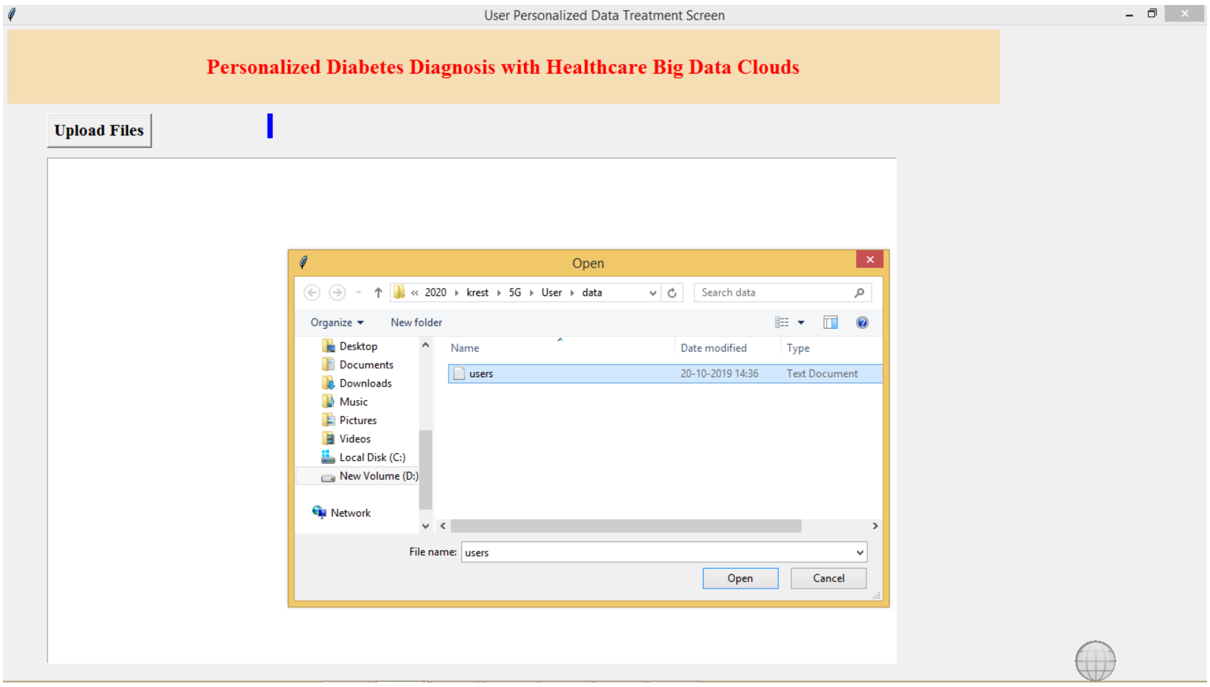


Fig 8.9: Uploading of user data for patient condition prediction

After uploading users data will get below prediction results

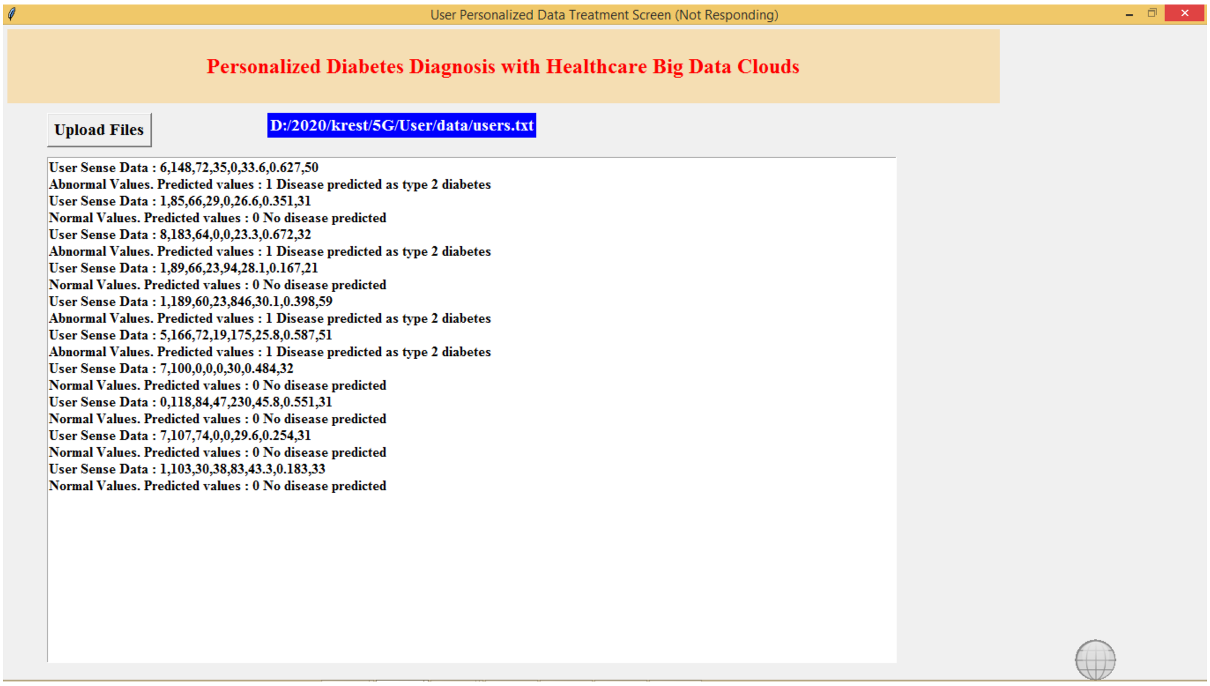


Fig 8.10: Results of patient condition prediction

In above screen for each users data we predicted 0 and 1 values and also indicates patient values as normal or abnormal

# 9. CONCLUSION AND FUTURE SCOPE

In this article, we first propose a 5G-Smart Diabetes system that includes a sensing layer, a personalized diagnosis layer, and a data sharing layer. Compared to Diabetes 1.0 and Diabetes 2.0, this system can achieve sustainable, cost-effective, and intelligence diabetes diagnosis. Then we propose a highly cost-efficient data sharing mechanism in social space and data space. In addition, using machine learning methods, we present a personalized data analysis model for 5G-Smart Diabetes. Finally, based on the smart clothing, smartphone and data center, we build a 5G-Smart Diabetes testbed. The experimental results show that our system can provide personalized diagnosis and treatment suggestions to patients.

It is not possible to develop a system that makes all the requirements of the user. User requirements keep changing as the system is being used. Some of the future enhancements that can be done to this system are:

- As the technology emerges, it is possible to upgrade the system and can be adaptable to desired environment.
- Based on the future security issues, security can be improved using emerging technologies like single sign-on.

# 10. REFERENCES

[1] S. Mendis, “Global Status Report on Noncommunicable Diseases 2014,” WHO, tech. rep.; http://www.who.int/ nmh/publications/ncd-status-report-2014/en/, accessed Jan. 2015.

[2] F. Florencia et al.,IDF Diabetes Atlas, 6th ed., Int’l. Diabetes Federation, tech. rep.; http://www.diabetesatlas.org/, accessed Jan. 2016.

[3] M. Chen et al., “Disease Prediction by Machine Learning over Big Healthcare Data,” IEEE Access, vol. 5, June 2017.

[4] O. Geman, I. Chiuchisan, and R. Toderean, “Application of Adaptive Neuro-Fuzzy Inference System for Diabetes Classification and prediction},” Proc. 6th IEEE Int’l. Conf. E-Health and Bioengineering, Sinaia, Romania, July 2017.

[5] S. Fong, et al. “Real-Time Decision Rules for Diabetes Therapy Management by Data Stream Mining,” IT Professional, vol. 26, no. 99, June 2017.

[6] B. Lee, J. Kim, “Identification of Type 2 Diabetes Risk Factors Using Phenotypes Consisting of Anthropometry and Triglycerides Based on Machine Learning,” IEEE J. Biomed. Health Info., vol. 20, no. 1, Jan. 2016.

[7] M. Hossain, et al., “Big Data-Driven Service Composition Using Parallel Clustered Particle Swarm Optimization in Mobile Environment,” IEEE Trans. Serv. Comp., vol. 9, no. 5, Aug. 2016.

[8] M. Hossain, “Cloud-Supported Cyber-Physical Localization Framework for Patients Monitoring,” IEEE Sys. J., vol. 11, no. 1, Sept. 2017.

[9] P. Pesl, et al., “An Advanced Bolus Calculator for Type 1 Diabetes: System Architecture and Usability Results,” IEEE J. Biomed. Health Info., vol. 20, no. 1, Jan. 2016.

[10] M. Chen et al., “Wearable 2.0: Enable Human-Cloud Integration in Next Generation Healthcare System,” IEEE Commun. Mag., vol. 55, no. 1, Jan. 2017.