A Supervised Learning Based Prediction of Type 2 Diabetes

Ramavath Vinod Kumar, Ramavath Vinod Kumar

Department of Computer Science and Engineering

Sree Dattha Group of Institutions, Hyderabad, Telangana, India.

ABSTRACT

This project uses today's 5G technology to monitor condition of diabetic patients with low cost. Now-a-days many peoples are suffering with diabetic disease due to work stress or unhealthy lifestyles and peoples will not know about the current health condition till symptoms appear or diagnosis through medical check-up and the condition of disease will be severe by that time and there is no possible way to get that intimation prior. 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.

Keywords: Prediction of diabetes, diabetes-1, diabetes-2, supervised learning.

1. INTRODUCTION

Diabetes is an extremely common chronic disease from which nearly 8.5 percent of the world population suffer; 422 million people worldwide must 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, with more teenagers and youth becoming susceptible to diabetes as well. Because diabetes has a huge impact on global well-being and economy, it is urgent to improve methods for the prevention and treatment of diabetes. 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 multidimensional physiological indicators of patients suffering from diabetes.
- 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 [2].
- There are no continuous suggestions for the prevention and treatment of diabetes and corresponding supervision strategies.

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, 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 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, 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 testbed.

2. LITERATURE SURVEY

Chen et al. proposed the 5G-Smart Diabetes system, which combined 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 this work presented the data sharing mechanism and personalized data analysis model for 5G-Smart Diabetes. Finally, this work builds a 5G-Smart Diabetes testbed that includes smart clothing, smartphone, and big data clouds. The experimental results showed that the system can effectively provide personalized diagnosis and treatment suggestions to patients.

Rghioui et al. presented an intelligent architecture for monitoring diabetic patients by using machine learning algorithms. The architecture elements included smart devices, sensors, and smartphones to collect measurements from the body. The intelligent system collected the data received from the patient and performed data classification using machine learning to make a diagnosis. The proposed prediction system was evaluated by several machine learning algorithms, and the simulation results demonstrated that the sequential minimal optimization (SMO) algorithm gives superior classification accuracy, sensitivity, and precision compared to other algorithms.

Venkatachalam et al. motivated to develop a diabetes motoring system for patients using IoT device in their body which monitors their blood sugar level, blood pressure, sport activities, diet plan, oxygen level, ECG data. The data are processed using feature selection algorithm called as particle swarm optimization and transmitted to nearest edge node for processing in 5G networks. Secondly, data are

processed using DBN Layer. Thirdly, this work shared the diagnosed data output through the wireless communication such as LTE/5G to the patients connected through the edge nodes for further medical assistance. The patient wearable devices are connected to the social network. The Result of this proposed system is evaluated with some existing system. Time and Performance outperform than other techniques.

Prakash et al. introduced a neural network-based ensemble voting classifier to predict accurately the diabetes in the patients via online monitoring. The study consists of Internet of Things (IoT) devices to monitor the instances of the patients. While monitoring, the data are transferred from IoT devices to smartphones and then to the cloud, where the process of classification takes place. The simulation is conducted on the collected samples using the python tool. The results of the simulation show that the proposed method achieves a higher accuracy rate, higher precision, recall, and f-measure than existing state-of-art ensemble models.

Tsoulchas et al. proposed a model to monitor the health of people with diabetes melitus, a disease with high incident rates mainly at the elderly but also in younger people. Specifically, a study about the existing medically approved technologies for continuous measurement of diabetes is described. Subsequently, the model for monitoring patient's blood glucose levels is described. Whenever a patient's blood glucose levels are Low or High, the model triggers an alarm to a Cloud infrastructure in order remote medical staff to provide immediate cure to the patient. Furthermore, to assure the immediate response of the remote medical staff, the proposed model is deployed upon a 5G wireless network architecture.

Huang et al. proposed a 5G-based Artificial Intelligence Diabetes Management architecture (AIDM), which can help physicians and patients to manage both acute complications and chronic complications. The AIDM contains five layers: the sensing layer, the transmission layer, the storage layer, the computing layer, and the application layer. We build a test bed for the transmission and application layers. Specifically, this work applied a delay-aware RA optimization based on a double-queue model to improve access efficiency in smart hospital wards in the transmission layer. In application layer, this work builds a prediction model using a deep forest algorithm.

3. PROPOSED SYSTEM

In proposed work, 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 these three algorithms will be merging inside Ensemble Algorithm to get better accuracy and prediction.

1) 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.

2) Smartness: this technique will be considered as smart as its require no human effort to inform patient about current condition.

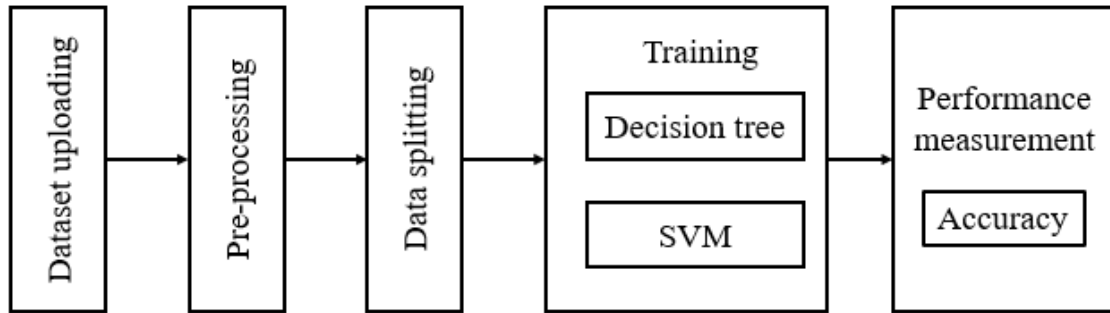Fig. 1: Block diagram of cloud application.

Here we have designed 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 sent 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.
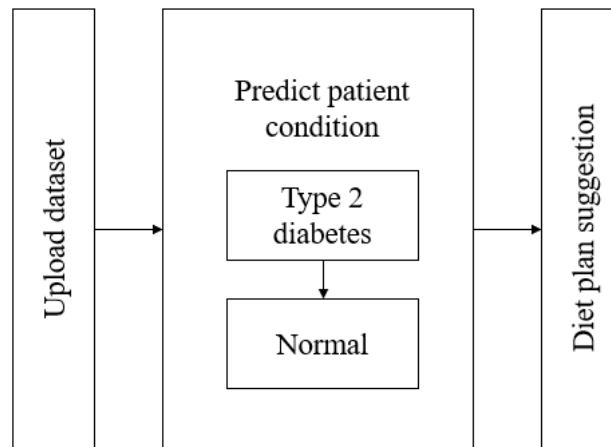
Fig. 2: Block diagram of user application.

3.1 Dataset

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 value 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 these 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

3.2 Introduction of Decision Tree Algorithm

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node represents the outcome.
- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
- The decisions or the test are performed based on features of the given dataset.
- It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
- It is called a decision tree because, like a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure.
- To build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm.
- A decision tree simply asks a question and based on the answer (Yes/No), it further split the tree into subtrees.
- Below diagram explains the general structure of a decision tree:

Features of Decision Tree Learning

- Method for approximating discrete-valued functions (including boolean)
- Learned functions are represented as decision trees (or if-then-else rules)
- Expressive hypotheses space, including disjunction
- The decision tree is robust to noisy data
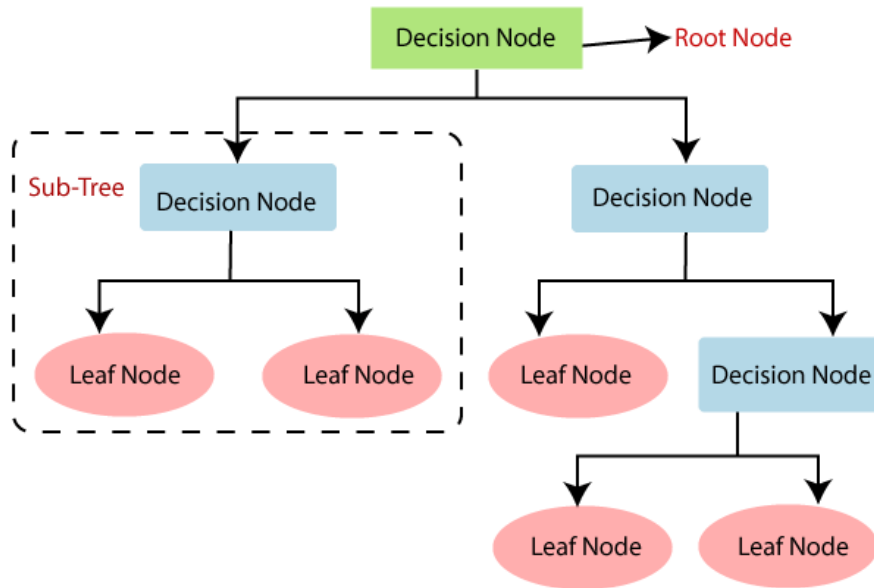
Decision Tree Representation

A decision tree is an arrangement of tests that provides an appropriate classification at every step in an analysis.

"In general, decision trees represent a disjunction of conjunctions of constraints on the attribute-values of instances. Each path from the tree root to a leaf corresponds to a conjunction of attribute tests, and the tree itself to a disjunction of these conjunctions".

More specifically, decision trees classify instances by sorting them down the tree from the root node to some leaf node, which provides the classification of the instance. Each node in the tree specifies a test of some attribute of the instance, and each branch descending from that node corresponds to one of the possible values for this attribute.

An instance is classified by starting at the root node of the decision tree, testing the attribute specified by this node, then moving down the tree branch corresponding to the value of the attribute. This process is then repeated at the node on this branch and so on until a leaf node is reached.

Diagram

- Each nonleaf node is connected to a test that splits its set of possible answers into subsets corresponding to different test results.
- Each branch carries a particular test result's subset to another node.
- Each node is connected to a set of possible answers.
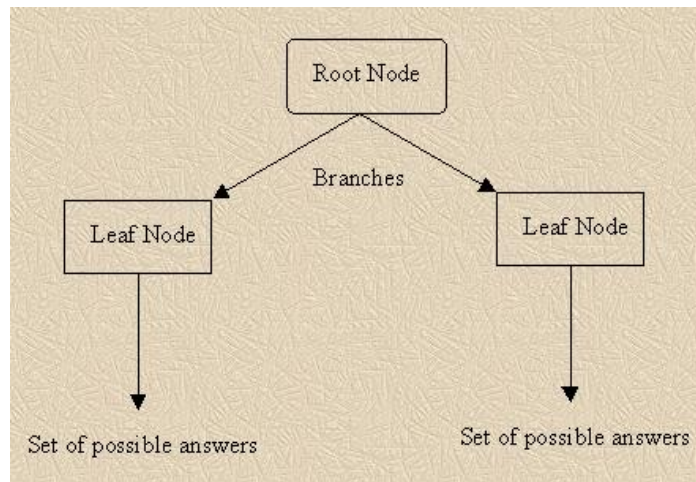
3.3 Support Vector Machine Algorithm

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:
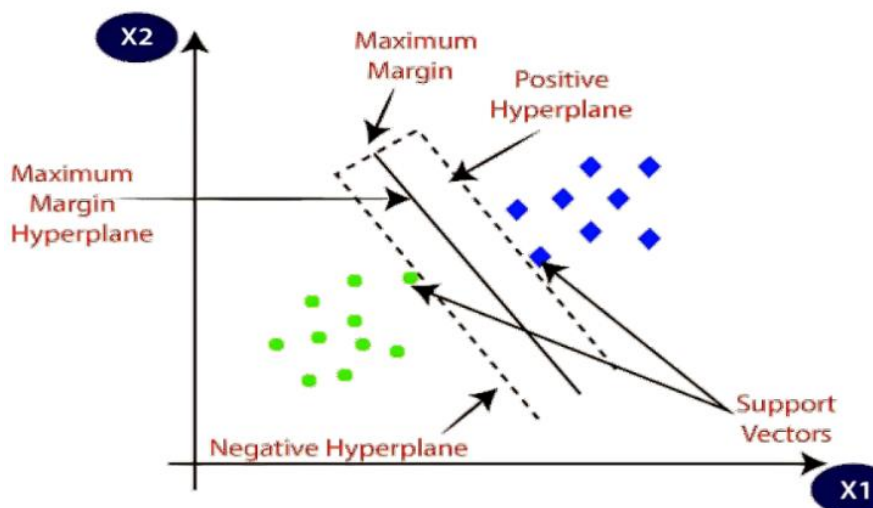
Applications

- Face recognition
- Weather prediction
- Medical diagnosis
- Spam detection
- Age/gender identification
- Language identification
- Sentimental analysis

- Authorship identification

3.3 Advantages of proposed system

- SVM works relatively well when there is a clear margin of separation between classes.
- SVM is more effective in high dimensional spaces.
- Decision tree is Simple to Understand for Coders.
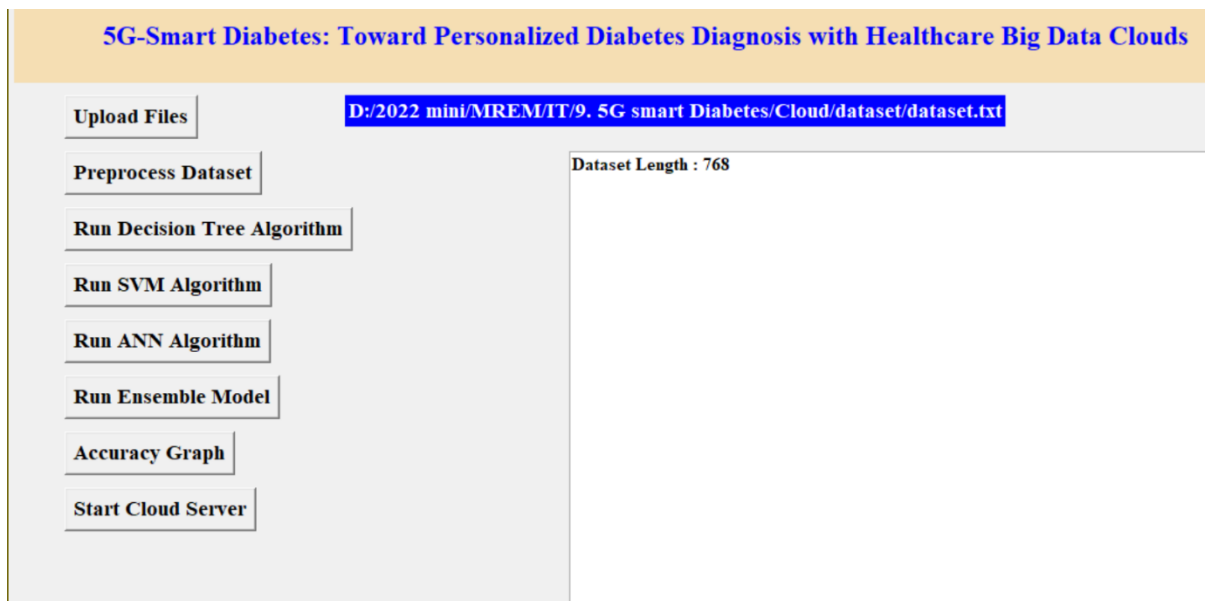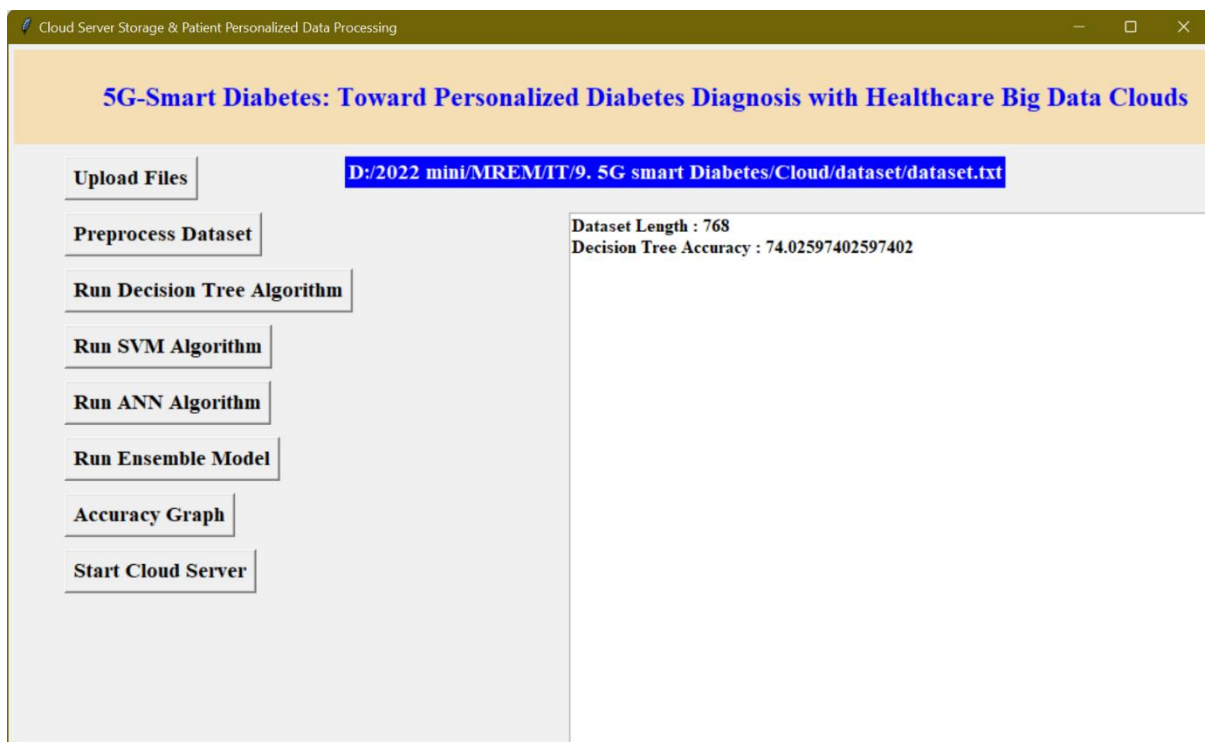- Missing Values Aren't an Issue

4. RESULTS
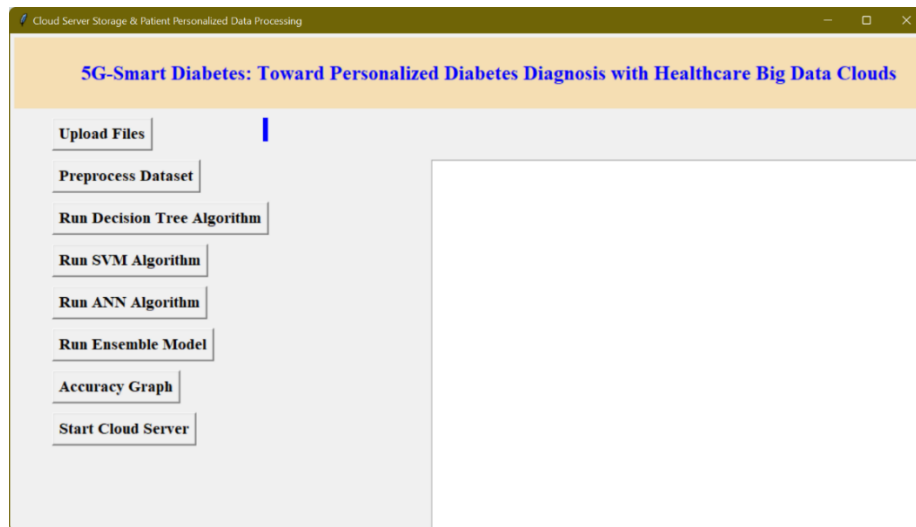
In above screen click on 'Upload Files' button to upload diabetes dataset
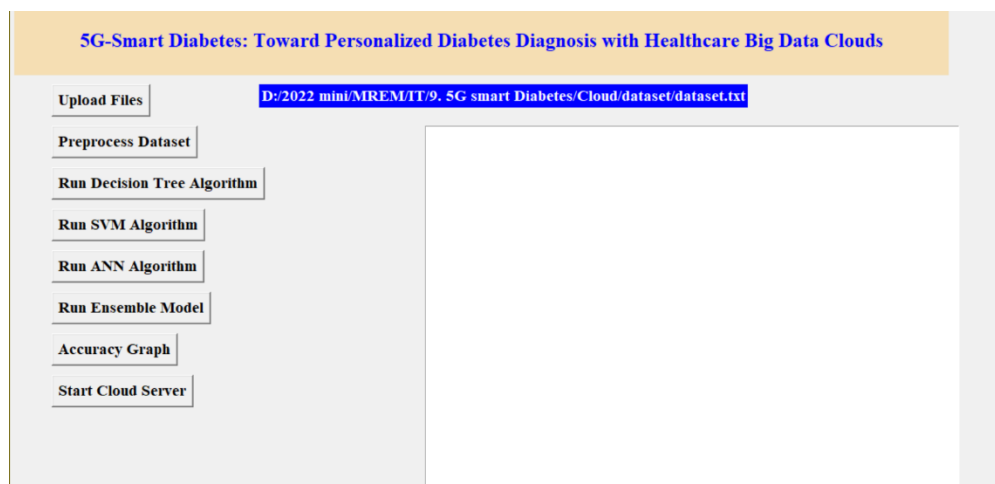
After uploading dataset click on 'Pre-process Dataset' button to clean dataset

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

Similarly run other buttons to build models with algorithms
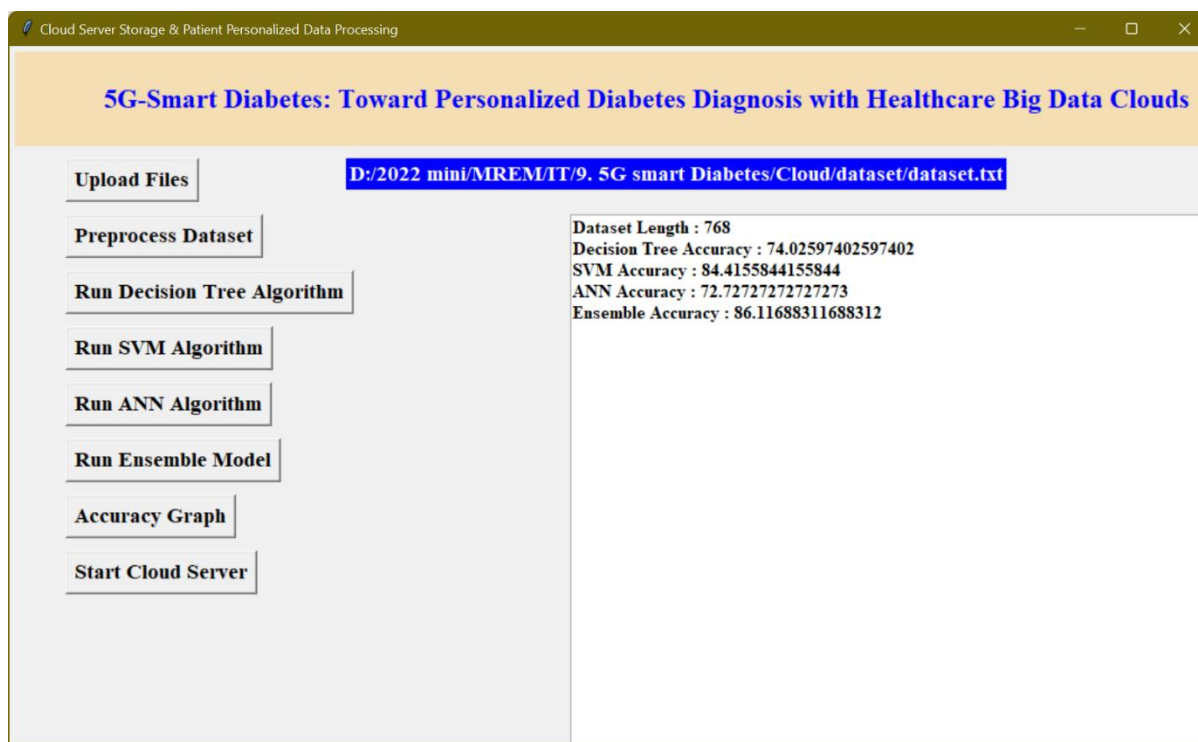
In above screen we got accuracy for all algorithms, now click on 'Accuracy Graph' button to get accuracy of all algorithms
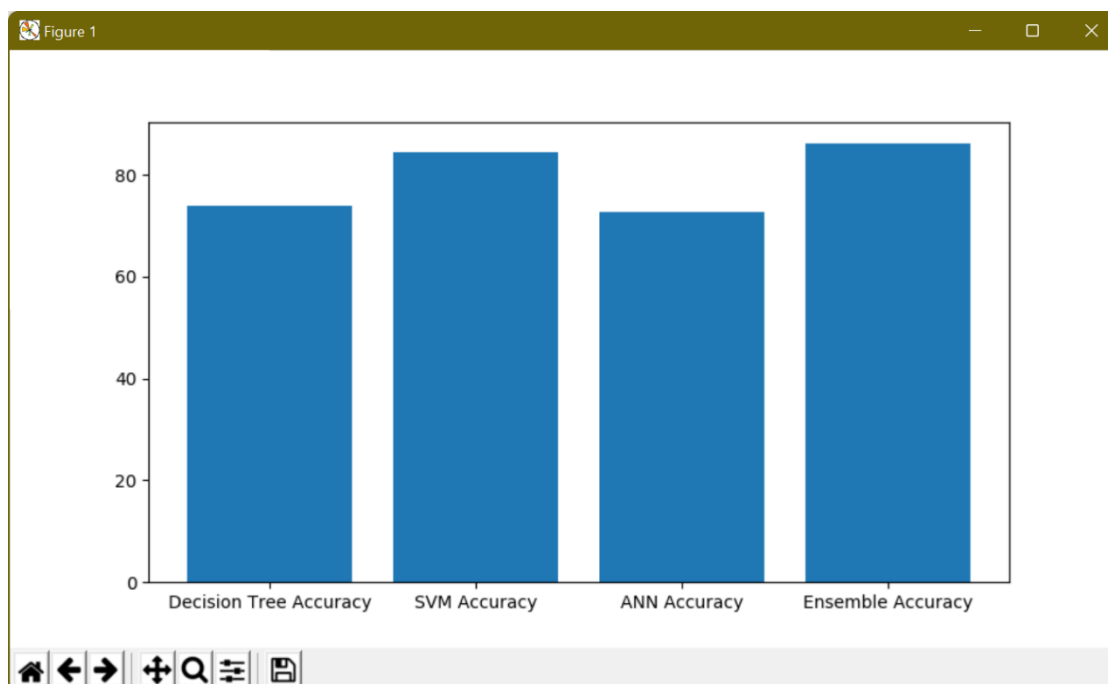
In above screen graph x-axis represents algorithm name and y-axis represents accuracy values.

Now click on 'Start Cloud Server' button to start server and this server will receive data from user and predict disease details. and now double clicks on 'run.bat' file from User folder to start User sensing application and click on 'Upload Files' button to upload test file and to predict patient condition. After uploading users data will get below prediction results

Personalized Diabetes Diagnosis with Healthcare Big Data Clouds

Upload Files   D:/2022 mini/MREM/IT/9. 5G smart Diabetes/User/data/users.txt

User Sense Data : 6,148,72,35,0,33.6,0.627,50
Abnormal Values. Predicted values : 1 Disease predicted as type 2 diabetes
User Sense Data : 1,85,66,29,0,26.6,0.351,31
Normal Values. Predicted values : 0 No disease predicted
User Sense Data : 8,183,64,0,0,23.3,0.672,32
Abnormal Values. Predicted values : 1 Disease predicted as type 2 diabetes
User Sense Data : 1,89,66,23,94,28.1,0.167,21
Normal Values. Predicted values : 0 No disease predicted
User Sense Data : 1,189,60,23,846,30.1,0.398,59
Abnormal Values. Predicted values : 1 Disease predicted as type 2 diabetes

Diet Plan Details

Monday
 Breakfast: One poached egg and half a small avocado spread on one slice of Ezekiel bread, one orange. Total carbs: Approximately 39

 Lunch: Mexican bowl: two-thirds of a cup low-sodium canned pinto beans, 1 cup chopped spinach, a quarter cup chopped tomatoes, a qua
bell peppers, 1 ounce (oz) cheese, 1 tablespoon (tbsp) salsa as sauce. Total carbs: Approximately 30.

 Snack: 20 1-gram baby carrots with 2 tbsp hummus. Total carbs: Approximately 21.

 Dinner: 1 cup cooked lentil penne pasta, 1.5 cups veggie tomato sauce (cook garlic, mushrooms, greens, zucchini, and eggplant into it), 2
nd lean turkey. Total carbs: Approximately 35.

In above screen for each user data, we predicted 0 and 1 values and also indicates patient values as normal or abnormal. Further, it also suggests the diet plan.

5. CONCLUSION AND FUTURE SCOPE

In this article, we at first propose a 5G-Smart Diabetes structure that joins a distinguishing layer, an altered end layer, and a data sharing layer. Appeared differently in relation to Diabetes 1.0 and Diabetes 2.0, this system can achieve supportable, viable, and understanding diabetes assurance. By then we propose a very financially savvy data sharing framework in social space and data space. Besides, using AI procedures, we present a tweaked data examination model for 5G- Smart Diabetes. Finally, considering the shrewd dress, wireless and server ranch, we gather a 5G-Smart Diabetes

testbed. The preliminary outcomes exhibit that our structure can give tweaked finding and treatment proposals to patients.

Future Scope

This work extended with an intelligent architecture for monitoring diabetic patients by using machine learning algorithms. The architecture elements included smart devices, sensors, and smartphones to collect measurements from the body. The intelligent system collected the data received from the patient and performed data classification using machine learning in order to make a diagnosis. The proposed prediction system was evaluated by several machine learning algorithms, and the simulation results demonstrated that the sequential minimal optimization (SMO) algorithm gives superior classification accuracy, sensitivity, and precision compared to other algorithms.

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] 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, pp. 39--46.

[3] M. Chen et al., "Disease Prediction by Machine Learning over Big Healthcare Data," IEEE Access, vol. 5, June 2017, pp. 8869--79

[4] M. Chen, J. Yang, J. Zhou, Y. Hao, J. Zhang and C. -H. Youn, "5G-Smart Diabetes: Toward Personalized Diabetes Diagnosis with Healthcare Big Data Clouds," in IEEE Communications Magazine, vol. 56, no. 4, pp. 16-23, April 2018, doi: 10.1109/MCOM.2018.1700788.

[5] Rghioui A, Lloret J, Sendra S, Oumnad A. A Smart Architecture for Diabetic Patient Monitoring Using Machine Learning Algorithms. Healthcare. 2020; 8(3):348. https://doi.org/10.3390/healthcare8030348.

[6] Venkatachalam, K., Prabu, P., Alluhaidan, A.S. et al. Deep Belief Neural Network for 5G Diabetes Monitoring in Big Data on Edge IoT. Mobile Netw Appl 27, 1060–1069 (2022). https://doi.org/10.1007/s11036-021-01861-y.

[7] E P, Prakash et al. "Implementation of Artificial Neural Network to Predict Diabetes with High-Quality Health System." Computational intelligence and neuroscience vol. 2022 1174173. 30 May. 2022, doi:10.1155/2022/1174173.

[8] V. Tsoulchas, N. Tsolis, E. Zoumi, E. Skondras and D. D. Vergados, "Health Monitoring of People with Diabetes using IoT and 5G Wireless Network Infrastructures," 2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA, 2020, pp. 1-6, doi: 10.1109/IISA50023.2020.9284388.

[9] R. Huang, W. Feng, S. Lu, T. shan, C. Zhang, Y. Liu, "An artificial intelligence diabetes management architecture based on 5G", Digital Communications and Networks, 2022, ISSN 2352-8648, https://doi.org/10.1016/j.dcan.2022.09.004.