JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

An Ensemble learning framework for Accurate Diabetes prediction

Premalatha V
Computer Science and Engineering
Sasi Institute of Technology and Engineering
Tadepalligudem

yaswanth G
Computer Science and Engineering
Sasi Institute of Technology and Engineering
Tadepalligudem

Amruth Ram P
Computer Science and Engineering
Sasi Institute of Technology and Engineering
Tadepalligudem

Sravani CH
Computer Science and Engineering
Sasi Institute of Technology and Engineering
Tadepalligudem

Vijay Prabhas G
Computer Science and Engineering
Sasi Institute of Technology and Engineering
Tadepalligudem

Abstract—**Still a major worldwide health issue, diabetes affects many people who are unaware of their condition until serious symptoms appear. Sedentary lifestyles, bad eating habits, and stress all help to fuel diabetes, which calls for early detection and proactive control. With an eye on preventing major complications and lowering hospitalisations, this work aims to differentiate between Type 1 and Type 2 diabetes by means of monitoring and alerting.**

Aiming at patient datasets, the suggested solution uses sophisticated data analytics methods combining Decision Trees, Support Vector Machines (SVM), and Artificial Neural Networks (ANN). The method improves early detection and correct diabetes classification by using machine learning models, hence enabling prompt intervention. The initiative comprises two linked programs: the User Application and the Cloud Application. While the User Application allows patient engagement, hence allowing people to obtain their health insights, get notifications, and monitor their progress, the Cloud Application handles data storage, model training, and real-time analysis.

Its capacity to offer individualised forecasts and customised food regimens depending on a person's health data is a main feature of the system. The system provides dynamic recommendations to assist efficient diabetes control by means of trend and pattern analysis inside patient datasets. Cloud-based data sharing integration increases accessibility and guarantees that both patients and medical professionals can make educated judgements. This initiative seeks to enable people with proactive monitoring tools, hence increasing diabetes control, lowering complications, and raising general quality of life.

Keywords: *Diabetes Prediction, tkinter Framework, Personalised Healthcare,svm,ANN*

I. INTRODUCTION

Diabetes is a chronic health condition affecting millions worldwide, often going undetected due to stress, unhealthy lifestyles, and delayed symptom onset. Without timely diagnosis and management, it can lead to severe complications and hospitalization. This project focuses on distinguishing between Type 1 and Type 2 diabetes, with a primary emphasis on monitoring and alerting for Type 1 diabetes to prevent critical health conditions. By leveraging data analytics and machine learning algorithms such as Decision Trees, SVM, and ANN, the system aims to provide accurate predictions and proactive health recommendations. The project incorporates two applications—Cloud and User, ensuring efficient data sharing, model training, and real-time patient monitoring. Additionally, it offers personalized predictions and diet plans, empowering individuals with better diabetes management and reducing medical emergencies.

II. LITERATURE SURVEY

a) Communicating While Computing: Distributed mobile cloud computing over 5G heterogeneous networks: *https://ieeexplore.ieee.org/document/6923537*

Mobile data traffic is expected to double annually from 2010 to 2020, a 1,000-fold increase. This tremendous development demands vast wireless network bandwidth. As data traffic soars, we use smartphones, tablets, and computers for entertainment, health care, business, social media, travel, news, and more. This lifestyle's unprecedented wireless traffic growth is not matched by mobile device battery longevity [3]. The energy gap between complicated apps and mobile devices is widening. Allowing mobile devices to transfer energy-intensive processes to nearby stationary servers may fix this. Cyberforaging [4] or compute offloading [5], [6] has been studied for a long time. Computing offloading has increased due to cloud computing (CC), which provides resources on demand. All three are called infrastructure, platform, and software as a service. Virtualisation, which isolates and protects programs and data, is a key component of CC. VMs may scale on demand, improving system computing efficiency. MCC accesses cloud services using mobile phones [5]. Radio access energy consumption and wide area network latency to the cloud provider restrict today's MCC. Mobile users at macrocellular network edges have poor power consumption and WAN latency management. Due to millisecond contact times in 5G networks, notably the tactile Internet [10], near-future MCC must have strict latency control. To meet this constraint, the entire service chain must be rethought, from physical to virtual.

b) Mobile-Edge Computing Architecture: The role of MEC in the Internet of Things

https://ieeexplore.ieee.org/document/7574435

5G networks are made possible by the new technology known as mobile-edge computing (MEC). Inspired by the massive proliferation of the Internet of Things, MEC will manage several essential 5G purposes that are compatible with 4G networks. This article will provide the MEC architecture and framework, which were established by the ETSI MEC ISG standards group. We illustrate MEC in action, with an emphasis on the Internet of Things (IoT) as it is a vital component of 5G. Finally, we will go over the main points about the advantages and disadvantages of MEC moving towards 5G.

c) A Survey on Mobile Edge Computing: The Communication Perspective:

https://ieeexplore.ieee.org/document/8016573

Due to the proliferation of connected devices and the advent of 5G networks, mobile computing has moved away from cloud computing and towards mobile edge computing in recent years. Mobile Edge Computing (MEC) enables resource-constrained mobile devices to efficiently execute computing-intensive and latency-sensitive applications by offloading computation, network management, and storage to network edge locations, such as base stations and access points.In order to solve the main problems with 5G, MEC claims to significantly decrease mobile energy consumption and latency. Businesses and universities are working on MEC technology because of all the good things it can do. Several novel ideas, including computation offloading and network topologies, have emerged from MEC research, which seeks to integrate wireless communications with mobile computing in an efficient manner. Recent work from MEC on the topic of radio-and-computational resource management integration is summarised in this article. Additionally, we take a look at the problems, obstacles, mobility management, green MEC, privacy-aware MEC, cache-enabled MEC, and MEC system implementation. These upgrades will facilitate MEC's transition from theory to practice. We wrap off by discussing efforts to standardise MEC and common circumstances for its deployment.

d) An Advanced Bolus Calculator for Type 1 Diabetes: System Architecture and Usability Results

https://pubmed.ncbi.nlm.nih.gov/26259202/#:~:text=This%20paper%20presents%20the%20architecture%20and%20initial%20usability,various%20diabetes%20scenarios%20and%20automatically%20adjusting%20its%20param

This study details the design and preliminary usability testing of ABC4D, a sophisticated insulin bolus calculator for diabetes that can recognise diverse diabetes circumstances and adapt its settings dynamically to give individualised insulin recommendations. A patient platform accessible through a smartphone enables the manual entry of glucose levels as well as factors impacting blood glucose levels, such as the carbohydrate content of meals and physical activity, and then offers suggestions for insulin boluses in real-time. The automated bolus calculator's parameter adjustments are overseen by a clinical revision platform. The system uses an algorithm for bolus calculations based on case-based reasoning, which learns from new data and improves its recommendations for insulin boluses based on similar past occurrences (cases). Analysis of system was used to examine the usability of ABC4D. At the conclusion of the study, all participants were asked to fill out a usability questionnaire in order to offer more comments on ABC4D. A total of 103 ± 28 insulin suggestions were approved out of 115 ± 21 that the participants requested. A clinical expert authorised 723 (or 96%) of the 754 case revisions indicated by the clinical revision program, and the patient platform was modified accordingly.

e) Green and Mobility-Aware Caching in 5G Networks:

https://ieeexplore.ieee.org/document/8067654

The proliferation of mobile devices has led to an increase in mobile traffic as well as requests for mobile content. Efficient mobile traffic management during peak hours in 5G networks can be achieved through small cell base stations (SBSs) and wireless device-to-device (D2D) network caching. Most existing research focuses on caching content in SBSs and mobile devices, assuming users can fully retrieve requested data. However, user mobility and interaction time unpredictability have been largely overlooked. Optimizing caching by leveraging user mobility remains a complex challenge. This paper addresses this issue by proposing a caching placement strategy that enhances the cache hit ratio by utilizing user mobility for cache deployment on SBSs and mobile devices. The caching placement problem is formulated as an integer programming model, and submodular optimization is employed to derive an optimal solution. Additionally, energy efficiency is improved by optimizing the transmission power of SBSs and mobile devices for serving cached content. Simulation results demonstrate that the proposed strategy outperforms existing approaches in both cache hit ratio and energy efficiency.

III. METHODOLOGY

A) Proposed System

The proposed system aims to monitor and predict Type 1 diabetes using data analytics and machine learning, integrating Decision Trees, SVM, and ANN algorithms for accurate classification and personalized health recommendations. The system consists of two applications: a **Cloud Application**, which securely stores patient data, trains predictive models, and allows medical professionals to access insights, and a **User Application**, which collects real-time glucose levels from sensors, provides alerts for abnormal trends, and offers personalized diet and lifestyle suggestions. By leveraging AI-driven predictions and real-time monitoring, the project enhances diabetes management, reduces hospitalization risks, and improves patient well-being through proactive intervention.

B) System Architecture

The extended Smart Diabetes system integrates advanced ensemble learning techniques to enhance diabetes diagnosis and management. Unlike traditional methods, this system leverages a combination of SVM and DECISION TREE ,ANN models, which effectively process both spatial and temporal features from real-time physiological data. Our proposed approach enhances prediction accuracy, allowing for early detection and personalized treatment recommendations. the system enables continuous health monitoring with seamless data transmission to cloud-based healthcare platforms.

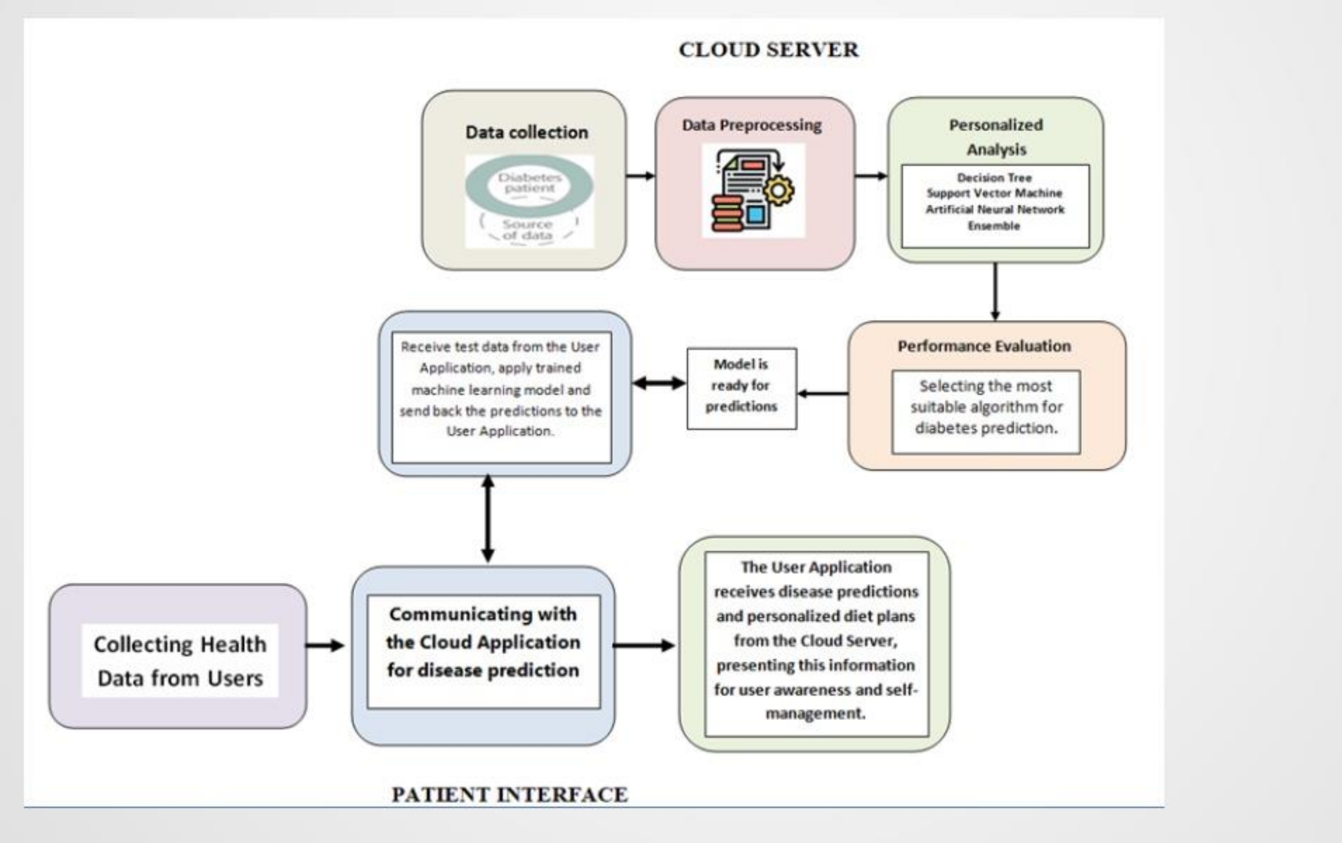

Fig 1:Architecture

C) MODULES

a) Data Loading

This module is responsible for importing the dataset into the system. The dataset includes essential diabetes-related information such as glucose levels, insulin levels, age, BMI, and other physiological parameters. Proper data loading ensures that all required data is available for further processing and analysis, forming the foundation for accurate diabetes prediction.

b) Data Preprocessing

Data preprocessing involves cleaning, transforming, and organizing the dataset to ensure consistency and accuracy. This step includes handling missing values, removing duplicate records, and normalizing the data. Proper preprocessing enhances the efficiency of the machine learning models, ensuring high-quality input for accurate predictions in diabetes diagnosis.

c) Data Visualization

This module presents processed data through various graphical representations, such as charts, histograms, and correlation heatmaps. Visualizing the data helps in identifying trends, patterns, and anomalies, making it easier for healthcare professionals to analyze diabetes risks and develop effective treatment plans.

d) Extra Tree Feature Selection

The Extra Tree ensemble method is used to select and rank the most important features from the dataset. By reducing the dimensionality of the data, this technique eliminates irrelevant attributes, improving model efficiency and prediction accuracy. Feature selection ensures that only the most critical diabetes-related factors contribute to the diagnosis.

e) Splitting Data into Train & Test

This module divides the dataset into two subsets: training and testing data. The training data is used to build and train machine learning models, while the testing data evaluates model performance. Proper data splitting ensures that the system generalizes well to new, unseen data, improving reliability in real-world diabetes diagnosis.

f) Model Generation

This module focuses on building and training different ML and DL models, including ANN, SVM, DT, and an Ensemble Learning Algorithm. Each diabetes prediction model is trained and evaluated using recall, accuracy, precision, and F1-score to discover the best one.

g) User Signup & Login

A secure authentication system is integrated into the platform, allowing users to register and log in. This module ensures data privacy and personalized access to diabetes monitoring. A Flask-based framework is used to implement a user-friendly interface for managing user accounts securely.

h) User Input

Users can enter their real-time health data, such as glucose levels, heart rate, and dietary habits, into the system. This module processes user inputs and prepares them for analysis, enabling personalized diabetes predictions and recommendations based on individual health parameters.

i) Prediction

The final prediction is displayed to the user based on the trained model's analysis. The system provides insights into diabetes risk levels, personalized treatment suggestions, and preventive measures. By using an ensemble method, the system enhances accuracy and reliability in diabetes detection, ensuring better health management.

D) Algorithms

a) ANN (Artificial Neural Network):

Through the use of interconnected layers of nodes, ANNs are able to simulate the way the brain processes data. In order to gather intricate patterns of diabetes data, this study employs ANN. ANN learns from the dataset and uses patient attributes to predict diabetes. It is an excellent option for enhancing system diabetes forecasts due to its ability to recognise complicated relationships and patterns.

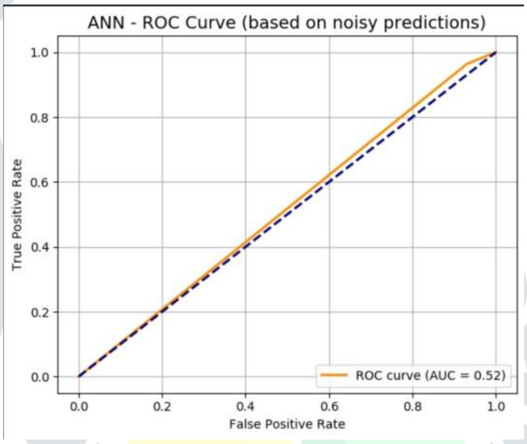

Fig 2:ANN ROC graph

b) SVM

SVM does both classification and regression. SVM determines the optimal hyperplane for data point classification. Optimal distance between data points of different classes is maximised by selecting the hyperplane with the highest margin. By capturing complex dataset relationships, this project correctly detects diabetic conditions, making it a good fit using SVM. Its capacity for binary categorisation and high-dimensional space lend credence to the project's aim of precise diabetes prediction.

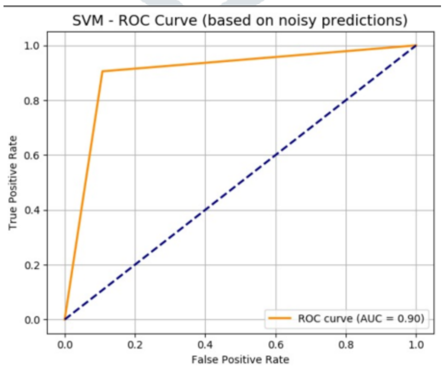

Fig 3:SVM ROC graph

c) DECISION TREE

Machine learning algorithms known as decision trees use a hierarchical representation of decisions and their outcomes to form conclusions. Core nodes reflect feature-based decisions and leaf nodes indicate outcomes in a tree structure that is created by recursively splitting the dataset into subgroups based on the most relevant attributes. An interpretable diabetes prediction model that aids in patient categorisation based on physiological factors has been developed in this study using the

Decision Tree algorithm. Because of its openness, the system may be utilised to examine the parameters for diabetes diagnosis prediction.

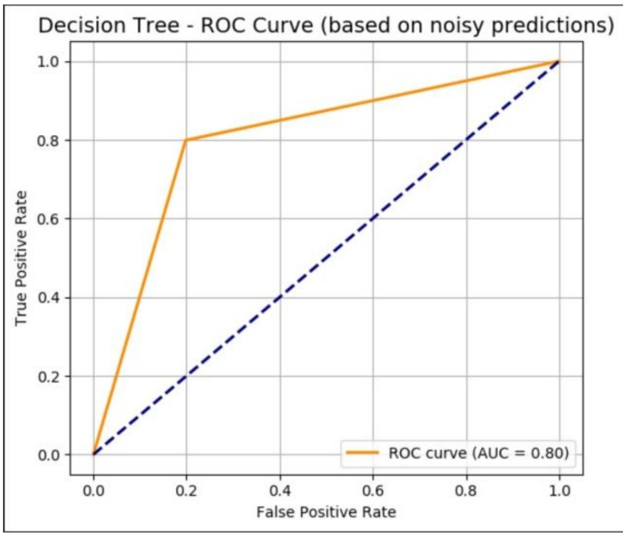

Fig 4:Decision Tree ROC graph

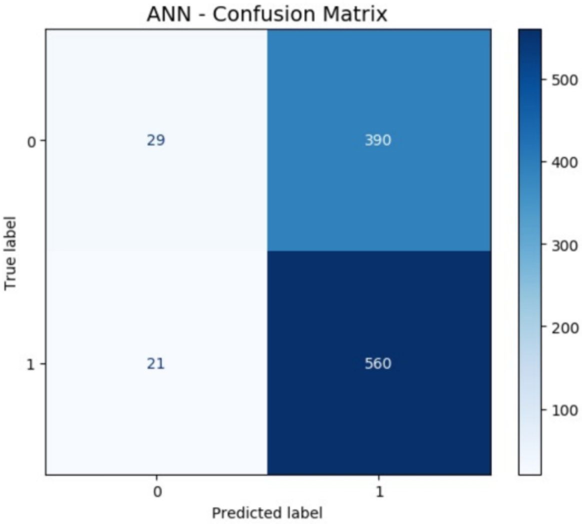

Fig 5:ANN Confusion Matrix

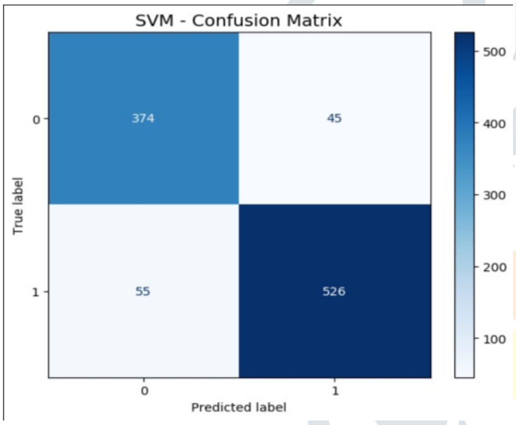

Fig 6:SVM confusion Matrix

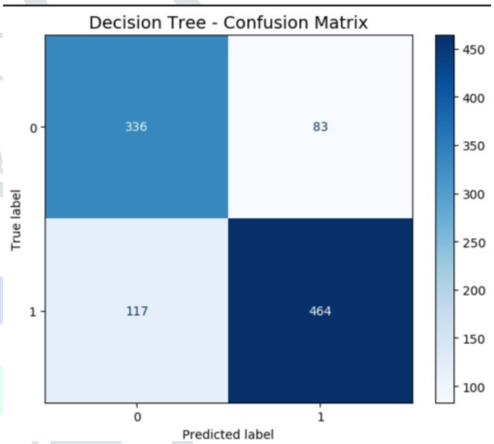

Fig 7:Decision Tree Confusion Matrix

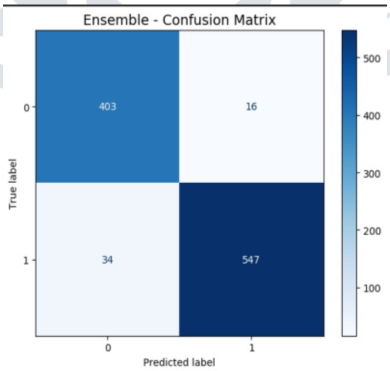

Fig 489:Ensemble Confusion Matrix

d) ENSEMBLE

To increase precision and robustness, machine learning ensembles pool algorithm predictions. The Ensemble Algorithm is employed in this study to improve the accuracy of diabetes prediction by combining the results of Decision Tree, SVM, and ANN. With the help of ensemble, the system is made stronger, biases are reduced, and diabetes predictions are improved, resulting in a more comprehensive and effective monitoring system.

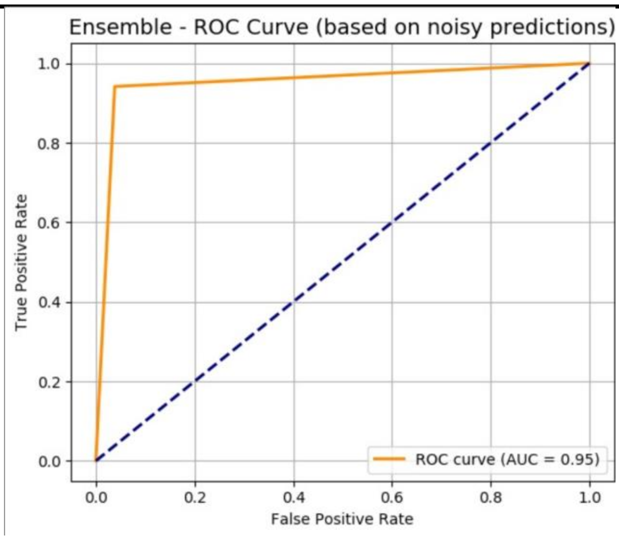

Fig 9:Ensemple Roc Graph

e)About Dataset

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

IV. EXPERIMENTAL RESULTS

Accuracy: Find out how reliable a test is by comparing real positives and negatives. Following mathematical:
Accuracy = (TP + TN) /(TP + TN + FP + FN)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying using the one that follows:

Precision = (TP)/ (TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by percent of correctly anticipated positive observations relative to total positives.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: An accurate machine learning model has a high F1 score. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

MAP: Information retrieval system performance is measured by MAP, which stands for Mean Average Precision. It finds the mean precision for all classes or queries. While accuracy measures the validity of results, precision determines the mean accuracy for all queries. MAP evaluates the system's performance by averaging the AP scores across all queries or classes.

$$\text{MAP} = \frac{1}{N}\sum_{i=1}^{N} \text{AP}_i$$

Model	**Accuracy**	**F1-Score**	**Precision**	**Recall**	**ROC-AUC**
ANN	**0.589**	**0.727**	**0.589**	**0.962**	**0.52**
SVM	**0.900**	**0.901**	**0.921**	**0.905**	**0.90**
Decision Tree	**0.800**	**0.775**	**0.848**	**0.799**	**0.80**
Ensemble	**0.926**	**0.930**	**0.971**	**0.941**	**0.95**

The **Ensemble model** has the highest **ROC-AUC (0.95)** and overall best classification performance.

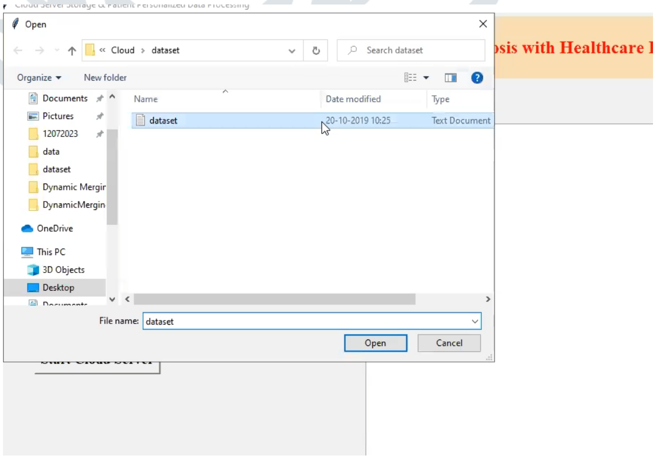

Fig.10: upload dataset

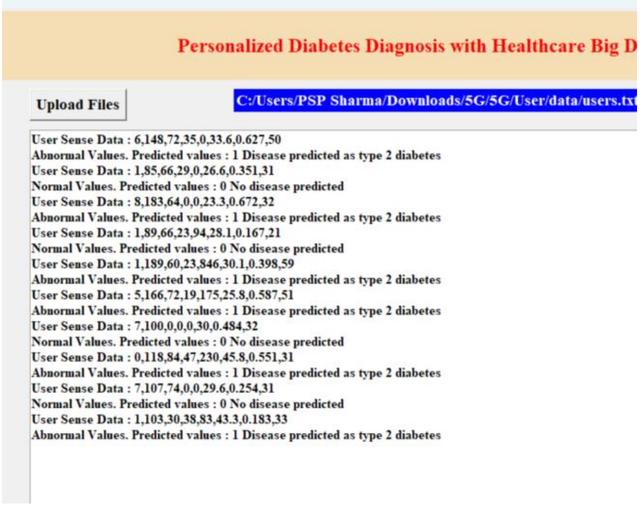

Fig.11. predicted results

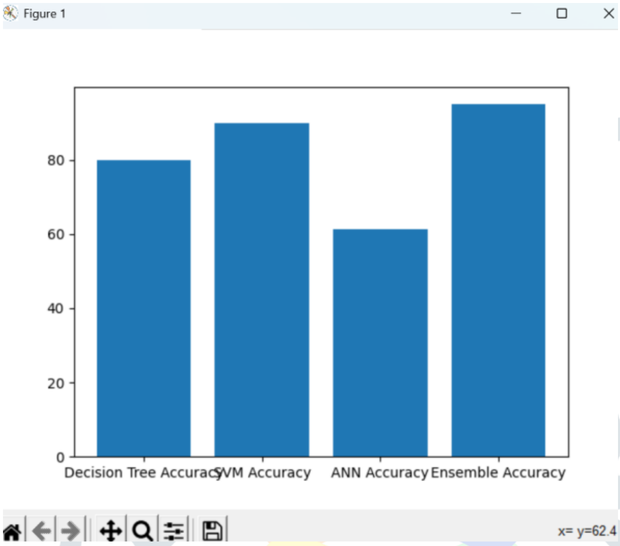

Fig.12. accuracy graph

V. CONCLUSION

In conclusion, the extended Smart Diabetes system revolutionizes diabetes management by integrating advanced ensemble learning techniques with real-time health monitoring. By leveraging Decision Trees, SVM, and ANN models the system ensures seamless data transmission, early detection, and personalized treatment recommendations. This innovative approach enhances prediction accuracy, reduces hospitalization risks, and empowers individuals with proactive diabetes management, ultimately improving overall health outcomes and quality of life.

REFERENCES

[1] Sergio Barbarossa, Stefania Sardellitti, and Paolo Di Lorenzo. "Communicating While Computing: Distributed Mobile Cloud Computing over 5G Heterogeneous Networks", IEEE Signal Processing Magazine 31.6 (2014): 45-55.

[2] Dario Sabella, et al. "Mobile-edge computing architecture: The Role of MEC in the Internet of Things", IEEE Consumer Electronics Magazine 5.4 (2016): 84-91.

[3] Peter Corcoran, and Soumya Kanti Datta. "Mobile-Edge Computing and the Internet of Things for Consumers: Extending Cloud Computing and Services to the Edge of the network", IEEE Consumer Electronics Magazine 5.4 (2016): 73-74.

[4] Yuyi Mao, Changsheng You, Jun Zhang, Kaibin Huang and Khaled B. Lataief, "A Survey on Mobile Edge Computing: The Communication Perspective", IEEE Communications Surveys & Tutorials. IEEE, 2017.

[5] Nasir Abbas, Yan Zhang, Amir Taherkordi, et al. "Mobile Edge Computing: A Survey", IEEE Internet of Things Journal 5.1 (2018): 450-465.

[6] Shankar Lal, Tarik Taleb, and Ashutosh Dutta. "NFV: Security Threats and Best Practices", IEEE Communications Magazine, Vol. 55, Issue 8, pp 211–217. IEEE, 2017.

[7] Hongwei Li, Rongxing Lu, Jelena Misic, et al., "Security and Privacy of Connected Vehicular Cloud Computing", IEEE Network, Vol. 32, Issue 3, pp. 4-6. IEEE, June 2018

[8] A. Klein, C. Mannweiler, J. Schneider and H. D. Schotten, "Access Schemesfor Mobile Cloud Computing", in 2010 Eleventh International Conferenceon Mobile Data Management, Kansas City, MO, USA, 2010, pp. 387-392.