**An Integrated Approach for Detecting Diabetes and Recommending Diet Plans in Healthcare Big Data Clouds using Ensemble Framework**

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 using below techniques:

1. Cost: this technique requires no cost compared 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 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 consider as smart as its require no human effort to inform patient about current condition.

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.

Using diabetes data as dataset and below are 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 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

**Diagram, schematic

Description automatically generated**

Fig. 1: Block diagram of cloud application.

**Diagram, schematic

Description automatically generated**

Fig. 2: Block diagram of user application.

**RESULTS AND DISCUSSION**

**Dataset description**

The dataset includes various attributes or features that are commonly used in diabetes prediction. Here's a description of each attribute:

* Pregnancies: This attribute represents the number of times a person has been pregnant. It is a discrete numerical value.
* Glucose: This measures the plasma glucose concentration after a certain period of fasting. It is a continuous numerical value and is often measured in milligrams per deciliter (mg/dL).
* BloodPressure: This feature refers to the diastolic blood pressure of an individual, measured in millimeters of mercury (mmHg). Diastolic blood pressure is the pressure in the arteries when the heart is at rest between beats.
* SkinThickness: This indicates the thickness of the skinfold at a particular site on the body. It is often measured in millimeters and can be used as an indicator of body fat.
* Insulin: This value represents the level of insulin in the blood. Insulin is a hormone that helps regulate blood sugar levels. The measurement is typically given in microunits per milliliter (μU/mL).
* BMI (Body Mass Index): BMI is a calculated value that reflects a person's body fat based on their weight and height. It is a numerical value and is commonly used to assess whether a person is underweight, normal weight, overweight, or obese.
* DiabetesPedigreeFunction: This attribute is a numerical value that represents the genetic influence of diabetes based on family history. It provides an estimation of the risk of diabetes based on genetics.
* Age: This indicates the age of the individual in years. It is a discrete numerical value.
* Outcome: This attribute represents the target variable or class label that indicates whether an individual has diabetes or not. It's a binary attribute, where 1 could indicate the presence of diabetes and 0 could indicate the absence of diabetes.

Figure 1 depicts the graphical user interface (GUI) application of a system designed for detecting diabetes and recommending personalized diet plans. The user interface includes input fields for entering relevant data (such as medical information), buttons to trigger actions like data prediction and diet plan recommendation, and visualizations for displaying results. This application serves as the front-end interaction point for users to input their data, receive predictions, and get diet recommendations.

A screenshot of a computer

Description automatically generated

Figure 1: User interface application of proposed diabetes detection and diet plan recommendation system.

A screenshot of a computer

Description automatically generated

Figure 2: Illustrating the data length after performing data preprocessing.

Figure 2 demonstrates the effect of data preprocessing on the dataset's length. Data preprocessing involves various operations like cleaning, transforming, and encoding data to make it suitable for analysis or modeling. It also shows a comparison between the original raw data length and the length after preprocessing, showcasing how certain data points may have been removed or modified during the process. Figure 3 displays a comparison of training accuracy results between existing and proposed ensemble classifier models on the user interface. It presents the training accuracy values achieved by the different classifier models, which helps the users to understand the performance improvement brought by the proposed ensemble models.

A screenshot of a computer

Description automatically generated

Figure 3: User interface application with obtained results of training accuracy using existing and proposed ensemble classifier models.

Figure 4 highlights the comparison of accuracy performance between the existing and proposed ensemble models for diabetes prediction. It emphasizes the practical impact of the proposed ensemble model's accuracy improvement. Figure 5 represents a screen within the user interface that allows users to upload their personalized test data. The screen includes fields for uploading files of test dataset. It is crucial for the prediction process, as the uploaded data will be used to determine the patient's health condition.

Chart, bar chart

Description automatically generated

Figure 4: Performance comparison of accuracy obtained using existing and proposed ensemble model for diabetes prediction.

Now ‘Start Cloud Server’ and this server will receive data from user and predict disease details.

Now run application in User folder to start User sensing application and upload test file to predict patient condition.

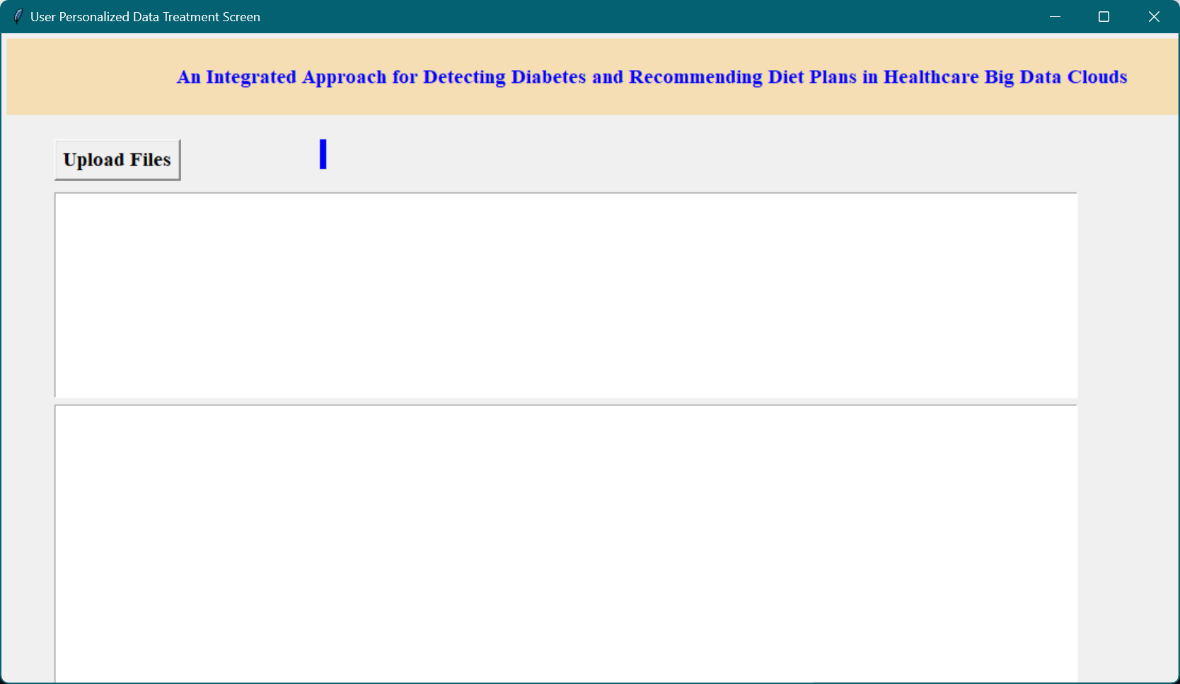


Figure 5: User personalized data treatment screen for uploading the test data.

Figure 6 displays the outcomes of the prediction process after the user has uploaded their data. It shows predicted health condition details, such as whether the patient is diagnosed with diabetes or not. Additionally, it also includes details of the recommended diet plan tailored to the user's health condition. Figure 7 shows a sample image of a diet plan that the system recommends to users based on their health condition. The image could include meal suggestions, portion sizes, and dietary guidelines. This visual aid helps users visualize the diet plan and understand what kinds of foods they should be consuming.

A screenshot of a computer

Description automatically generated

Figure 6: Obtained predictions after uploading users’ data along with diet plan suggestion details.

A picture containing text

Description automatically generated

Figure 7: Sample diet plan image.