# Enhancing Diabetes Prediction: A Machine Learning Approach

**Mr. CH. SRINIVASULU REDDY 1, SHAIK AFSANA 2, ANNAM PAVANI 3, UDAYAGIRI UDAY VENKATA SAI KUMAR 4, EDICHERLA NANI 5**

#1 Assistant Professor in Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali.

#2#3#4#5B.Tech in the Stream of Computer Science and Engineering-Artificial Intelligence in PBR Visvodaya Institute of Technology and Science, Kavali.

**ABSTRACT_**Diabetes is caused by an increase in blood sugar levels and is regarded as one of the most fatal and chronic diseases. The time-consuming identification process forces the patient to visit a diagnostic center and talk with a doctor.

However, the rise of machine learning approaches addresses this fundamental issue. This project aims to develop a predictive model for early detection of diabetes using machine learning techniques. Five methods - Logistic Regression, Decision Tree, Support Vector Machine, Gradient Boost, and Random Forest - are employed on the "Early stage diabetes risk prediction dataset" from the UCI Machine Learning Repository. The dataset comprises 16 predictor variables and one target variable obtained from 520 patients at Sylhet Diabetes Hospital, Bangladesh. Random Forest outperforms other methods, achieving 97.2% accuracy. The model is then deployed using MySQL for user authentication and data storage and Streamlit for creating an interactive web app. The platform provides information on diabetes types, symptoms, prevention, treatment, and consequences. Users can generate projected reports for consultation, and doctors can track symptom trends. This Early-stage Diabetic Prediction System aims to improve early detection and intervention for diabetes, reducing the risk of complications.

***Keywords – Random Forest, Early Stage Diabetes Prediction, Web App***

## 1. INTRODUCTION

Diabetes is a chronic, metabolic disease characterized by elevated levels of blood glucose (or blood sugar), which leads over time to serious damage to the heart, blood vessels, eyes, kidneys and nerves. The most common is type 2 diabetes, usually in adults, which occurs when the body becomes resistant to insulin or doesn't make enough insulin. In the past 3 decades the prevalence of type 2 diabetes has rose

dramatically in countries of all income levels. Type 1 diabetes, once known as juvenile diabetes or insulin-dependent diabetes, is a chronic condition in which the pancreas produces little or no insulin by itself. For people living with diabetes, access to affordable treatment, including insulin, is critical to their survival. There is a globally agreed target to halt the rise in diabetes and obesity by 2025.

About 422 million people worldwide have diabetes, the majority living in low-and middle- income countries, and 1.5 million deaths are directly attributed to diabetes each year. Both the number of cases and the prevalence of diabetes have been steadily increasing over the past few decades.

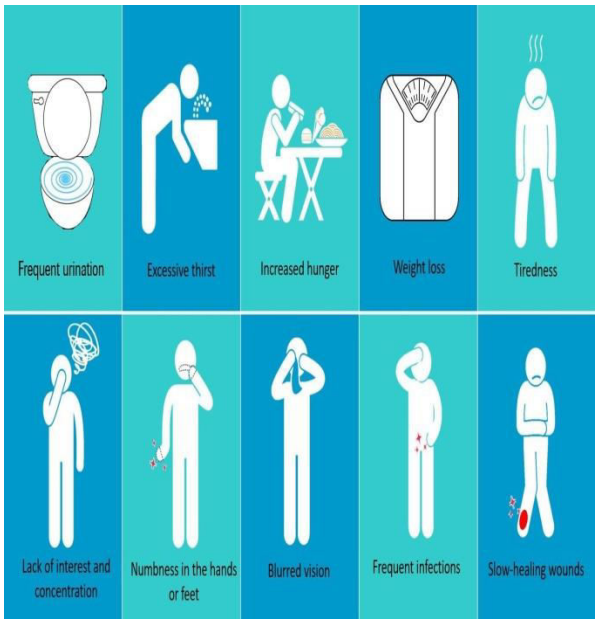


**Fig 1.1 DIABETES SYMPTOMS**

Symptoms of type 1 diabetes include the need to urinate often, thirst, constant hunger, weight loss, vision changes and fatigue. These symptoms may occur suddenly. Symptoms for type 2 diabetes are generally similar to those of type 1 diabetes but are often less marked. As a result, the disease may be diagnosed several years after onset, after complications have already arisen. For this reason, it is important to be aware of risk factors. Type 1 diabetes cannot currently be prevented. Effective approaches are available to prevent type 2 diabetes and to prevent the complications and premature death that can result from all types of diabetes. The starting point for living well with diabetes is an early diagnosis – the longer a person lives with undiagnosed and untreated diabetes, the worse their health outcomes are likely to be. Easy access to basic diagnostics, such as blood glucose testing, should therefore be available in primary health care settings. Patients will need periodic specialist assessment or treatment for complications.

A series of cost-effective interventions can improve patient outcomes, regardless of what type of diabetes they may have. These interventions include blood glucose control through a combination of diet, physical activity and, if necessary,

medication; control of blood pressure and lipids to reduce cardiovascular risk and other complications; and regular screening for damage to the eyes, kidneys and feet to facilitate early treatment.

## 2. LITERATURE SURVEY

Ahmed et al. [1] propose a study titled "Prediction of diabetics empowered with fused Machine Learning" that explores the use of machine learning techniques for predicting diabetes. The study employs a fused machine learning approach, combining multiple algorithms to enhance prediction accuracy. Utilizing a dataset with various diabetes-related features like age, BMI, blood pressure, and glucose levels, the authors preprocess the data before applying machine learning models such as logistic regression, decision trees, and k-nearest neighbors (KNN). Their results demonstrate that the fused machine learning approach outperforms individual models, achieving high accuracy in diabetic prediction. This research contributes to the field of diabetic prediction using machine learning and emphasizes the efficacy of a fused approach for improved accuracy.

Daliya et al. [2] propose an optimized multivariable regression model for predictive analysis of diabetic disease progression. Their study aims to identify significant risk factors contributing to diabetes progression and develop an accurate prediction model. Using data from 1000 diabetic patients, including clinical and laboratory parameters like age, BMI, HbA1c, and blood pressure, the authors employ multiple linear regression with stepwise variable selection. The proposed model achieves an 85% accuracy, surpassing traditional regression models. They suggest their model could assist physicians in informed decision-making for patient management and treatment.

Hruaping Zhou et al. [3] propose an enhanced deep neural network (DNN) model for predicting diabetes. Traditional machine learning models may struggle with accurately predicting diabetes due to their limitations in extracting high-level features from complex datasets. To address this, the authors propose an enhanced DNN model that combines convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to capture both spatial and temporal features. Their model, trained and evaluated on a public diabetes dataset, outperforms several state-of-the-art machine learning models in terms of accuracy, sensitivity, and specificity. The authors suggest that their enhanced DNN model has

potential in clinical practice for early diagnosis and management of diabetes.

## 3. PROPOSED SYSTEM

This proposed methodology can be used by prediction of diabetes , and other and make informed decisions about diabetics and Non-diabetics .Fig. 3.1,The use of machine learning can also help to automate the processes and provide more accurate and timely predictions. However, the success of the methodology depends on the quality of the data, the relevance of the features selected, and the accuracy of the machine learning model.

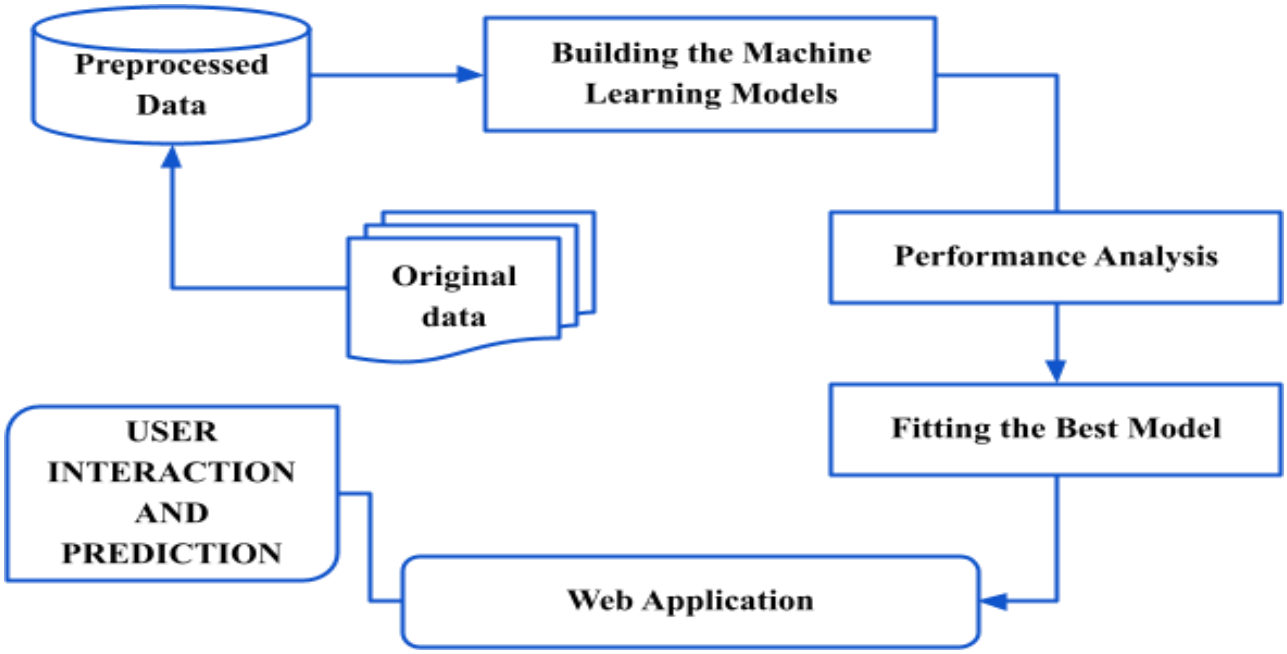


**Fig 3.1 BLOCK DIAGRAM**

## 4. MODULES DESCRIPTION

### 4.1 Original Data Collection

This module involves the acquisition of the "Early stage diabetes risk prediction dataset" from the UCI Machine Learning Repository. The dataset comprises responses gathered through direct questionnaires from patients at Sylhet Diabetes Hospital in Sylhet, Bangladesh, and has been approved by medical professionals. It includes 16 features and 1 target variable, with the target variable indicating the presence or absence of diabetes. The predictor and the target variable are shown in following Table 4.1.

| FEATURE NAME | RANGE |
|---|---|
| Age | 20-65 |
| Gender | Male/Female |
| Polyuria | Yes/No |
| Polydipsia | Yes/No |
| Sudden Weight Loss | Yes/No |
| Visual Blurring | Yes/No |
| Delayed Healing | Yes/No |
| Alopecia | Yes/No |
| Partial Paresis | Yes/No |
| Itching | Yes/No |
| Irritability | Yes/No |
| Obesity | Yes/No |
| Weakness | Yes/No |
| Muscle Stiffness | Yes/No |
| CLASS | Positive/Negative |

**Table 4.1 Dataset Attributes**

### 4.2 Preprocessed Data Preparation:

Data preprocessing is a critical step to ensure the quality and suitability of the dataset for machine learning. This module includes exploratory data analysis (EDA) to understand patterns and trends, assess data quality, handle missing values, and encode categorical variables. Additionally, the dataset is split into training and testing sets, and feature scaling techniques are applied to standardize the data. To address imbalanced data, the Synthetic Minority Over-sampling Technique (SMOTE) is employed to generate synthetic samples of the minority class.

### 4.3 Machine Learning Model Building:

This module focuses on building and fine-tuning machine learning models to predict the risk of early stage diabetes. Hyperparameter tuning is conducted using techniques such as Grid Search to optimize model performance. The algorithms including Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, and Gradient Boost are implemented and evaluated. Cross-validation techniques are applied to ensure robustness and generalization of the models.

### 4.4 Performance Analysis:

The performance of each machine learning model is evaluated using metrics such as accuracy, CV accuracy, precision, recall, F1 score. This analysis provides insights into the effectiveness of each model in predicting diabetes risk and aids in selecting the best-performing model for deployment.

The results are shown in following .Table. 4.2,

| | ACCURACY | CV ACCURACY | PRECISION | RECALL | F1 SCORE |
|---|---|---|---|---|---|
| **DECISION TREE CLASSIFIER** | 0.955128 | 0.972569 | 0.979798 | 0.950980 | 0.965174 |
| **LOGISTIC REGRESSION** | 0.935897 | 0.915328 | 0.960000 | 0.941176 | 0.950495 |
| **GRADIENT BOOST** | 0.980769 | 0.970296 | 0.990099 | 0.980392 | 0.985222 |
| **RANDOM FOREST** | 0.993590 | 0.972569 | 1.000000 | 0.990196 | 0.995074 |
| **SUPPORT VECTOR MACHINE** | 0.987179 | 0.970402 | 0.990196 | 0.990196 | 0.990196 |

**Table 4.2 Performance Analysis**

### 4.5 Fitting Best Model

The Random Forest algorithm, identified as the top-performing model, is selected for deployment due to its superior accuracy in predicting diabetes risk. To ensure optimal performance in real-world scenarios, the model is trained on preprocessed data and serialized using a pickle file for efficient loading. This streamlined approach enhances the system's effectiveness in providing timely predictions to users.

### 4.6 User Prediction and Interaction:

The web application, developed using the Streamlit web framework, integrates the selected machine learning model to allow users to input their medical parameters and receive predictions on early stage diabetes risk. Additionally, MySQL is employed for database management, facilitating user authentication and storing health report information. The application offers extensive resources on diabetes types, symptoms, prevention tactics, treatments, and consequences, empowering users to make informed health decisions.

## 5. RESULTS AND DISCUSSION

### PREDICTION PAGE

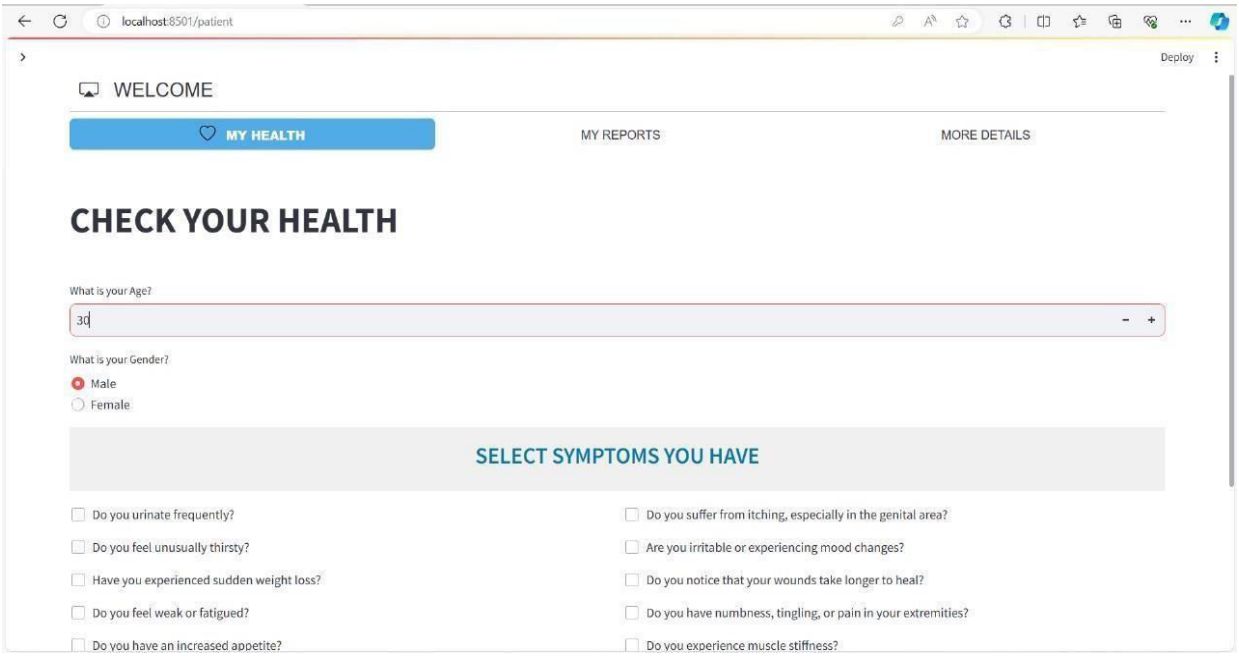


**Fig 5.1 Prediction Page**

### DIABETES NOT DETECTED

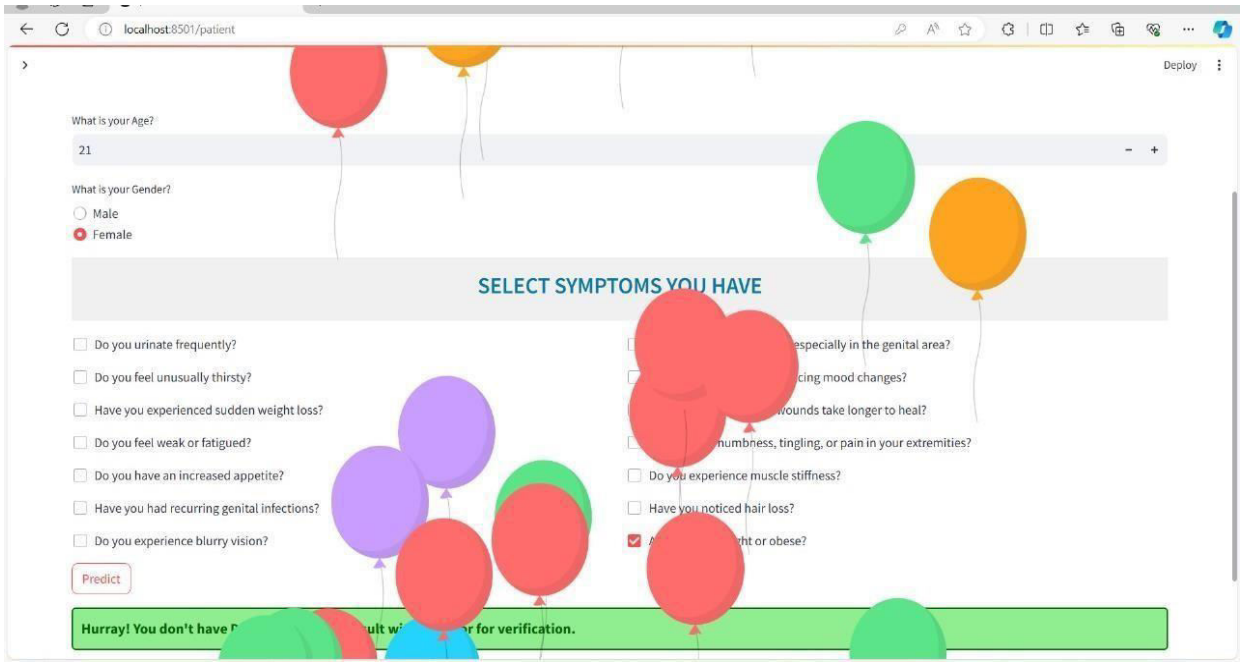


**Fig 5.2 Diabetes Negative Case**

## DIABETES DETECTED

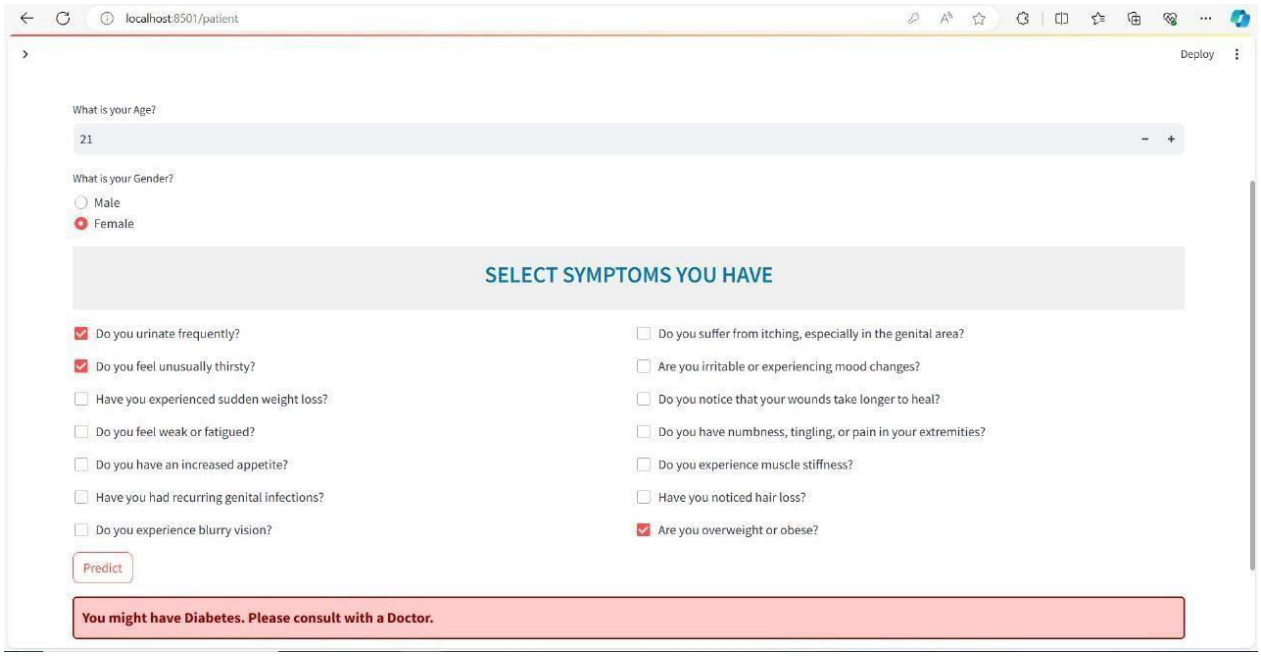


**Fig 5.3 Diabetes Positive Case**

## USER REPORTS

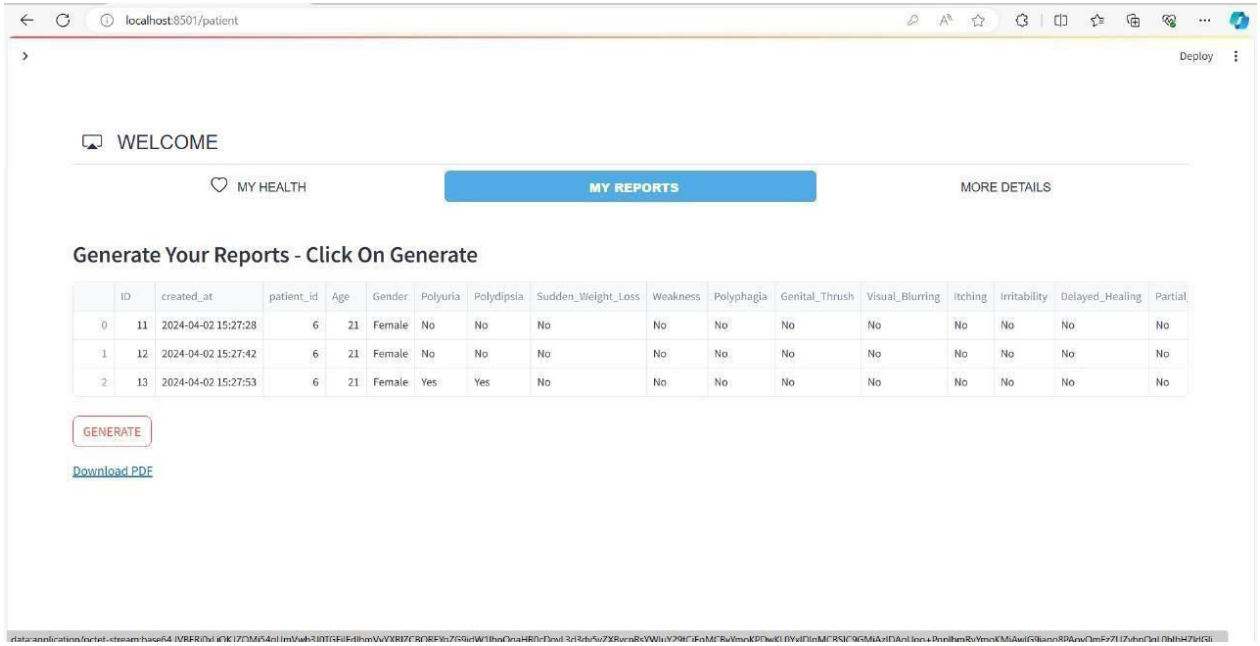


**Fig 5.4 User Reports History**

## 6. CONCLUSION

The Early-stage Diabetic Prediction System is a comprehensive system that uses machine learning techniques to forecast an individual's risk of acquiring diabetes based on their symptoms. The system analyzes a collection of diabetic symptoms and use a variety of machine learning techniques to detect patterns and correlations between the symptoms and the likelihood of getting diabetes.

To ensure a consistent user experience, a dedicated platform for diabetic conditions has been developed, which covers a variety of information such as diabetes types, symptoms, preventive and treatment advice, and diabetes consequences. Allows users to produce anticipated reports for treatment, and doctors to see prior trends in the user's symptoms. The Early-stage Diabetic Prediction System aims to enhance diabetes identification, allowing for early intervention and treatment, lowering the risk of diabetic complications. We seek to empower people to take control of their health and make educated lifestyle choices by giving them access to pertinent information and tools. To summarize, our Diabetic Prediction system and the associated platform for diabetic disorders are a comprehensive solution that uses machine learning techniques to forecast the possibility of acquiring diabetes and gives users with access to relevant information and resources.

## 7. REFERENCES

[1] Ahmed,"Prediction of diabetics empowered with fused Machine Learning", 2022 International Research Journal of Modernization in Engineering Technology and Science, Student, Department Of Computer Engineering, Pune Institute Of Computer Technology, Pune, India.

[2] Daliya, "An Optimized Multivariable Regression Model for Predictive Analysis of Diabetic Disease Progression",2021 Department of Electronics and Communication Engineering, Amrita School of Engineering, AmritaVishwa Vidyapeetham, Bengaluru, India.

[3] HruapingZhou, Raushan Myrzashova and Ruiz Heng, “An enhanced deep neural network (DNN)model for predicting diabetes.”

**AUTHORS PROFILES**

**Mr. Ch. SRINIVASULU REDDY** working as Assistant Professor in Department of CSE , PBR Visvodaya Institute of Technology and Science. He has 20 years of Teachingexperience.

**UDAYAGIRI UDAY VENKATA SAI KUMAR**, B. Tech in the stream of Computer Science and Engineering - Artificial Intelligence in PBR Visvodaya Institute of Technology and Science, Kavali.

**SHAIK AFSANA**, B.Tech in the stream of Computer Science and Engineering- Artificial Intelligence in PBR Visvodaya Institute of Technology and Science, Kavali.

**EDICHERLA NANI**, B. Tech in the stream of Computer Science and Engineering - Artificial Intelligence in PBR Visvodaya Institute of Technology and Science, Kavali.

**ANNAM PAVANI**, B.Tech in the stream of Computer Science and Engineering - Artificial Intelligence in PBR Visvodaya Institute of Technology and Science, Kavali.