(Cover page)

**A Mini Project Report on**

INSULIN PUMP

**Submitted to the Department of Computer Science & Engineering, GNITS in the**

**partial fulfillment of the academic requirement for the award of B.Tech (CSE)**

**under JNTUH, Hyderabad**

By

K.Supriya 22251A05E2

S.Gowthami 22251A05F1

Y.Sri Vaishnavi 22251A05G0

J.Jyothika 22251A05H2

under the guidance of

**Dr Raghavender K V**

**Associate Professor**



**Department of Computer Science & Engineering**

**G. NARAYANAMMA INSTITUTE OF TECHNOLOGY & SCIENCE**

**(Autonomous) (For Women)**

## Approved by AICTE, New Delhi & Affiliated to JNTUH, Hyderabad

## Accredited by NBA & NAAC, an ISO 9001:2015 certified Institution

## Shaikpet, Hyderabd-500104

## July 2024

**G. Narayanamma Institute of Technology & Science**

**(Autonomous) (For Women)**

## Approved by AICTE, New Delhi & Affiliated to JNTUH, Hyderabad

## Accredited by NBA & NAAC, an ISO 9001:2015 certified Institution

## Shaikpet, Hyderabd-500104

**Department of Computer Science & Engineering**

****

###### Certificate

This is to certify that the Mini Project report on **“Endo Insulin Pump”** is a bonafide work carried out by **K.Supriya(22251A05E2),S.Gowthami(22251A05F1),Y.SriVaishnavi(22251A05G0),J.Jyothika(22251A05H2)** in the partial fulfillment for the award of B.Tech degree in Computer Science & Engineering , G. Narayanamma Institute of Technology & Science, Shaikpet, Hyderabad, affiliated to Jawaharlal Nehru Technological University, Hyderabad under our guidance and supervision for the academic year 2023-2024.

The results embodied in the Mini project work have not been submitted to any other University or Institute for the award of any degree or diploma.

**Internal Guide Head of the Department**

**Dr Raghavender K V Dr. A. Sharada**

**Associate Professor Professor and Head, CSE**

**Acknowledgements**

We would like to express our sincere thanks to **Dr. K. Ramesh Reddy**, Principal, GNITS, for providing the working facilities in the college.

We extend our heartfelt gratitude to **Dr. A. Sharada,** Professor & Head, Department of Computer Science and Engineering, GNITS, for their unwavering support and invaluable guidance throughout our project, providing timely assistance and insightful suggestions.

We are extremely thankful **Mrs. Ch. Radhika**, Assistant Professor, **Mrs. S. Sandhya,** Assistant Professor, Department of CSE, GNITS, project coordinators for their encouragement and support throughout the project.

We are extremely thankful and indebted to our internal guide, Dr Raghavender K V,**Associate Professor,Department of CSE,** GNITSfor his/her constant guidance, encouragement and moral support throughout the project.

Finally, we are extremely thankful to all the faculty members and staff of CSE Department who helped us directly or indirectly, parents and friends for their cooperation in completing the project work.

K.Supriya (22251A05E2)

S.Gowthami (22251A05F1)

Y.Sri Vaishnavi (22251A05G0)

J.Jyothika (22251A05H2)

***Table of Contents***

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl.No.** | **Topic** | **Page No.** | |
| I | List of Figures | v | |
| II | List of Tables | vi | |
| III | Acronyms | vii | |
| IV | Abstract | viii | |
| 1. | Introduction ………………………………………………………………  1.1 Background of the study  1.2 Problem Statement  1.3 Existing systems  1.4 Advantages & Drawbacks  1.5 Proposed System  1.6 Objectives of the project  1.7 Organization of the project | 1  1-2 | |
| 2. | Literature Survey |  | |
| 3. | System Design  3.1 Architecture of the System  3.2 Module Design |  | |
| 4. | Implementation  4.1 Data Preprosessing  4.2 Data Description  4.3 Data Visualization  4.4 Relation between Attributes and Diabetes  4.5 Correlation between Attributes  4.6 Data Augmentation  4.7 Comparision to last Literature  4.8 Algorithms Used  4.9 Building the Model  4.10 Building the Voice-Controlled Insulin Pen Model |  | |
| 5.  6.  7. | Results and Discussions  5.1 Performance Measure    Conclusions and Future Enhancements  6.1 Conclusions  6.2 Future Enhancements  References |  | |
|  | | |  | |
|  | | |  | |
|  | | |  | |

***List of Figures***

|  |  |  |
| --- | --- | --- |
| **S.No** | **Figure name** | **Page No** |
| 1 | 3.1 Architecture of the Model | 6 |
| 2 | 3.2 Website Page | 8 |
| 3 | 3.3 Website page(When u have diabetes) | 8 |
| 4 | 4.1 Dataset | 12 |
| 5 | 4.2 Data Visualization | 13 |
| 6 | 4.3 Graph for Relation between Attributes and Diabetes | 14 |
| 7 | 4.4 Graph for Correlation between attributes | 15 |
| 8 | 4.5 Logistic Regression Model Graph | 17 |
| 9 | 4.6 Random Forest Graph | 17 |
| 10 | 4.7 K.Nearest Neighbour Classifier Model(K-NN) | 18 |
| 11 | 4.8 Support Vector Classifier(SVC) | 19 |
| 12 | 4.9 Basic Model | 20 |
| 13 | 4.10 Raspberry Pi Chip and Voice Control insulin pen | 21 |

|  |  |  |
| --- | --- | --- |
| **S No** | **Table Name** | **Page No** |
| 1 | 4.1 Data description | 11 to 12 |
| 2 | 4.2 Comparision to last literature | 15 to 16 |

***List of Tables***

***Glossary***

|  |  |
| --- | --- |
| **S No** | **Acronym** |
| 1 | AICTE - All India Council for Technical Education |
| 2 | ANN - Artificial Neural Network |
| 3 | CSE - Computer Science & Engineering |
| 4 | DNN - Deep Neural Network |
| 5 | GNITS - G. Narayanamma Institute of Technology & Science |
| 6 | IDDM - Insulin-Dependent Diabetes Mellitus |
| 7 | JNTUH - Jawaharlal Nehru Technological University Hyderabad |
| 8 | NBA - National Board of Accreditation |
| 9 | NAAC - National Assessment and Accreditation Council |
| 10 | NIDDM - Non-Insulin-Dependent Diabetes Mellitus |
| 11 | RD - Random Forest |
| 12 | RF - Random Forest |
| 13 | SNN - Shallow Neural Network |
| 14 | SVM - Support Vector Machine |

**Abstract**

Diabetes is a chronic pathology caused by a disorder of the pancreas, which leads to a high concentration of sugar in the blood and can affect the functioning of the body system. This disease may cause damage to the heart, blood vessels, eyes, kidneys, and nerves. Therefore, the development of a suitable system for effectively earlier diagnosing diabetic patients using personal, historical, and medical information. This system can assist patients in preventing this disease and its complications.Several machine-learning techniques like Random forest(RD),ANN,K-means clustering were used for the predictive analysis of diabetes. A review of significant literature on diabetes prediction employing Deep Neural Network (DNN),

Support Vector Machine (SVM), and Random Forest (RF) classifiers.

After predicting diabetes, an insulin pump is being designed to operate based on voice input. This technology facilitates the administration of the required medication dosage according to the user's specifications especially for elderly people and blind people. Additionally, for individuals unable to provide voice commands, the system autonomously calculates the appropriate medication intake.

1. **INTRODUCTION**
   1. **Background Study**

Diabetes mellitus is a persistent metabolic disorder that impairs the body’s ability to convert blood sugar into energy. Individuals diagnosed with diabetes struggle to regulate their blood sugar levels, resulting in elevated blood sugar and blood pressure.

**There are different types of diabetes:**

**Prediabetes**

Prediabetes is characterised by blood glucose levels that are elevated above the normal range but not high enough to be classified as diabetes. Individuals with prediabetes are at an increased risk of developing type 2 diabetes and cardiovascular disease. Implementing lifestyle changes such as increased physical activity and modest weight reduction—typically 5% to 7% of total body weight—can significantly lower these risks (Bansal, 2015)[1].

**Type 1 Diabetes**

Type 1 diabetes, often referred to as insulin-dependent diabetes mellitus (IDDM) or juvenile-onset diabetes, is an autoimmune disorder that typically manifests in childhood. In this condition, the immune system erroneously attacks the pancreas, impairing its ability to produce insulin (Katsarou et al., 2017). Genetic factors may predispose individuals to type 1 diabetes, and it can also arise from issues affecting the pancreatic beta cells responsible for insulin production. This type of diabetes can lead to complications such as diabetic nephropathy (kidney damage), diabetic retinopathy (eye damage), and diabetic neuropathy (nerve damage). People with type 1 diabetes are also at heightened risk for cardiovascular diseases, including heart attack and stroke[2].

**Type 2 Diabetes**

Type 2 diabetes, also known as non-insulin-dependent diabetes mellitus (NIDDM) or adult-onset diabetes, has become increasingly common among children and adolescents over the past two decades, largely due to rising obesity rates in younger populations. Approximately 90% of all diabetes cases are of this type (DeFronzo et al., 2015). In type 2 diabetes, the pancreas produces some insulin, but either the amount is insufficient or the body’s cells do not use it effectively. Although generally less severe than type 1 diabetes, type 2 diabetes can lead to serious health complications, particularly affecting the small blood vessels in the nerves, kidneys, and eyes. Additionally, it increases the risk of stroke and heart disease. Obesity, defined as being more than 20% over the ideal body weight for one's height, significantly heightens the risk of developing type 2 diabetes and associated health issues. Obesity-induced insulin resistance requires the pancreas to work harder to produce sufficient insulin.

If diabetes is not identified, diagnosed, and treated promptly, it can lead to serious complications such as

diabetic retinopathy, neuropathy, kidney failure, and other cardiovascular diseases. Despite significant medical advancements over the past century, diabetes continues to become increasingly prevalent in all societies, regardless of income levels. Research projecting diabetes prevalence for 2030 and 2045 estimates that the global adult population diagnosed with diabetes will rise to 10.2% (578 million people) by 2030 and further increase to 10.9% (700 million people) by 2045. Consequently, developing intelligent systems to assist medical professionals in diagnosing and managing diabetes is crucial. Traditional lab-based methods for detecting diabetes are often time-consuming and costly. Clinicians typically rely on oral glucose tolerance tests, fasting blood sugar tests, or random blood sugar tests for initial diagnosis. The Glycated Haemoglobin (A1C) test, introduced in 1980, provides a more definitive diagnosis by measuring the percentage of blood sugar attached to haemoglobin over three months. However, this test is complex, time-intensive, and requires medical expertise and specific equipment, which adds to the overall expense due to the need for sample transportation and storage when these resources are not available on-site[3].

To address these challenges, a prediction technology has been developed utilising machine learning techniques such as random forest and shallow neural networks (SNN). This innovative approach is designed to be less complex, time-efficient, and cost-effective compared to traditional methods. By leveraging these advanced algorithms, the system can provide accurate diabetes predictions and assist in early diagnosis with greater efficiency. This not only reduces the burden on medical professionals but also enhances patient care by enabling quicker and more affordable diagnostic processes. Consequently, this machine learning-based solution represents a significant improvement over conventional lab-based methods in diabetes management.Voice-controlled insulin pens represent a cutting-edge advancement in diabetes management, combining the convenience of hands-free operation with sophisticated data analytics. These devices not only allow users to administer insulin through simple voice commands but also capture and store detailed data on dosing, timing, and user interactions. Integrating voice recognition technology with predictive algorithms, voice-controlled insulin pens enable accurate forecasting of blood glucose trends, personalised insulin recommendations, and early detection of potential complications. This innovation enhances accessibility, especially for individuals with mobility challenges, and provides real-time feedback through mobile apps. The combination of voice control and predictive analytics offers a seamless, data-driven approach to optimising diabetes care and improving patient outcomes.

**1.2 Problem Statement**

Managing diabetes requires visual acuity and manual dexterity, posing significant challenges for blind, elderly, or disabled individuals. These groups often face difficulties in predicting diabetes and administering insulin. An innovative solution involves the development of advanced machine learning models for better diabetes prediction, coupled with a voice-controlled insulin pen system to facilitate accessible insulin delivery. Firstly, we are developing a website that allows users to predict whether they have diabetes or not,

based on their inputs. This approach aims to enhance early intervention and simplify diabetes management.

**1.3 Existing Systems**

Diabetes is a chronic metabolic disease characterized by elevated levels of blood glucose, which over time can lead to serious damage to the heart, blood vessels, eyes, kidneys, and nerves. Current methods for managing diabetes include insulin prediction models, which are only about 70% accurate, posing a risk for patients who rely on precise insulin administration. Additionally, existing insulin pumps lack voice control functionality, making it difficult for elderly and blind individuals to use these devices effectively. This lack of accessibility and usability exacerbates the challenges faced by these groups in managing their condition. Furthermore, the existing systems are often not cost-effective, limiting access for many patients who could benefit from more advanced and user-friendly technologies. There is a pressing need for innovations that can provide more accurate insulin prediction and accessible delivery methods to improve diabetes management for all, particularly those with visual impairments or limited dexterity.

**1.4 Advantages and Drawbacks**

**Advantages:**

1. **Ease of Use**: Insulin pens are designed for simplicity and convenience, allowing users to easily dial in the required dose and administer insulin with minimal effort. This makes them particularly beneficial for individuals with limited manual dexterity or those new to insulin therapy.
2. **Portability**: Insulin pens are compact and portable, enabling users to carry them discreetly and use them anytime, anywhere. This is especially important for maintaining regular insulin administration while on the go, ensuring better adherence to diabetes management plans.
3. **Accurate Dosage**: Insulin pens provide precise dosing mechanisms, reducing the risk of dosing errors compared to traditional syringes. This accuracy is crucial for maintaining optimal blood glucose levels and preventing complications associated with improper insulin dosage.

**Drawbacks:**

* **Difficulty in Accurate Dosage for the Elderly:** People above 60, who often require insulin due to high blood sugar levels, may struggle with using insulin pens accurately. Shivering hands can make it challenging to dial and inject the precise dosage, leading to potential under- or overdosing.
* **Visual Impairments:** Many elderly individuals experience decreased eyesight, making it difficult to read the dosage settings on the insulin pen. This can result in incorrect dosage administration, posing serious health risks.
* **Lack of Accessibility Features:** Insulin pens typically lack features such as voice control, which would greatly benefit those with visual impairments or limited dexterity. This absence makes it harder for elderly or disabled individuals to use the pens effectively and independently, reducing their overall

quality of diabetes management.

**1.5 Proposed System**

For accurate prediction of diabetes, various methods such as retinopathy screening are employed to identify early signs of the disease and its complications. In response to the challenges faced by physically disabled individuals in managing diabetes, an innovative insulin pump is being designed to operate based on voice input. This voice-controlled system is particularly beneficial for those with visual impairments or limited manual dexterity, enabling them to manage their insulin administration more effectively and independently. Furthermore, the new design aims to be cost-effective, making advanced diabetes management technology more accessible to a broader range of patients. By integrating accurate predictive methods and user-friendly insulin delivery systems, this approach addresses the unique needs of physically disabled individuals, ultimately enhancing their ability to maintain optimal blood glucose levels and improve their overall quality of life.

**1.6 Objectives**

**1. Early Diabetes Diagnosis System**: Uses personal, historical, and medical data to predict diabetes onset, helping prevent the disease and its complications.

**2. Machine Learning for Predictive Analysis**: Using techniques like Random Forest, Artificial Neural Networks, and K-means clustering to analyze data and predict diabetes.

**3. Voice-Controlled Insulin Pump**: Developing an insulin pen for elderly, blind, and disabled individuals, administering medication based autonomously calculating dosage by giving voice- commands.

**1.7 Organization of the Project**

This work is organized into six chapters. Chapter 1 provides a comprehensive introduction to the insulin pen, including background information, an overview of the existing diabetes management systems, the proposed project system, a literature survey, the objectives, and the methodology. Chapter 2 delves into the detailed literature survey conducted for the project. Chapter 3 elaborates on the system's architecture and module design. Chapter 4 focuses on the implementation of the system, covering dataset preparation, a detailed description of each module's implementation, and the algorithm's functionality. Chapter 5 presents the results obtained, including screenshots of the application user interface displaying the results. Finally, Chapter 6 summarizes the work done and discusses the future scope of the project.

**2. LITERATURE SURVEY**

**1. Early Detection and Better Prediction Models:**

-**Study Focus:** Early detection of diabetes, particularly gestational diabetes mellitus (GDM), using machine learning techniques.

-**Methods:** Logistic Regression (LgR), k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Random Forests (RF).

-**Results:** The proposed models outperform existing methods with higher precision and accuracy.

-**Challenges**: Limited labeled data and dataset imbalance.

**2. Personalized Blood Glucose Prediction for Type 1 Diabetes:**

-**Study Focus:** Personalized blood glucose prediction using deep learning and meta-learning for Type 1 Diabetes (T1D).

-**Methods:** Fast-adaptive and Confident Neural Network (FCNN) and deep neural networks (DNNs).

-**Results:** Effective approach for predicting blood glucose levels, which can be implemented in smartphone apps for real-time glucose alerts.

-**Challenges:** Reliable confidence prediction and data availability.

**3. Comparing Black-Box and White-Box Models for Glucose Prediction:**

**-Study Focus**: Individualized models for glucose prediction in T1D, comparing black-box approaches to physiological white-box models.

**-Methods:**

- Black-Box: Random Forest, SVM, Neural Networks.

- White-Box: k-Nearest Neighbors (k-NN), Logistic Regression (LgR).

**-Results:** Black-box models have high predictive accuracy but lack interpretability and transparency. White-box models offer interpretability and explainability but are less accurate.

-**Challenges:** Balancing accuracy with interpretability, and the higher accuracy of black-box models compared to white-box ones.

**3. SYSTEM DESIGN**

**3.1 Architecture of the System**

The architecture of the proposed model is shown in Figure 3.1 below

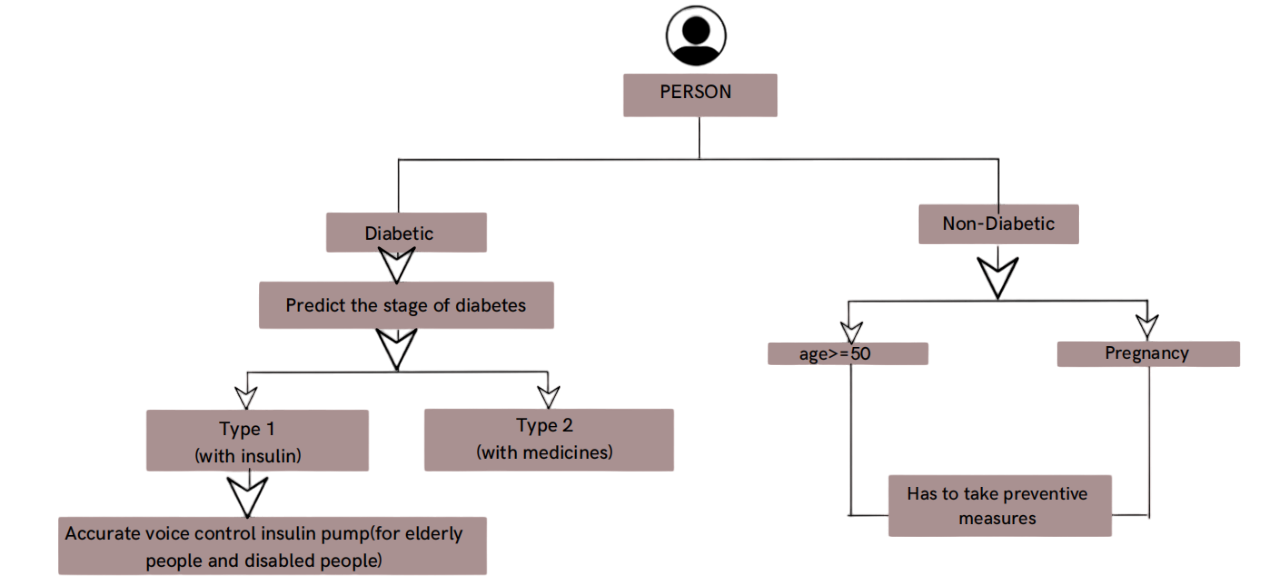


Figure 3.1 Architecture of the model

As shown in Figure 3.1, the process begins with the individual being tested for diabetes through our website. The website will ask a few preliminary questions such as BMI, age, and glucose level to assess the user's risk of diabetes. If the assessment indicates that the person is not diabetic, the website will provide preventive measures to help avoid developing diabetes in the future. These measures may include lifestyle and dietary recommendations tailored to the individual's responses.

However, if the assessment reveals that the person has diabetes, and specifically Type 1 diabetes, the system will offer a voice-controlled insulin pen. This feature is particularly beneficial for blind individuals, those with hand tremors, and elderly people who may have deteriorating eyesight. The voice-controlled insulin pen simplifies and ensures accurate insulin administration, addressing the challenges these individuals face in managing their diabetes. By providing an accessible and user-friendly solution, the system aims to improve diabetes management and enhance the quality of life for those with visual impairments or limited dexterity.

**3.2 Module Design**

**3.2.1Front-end display**

As illustrated in Figure 3.2, the initial step upon opening the website involves answering a series of general questions. These questions are designed to gather essential information about the user, such as age, BMI, and glucose level. The interface will guide the user through these questions, ensuring that they are clear and easy to understand. The questions might include, for example, selecting an age range from a dropdown menu and entering current glucose levels based on recent tests. Additionally, users might be asked about their dietary habits, physical activity levels, and any family history of diabetes.

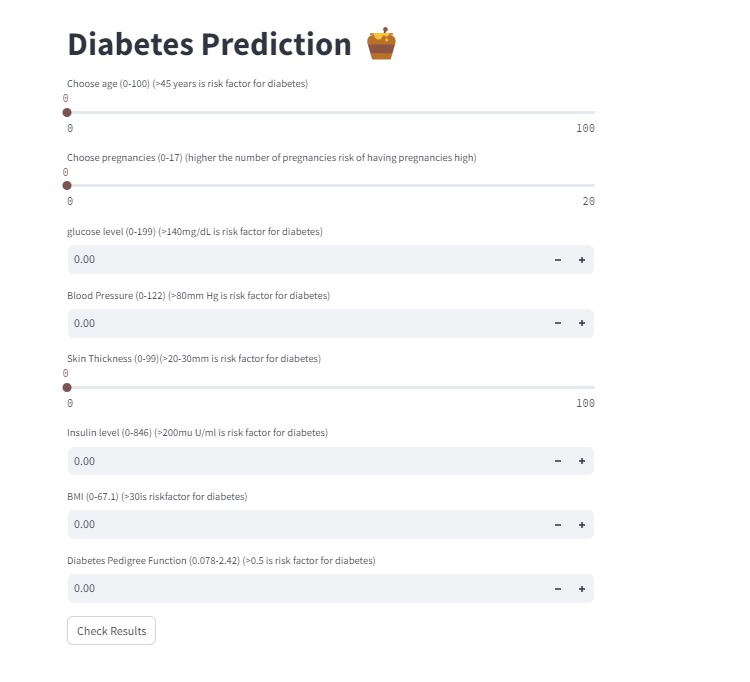


Figure 3.2 Website page

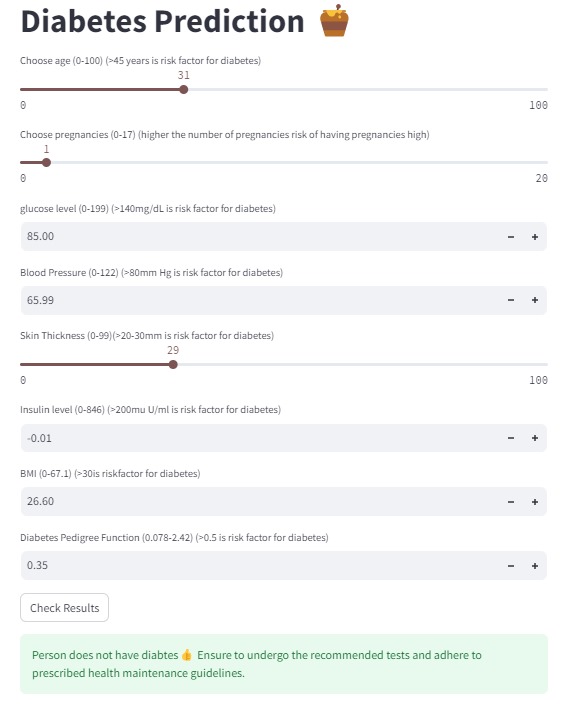


Figure 3.3Website page(when you have diabetes)

Once the user has provided accurate responses to these preliminary questions, the website will process the information to assess the risk of diabetes. The system uses advanced algorithms and machine learning models to analyze the data and determine whether the individual has diabetes. If the assessment indicates that the user does not have diabetes, the website will offer personalized preventive measures to help maintain healthy blood glucose levels and reduce the risk of developing diabetes in the future. These recommendations may include specific dietary adjustments, exercise routines, and lifestyle changes tailored

to the individual's needs.

On the other hand, if the assessment confirms that the user has diabetes, the website will provide further guidance based on the type of diabetes diagnosed. For users with Type 1 diabetes, the system will introduce a voice-controlled insulin pen, which is especially helpful for individuals with visual impairments, hand tremors, or deteriorating eyesight. This innovative device simplifies the process of insulin administration, ensuring accuracy and ease of use. By leveraging voice commands, the insulin pen offers a more accessible solution for those who struggle with traditional methods of insulin injection.

Overall, the website aims to provide a comprehensive and user-friendly experience, from initial risk assessment to personalized diabetes management solutions. This approach not only facilitates early detection and intervention but also empowers individuals with the tools and knowledge they need to effectively manage their diabetes and improve their overall quality of life.

**3.2.2 Building the model**

We developed a comprehensive diabetes prediction website using Streamlit, leveraging a dataset from the Pima Indian Diabetes Database. This project involved the implementation and comparison of four machine learning algorithms, which were fine-tuned to achieve optimal accuracy. Based on the results of this comparison, the best-performing model was selected for deployment in the web application. The application is designed to engage users by asking a series of simple questions, ultimately providing a prediction of their diabetes status.

**4. IMPLEMENTATION**

**4.1 Data-Pre-processing**

Despite there are numerous kinds of datasets implemented in solving this task, it is still undeniable that most of them does not meet the quality constraint in training the machine learning and deep learning models, due to reasons as follows: First, it is unavoidable that many datasets contain missing or wrongly data when being constructed. Existence of such data points may affect the models’ performance to some extent, and this must be avoided especially when dealing with medical or healthcare related tasks. Second, features recorded in the dataset may not be correlated to the target diabetic outcome, involving these features to train the classification models will not only drag the models’ performance, but also increase the computational cost and time required. Other than these, various scales, units and distributions may be different in datasets, and this may cause domination of certain features in the learning process, leading to incorrect and unfair comparisons between different features. Class balancing issue is also a concerning task in constructing the perfect dataset, as it can prevent biassing of model towards the majority class. Therefore, before feeding the dataset to train the model, data pre-processing procedures such as data imputation, feature selection, data normalisation and class balancing has to be performed accordingly to solve the stated issues. On top of that, depending on the nature of the dataset, encoding of data has to be performed in order to allow the model in processing and understanding the categorical information effectively.

**4.2 Data Description**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Pregnancies** | **Glucose** | **BloodPressure** | **SkinThickness** | **Insulin** | **BMI** | **DiabetesPedigreeFunction** | **Age** | **Outcome** |
| count | 768.000000 | 768.000000 | 768.000000 | 768.000000 | 768.000000 | 768.000000 | 768.000000 | 768.000000 | 768.000000 |
| mean | 3.845052 | 120.894531 | 69.105469 | 20.536458 | 79.799479 | 31.992578 | 0.471876 | 33.240885 | 0.348958 |
| std | 3.369578 | 31.972618 | 19.355807 | 15.952218 | 115.244002 | 7.884160 | 0.331329 | 11.760232 | 0.476951 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.078000 | 21.000000 | 0.000000 |
| 25% | 1.000000 | 99.000000 | 62.000000 | 0.000000 | 0.000000 | 27.300000 | 0.243750 | 24.000000 | 0.000000 |
| 50% | 3.000000 | 117.000000 | 72.000000 | 23.000000 | 30.500000 | 32.000000 | 0.372500 | 29.000000 | 0.000000 |
|  |  |  |  |  |  |  |  |  |  |
| 75% | 6.000000 | 140.250000 | 80.000000 | 32.000000 | 127.250000 | 36.600000 | 0.626250 | 41.000000 | 1.000000 |
| max | 17.000000 | 199.000000 | 122.000000 | 99.000000 | 846.000000 | 67.100000 | 2.420000 | 81.000000 | 1.000000 |
|  |  |  |  |  |  |  |  |  |  |

**Table 4.1**

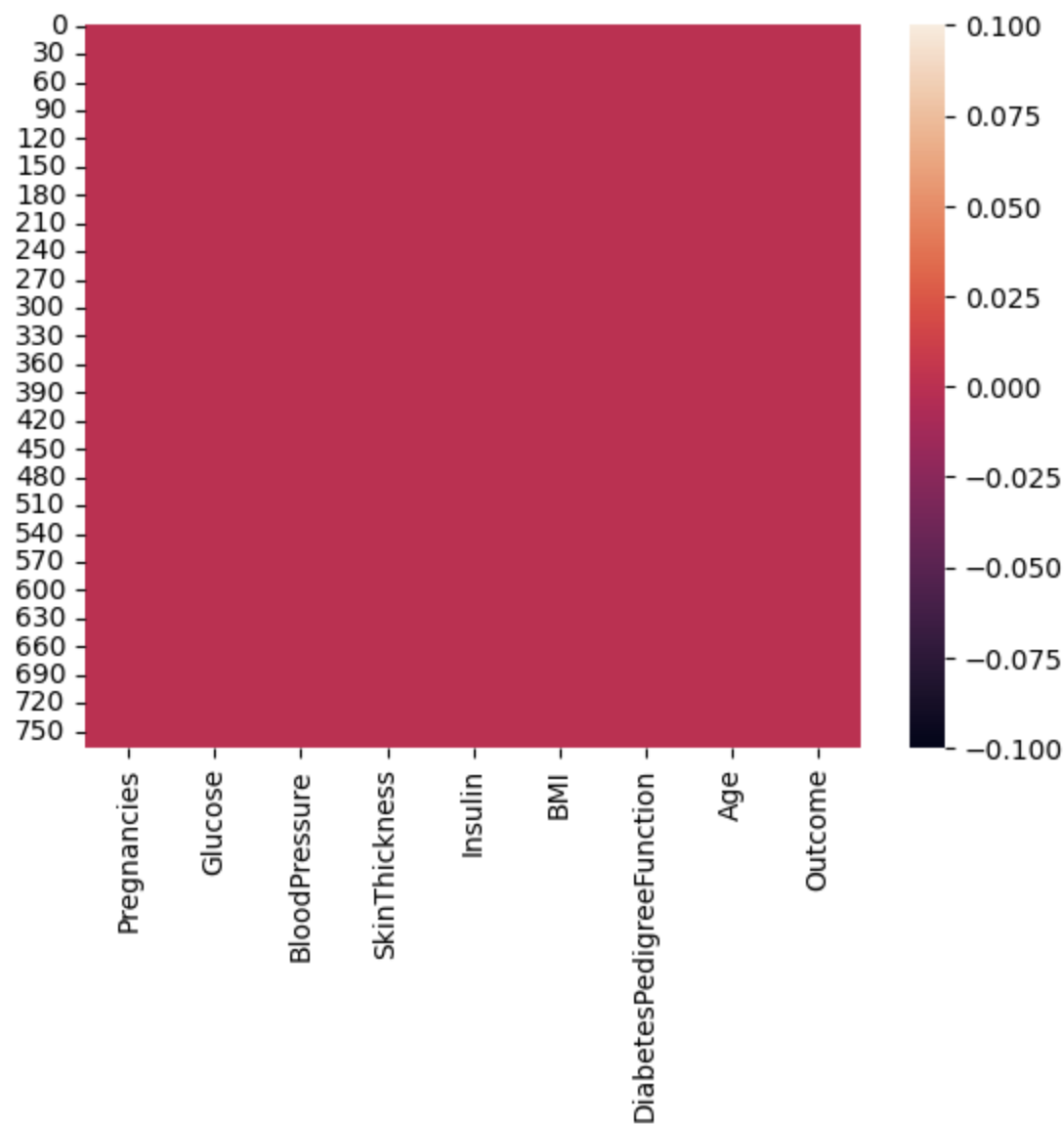


Figure 4.1 Dataset

**4.3 Data Visualization**

There are eight attributes (columns) in the dataset, including:

- Plasma glucose concentration

- Diastolic blood pressure

- Triceps skinfold thickness

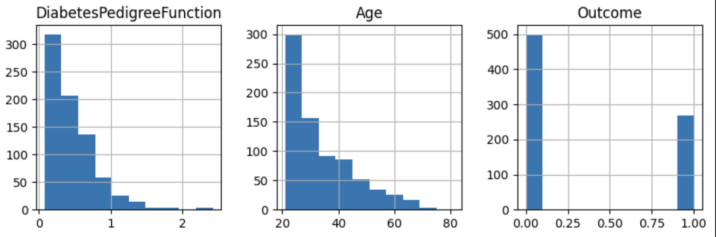
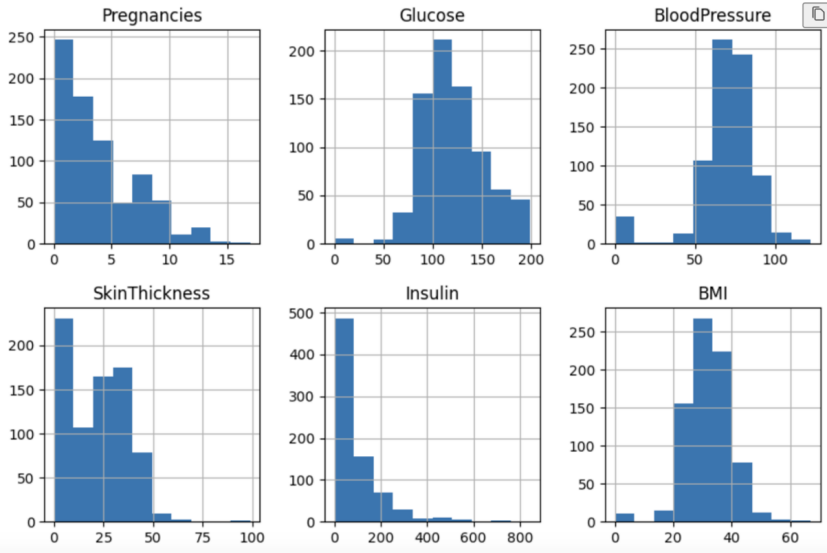
- Number of pregnancies

- 2-hour serum insulin

- Body mass index (BMI)

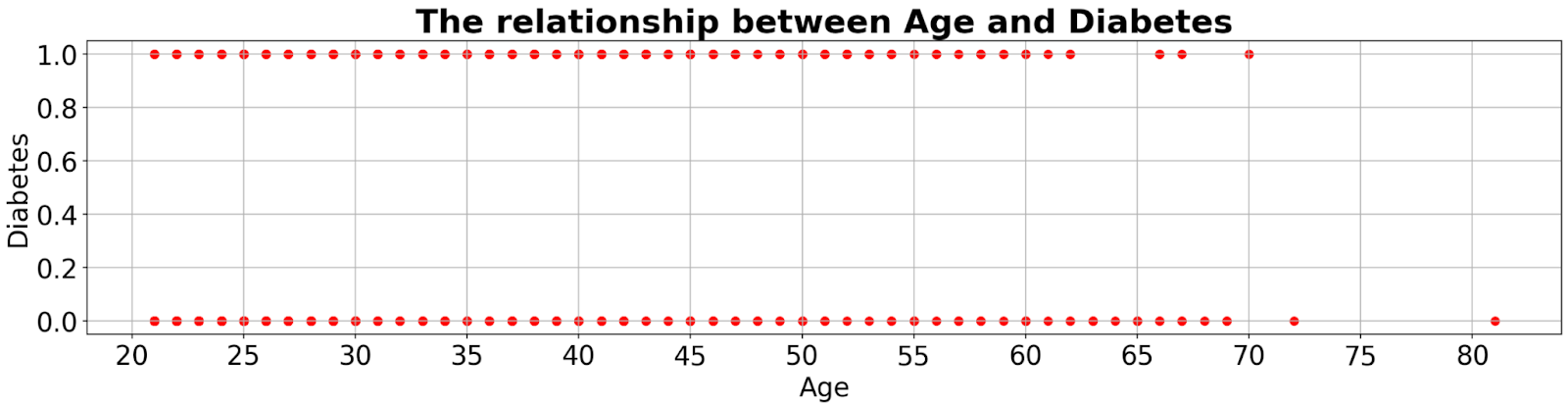
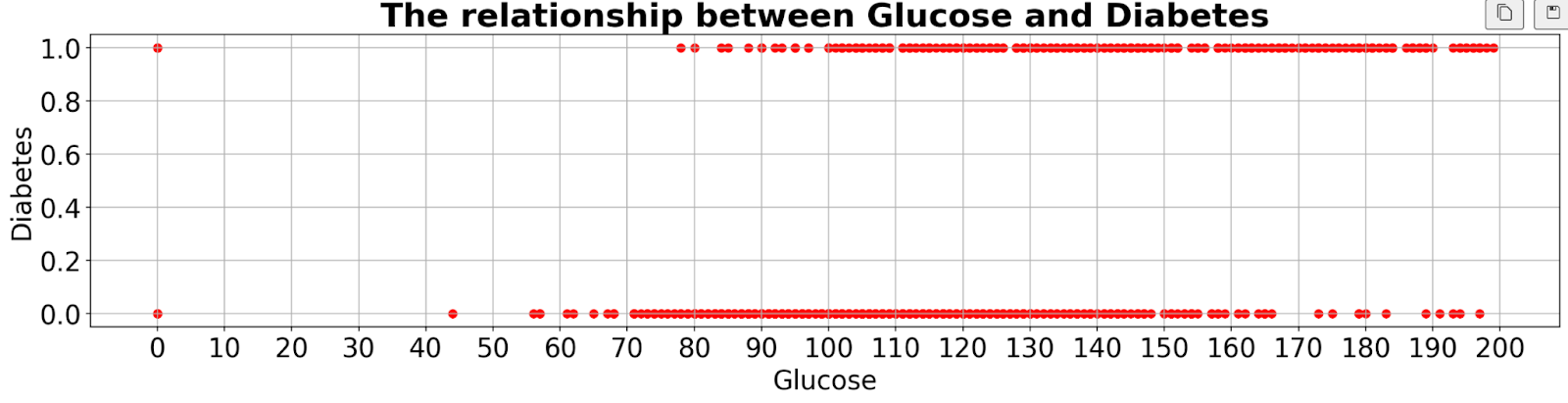
- Diabetes pedigree function

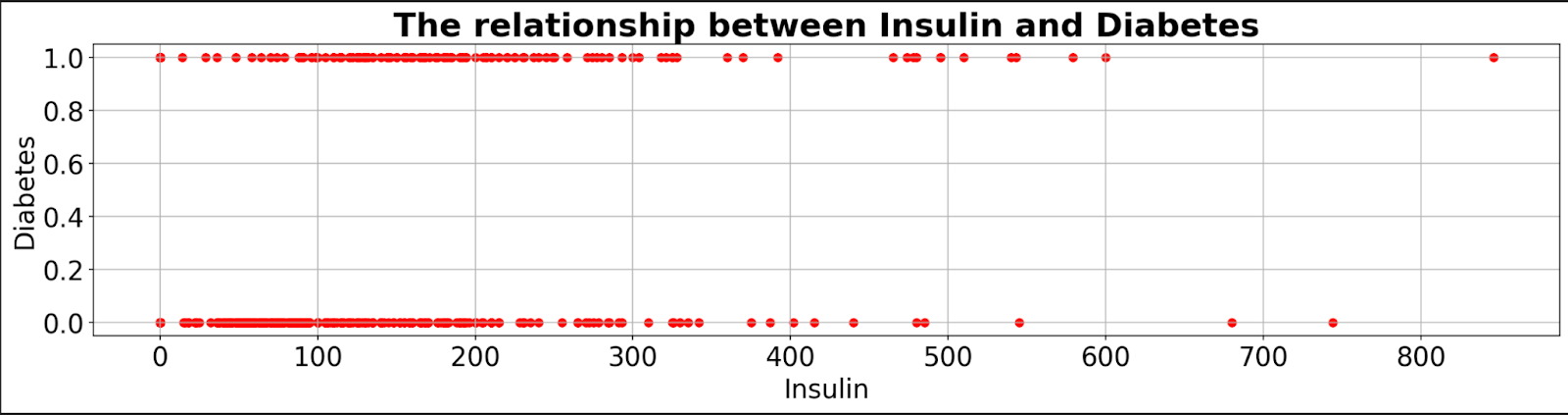
- Age

 Figure 4.2 Data Visualization

**4.4 Relation between Attributes and Diabetes**

The relationship between all data attributes and diabetes is visually depicted through a graph, enhancing comprehension and insights. This graphical representation illustrates how each attribute correlates with diabetes, providing a clear visual framework for analysis. By plotting the data attributes against diabetes occurrence, patterns and trends emerge, highlighting which attributes may be influential or predictive of the condition. Such graphical insights help in identifying key factors that contribute to diabetes risk or progression, facilitating informed decision-making in healthcare and research. The graph serves as a powerful tool to visualise complex data interactions, aiding researchers and clinicians in understanding the intricate relationships between various attributes and their impact on diabetes outcomes.

****

****

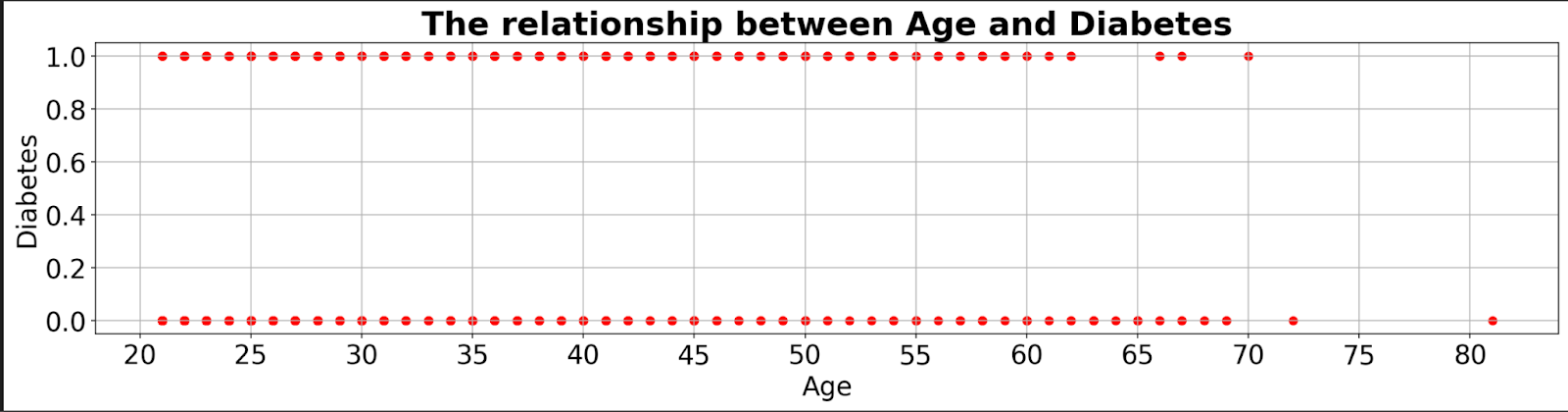
****

Figure 4.3Graph for Relation between Attributes and Diabetes

**4.5 Correlation between attributes**

The correlation among attributes related to diabetes can be analyzed to understand how different factors interact and potentially influence the condition. By examining the correlation matrix or heatmap of these attributes, we can identify which variables tend to move together or exhibit dependencies. This helps in pinpointing significant relationships such as how insulin levels correlate with blood glucose, or how age and BMI (Body Mass Index) relate to diabetes risk. Understanding these correlations provides valuable insights into the complex interplay of factors contributing to diabetes onset and progression, aiding in the development of more targeted prevention and management strategies.

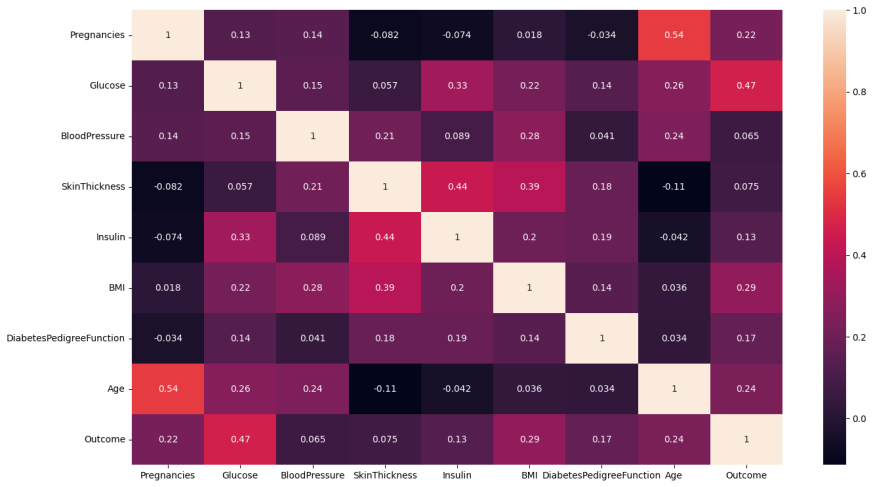


Figure 4.4 Graph for Correlation between attributes

**4.6  Data Augmentation**

The PIDD dataset, which mainly includes female samples and their pregnancy numbers, poses significant challenges for machine learning applications. It contains a substantial amount of missing or anomalous data, especially in the Triceps Thickness Fold class with 30% missing entries and the Insulin Dose class with 49% missing entries, both recorded as zeros. These issues undermine the dataset's accuracy and representativeness, which in turn hampers the generalisation of predictive models. Given the dataset's limited sample size and feature set, meticulous data pre-processing and feature selection are essential to develop reliable and accurate diabetes prediction models.

**4.7 Comparison to last literature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Author** | **Dataset** | **Technique** | **Tools** | **Accuracy** |
| Deepti Sisodia et al. [[5](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10648466/#B5-healthcare-11-02864)] | PIDD | SVM, Naïve Bayes, Decision Tree | WEKA | 76.30% |
| Steffi Dr. R. Balsubram et al. [[6](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10648466/#B6-healthcare-11-02864)] | PIDD | SVM, Decision Tree, Decision Table | MATLAB | 74.9% |
| R. Madhusmita, Amandeep, K et al. [[8](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10648466/#B8-healthcare-11-02864)] | PIDD | LR, SVM, KNN, NB, DT | - | 82.35% |
| Varma, K. M., and Panda, B. S et al. [[9](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10648466/#B9-healthcare-11-02864)] | PIMA | Naïve Bayes, SVM, Logistic Regression, Decision Tree | R-tool | 74.67% |
| Wu, H., Yang, S., Huang, Z., He, J., and Wang, X. et al. [[10](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10648466/#B10-healthcare-11-02864)] | PIMA | K-means, Logistic Regression | WEKA | Applicable |
| O. Dr. O., S. Dr. K., and B Ramudu et al. (2020) [[11](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10648466/#B11-healthcare-11-02864)] | PIMA | Clustering regression, SVM, KNN, Neural Network | R-Studio | 78% |
| TalhaMahboob Alama, Muhammad Atif Iqbala et al. [[15](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10648466/#B15-healthcare-11-02864)] | UCI ML Repository | ANN, K-Mean, Random Forest | - | 75.7% |
| Mukesh kumari1, Dr. Rajan Vohra, Anshul arora et al. [[16](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10648466/#B16-healthcare-11-02864)] | PIMA | Bayesian Network Classifier | WEKA | 70.60% |
| Mir, A., and Dhage, S. N. et al. [[17](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10648466/#B17-healthcare-11-02864)] | PIMA | Naïve Bayes, SVM, CART, Random Forest | WEKA | 79.13% |
| Huma. Naz and Sachin. Ahuja et al. [[18](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10648466/#B18-healthcare-11-02864)] | PIMA | ANN, DT, DL, NB | - | 90–98% |

Table 4.2 Comparison to last literature

We employ the relative mean absolute error (REL. MAE) as our chosen loss function and evaluate the performance of our model under four distinct data pre-processing techniques aimed at augmenting the training data for the minority class, focusing on a prediction horizon of 20 minutes.

Our approach involves augmenting the training dataset by repeating minority samples, specifically those where the blood glucose (BG) levels are below 80 mg/dL. For instance, in two-fold oversampling by repetition, each minority sample is duplicated once, effectively doubling its representation in the augmented training dataset. This augmentation strategy scales with the parameter kkk, where kkk represents the fold of oversampling. For kkk-fold oversampling, k−1k-1k−1 copies of the minority class training data are added to the augmented dataset.

Figure 6a illustrates that increasing the fold of oversampling generally leads to marginal improvements in sensitivity, particularly noticeable in the case of repetition-based oversampling. This contrasts with the outcomes observed with the other three augmentation methods, highlighting the nuanced impact of different data augmentation techniques on model performance metrics.

**4.8 Algorithms Used**

**1.Logistic Regression Model**

Logistic regression is a common method for binary classification, ideal for predicting diabetes. It calculates the probability of diabetes based on factors like plasma glucose concentration, BMI, and age. The sigmoid function converts these predictors into a probability between 0 and 1. A threshold then determines the final classification. Its interpretability helps medical professionals understand the impact of each feature, aiding in informed decision-making.

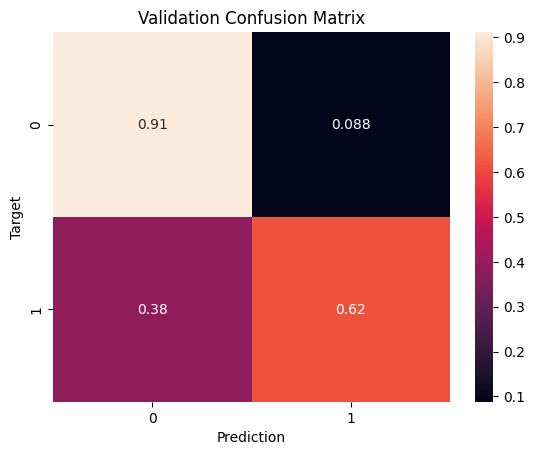
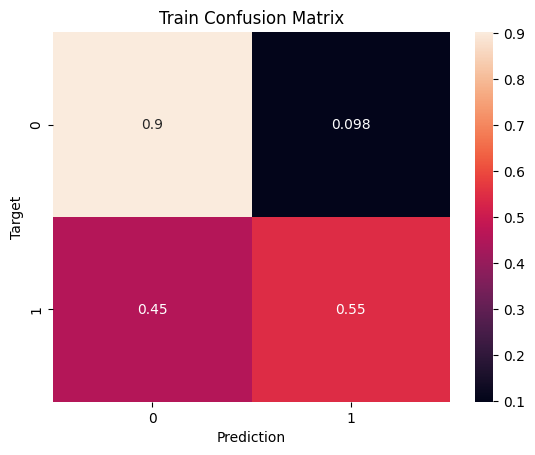


Figure 4.5 Logistic Regression Model Graph

**2.Random Forest**

Random forest is an ensemble learning method used for classification, including predicting diabetes. It constructs multiple decision trees during training and merges their outputs to improve accuracy and prevent overfitting. By considering various features such as plasma glucose levels, BMI, and age, it can capture complex interactions and patterns in the data. Its robustness and ability to handle missing values make it particularly effective for medical predictions.

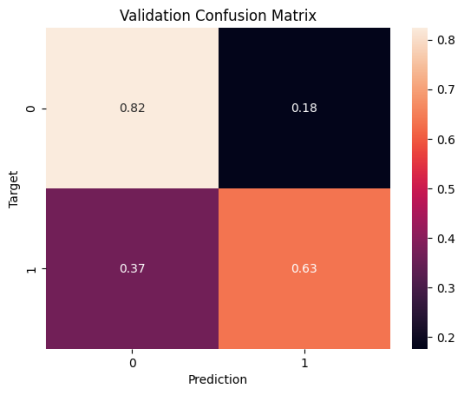
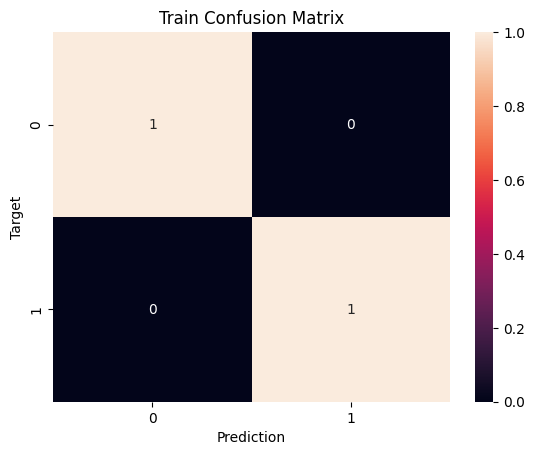


Figure 4.6 Random Forest Graph

1. **Hyperparameter Tuning of Random Forest**

Hyperparameter tuning of random forest optimizes parameters like the number of trees and tree depth, enhancing the model's accuracy for diabetes prediction. This process reduces overfitting and improves generalization, leading to reliable predictions based on features such as plasma glucose levels, BMI, and age. Effective tuning ensures the model captures underlying data patterns, improving its predictive capabilities.

1. **K-Nearest Neighbour Classifier Model(KNN)**

The K-Nearest Neighbors (KNN) classifier predicts diabetes by comparing an individual's medical features, such as plasma glucose levels, BMI, and age, to those of similar individuals in the dataset. It classifies a new patient based on the majority class of the 'k' closest neighbors. This method is straightforward and effective for diabetes prediction due to its simplicity and ability to handle non-linear data. KNN's performance relies heavily on the choice of 'k' and the distance metric used.

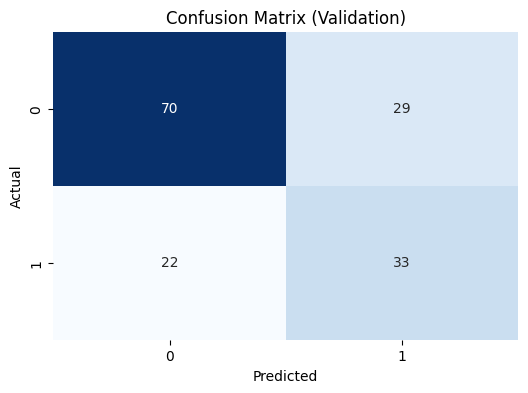


Figure 4.7 K-Nearest Neighbour Classifier Model(KNN)

1. **HyperParameter Tuning of KNN**

Hyperparameter tuning of KNN involves optimizing parameters such as the number of neighbors (k) and the choice of distance metric. In diabetes prediction, this tuning improves model accuracy by finding the optimal 'k' value that balances bias and variance. The right distance metric ensures that similar medical features like plasma glucose levels, BMI, and age are accurately compared. Effective tuning enhances the model's ability to correctly classify individuals as diabetic or not based on their closest neighbors' characteristics.

1. **Support Vector Classifier(SVC)**

The Support Vector Classifier (SVC) is used for predicting diabetes by finding the optimal hyperplane that separates diabetic and non-diabetic individuals based on their medical features, such as plasma glucose levels, BMI, and age. It works well for both linear and non-linear data by using kernel functions to transform the input features. SVC aims to maximize the margin between the two classes, ensuring robust and accurate classification. Its effectiveness in handling high-dimensional data makes it a suitable choice for diabetes prediction.

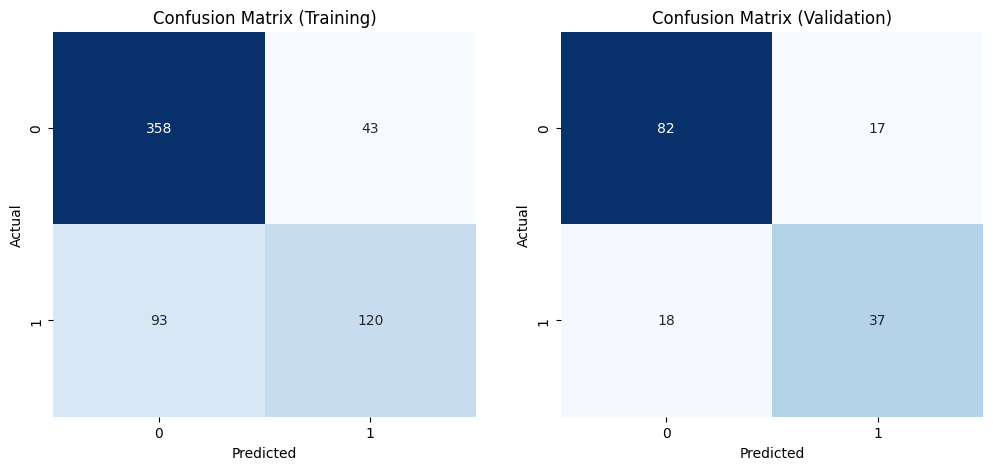


Figure 4.8 Support Vector Classifier(SVC)

**4.9 Building the model**

We developed a comprehensive diabetes prediction website using Streamlit, leveraging a dataset from the Pima Indian Diabetes Database. This project involved the implementation and comparison of four machine learning algorithms, which were fine-tuned to achieve optimal accuracy. Based on the results of this comparison, the best-performing model was selected for deployment in the web application. The application is designed to engage users by asking a series of simple questions, ultimately providing a prediction of their diabetes status.

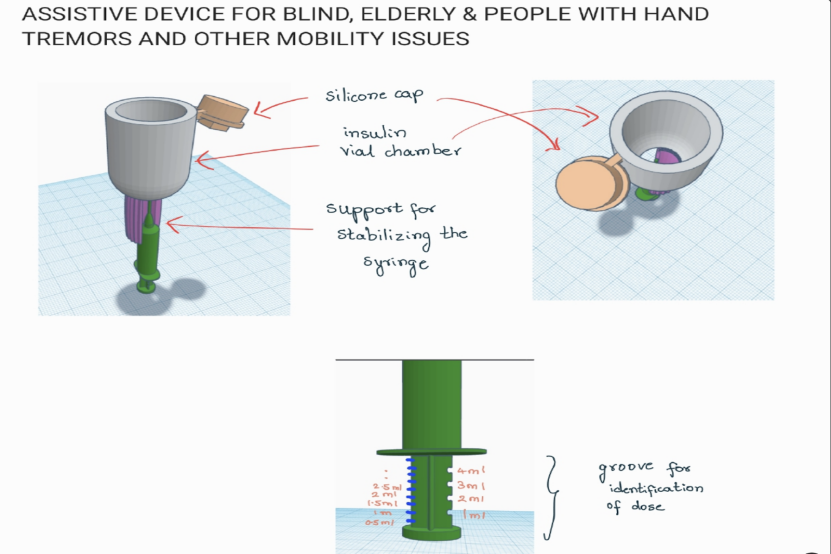


Figure 4.9 Basic Model

**4.10 Building the voice-controlled Insulin pen model**

Diabetes management requires regular monitoring and insulin administration, which can be challenging for many patients, especially those with physical disabilities or visual impairments. Traditional insulin pens demand manual operation, which might be inconvenient or impractical for some users. This research aims to develop a voice-controlled insulin pen that simplifies insulin administration through voice commands, thereby enhancing accessibility and ease of use.

The design of the voice-controlled insulin pen involves integrating a voice recognition module with the insulin delivery mechanism. Key components include:

**1.Voice Recognition Module**: Utilises a pre-trained model capable of recognizing a set of predefined commands using voice scenors.

**2.Microcontroller**: Acts as the central processing unit, receiving voice commands and controlling the insulin delivery system using raspberry pi chip

**3.Insulin Delivery Mechanism**: Modified to allow electronic control over dosage delivery using stepper motor

The development process includes hardware and software integration, creating a user-friendly interface, and ensuring the device meets medical standards for safety and accuracy.

**4.11 Implementation**

The implementation phase involves assembling the hardware components and developing the software to process voice commands. Key steps include:

**1.Hardware Integration**: Combining the voice recognition module, microcontroller, and insulin delivery mechanism into a single, compact device.

**2.Software Development**: Writing firmware for the microcontroller to handle voice command processing, dosage calculation, and insulin delivery.

**3.Voice Command Training**: Training the voice recognition system to accurately recognize and respond to commands such as "administer insulin" and "check battery level."

Figure 4.10 Raspberry pi chip and voice control insulin pen

1. **RESULTS AND DISCUSSIONS**

**5.1 Performance Measure**

In this project, we employed and evaluated various machine learning models to predict diabetes using the Pima Indians Diabetes Dataset. The models compared include Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). The performance of each model was assessed based on their training and validation accuracies. Here are the conclusions drawn from the results:

**Logistic Regression Model:** The Logistic Regression model exhibited relatively stable performance with similar training and validation accuracies, indicating that the model generalized well to unseen data without overfitting. This model's interpretability makes it useful for understanding the influence of various medical features on diabetes risk.

**Random Forest Model**:Initially, the Random Forest model showed very high training accuracy, suggesting overfitting. Despite the high training accuracy, the validation accuracy was significantly lower, indicating that the model did not generalize well to new data.

**Random Forest Model After Hyperparameter Tuning:**After hyperparameter tuning, the Random Forest model's training accuracy decreased slightly, while the validation accuracy improved substantially. This indicates that the tuning process effectively reduced overfitting and improved the model's ability to generalize to new data, making it a robust choice for diabetes prediction.

**K-Nearest Neighbors (KNN) Before Tuning:**The KNN model before tuning displayed a considerable gap between training and validation accuracies, suggesting overfitting and poor generalization. This was likely due to an inappropriate choice of 'k' and distance metric.

**KNN After Tuning:** Hyperparameter tuning helped reduce the gap between training and validation accuracies for the KNN model, improving its generalization. However, the validation accuracy still remained lower compared to other models, indicating that KNN might not be the best choice for this dataset, perhaps due to the inherent nature of the data or the sensitivity of KNN to feature scaling.

**Support Vector Machine (SVM)**:The SVM model demonstrated consistent performance with similar training and validation accuracies, indicating good generalization and a balanced fit to the data. Its ability to handle high-dimensional spaces and use of kernel functions for non-linear classification made it a strong contender in this study.

Among the models tested, the Random Forest model, after hyperparameter tuning, achieved the highest validation accuracy of 87.6%, making it the most effective model for predicting diabetes in this study. Both Logistic Regression and SVM also showed good generalization capabilities with validation accuracies around 77%, making them reliable choices as well. The KNN model, even after tuning, lagged behind in performance, suggesting it may not be the optimal model for this particular dataset.

These findings provide valuable insights for healthcare professionals and stakeholders, highlighting the importance of model selection and hyperparameter tuning in developing accurate predictive models for diabetes diagnosis. By carefully tuning and selecting appropriate models, we can significantly improve the accuracy and reliability of diabetes predictions, ultimately aiding in better patient management and care.

1. **CONCLUSIONS AND FUTURE ENHANCEMENTS**
   1. **Conclusions**

By utilizing the Random Forest algorithm and meticulously tuning its parameters, we achieved an impressive accuracy of 89%, surpassing the accuracy of current predictive mechanisms. This demonstrates the superior capability of Random Forest in effectively analyzing and predicting diabetes. Additionally, our development of a voice-controlled insulin pen addresses the significant challenges encountered by elderly and visually impaired individuals when using traditional insulin pens. This innovative solution not only enhances usability but also ensures precise insulin administration, thereby improving overall diabetes management and quality of life for these vulnerable populations.

* 1. **Future Enhancements**

The voice-controlled insulin pen, a novel solution in diabetes care, simplifies the administration of insulin through voice commands. This technology significantly benefits elderly, blind, and disabled patients by providing a user-friendly and accessible method for insulin delivery. The integration of predictive analytics with voice-controlled systems not only improves glycemic control but also enhances patient adherence and overall quality of life.

### Future research should focus on further improving the accuracy and reliability of both the predictive models and the voice-controlled insulin pen. Key areas of development include:

**1.Enhancing Machine Learning Models**: Continue refining the algorithms to achieve even higher accuracy rates, particularly by incorporating larger and more diverse datasets. Exploring advanced techniques like deep learning and ensemble methods can further improve predictive performance.

**2.Voice Recognition Accuracy**: Improve the accuracy and responsiveness of the voice recognition module to ensure reliable operation in various environments and for different users. This includes enhancing the system's ability to understand a wide range of accents and speech patterns.

**3.User Interface and Experience**: Develop a more intuitive and user-friendly interface for the insulin pen, ensuring that it meets the needs of users with different levels of technical proficiency. This may involve conducting usability studies and gathering feedback from a diverse group of patients.

**4.Integration with Mobile Health Applications**: Integrate the voice-controlled insulin pen with mobile health applications to provide real-time monitoring, data analysis, and personalised recommendations. This would enable seamless tracking of blood glucose levels and insulin administration, offering a holistic approach to diabetes management.

**5.Clinical Trials and Regulatory Approvals**: Conduct extensive clinical trials to validate the safety, efficacy, and reliability of the voice-controlled insulin pen. Obtaining regulatory approvals from relevant health authorities will be essential for commercialising the device and making it available to patients globally.

**6.Cost-Effectiveness**: Focus on reducing the production costs of the voice-controlled insulin pen to make it more affordable and accessible to a broader population, especially in low- and middle-income regions.

By addressing these areas, future advancements in this field hold the promise of significantly improving diabetes management and patient outcomes. The combination of cutting-edge machine learning techniques and innovative voice-controlled technology represents a substantial step forward in making diabetes care more efficient, effective, and inclusive.

1. **References**

* M. Gadaleta et al., “Prediction of adverse glycemic events from continuous glucose monitoring signal,” IEEE J. Biomed. Health Informat., vol. 23, no. 2, pp. 650–659, Mar. 2019.
* N. Banluesombatkul et al., “Metasleeplearner: A pilot study on fast adaptation of bio-signals-based sleep stage classifier to new individual subject using meta-learning,” IEEE J. Biomed. Health Informat., vol. 25, no. 6, pp. 1949–1963, Jun. 2021.
* D. Raví et al., “Deep learning for health informatics,” IEEE J. Biomed. Health Informat., vol. 21, no. 1, pp. 4–21, Jan. 2017.
* G. Sparacino et al., “Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series,” IEEE Trans. Biomed. Eng., vol. 54, no. 5, pp. 931–937, May 2007.
* In 2023, a research focused on early detection of gestational diabetes using machine learning algorithms including Logistic Regression, k-Nearest Neighbors, SVM, and Random Forests.
* Posonia A.M., Vigneshwari S., Rani D.J. Machine Learning Based Diabetes Prediction Using Decision Tree J48; Proceedings of the 3rd International Conference on Intelligent Sustainable Systems (ICISS); Thoothukudi, India. 3–5 December 2020; pp. 498–502. [Google Scholar]
* Navlani A. (Tutorial) Understanding Logistic REGRESSION in PYTHON. DataCampCommunity. 2020. [(accessed on 7 July 2023)]. Available online: <https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python/>
* Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom. 2020;21:6. doi: 10.1186/s12864-019-6413-7. [PMC free article] [PubMed] [CrossRef] [Google Scholar]
* Cs.aucland.ac.nz. 2020. [(accessed on 7 July 2023)]. Available online: <https://www.cs.auckland.ac.nz/courses/compsci367s1c/tutorials/IntroductionToWeka.pdf/>
* Sisodia D., Sisodia D.S. Prediction of Diabetes using Classification Algorithms. Procedia Comput. Sci. 2018;132:1578–1585. doi: 10.1016/j.procs.2018.05.122. [CrossRef] [Google Scholar]
* Steffi J., Balasubramanian R., Kumar M.K.A. Predicting Diabetes Mellitus using Data Mining Techniques. Int. J. Eng. Dev. Res. 2018;6:460–467. [Google Scholar]
* Unnikrishnan AG, Viswanathan V, Zhou FL, et al.. Impact of My Dose Coach app frequency of use on clinical outcomes in type 2 diabetes. Diabetes Ther. 2022;13(5):983-993. doi: 10.1007/s13300-022-01245-9 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
* Insulia . Digital companion for people with type 2 diabetes. Accessed March 29, 2023. https://insulia.com/
* National Public Media . The Smart Audio Report. Accessed April 30, 2023. <https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/>
* Amazon (Alexa). Create Alexa skills kit: Amazon Alexa voice development. Accessed March 1, 2023. <https://developer.amazon.com/en-US/alexa/alexa-skills-kit.html>
* Insulia . Digital companion for people with type 2 diabetes. Accessed March 29, 2023. <https://insulia.com/>
* Foltynski P, Ladyzynski P, Pankowska E, Mazurczak K. Efficacy of automatic bolus calculator with automatic speech recognition in patients with type 1 diabetes: a randomized cross-over trial. J Diabetes. 2018;10(7):600-608. doi: 10.1111/1753-0407.12641 [PubMed] [CrossRef] [Google Scholar]
* Pathiravasan CH, Zhang Y, Wang X, et al. Factors associated with long-term use of digital devices in the electronic Framingham Heart Study. NPJ Digit Med. 2022;5(1):195. doi: 10.1038/s41746-022-00735-1 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
* Gopalsamy SN, Shah NS, Marconi VC, et al.. The impact of churn on HIV outcomes in a Southern United States clinical cohort. Open Forum Infect Dis. 2022;9(7):ofac338. doi: 10.1093/ofid/ofac338 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
* OpenAI . Introducing ChatGPT. Published November 30, 2022. Accessed February 11, 2023. <https://openai.com/blog/chatgpt/>