# MINI PROJECT REPORT

*On*

**GLUCO-SENSE AI**

*Submitted in partial fulfilment for the award of*

*degree Of*

***Master of Computer Applications***

*By*


**HARI C S**

**(MLM2024MCA-2029)**


Under the Guidance of


**Ms. SREELEKSHMI K S**

(Assistant Professor, Department. of Computer Applications)


**DEPARTMENT OF COMPUTER APPLICATIONS**

**MANGALAM COLLEGE OF ENGINEERING, ETTUMANOOR**

***(Affiliated to APJ Abdul Kalam Technological University)***

**OCTOBER 2025**

# VISION

To become a centre of excellence in computer applications,competent in the global ecosystem with technical knowledge,innovation with a sense of social commitment.

# MISSION

- To serve with state of the art education,foster advanced research and cultivate innovation in the field of computer applications.
- To prepare learners with knowledge skills and critical thinking to excel in the technological landscape and contribute positively to society.

## Program Educational Objectives

- PEO I :Graduates will possess a solid foundation and in-depth understanding of computer applications and will be equipped to analyze real-world problems, design and create innovative solutions, and effectively manage and maintain these solutions in their professional careers.
- PEO II: Graduates will acquire technological advancements through continued education, lifelong learning and research, thereby making meaningful contributions to the field of computing.
- PEO III: Graduates will cultivate team spirit, leadership, communication skills, ethics, and social values, enabling them to apply their understanding of the societal impacts of computer applications effectively.

## Program Specific Outcomes

- **PSO I**: Apply advanced technologies through innovations to enhance the efficiency of design development.
- **PSO II**: Apply the principles of computing to analyze, design and implement sustainable solutions for real world challenges.

# MAPPING OF PO-PSO-SDG

## 1. MAPPING WITH PROGRAM OUTCOMES (POs):-

| SL.NO | Pos ADDRESSED | RELEVANCE TO PROJECT |
|---|---|---|
| 1 | PO1 | applied core knowledge of mathematics, data science, and machine learning to build the diabetes prediction model. |
| 2 | PO3 | designed and developed a complete solution that addresses a real-world health problem, predicting diabetes and connecting patients with doctors. |
| 3 | PO4 | conducted research on existing diabetes datasets (like PIMA dataset) to understand influential medical parameters (e.g., glucose level, BMI, insulin) |

**<u>LIST OF PROGRAM OUTCOMES (POs):</u>**

**PO1 – Engineering Knowledge**: Apply knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to solve complex engineering problems.

**PO2 – Problem Analysis**: Identify, formulate, research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

**PO3 – Design/Development of Solutions**: Design solutions for complex engineering problems and design systems, components, or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.

**PO4 – Conduct Investigations of Complex Problems**: Use research-based knowledge and research methods including design of experiments, analysis, and interpretation of data, and synthesis of information to provide valid conclusions.

**PO5– Modern Tool Usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools, including prediction and modeling, to complex engineering activities with an understanding of the limitations.

**PO6 – The Engineer and Society**: Apply reasoning informed by contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to professional engineering practice.

**PO7 – Environment and Sustainability**: Understand the impact of professional engineering solutions in societal and environmental contexts and demonstrate knowledge of, and need for sustainable development.

**PO8 – Ethics** : Apply ethical principles and commit to professional ethics and responsibilities and norms of engineering practice.

**PO9 – Individual and Team Work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

**PO10 – Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

**PO11– Project Management and Finance**: Demonstrate knowledge and understanding of engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

**PO12 – Lifelong Learning**: Recognize the need for, and have the ability to engage in independent and life-long learning in the broadest context of technological change.

**2. MAPPING WITH PROGRAM SPECIFIC OUTCOMES (PSOs):**

| SL.NO | PSOs ADDRESSED | RELEVANCE TO PROJECT |
|---|---|---|
| 1 | **PSO 2** | applied core computing principles such as data analysis, algorithm design, and software implementation to solve a real-world health problem early diabetes detection |

**LIST OF PROGRAM SPECIFIC OUTCOMES (PSOs):**

**PSO 1**: Apply advanced technologies through innovations to enhance the efficiency of design development.

**PSO 2**: Apply the principles of computing to analyze, design and implement sustainable solutions for real world challenges.

**3. MAPPING WITH SUSTAINABLE DEVELOPMENT GOALS (SDGs):**

| SDG NO | SDGs ADDRESSED | RELEVANCE TO PROJECT |
|---|---|---|
| SDG 3 | **Good Health and Well-Being** | Ensure healthy lives and promote well-being for all at all ages. |
| SDG 9 | **Industry, Innovation, and Infrastructure** | It promotes innovation through the application of AI and ML technologies in healthcare systems |

**SUSTAINABLE DEVLOPMENT GOALS (SDGs):**

**SDG 1 – No Poverty**-End poverty in all its forms everywhere.

**SDG 2 – Zero Hunger**-End hunger, achieve food security and improved nutrition, and promote sustainable agriculture.

**SDG 3 – Good Health and Well-Being**-Ensure healthy lives and promote well-being for all at all ages.

**SDG 4 – Quality Education**-Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.

**SDG 5 – Gender Equality**-Achieve gender equality and empower all women and girls.

**SDG 6 – Clean Water and Sanitation**-Ensure availability and sustainable management of water and sanitation for all.

**SDG 7 – Affordable and Clean Energy**-Ensure access to affordable, reliable, sustainable, and modern energy for all.

**SDG 8 – Decent Work and Economic Growth**-Promote sustained, inclusive, and sustainable economic growth, full and productive employment, and decent work for all.

**SDG 9 – Industry, Innovation, and Infrastructure**-Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation.

**SDG 10 – Reduced Inequality**-Reduce inequality within and among countries.

**SDG 11 – Sustainable Cities and Communities**-Make cities and human settlements inclusive, safe, resilient, and sustainable.
**SDG 12 – Responsible Consumption and Production**-Ensure sustainable consumption and production patterns.
**SDG 13 – Climate Action**-Take urgent action to combat climate change and its impacts.
**SDG 14 – Life Below Water**-Conserve and sustainably use the oceans, seas, and marine resources.
**SDG 15 – Life on Land** -Protect, restore, and promote sustainable use of terrestrial ecosystems, manage forests sustainably, combat desertification, halt and reverse land degradation, and halt biodiversity loss.
**SDG 16 – Peace, Justice, and Strong Institutions**- Promote peaceful and inclusive societies, provide access to justice for all, and build effective, accountable, and inclusive institutions.
**SDG 17 – Partnerships for the Goals** -Strengthen the means of implementation and revitalize the global partnership for sustainable development.

**MANGALAM COLLEGE OF ENGINEERING, ETTUMANOOR**

**DEPARTMENT OF COMPUTER APPLICATION**

**OCTOBER 2025**

विश्वं शास्त्रे प्रतिष्ठितम्।

# ***DECLARATION***

*I hereby certify that the work which is being presented in the project entitled "GLUCO-SENSE AI" submitted in the* ***DEPARTMENT OF COMPUTER APPLICATIONS*** *is an authentic record of my own work carried under the supervision of* ***Ms. SREELEKSHMI K S, ASSISTANT PROFESSOR*** *This study has not been submitted to any other institution or university for the award of any other degree. This report has been checked for plagiarism by the college and the similarity index is within permissible limits set by the college.*

*Name & Signature of Student*

*Date:*

*Place:*

**MANGALAM COLLEGE OF ENGINEERING, ETTUMANOOR**

**DEPARTMENT OF COMPUTER APPLICATION**

**OCTOBER 2025**

विश्वं शास्त्रे प्रतिष्ठितम्।

## ***CERTIFICATE***

*This is to certify that the Project titled "**GLUCO-SENSE AI**" is the bonafide record of the work done by **HARI C S (MLM2024MCA-2029)** of MCA in Computer Science towards the partial fulfilment of the requirement for the award of **MASTER OF COMPUTER APPLICATIONS by APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**, during the academic year 2025-26.*

**Internal Examiner**

**Project Coordinator**

**Ms. Banu Sumayya S**

**Assistant Professor**

**Department of Computer Applications**

**Project Guide**

**Ms. SREELEKSHMI K S**

**Assistant Professor**

**Department of Computer Applications**

**Head of the Department**

**Ms. Divya S B**

**Associate Professor**

**Department of Computer Applications**

# ACKNOWLEDGEMENT

I am greatly indebted to the authorities of Mangalam College of Engineering for providing the necessary facilities to successfully complete my Project on the topic "Gluco-Sense AI".

I express my sincere thanks to **Dr. Vinodh P Vijayan**, Principal, Mangalam College of Engineering for providing the facilities to complete my Project successfully.

I thank and express my solicit gratitude to **Ms. Divya S B,** HOD, Department of Computer Applications, Mangalam College of Engineering, for her invaluable help and support which helped me a lot in successfully completing this Project work.

I express my gratitude to my Internal Guide, **Ms. SREELEKSHMI K S**, Assistant professor, Department of Computer Applications for the suggestions and encouragement which helped in the successful completion of my Project.

Furthermore, I would like to acknowledge with much appreciation the crucial role of the faculties especially Project coordinator, **Ms. Banu Sumayya S**, Department of Computer Applications, Mangalam College of Engineering, who gave the permission to use all the required equipment and the necessary resources to complete the presentation & report.

**HARI C S (MLM2024MCA-2029)**

# ABSTRACT

Diabetes is a widespread chronic disease that can lead to serious health complications if not detected and treated early. The objective of **GLUCO-SENSE AI** is to provide an intelligent and efficient system for **early diabetes prediction** using advanced artificial intelligence techniques. The system analyses user health data such as glucose levels, BMI, blood pressure, and other medical parameters to accurately predict the likelihood of diabetes.

The project is structured into four key modules: **data collection**, **preprocessing**, **feature extraction**, and **classification**. In the classification phase, **Support Vector Machine (SVM)**, **Logistic Regression**, **K-Nearest Neighbors (KNN)**, and **Random Forest** algorithms are employed to enhance prediction accuracy and reliability. Furthermore, **GLUCO-SENSE AI** integrates a **doctor booking system** that allows users with a positive prediction to instantly schedule consultations with healthcare professionals for further evaluation and treatment.

By combining machine learning with healthcare automation, **GLUCO-SENSE AI** aims to provide timely detection, personalized medical assistance, and improved patient outcomes. This project demonstrates how AI-driven prediction and medical connectivity can revolutionize diabetes management and contribute to the advancement of intelligent healthcare solutions.



**Keywords:** ***SVM, Logistic Regression, KNN, Random Forest, Classification, Feature extraction, Data preprocessing, Data collection***


| | |
|---|---|
| **Mapping with Sustainable Development Goals** | SDG 3 - Good Health and Well-Being<br>SDG 9 - Industry, Innovation, and Infrastructure |

# LIST OF FIGURES

# LIST OF FIGURES

# LIST OF ABBREVATIONS

| ABBREVIATION | FULL FORM |
|---|---|
| CNN | Convolutional Neural Network |
| NLP | Natural Language Processing |
| RL | Reinforcement Learning |
| AI | Artificial Intelligence |
| DL | Deep Learning |
| SVM | Support Vector Machine |
| KNN | K-Nearest Neighbors |
| RF | Random Forest |
| LR | Logistic Regression |
| PHQ-9 | Patient Health Questionnaire-9 |
| DFD | Data Flow Diagram |
| EHR | Electronic Health Record |

# CHAPTER 1
# INTRODUCTION

## 1 Background

Diabetes mellitus is a chronic metabolic disorder that affects millions of individuals globally, posing serious health risks if not diagnosed and managed in time. The condition occurs when the body is unable to properly regulate blood glucose levels, leading to long-term complications such as cardiovascular diseases, kidney failure, neuropathy, and vision loss. Early detection and timely medical intervention are crucial for preventing or delaying these complications. Traditional diagnostic methods often require regular clinical testing, which can be time-consuming, costly, and inaccessible for people in remote areas. To address these challenges, there is a growing interest in leveraging **artificial intelligence (AI)** and **machine learning (ML)** technologies to develop automated and reliable diabetes prediction systems.

**GLUCO-SENSE AI** is an intelligent diabetes prediction system that utilizes **machine learning algorithms** to analyze user health data and predict the likelihood of diabetes with high accuracy. The system focuses on integrating advanced AI models to enhance prediction reliability and provide accessible healthcare support. The key health parameters considered include glucose level, BMI, age, blood pressure, and other physiological indicators. The system is designed around four major modules: **data collection**, **preprocessing**, **feature extraction**, and **classification**.

In the classification module, **Support Vector Machine (SVM)**, **Logistic Regression**, **K-Nearest Neighbors (KNN)**, and **Random Forest** algorithms are implemented to ensure accurate and efficient prediction results. Each of these algorithms plays a crucial role in analyzing complex datasets and identifying patterns that distinguish diabetic from non-diabetic individuals. By combining multiple algorithms, **GLUCO-SENSE AI** achieves higher precision, adaptability, and robustness in prediction outcomes.

Beyond prediction, **GLUCO-SENSE AI** incorporates an innovative **doctor booking system** that connects users who receive a positive diabetes prediction directly to qualified healthcare professionals. This integration ensures that users can take immediate action by consulting doctors for further diagnosis, treatment, and lifestyle guidance. The seamless combination of AI-based prediction and real-time medical consultation makes the system both preventive and actionable.

Furthermore, the system emphasizes **personalized healthcare** by utilizing user-specific data and patterns to tailor recommendations and interventions. This approach acknowledges that diabetes risk and progression vary across individuals due to genetic, environmental, and lifestyle factors. Through continuous data analysis and model optimization, the system adapts to individual user profiles, ensuring accurate and personalized predictions.

By integrating advanced machine learning techniques and a patient-centric approach, **GLUCO-SENSE AI** represents a significant step forward in predictive healthcare. The project demonstrates how AI can transform traditional medical diagnosis into a proactive, data-driven process, enhancing accessibility, accuracy, and efficiency. Ultimately, **GLUCO-SENSE AI** aims to reduce the prevalence and severity of diabetes by enabling early detection, timely intervention, and improved health outcomes for individuals worldwide.

### 1.1 Introduction

The growing global incidence of **diabetes** highlights the urgent need for innovative, data-driven approaches to its **early detection and prevention**. Traditional diagnostic methods rely heavily on clinical testing and periodic health checkups, which may not always be accessible or affordable for everyone, especially in remote or underdeveloped regions. Moreover, many individuals remain undiagnosed due to a lack of awareness or delayed testing, leading to severe health complications such as cardiovascular disease, neuropathy, and kidney failure. To address these challenges, the proposed project—**GLUCO-SENSE AI**—introduces a novel framework that integrates artificial intelligence and machine learning to provide a fast, accurate, and accessible system for **diabetes prediction and medical consultation**.

**GLUCO-SENSE AI** utilizes advanced **machine learning algorithms**, namely **Support Vector Machine (SVM)**, **Logistic Regression**, **K-Nearest Neighbors (KNN)**, and **Random Forest**, to analyze user health data such as glucose levels, BMI, age, blood pressure, and insulin levels. By applying these algorithms, the system identifies patterns and relationships within the data that can indicate a user's likelihood of developing diabetes. The project's architecture is systematically divided into four major modules—**data collection**, **preprocessing**, **feature extraction**, and **classification**—ensuring a structured and reliable analytical process.

In addition to prediction, **GLUCO-SENSE AI** includes an integrated **doctor booking system** that allows users who receive a positive prediction to directly schedule an appointment with a certified healthcare professional. This feature bridges the gap between early detection and medical intervention, ensuring timely treatment and professional guidance. The incorporation of this functionality emphasizes the project's commitment to not only identifying potential diabetes risks but also enabling users to take immediate, informed action.

Furthermore, **GLUCO-SENSE AI** emphasizes **personalized healthcare** by tailoring recommendations and predictive models based on each user's data profile. This adaptive capability acknowledges the variability in how diabetes develops across individuals due to genetic, lifestyle, and environmental factors. By combining predictive analytics with user-specific insights, the system supports proactive and preventive healthcare management.

The project's design ensures methodological rigor and real-world applicability through systematic data handling and algorithmic transparency. By translating machine learning outputs into actionable medical insights, **GLUCO-SENSE AI** enhances the effectiveness and accessibility of early diabetes detection.

Ultimately, **GLUCO-SENSE AI** aspires to **revolutionize the healthcare landscape** by making diabetes prediction **accurate, timely, and personalized**. It demonstrates how artificial intelligence can transform traditional healthcare into a proactive, technology-driven process—reducing the burden of undiagnosed diabetes, improving patient outcomes, and advancing global health equity.

### 1.3 Problem Statement

In today's healthcare landscape, the increasing prevalence of **diabetes** highlights the urgent need for improved diagnostic and intervention strategies. Traditional diagnostic methods often rely on periodic medical checkups and laboratory testing, which can be expensive, time-consuming, and inaccessible for individuals in remote or underprivileged regions. Moreover, a lack of awareness and the asymptomatic nature of early-stage diabetes frequently lead to delayed diagnosis and treatment. This delay increases the risk of severe complications such as heart disease, kidney failure, and neuropathy.

The central problem addressed by **GLUCO-SENSE AI** is the **absence of an intelligent, accessible, and early detection system** that can efficiently analyze patient data and identify potential diabetes cases before critical stages develop. Existing methods lack automation, personalization, and immediate post-diagnosis guidance. The integration of **artificial intelligence** and **machine learning algorithms** presents a transformative solution to these challenges, providing a **data-driven, accurate, and user-friendly** system for diabetes prediction. By combining predictive modeling with a built-in **doctor booking system**, **GLUCO-SENSE AI** aims to make early detection and medical consultation **timely, personalized, and universally accessible**, thereby reducing the global burden of diabetes and improving patient outcomes.

### 1.4 Motivation

The motivation behind developing **GLUCO-SENSE AI** stems from the growing global health crisis posed by diabetes, which continues to affect millions of people across all age groups. Many individuals remain undiagnosed until advanced stages due to the limitations of traditional diagnostic methods, lack of regular screenings, and inadequate access to healthcare services. This project is driven by the urgent need to **revolutionize diabetes detection** through automation, accuracy, and accessibility.

The system leverages **machine learning algorithms** such as **Support Vector Machine (SVM)**, **Logistic Regression**, **K-Nearest Neighbors (KNN)**, and **Random Forest** to analyze complex medical data and identify patterns indicative of diabetes risk. This approach allows for **precise and reliable predictions** based on a combination of health metrics like glucose level, BMI, blood pressure, and age.

Beyond prediction, **GLUCO-SENSE AI** focuses on **personalized healthcare** by tailoring recommendations based on individual patient profiles. The integration of a **doctor booking module** ensures that users who are predicted to be at risk can immediately connect with healthcare professionals for consultation and treatment. This proactive approach bridges the gap between prediction and medical intervention, transforming healthcare from reactive to preventive.

Moreover, the project is motivated by the vision of **democratizing healthcare**, making accurate diabetes prediction accessible to all, regardless of location or economic status. By integrating AI with user-friendly digital platforms, **GLUCO-SENSE AI** enables continuous monitoring, early warning, and timely intervention—ultimately contributing to improved global health outcomes and a reduction in diabetes-related complications.

### 1.5 Scope

The scope of **GLUCO-SENSE AI** is extensive and transformative, encompassing the integration of advanced artificial intelligence with healthcare to enable **early, accurate, and personalized diabetes prediction**. At its foundation, the system is designed to process diverse medical datasets, perform feature extraction, and apply multiple classification algorithms to achieve high accuracy and reliability in predictions.

The project's four core modules—**data collection**, **preprocessing**, **feature extraction**, and **classification**—form a structured framework that ensures a systematic approach to diabetes analysis. Through these modules, the system can efficiently handle raw patient data, identify significant medical indicators, and predict diabetes likelihood with precision.

Furthermore, the inclusion of a **doctor booking system** significantly broadens the project's scope by transforming it from a mere diagnostic tool into a **comprehensive healthcare platform**. Users can seamlessly transition from prediction to professional consultation, ensuring timely medical support.

The scope also extends to **personalized healthcare management**, where the system adapts to individual user data, offering tailored insights and preventive recommendations. With potential for integration into **mobile applications and cloud platforms**, **GLUCO-SENSE AI** envisions a future where diabetes detection and consultation are **accessible anytime, anywhere**.

Ultimately, the scope of **GLUCO-SENSE AI** extends beyond detection—it aims to **redefine preventive healthcare** by combining intelligent prediction, accessibility, and personalized care. The project aspires to significantly reduce the global burden of diabetes and pave the way for a future where **AI-driven health systems empower individuals to take control of their well-being**.

# CHAPTER 2
# LITERATURE REVIEW

## 2.1 Diabetes Prediction Using Machine Learning and Explainable AI Techniques [Isfafuzzaman Tasin et al., 2022][1]

The paper titled *"Diabetes Prediction Using Machine Learning and Explainable AI Techniques"* presents a comprehensive framework for predicting diabetes using multiple machine learning models and explainable AI tools. The study focuses on improving prediction accuracy, data interpretability, and accessibility through the integration of advanced algorithms and real-time applications.

Key Aspects:

- Machine Learning Techniques:

The research utilizes several machine learning algorithms including Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Decision Tree, AdaBoost, Bagging, Voting Classifier, and Extreme Gradient Boosting (XGBoost). Hyperparameter tuning was performed using GridSearchCV to optimize performance and prevent overfitting.

- Explainable Artificial Intelligence (XAI):

A major contribution of the study is the integration of LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) frameworks. These tools were used to interpret how the model predicts diabetes and identify which health features—such as glucose level, BMI, age, and insulin—have the most significant impact on prediction outcomes. This enhances the model's transparency and trustworthiness.

- Dataset and Data Fusion:

The study employs two datasets: the Pima Indian Diabetes Dataset and a private dataset (RTML) collected from 203 female patients in Bangladesh. Missing insulin values in the private dataset were predicted using a semi-supervised XGBoost regression model, and both datasets were merged to improve robustness and generalization.

- Data Preprocessing and Balancing:

The data underwent several preprocessing steps, including mean value imputation, mutual information-based feature selection, and Min–Max normalization. To address the problem of class imbalance, SMOTE (Synthetic Minority Oversampling Technique) and ADASYN (Adaptive Synthetic Sampling) were employed, ensuring equal representation of diabetic and non-diabetic samples.

- Performance Evaluation:

The study evaluates models based on accuracy, precision, recall, F1-score, and AUC (Area Under the ROC Curve). Among all tested classifiers, the XGBoost model combined with ADASYN achieved the best performance, with 81% accuracy, an F1-score of 0.81, and an AUC of 0.84.

- Practical Implementation:

To enhance accessibility and usability, the best-performing model was deployed as both a web application and an Android smartphone application. These platforms allow users to input health parameters and receive instant diabetes predictions, making early detection more convenient and accessible.

- Explainability and Domain Adaptation:

The system was further validated through a domain adaptation technique, demonstrating its ability to generalize across datasets with 96% accuracy on the private RTML dataset. The use of explainable AI (LIME and SHAP) ensured clear understanding of how different features influence predictions, improving confidence in AI-assisted healthcare.

### 2.2 Diabetes Prediction Using Machine Learning [KM Jyoti Rani, 2020][2]

Diabetes is a chronic disease with the potential to cause a worldwide health crisis, affecting 382 million people globally, with numbers projected to rise to 592 million by 2035. It is caused by high blood glucose levels, which produce symptoms such as frequent urination, increased thirst, and increased hunger, and can lead to serious complications including blindness, kidney failure, amputations, heart failure, and stroke. Normally, the pancreas releases insulin to allow glucose to enter cells for energy, but in diabetes, this system malfunctions. The most common types are Type 1 and Type 2 diabetes, while gestational diabetes occurs during pregnancy, with other less common types also existing. Machine learning offers a way to predict diabetes early with higher accuracy by analyzing patient data using algorithms such as K-Nearest Neighbors (KNN), Logistic Regression, Random Forest, Support Vector Machine (SVM), and Decision Tree. The accuracy of each algorithm is evaluated, and the model with the best performance is used for predicting diabetes.

Key Aspects:

- **Machine Learning Techniques:**

The study employs multiple machine learning algorithms, including K-Nearest Neighbors (KNN), Logistic Regression, Random Forest, Support Vector Machine (SVM), and Decision Tree, to predict diabetes.

These algorithms analyze patient data to identify patterns associated with diabetes.

- **Dataset and Data Analysis:**

The study uses a dataset containing 2000 cases from Kaggle, with patient measures such as blood tests and health indicators.

Data preprocessing and classification techniques are applied to ensure accurate predictions.

- **Automated Diabetes Prediction:**

The system automates the identification of diabetes risk by evaluating input features against trained models.

Accuracy of each algorithm is calculated, and the best-performing model is selected for prediction.

- **Enhanced Accuracy and Model Evaluation:**

Different algorithms are compared using metrics such as accuracy, sensitivity, and specificity.

The methodology includes cross-validation approaches and handling of dataset issues like class imbalance to improve predictive performance.

- **Types and Causes of Diabetes:**

Covers Type 1, Type 2, and gestational diabetes, highlighting genetic and lifestyle factors as causes.

Explains the mechanism of insulin deficiency or resistance leading to high blood glucose.

- **Symptoms and Health Impact:**

Identifies common symptoms: frequent urination, increased thirst, fatigue, weight loss, blurred vision, mood swings, confusion, and frequent infections.

Highlights the serious health consequences of diabetes, including blindness, kidney failure, heart disease, and stroke.

- **Proactive Health Intervention:**

By accurately predicting diabetes early, the model aims to facilitate timely medical intervention and reduce health risks associated with the disease.

### 2.3 Machine learning based diabetes prediction and development of smart web application [Nazin Ahmed , Rayhan Ahammed , Md. Manowarul Islam (2022)][3]

Diabetes is a common disease affecting individuals worldwide and increases the risk of long-term complications such as heart disease and kidney failure. Early detection can enable people to live longer and healthier lives. This study aims to identify effective machine-learning-based classifier models for detecting diabetes using clinical data. The algorithms trained in this work include Decision Tree (DT), Naive Bayes (NB), K-Nearest Neighbor (KNN), Random Forest (RF), Gradient Boosting (GB), Logistic Regression (LR), and Support Vector Machine (SVM). Efficient preprocessing techniques, including label encoding and normalization, were applied to improve model accuracy, and various feature selection approaches were used to identify and prioritize key risk factors. Extensive experiments were conducted using two different datasets to analyze model performance, and results were compared with recent studies, showing improved accuracy ranging from 2.71% to 13.13% depending on the dataset and algorithm. The highest-performing machine learning model was further developed and integrated into a web application using the Python Flask framework. The study demonstrates that an appropriate preprocessing pipeline combined with ML-based classification can predict diabetes accurately and efficiently.

Key Aspects:

- **Prevalence and Impact of Diabetes:**

Diabetes mellitus is a common and critical disease, affecting 8.5% of adults globally and causing 1.6 million deaths annually.Long-term complications include heart disease, stroke, kidney failure, nerve damage, and eye issues.Major contributing factors: obesity, age, lack of exercise, lifestyle, hereditary factors, high blood pressure, and poor diet.

- **Role of Machine Learning:**

Supervised and unsupervised ML methods are applied to detect diabetes early and reduce diagnostic costs.Predictive models learn from historical medical datasets to make accurate decisions on new patient data.Early detection through ML enables timely intervention, minimizing harmful effects.

- **Datasets and Data Preprocessing:**

Multiple datasets, including the PIMA Indian dataset and other clinical datasets, are used for robustness.Preprocessing techniques applied include outlier removal (IQR), missing value handling (mean imputation), data standardization, and label encoding.Correlation-based and other feature selection methods (chi-square, correlation) are used to identify important attributes.

- **Machine Learning Algorithms Used:**

Algorithms applied include Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Gradient Boosting (GB), and K-Nearest Neighbor (KNN).Performance of each algorithm is evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC curve.

- **Model Construction and Evaluation:**

Data split into 80% training and 20% testing sets to build predictive models.Algorithm performance is analyzed on both full and reduced feature sets.The best-performing algorithm is selected for deployment.

- **Web Application Integration:**

The selected ML model is integrated into a Flask-based web application.Users can submit their clinical parameters via the application to receive diabetes predictions in real time.

- **Research Contributions:**

Training and evaluation of multiple ML algorithms across several datasets.Identification of key features contributing to diabetes prediction.Development of an accessible web application for end-users.Demonstration of improved predictive performance compared to previous studies.

### 2.4 Research on Diabetes Prediction Method Based on Machine Learning [Jingyu Xue, Fanchao Min and Fengying Ma (2020)][4]

Diabetes mellitus (DM) is a metabolic disease characterized by high blood sugar. The main clinical types are type 1 diabetes and type 2 diabetes. Now, the proportion of young people suffering from type 1 diabetes has increased significantly. Type 1 diabetes is chronic when it occurs in childhood and adolescence, and has a long incubation period. The early symptoms of the onset are not obvious, which may lead to failure to detect in time and delay treatment. Long-term high blood sugar can cause chronic damage and dysfunction of various tissues, especially eyes, kidneys, heart, blood vessels and nerves. Therefore, the early prediction of diabetes is particularly important. In this paper, we use supervised machine-learning algorithms like Support Vector Machine (SVM), Naive Bayes classifier and LightGBM to train on the actual data of 520 diabetic patients and potential diabetic patients aged 16 to 90. Through comparative analysis of classification and recognition accuracy, the performance of support vector machine is the best.

Key Aspects:

**• Machine Learning Algorithms:**

The study explores supervised machine learning algorithms for diabetes prediction, with a primary focus on Support Vector Machine (SVM) for its robustness in binary classification tasks. Additionally, Naive Bayes classifier and LightGBM are utilized to compare performance, leveraging their strengths in probabilistic modeling and gradient boosting on decision trees, respectively, to handle symptom-based data effectively.

**• Diabetes Prediction:**

The core objective is the early detection and prediction of diabetes mellitus, particularly type 1 and type 2, using patient symptoms to identify risks before complications arise. Machine learning models are trained to analyze patterns in data from diabetic and potential patients, enabling timely intervention and reducing long-term health impacts like organ damage.

**• Dataset Integration:**

The research incorporates a dataset from the UCI repository, comprising 520 samples from patients aged 16-90 at Sylhet Diabetes Hospital in Bangladesh. This includes 17 attributes such as age, gender, polyuria, polydipsia, sudden weight loss, and others, with binary values for symptoms and a class label for positive/negative diabetes outcomes, facilitating comprehensive data mining.

**• Literature Review:**

The paper includes a review of diabetes etiology, global prevalence statistics from WHO (e.g., rise from 108 million cases in 1980 to 422 million in 2014), and prior applications of ML in disease prediction. It highlights environmental factors like viral infections and references studies on SVM in neurological and diabetes diagnostics to contextualize the approach.

**• Performance Metrics:**

Evaluation relies on a confusion matrix to measure model effectiveness, categorizing outcomes into true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Metrics such as accuracy are computed, with an 80:20 train-test split revealing SVM's superior 96.54% accuracy compared to 93.27% for Naive Bayes and 88.46% for LightGBM

### 2.5 Diabetes Prediction Using Machine Learning and Explainable AI Techniques" (Tasin et al., 2022)[5]

Key Aspects:

The study on improving healthcare prediction of diabetic patients using KNN imputed features and tri-ensemble model is a comprehensive exploration of leveraging machine learning techniques to enhance diabetes detection accuracy by addressing data gaps. The research employs ensemble learning models, including a Tri-Ensemble voting classifier combining Extra Tree Classifier (ETC), Extreme Gradient Boosting (XGB), and Random Forest (RF), to predict diabetes risks from imputed datasets. The study not only delves into handling missing values through KNN imputation but extends its focus to comparative evaluations under scenarios with and without imputation, demonstrating superior performance metrics. By integrating preprocessing strategies and ensemble methods, the research aims to develop a robust approach that captures the complexities of diabetes datasets and improves predictive outcomes. The ultimate goal is to contribute to more accurate, timely, and effective interventions in diabetes care. The research employs ensemble learning models, including a Tri-Ensemble voting classifier combining Extra Tree Classifier (ETC), Extreme Gradient Boosting (XGB), and Random Forest (RF), to predict diabetes risks from imputed datasets. The study not only delves into handling missing values through KNN imputation but extends its focus to comparative evaluations under scenarios with and without imputation, demonstrating superior performance metrics.

### 2.6 Towards Transparent and Accurate Diabetes Prediction Using Machine Learning and Explainable Artificial Intelligence[6]

**Authors**: Pir Bakhsh Khokhar, Viviana Pentangelo, Fabio Palomba, Carmine Gravino (2025)

Diabetes mellitus is a growing global health problem, and the integration of artificial intelligence into clinical prediction systems offers a powerful approach for early diagnosis and prevention.

This study proposes a **transparent and accurate diabetes prediction framework** that combines **Machine Learning (ML)** models with **Explainable Artificial Intelligence (XAI)** tools to ensure both **high predictive accuracy** and **interpretability**.

The researchers trained multiple ML algorithms—**Random Forest (RF)**, **Extreme Gradient Boosting (XGBoost)**, **Support Vector Machine (SVM)**, **Logistic Regression (LR)**, and **K-Nearest Neighbors (KNN)**—on the **Behavioral Risk Factor Surveillance System (BRFSS)** dataset, which includes over **250,000 health records**. To handle **data imbalance**, the **Synthetic Minority Oversampling Technique (SMOTE)** was applied, ensuring balanced class representation and improved generalization.

Feature selection was conducted using **Mutual Information** and **Recursive Feature Elimination (RFE)**, leading to identification of the **five most influential predictors** of diabetes: **Body Mass Index (BMI)**, **Age**, **General Health**, **Income**, and **Physical Activity**. The best-performing model, **XGBoost**, achieved **92.50% test accuracy** and an **ROC-AUC score of 0.975**, outperforming baseline approaches.

To improve trust and transparency, **SHAP (SHapley Additive Explanations)** and **LIME (Local Interpretable Model-Agnostic Explanations)** were integrated to visualize and interpret model predictions. SHAP global interpretation revealed that **BMI** and **Age** had the strongest influence on diabetes risk, while local LIME explanations provided case-by-case reasoning for individual predictions.

The study concludes that integrating **explainable AI** not only enhances **model transparency** and **clinical reliability**, but also helps medical professionals understand the *why* behind predictions, enabling more informed and personalized health interventions.

Key Contributions:

Developed a diabetes prediction framework combining machine learning models with explainable AI tools. Utilized Synthetic Minority Oversampling Technique (SMOTE) for data preprocessing. Achieved a test accuracy of 92.50% and an ROC-AUC of 0.975. Identified BMI, Age, General Health, Income, and Physical Activity as significant predictors. Emphasized the importance of model interpretability in healthcare applications.

### 2.7 Towards Transparent Diabetes Prediction: Combining AutoML with Explainable AI[7]

**Authors**: R Hasan et al. (2024)

Recent advances in **Automated Machine Learning (AutoML)** have streamlined the process of model selection and hyperparameter tuning, but many AutoML-generated models lack transparency—an essential factor for clinical adoption. To bridge this gap, this study introduces a **hybrid framework that integrates AutoML with Explainable Artificial Intelligence (XAI)** for **diabetes risk prediction**.

The authors utilized the **PIMA Indian Diabetes dataset** and a **real-world clinical dataset from Dhaka Medical College**, containing demographic, physiological, and lifestyle variables. The framework employed **AutoGluon's AutoML** to automatically explore and optimize a wide range of models, including Gradient Boosting Machines, Random Forests, Neural Networks, and Logistic Regression, using **stacked ensemble learning** for improved performance.

The **best-performing ensemble model** achieved an **accuracy of 85.01%**, **F1-score of 0.84**, and **ROC-AUC of 0.89** on the test dataset. While AutoML handled the complex task of model tuning, interpretability was ensured through a multi-layered XAI approach combining **SHAP**, **LIME**, **Integrated Gradients**, **Attention Mechanisms**, and **Counterfactual Explanations**.

The **SHAP summary plots** revealed that **Glucose Level**, **BMI**, **Age**, and **Blood Pressure** were the most influential predictors, while counterfactual analysis enabled clinicians to simulate *"what-if"* scenarios — for example, exploring how modifying lifestyle factors could lower diabetes risk.

To make the model clinically accessible, the team developed an **interactive Streamlit-based web application**, allowing healthcare professionals to visualize model decisions, inspect feature contributions, and test hypothetical patient profiles.

**Key Contributions**:

Integrated Automated Machine Learning (AutoML) with Explainable AI techniques to enhance diabetes risk prediction. Achieved an ensemble model accuracy of 85.01% using AutoGluon's AutoML framework. Applied SHAP, LIME, Integrated Gradients, Attention Mechanism, and Counterfactual Analysis for model interpretability. Developed an interactive Streamlit application for clinicians to explore feature importance and test hypothetical scenarios.

## 2.8 Interactive Diabetes Risk Prediction Using Explainable Machine Learning[8]

**Author**: Udaya Allani (2025)

This work introduces an **interactive, explainable machine learning system** for assessing individual risk of diabetes, combining predictive modeling with transparency and user engagement. Leveraging the 2015 CDC BRFSS dataset, the study evaluates a suite of classification algorithms — Logistic Regression, Random Forest, XGBoost, LightGBM, K-Nearest Neighbors, and Neural Networks — under three data sampling strategies: original, SMOTE oversampling, and undersampling. Among these configurations, **LightGBM with undersampling** yielded the highest recall, making it most sensitive in identifying individuals at risk.

To enhance interpretability and trust, the system integrates **SHAP** and **LIME** to elucidate both global feature importance and local instance explanations, enabling users to see how individual features contribute to the risk score. Furthermore, **comorbidity insights** are derived through **Pearson correlation analysis**, highlighting how combinations of risk factors (e.g. high blood pressure, obesity, cholesterol) interact and amplify diabetes risk.

The front end is implemented in **Dash**, providing an interactive UI where users can input their health and behavioral parameters, view predicted risk, and explore interactive visual explanations and feature insights. Personalized suggestions and interpretations support improved health awareness and decision making.

This approach bridges the gap between advanced machine learning methods and accessible, transparent health risk tools, offering a practical step forward in trustworthy AI-based healthcare decision support.

**Key Contributions:**

Developed a web-based interactive health risk prediction tool for diabetes risk assessment. Evaluated models including Logistic Regression, Random Forest, XGBoost, LightGBM, KNN, and Neural Networks. Utilized SHAP and LIME for model explainability. Integrated comorbidity insights using Pearson correlation analysis. Implemented a Dash-based user interface for personalized health awareness.

### 2.9An Explainable Artificial Intelligence Software System for Real-Time Blood Glucose Monitoring[9]

**Authors**: PN Srinivasu et al. (2024)

Continuous monitoring of blood glucose is critical for effective diabetes management, but current methods are often invasive or lack interpretability. This study proposes an **Explainable Artificial Intelligence (XAI)** system for **real-time glucose monitoring**, combining **CNN (Convolutional Neural Networks)** and **Bi-LSTM (Bidirectional Long Short Term Memory)** models to analyze spectrogram representations of glucose signal data obtained via glucose oxidase (GOD) strips. The input signals are converted into time–frequency images and classified into **low**, **normal**, or **abnormal** glucose levels.

The system integrates **interpretability mechanisms (e.g. SHAP)** to explain the influence of input features (or spectrogram regions) on prediction outcomes, thereby enhancing trust and transparency. Real-time alerts are generated when the predicted class falls into abnormal ranges, supporting prompt medical intervention.

Experimental results using standard classification metrics (accuracy, ROC, confusion matrix) demonstrate that the proposed model reliably identifies elevated glucose states. The XAI design ensures that clinicians or users can understand why a certain alert was triggered. This work contributes toward merging advanced deep learning for physiological signal analysis with interpretable AI, advancing more user-centric, trustworthy health monitoring systems.

**Key Contributions**:

Designed an AI-based system for real-time blood glucose level monitoring. Incorporated explainable AI methods to interpret model predictions. Provided timely alerts for medication, enhancing patient self-management.

### 2.10 Explainable Artificial Intelligence-PREDICT: Development and Validation of an Early Prediction Tool for Diabetic Complications[10]

**Authors**: ZH Al-Hudaibi et al. (2025)

Diabetes mellitus presents a marked public health challenge, and its complications (such as nephropathy, retinopathy, neuropathy, cardiovascular disease, and diabetic foot) significantly worsen patient outcomes. This study proposes **XAI-PREDICT**, an **explainable AI** tool designed to predict the risk of five key diabetic complications up to 24 months ahead, utilizing **electronic health records** (EHRs) from 12 hospitals across Saudi Arabia. A retrospective cohort of 87,542 patients was analyzed. The dataset was partitioned into development and validation sets, with additional external validation from independent hospitals to assess generalizability.

Advanced machine learning models, integrated within an interpretability framework, were trained and optimized to predict each complication. Contextual features specific to Saudi healthcare and patient variables were incorporated to enhance relevancy and interpretability. The model achieved robust performance: AUC values of 0.89 for nephropathy, 0.86 for retinopathy, 0.84 for neuropathy, 0.88 for cardiovascular events, and 0.91 for diabetic foot.

Interpretability modules provide clinicians and stakeholders with insights into **which features** drive risk predictions, thereby supporting transparent, evidence-based decision making. XAI-PREDICT thus bridges high-performing AI models with practical clinical utility, facilitating early interventions and helping reduce the burden of diabetic complications.

**Key Contributions**:

Developed XAI-PREDICT, an AI-based tool for early detection of diabetic complications. Utilized electronic health records from 12 hospitals in Saudi Arabia. Achieved high predictive accuracy with AUC values ranging from 0.84 to 0.91 for various complications. Emphasized the importance of interpretability in clinical decision-making.

# CHAPTER 3

# PROPOSED SYSTEM

### 3.1 Users

The GLUCO-SENSE AI application has two main types of users: **regular users (patients)** and **admin (system administrator)**. **Registration**

When a user registers, their information is collected and validated. This data is typically stored in a database for future reference.

#### 3.1.1 User Registration and Login

- Registration: Users create an account by providing basic information such as name, email, and password.
- Login: Users authenticate with their credentials to access the application.

### 3.2 Diabetes Prediction

#### 3.2.1 User Input

After logging in, users provide health-related data such as:

- Age, weight, height
- Blood sugar levels (fasting and/or postprandial)
- Lifestyle details (exercise, diet, family history)

#### 3.2.2 Preprocessing

The system preprocesses the input to ensure data quality:

Normalization of numerical values. Handling missing or inconsistent data

#### 3.2.3 Feature Extraction

Relevant features for diabetes prediction are extracted, including:

- Blood sugar trends
- BMI

- Lifestyle patterns

### 3.2.4 Classification

Machine learning models (SVM, Logistic Regression, KNN, Random Forest) are used to classify the user's risk of diabetes:

- Non-diabetic
- Diabetic / At Risk

### 3.2.5 Displaying Result

The prediction result is displayed to the user. If the user is predicted to have diabetes, they are given the option to book a doctor appointment.

## 3.3 Doctor Appointment Booking

### 3.3.1 Booking Process

- Users select a doctor and choose a preferred date for consultation.
- The booking is recorded in the Booking History as "Pending Approval."

### 3.3.2 Admin Actions

- Admin can log in to view all pending bookings.
- Admin can **approve** or **remove** a booking.

### 3.3.3 Booking Status Update

- Once a booking is approved, the user can see the status updated in their **Booking History.**
- Users are notified about approved or rejected bookings.

# CHAPTER 4
# METHODOLOGY

The methodology for the GLUCO-SENSE AI project is structured to provide a seamless and effective approach for diabetes prediction and doctor appointment management, combining data-driven machine learning techniques with secure healthcare workflow integration. The process begins with user registration and authentication, ensuring that users create secure accounts with verified credentials. Once logged in, users provide health-related input data, including age, weight, height, blood sugar levels, and lifestyle information such as diet, exercise patterns, and family medical history. This data collection phase emphasizes accuracy, completeness, and ethical handling of sensitive health information, following privacy best practices.

Following data acquisition, preprocessing techniques are applied to ensure the uniformity and reliability of the input. Numerical values, such as blood glucose levels and body mass index, are normalized, while categorical variables, such as diet type or exercise habits, are encoded appropriately. Missing or inconsistent data is handled through imputation or validation checks, and noise reduction techniques are applied to minimize errors. Proper preprocessing ensures that the input data is clean, consistent, and suitable for feature extraction and model training.

Feature extraction constitutes a critical stage, where relevant health indicators are identified and extracted from the preprocessed data. These features include blood glucose trends, BMI, lifestyle patterns, and familial predisposition to diabetes. By selecting the most informative features, the dimensionality of the dataset is reduced while preserving the information necessary for accurate risk prediction. The extracted features are then input into the machine learning classification models, which include Support Vector Machines (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest. These algorithms are trained on labeled datasets to predict diabetes risk, categorizing users as non-diabetic or diabetic/at-risk. Ensemble methods are utilized where appropriate to improve prediction accuracy and robustness.

Once the classification is complete, the system presents the prediction results to the user. If the user is identified as diabetic or at risk, they are offered the option to book a consultation with a healthcare professional. The doctor appointment booking module allows users to select a preferred doctor and consultation date, which is recorded in the **Booking History** as pending approval. The system administrator can review all pending bookings and either approve or remove them, ensuring proper management and oversight. Approved bookings are updated in the user's history, providing real-time status notifications and enabling users to plan their consultation accordingly.

Continuous monitoring and iterative refinement are integral to the methodology, as user feedback and model performance metrics are collected to improve prediction accuracy and system usability. This includes analyzing misclassifications, retraining models with new data, and optimizing feature selection. Ethical considerations, such as data privacy, secure authentication, and encrypted storage, are prioritized throughout the process to maintain compliance with healthcare data standards.

The GLUCO-SENSE AI methodology integrates data collection, preprocessing, feature extraction, predictive modeling, and appointment management into a cohesive system. By combining advanced machine learning techniques with user-centric healthcare workflows, the methodology provides an effective, reliable, and accessible platform for early diabetes detection and professional consultation management. Through this systematic approach, the project aims to empower users with actionable insights into their health while ensuring smooth interaction with healthcare providers.

### 3.2.1 Data Collection and User Input

The first stage of the methodology involves data collection, where users provide their personal and health-related information upon registration. This includes age, weight, height, body mass index (BMI), blood glucose readings, lifestyle factors such as diet and exercise habits, and family medical history. Users log in to the system through secure authentication, and their inputs are captured via structured forms. Ensuring accurate, complete, and consistent data is critical, as it forms the foundation of the predictive model. The system incorporates validation checks to prevent erroneous entries, and privacy measures, such as encrypted storage and access control, are implemented to protect sensitive health information.

### 3.2.2 Data Preprocessing

Preprocessing ensures that the collected data is uniform, clean, and suitable for machine learning. Numerical features, including fasting and postprandial blood glucose levels, are normalized to a standardized scale to avoid bias in the predictive model. Categorical variables, such as dietary habits or exercise frequency, are encoded appropriately, and missing or inconsistent values are addressed using imputation methods or default handling strategies. Noise reduction algorithms are applied to minimize the effect of outliers or erroneous inputs. This preprocessing stage ensures that the dataset is reliable and that subsequent feature extraction and model training processes can perform optimally.

### 3.2.3 Feature Extraction and Engineering

Following preprocessing, the system extracts relevant features that capture critical information for predicting diabetes risk. These features include trends in blood glucose readings, BMI changes over time, lifestyle patterns, and hereditary risk factors. Feature engineering may also involve creating derived metrics, such as average glucose levels, standard deviation, or activity scores, which provide additional predictive power. By reducing dimensionality while retaining key information, feature extraction ensures that the machine learning models focus on the most informative aspects of the data.

### 3.2.4 Machine Learning Classification

The core predictive component of GLUCO-SENSE AI involves training machine learning models to classify users into risk categories. Algorithms employed include **Support Vector Machines (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest**, which are trained on labeled datasets of diabetic and non-diabetic patients. Ensemble methods may be utilized to combine the strengths of multiple models, enhancing overall prediction accuracy and robustness. The system evaluates performance metrics, such as accuracy, precision, recall, and F1-score, to ensure reliable predictions. Predictions are generated in real-time and presented to users as risk categories, such as “Low Risk” or “High Risk,” providing actionable insights into their health status.

### 3.2.5 Doctor Appointment Booking

For users identified as diabetic or at risk, the system offers an integrated appointment booking module. Users can select a doctor, specify a preferred date, and submit the booking request. These requests are initially stored in the **Booking History** as **pending approval**, allowing the system administrator to review and manage all appointments. Admin actions include approving or removing bookings, ensuring accurate scheduling and efficient healthcare workflow management. Once approved, users receive notifications and the booking status is updated in real-time, enabling them to track confirmed consultations.

### 3.2.6 System Security and Ethical Considerations

Security and privacy are central to GLUCO-SENSE AI methodology. User data is encrypted during transmission and storage, and secure authentication protocols prevent unauthorized access. Role-based access ensures that sensitive data is only accessible to authorized personnel, such as administrators or the respective users. Ethical considerations, including informed consent, confidentiality, and data privacy, are incorporated at every stage of the system to maintain compliance with healthcare regulations and best practices.

### 3.2.7 Iterative Improvement and Feedback

The methodology incorporates an iterative approach to optimize both the predictive models and system usability. User feedback on prediction results, usability, and booking experience is collected to refine system features. Misclassified cases and system performance metrics are analyzed to retrain models and improve accuracy over time. This continuous monitoring and refinement process ensures that GLUCO-SENSE AI adapts to new data, evolving health trends, and user needs, maintaining a high standard of service delivery.

### 3.2.8 Workflow Integration

The entire methodology integrates predictive analytics and appointment management into a cohesive workflow. Users progress seamlessly from logging in to entering health data, receiving prediction results, and booking consultations if needed. Admin oversight ensures that appointments are validated and efficiently managed, while machine learning models provide reliable and timely health risk predictions. This integrated workflow supports proactive diabetes management, empowering users to take immediate and informed actions toward their health.

### 3.2.9 Summary

In summary, the GLUCO-SENSE AI methodology combines meticulous data collection, preprocessing, feature engineering, machine learning classification, and secure appointment management. By leveraging advanced analytics and a user-centric design, the system delivers accurate diabetes predictions and streamlines interactions with healthcare providers. Continuous monitoring, feedback integration, and ethical data handling ensure that the platform balances.

# CHAPTER 5

# SYSTEM ARCHITECTURE

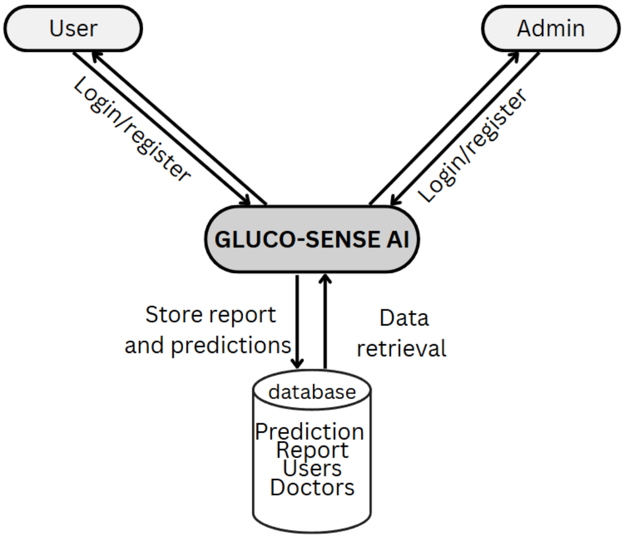


*FIGURE 5.1. System Architecture*

## 5.1 Dataset Collection

The system begins with the collection of user-provided health data, which includes demographic details, blood glucose readings, weight, height, BMI, lifestyle habits (diet and exercise), and family medical history. High-quality data acquisition is crucial, as the accuracy of the diabetes prediction model depends directly on the completeness and relevance of the collected information. Validation checks ensure that the input is accurate and consistent.

## 5.2 Data Preprocessing

Once collected, the data undergoes preprocessing to ensure uniformity and reliability. Numerical features, such as glucose levels and BMI, are normalized, while categorical variables like diet type or exercise frequency are encoded appropriately. Missing or inconsistent data is handled using imputation methods, and outliers are managed to reduce noise. Preprocessing ensures that the dataset is clean, standardized, and ready for feature extraction and predictive modeling, enhancing the performance of machine learning algorithms.

### 5.3 Feature Extraction

Feature extraction identifies the most relevant health indicators for predicting diabetes risk. Key features include blood glucose trends, BMI, age, family history, and lifestyle patterns. Derived features such as average glucose levels, glucose variability, and activity scores may also be calculated. Effective feature extraction reduces dimensionality while retaining critical information, making the dataset more suitable for machine learning classification.

### 5.4 Classification

Machine learning models are used to classify users' diabetes risk based on the extracted features. GLUCO-SENSE AI employs algorithms including **Support Vector Machines (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest**. Ensemble methods are also used to combine multiple model predictions, improving robustness and accuracy. The trained models categorize users as **non-diabetic** or **diabetic/at-risk**, providing actionable health insights.

### 5.5 Labels

Labels are applied to the training dataset to indicate whether a user is diabetic or non-diabetic. These labels allow the machine learning algorithms to learn patterns in the data associated with diabetes. Once trained, the models can predict the risk of diabetes for new users, enabling early detection and intervention.

### 5.6 Loss Function

The loss function quantifies the difference between predicted outcomes and actual labels during training. By minimizing the loss function, the models optimize internal parameters (weights and biases) to improve prediction accuracy. This process ensures that the system can reliably identify high-risk users based on their health data.

### 5.7 Training

The system is trained on a labeled dataset comprising real user data and historical health records. During training, model parameters are iteratively adjusted to minimize prediction errors. Proper training ensures that the models can generalize well to new user inputs, providing reliable diabetes predictions.

### 5.8 Prediction

After training, the models can process new user data to predict the likelihood of diabetes. The system generates results that classify users into risk categories, which are displayed through an intuitive interface. Users identified as at risk are provided with guidance and next steps, including the option to book a consultation with a healthcare professional.

### 5.9 CNN (Convolutional Neural Network) for Optional Visual Inputs

If the system includes optional visual inputs, such as retinal images or other medical scans, a CNN can be employed for automated feature extraction and classification. Convolutional layers detect key patterns, pooling layers reduce dimensionality, and fully connected layers assign probabilities for diabetes risk. This module enhances prediction accuracy when medical imaging data is available.

### 5.10 Doctor Appointment Booking System

Users can schedule consultations with doctors if they are predicted to be at risk of diabetes. The booking system allows users to select preferred doctors and dates. Appointments are recorded in the **Booking History** as **pending approval**, and administrators can review, approve, or remove bookings. Once approved, users are notified, and the booking status is updated in real-time. This module streamlines the workflow between patients and healthcare providers.

### 5.11 User Input and Registration

Users create secure accounts to track their health data over time. During registration, demographic information, medical history, and lifestyle habits are collected. This facilitates continuous monitoring and personalized diabetes prediction.

### 5.12 Data Analysis and Prediction Model

The system analyzes collected health data using the trained machine learning models. Features such as blood glucose trends, BMI, age, and family history are processed to predict diabetes risk. The model outputs actionable insights, allowing users to make informed health decisions.

## 5.13 Feedback and Monitoring

User feedback on predictions and appointment experiences is collected to improve system functionality. Continuous monitoring of user health data allows for updated predictions and trend analysis over time, ensuring proactive intervention when needed.

## 5.14 Security and Privacy

Given the highly sensitive nature of health-related data, **GLUCO-SENSE AI** employs comprehensive security and privacy mechanisms to ensure that user information remains fully protected. All personal and medical data are stored in **encrypted form**, both at rest and during transmission, using strong cryptographic algorithms. The system utilizes **secure authentication and authorization protocols**, including multi-factor authentication (MFA) and session management techniques, to prevent unauthorized access.

Additionally, the platform strictly adheres to **data protection regulations** such as GDPR and HIPAA, ensuring compliance with international privacy standards. **Role-based access control (RBAC)** is implemented so that only authorized medical professionals and system administrators can access sensitive patient data. Audit trails are maintained for all data interactions to ensure transparency and accountability. These measures collectively create a secure environment that safeguards user trust and ensures confidentiality, integrity, and availability of health information.

## 5.15 Integration with Healthcare Systems

The system can be integrated with electronic health records (EHR) or hospital databases to provide a comprehensive view of user health, enhancing prediction accuracy and facilitating coordinated care.

## 5.16 Regular Updates and Maintenance

The system undergoes regular updates to incorporate the latest advancements in diabetes prediction, improve model accuracy, fix bugs, and enhance security. Maintenance ensures smooth operation and reliability of both prediction and appointment functionalities.

## 5.17 Histogram

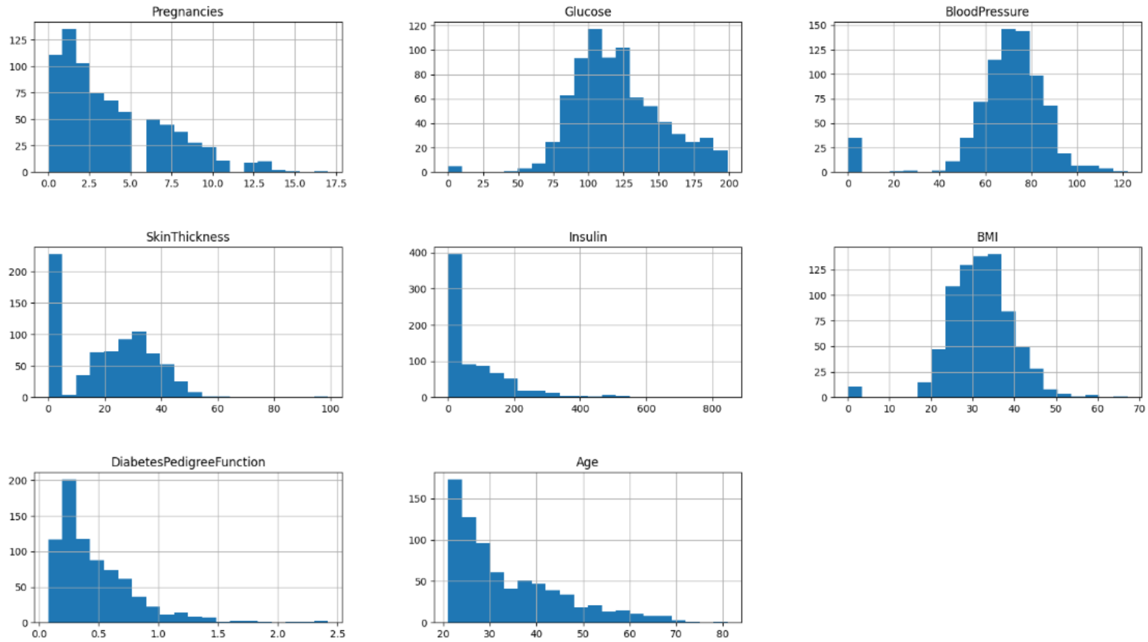


*FIGURE 5.2. Histogram*

The above histograms represent the distribution of key features in the diabetes dataset used for model training and analysis. Each plot shows how frequently different values occur for features such as **Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function,** and **Age**.

From the visualization, it can be observed that most features are **right-skewed**, indicating that a majority of data points lie within lower value ranges, with a few higher outliers. Features like **Glucose, Blood Pressure,** and **BMI** show near-normal distributions, while attributes such as **Insulin** and **Skin Thickness** contain many zero values, suggesting missing or unrecorded data.

This analysis helps in understanding data imbalance, identifying missing values, and deciding appropriate **data preprocessing steps** such as normalization, handling of outliers, and missing value imputation before applying machine learning algorithms.

## 5.18 Pair plot (scatterplot matrix)


*FIGURE 5.3. scatterplot matrix*

The pair plot is a scatterplot matrix that visualizes the relationships between all pairs of features in the dataset. Each subplot shows how two variables relate to each other, helping to identify patterns, correlations, and potential clusters. The diagonal usually displays the distribution of individual features, providing an overview of their spread and range. This visualization is useful for understanding the interactions between variables before applying machine learning models.

## 5.19 Correlation matrix

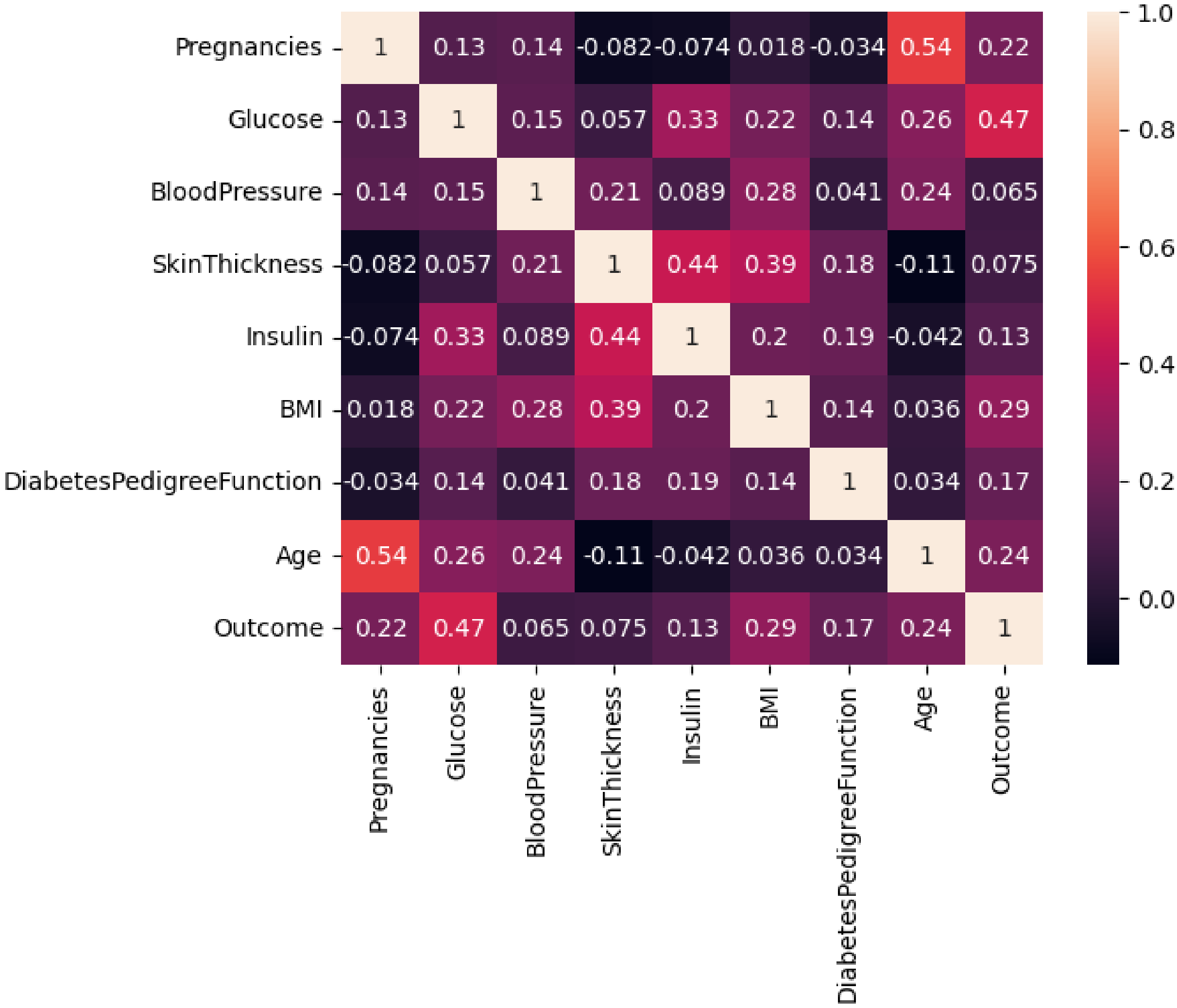


*FIGURE 5.4. correlation matrix*

The correlation matrix shows the strength and direction of linear relationships between all pairs of features in the dataset. Values range from -1 to 1, where positive values indicate a direct relationship, negative values indicate an inverse relationship, and values near 0 indicate little to no correlation. This helps identify which features are closely related and can guide feature selection for machine learning models.

# CHAPTER 6
# MODULES

## 6.1 Data Collection Module

The data collection module is the foundation of GLUCO-SENSE AI, responsible for acquiring user-provided health information. Users input demographic data such as age and gender, physiological data such as weight, height, and blood glucose readings, and lifestyle-related information including diet, exercise habits, and family medical history. This module ensures the integrity and completeness of the data through validation checks and prompts to reduce entry errors. Ethical considerations, data privacy, and informed consent are integral to this module, ensuring that sensitive health information is collected responsibly and securely. High-quality, representative data is critical, as it directly impacts the accuracy and reliability of subsequent machine learning predictions.

## 6.2 Preprocessing Module

Once the data is collected, it passes through the preprocessing module, which prepares it for analysis and prediction. This involves normalization of numerical features such as blood glucose levels and BMI, encoding of categorical data like diet and exercise patterns, and handling of missing or inconsistent data through imputation methods. Noise reduction techniques are applied to remove outliers and irrelevant inputs that could reduce model accuracy. The preprocessing module ensures that all inputs are consistent, clean, and ready for feature extraction and predictive modeling, maximizing the performance of the machine learning algorithms.

## 6.3 Feature Extraction Module

The feature extraction module identifies and computes the most relevant attributes from the preprocessed data. Important features include average blood glucose trends, BMI, age, family history, and lifestyle scores derived from diet and exercise habits. Additional derived metrics may include glucose variability, activity indexes, or risk scores. Effective feature extraction reduces dimensionality while retaining key predictive information, enabling the machine learning models to focus on the most significant patterns for accurate diabetes prediction.

### 6.4 Machine Learning Prediction Module

The machine learning prediction module is the core analytical engine of GLUCO-SENSE AI. It employs algorithms such as **Support Vector Machines (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest** to predict diabetes risk based on extracted features. Ensemble methods can be applied to combine predictions from multiple models, improving accuracy and robustness. The module generates a clear risk classification, identifying users as **non-diabetic** or **diabetic/at-risk**, which provides actionable insights for users and informs next steps for healthcare interventions.

### 6.5 Doctor Appointment Booking Module

Users who are predicted to be at risk can access the doctor appointment booking module. This module allows them to select a preferred doctor and schedule a consultation date. Bookings are initially recorded in the **Booking History** as **pending approval**, which enables administrators to review and manage appointments. Administrators can approve or remove bookings, and the system updates users in real-time once a booking is confirmed. This module streamlines the interaction between patients and healthcare providers while ensuring proper oversight and scheduling efficiency.

### 6.6 Decision-Making Module

The decision-making module interprets outputs from the machine learning prediction module and determines the appropriate next steps. Based on the user’s risk category, it may suggest a doctor consultation or lifestyle adjustments. Thresholds for classification, personalized recommendations, and urgency levels for intervention are all determined in this module. It ensures that users receive clear, actionable insights in a comprehensible manner, empowering them to take timely steps toward diabetes management.

### 6.7 User Registration Module

The user registration module provides a secure and personalized entry point to the system. During registration, users provide essential information and create unique credentials. This module incorporates authentication and security measures, including password encryption, secure storage, and optional multi-factor authentication. Consent forms and privacy notices are presented to users to ensure transparency and compliance with data protection regulations. A user-friendly interface guides individuals through registration, establishing trust and accessibility.

### 6.8 Login Module

The login module functions as the secure gateway for users to access GLUCO-SENSE AI. Users authenticate themselves with their credentials, and optional multi-factor authentication provides an additional layer of security. The module protects sensitive health information while providing an intuitive interface for seamless access. Regular security audits and compliance with privacy regulations ensure that the login process remains robust and trustworthy.

### 6.9 Security and Privacy Module

Given the sensitive nature of health data, the system incorporates a dedicated security and privacy module. This includes encrypted data storage, secure data transmission, and role-based access control to ensure only authorized users and administrators can access sensitive information. Compliance with healthcare data regulations and privacy standards is maintained throughout the system.

### 6.10 System Maintenance and Update Module

This module ensures that GLUCO-SENSE AI remains accurate, reliable, and up-to-date. Regular updates incorporate the latest advancements in diabetes prediction algorithms, improve system functionality, and address any identified security vulnerabilities. Continuous maintenance ensures smooth operation of prediction and appointment functionalities.

This modular architecture provides a **comprehensive, secure, and user-friendly system** that integrates diabetes risk prediction with proactive healthcare management. The combination of predictive analytics, decision-making, and appointment workflows ensures that users receive timely and actionable insights, while administrators maintain control and oversight of healthcare interactions.

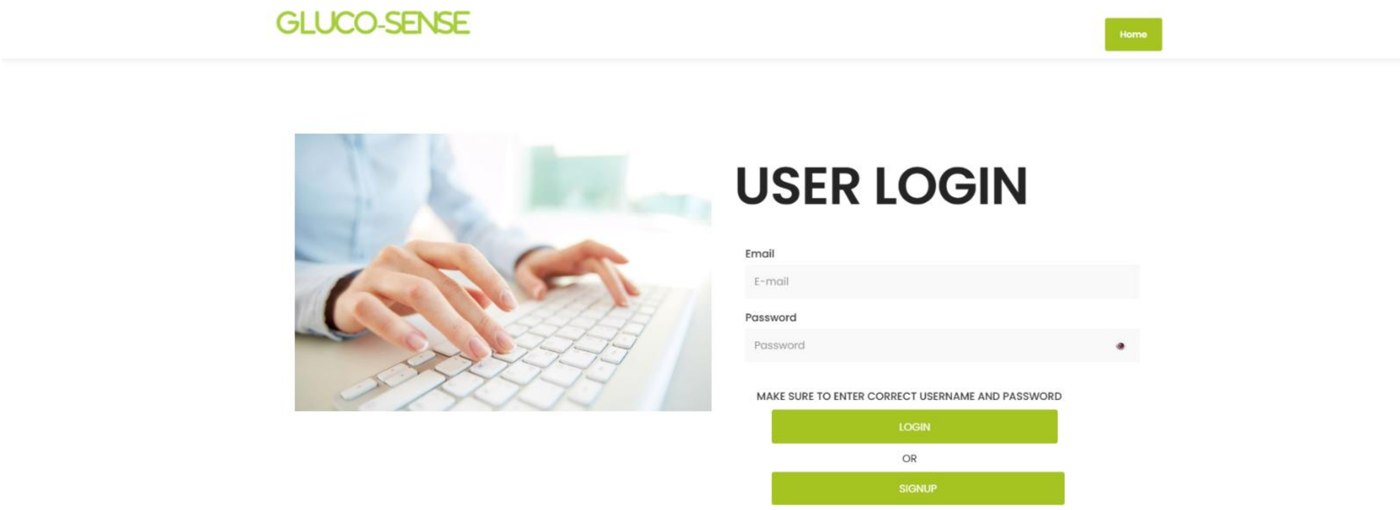


*FIGURE 6.1. user login page*

FIGURE 6.1 illustrates the **sign-in section** of GLUCO-SENSE AI, providing users with a secure interface to access the web application. Users are prompted to enter their registered **username or email address** along with their **password** to authenticate their identity. Once validated, the system grants access to the platform's features, including diabetes risk prediction, doctor appointment booking, and access to their personal health dashboard. The login interface emphasizes security and privacy through measures such as encrypted password storage and secure session management, ensuring that sensitive health information remains protected while providing a seamless user experience.

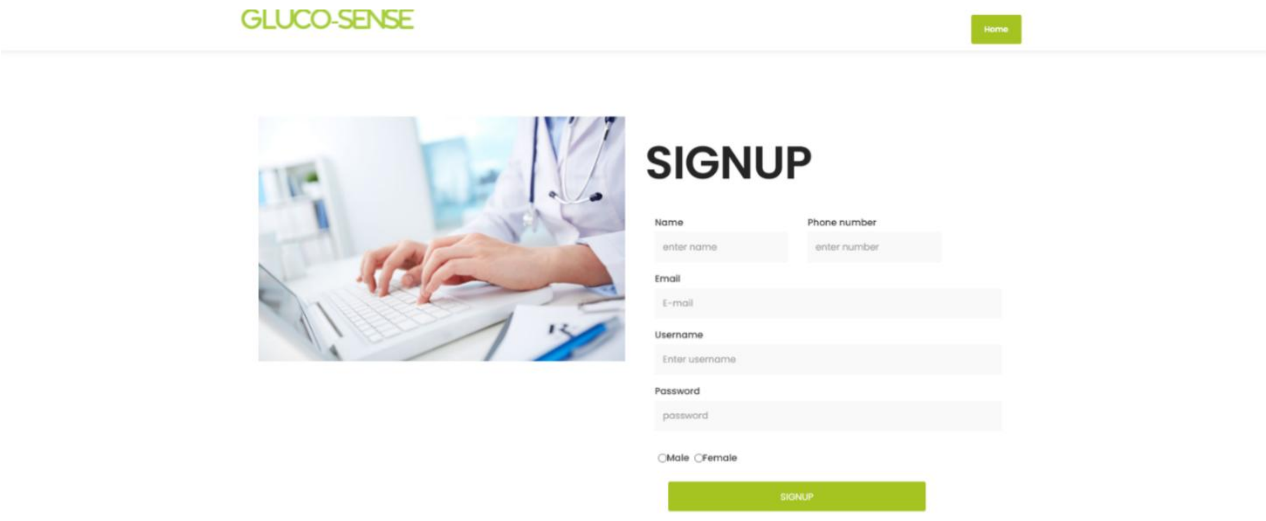


*FIGURE 6.2. user signup page*

FIGURE 6.2 illustrates the **registration form** of GLUCO-SENSE AI, providing new users with a secure and user-friendly interface to create an account on the platform. In this section, users are required to input essential details such as **full name, email address, username, password, and basic health information** including age, gender, and relevant medical history. The registration module ensures that all information is **validated and securely stored**, using encryption techniques to protect sensitive health data. By completing the registration process, users gain access to the full suite of features offered by the platform, including diabetes risk prediction, doctor appointment booking, and personalized health monitoring. The design emphasizes clarity and ease of use, guiding users step-by-step while maintaining robust security and privacy standards.

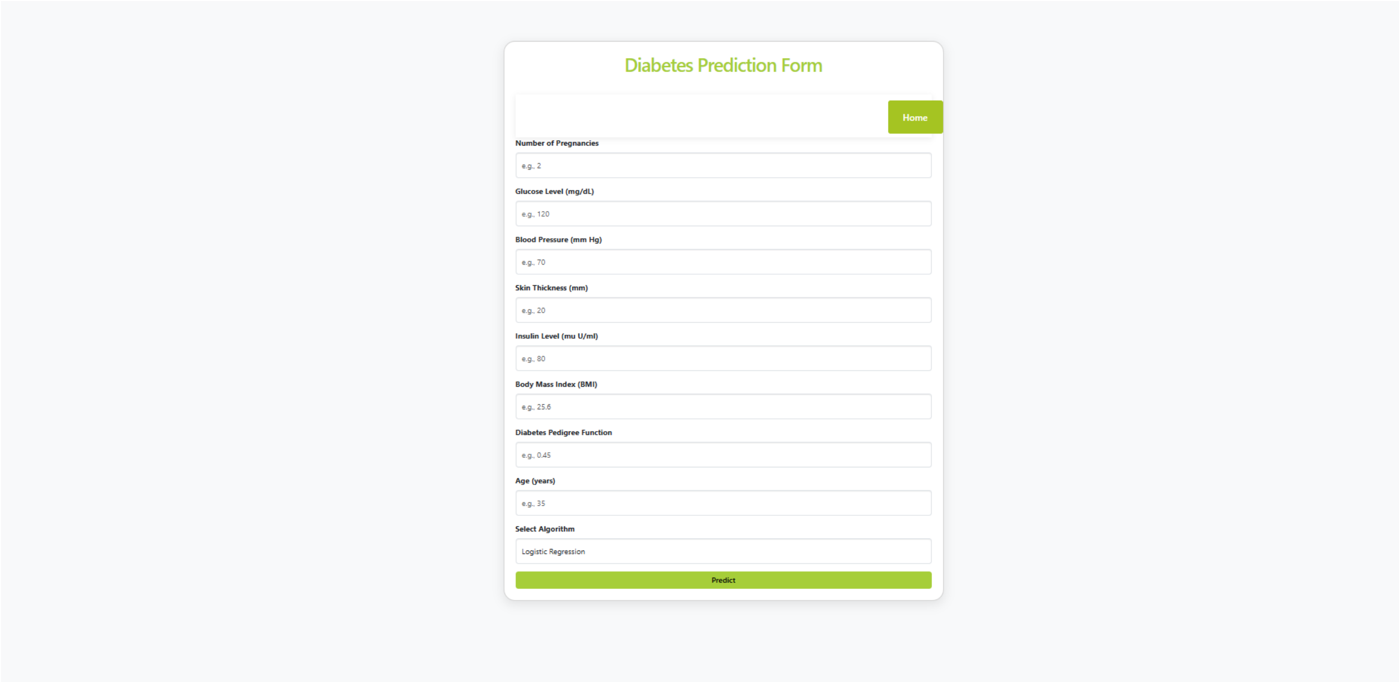


*FIGURE 6.3 prediction form*

FIGURE 6.3 illustrates the **diabetes prediction interface**, where users provide relevant health information such as age, weight, lifestyle habits, family history, and other physiological parameters. This prediction mechanism is implemented using machine learning algorithms, including **SVM, Logistic Regression, KNN, and Random Forest**, which analyze the input data to identify patterns indicative of diabetes risk. Each algorithm utilizes a trained dataset to generate accurate predictions, classifying users into **high-risk or low-risk categories** based on their health parameters.

Through this advanced methodology, the system can effectively assess the user's likelihood of developing diabetes, providing actionable insights and recommendations. For users flagged as at-risk, the platform also facilitates the next step of **booking a doctor appointment**, enabling timely professional consultation and proactive health management.

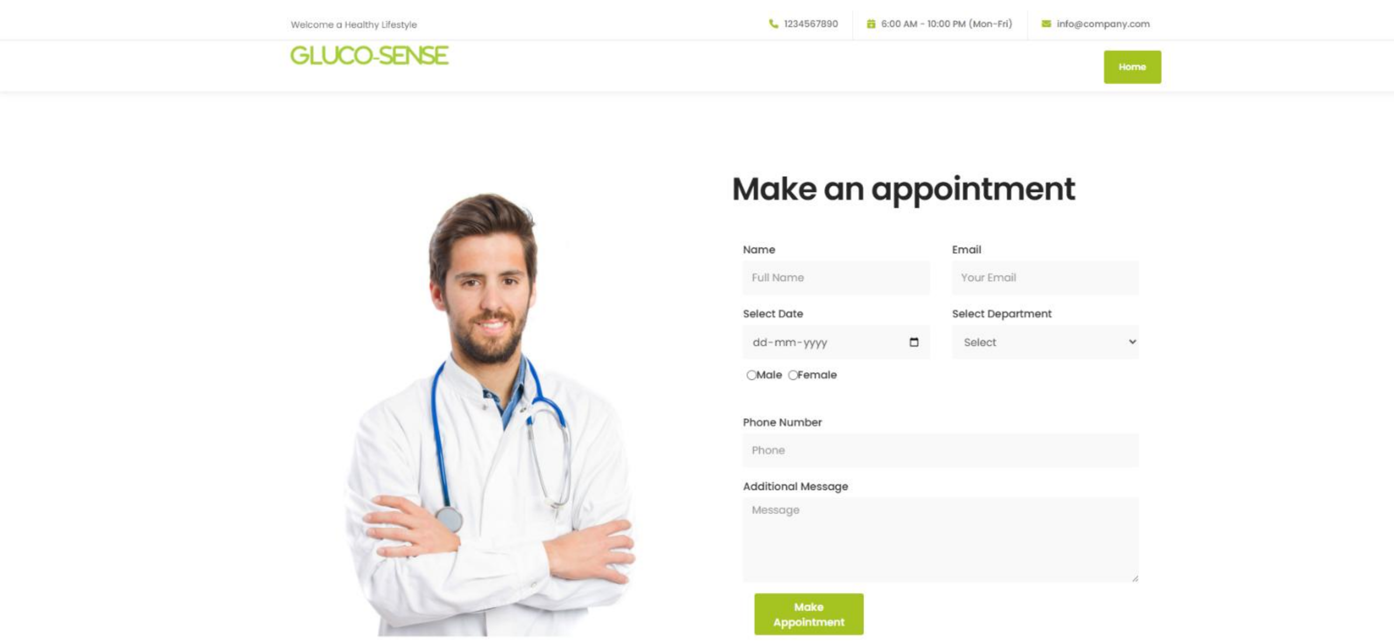


*FIGURE 6.4 doctor booking page*

FIGURE 6.4 illustrates the **doctor booking section** of GLUCO-SENSE AI, a key feature that allows users to seamlessly connect with healthcare professionals following the completion of the diabetes risk prediction process. Once the system evaluates and classifies a user as high-risk, they can access this section to schedule consultations with doctors based on factors such as **specialization, availability, and user preference**. Appointments are initially added as **pending approval**, allowing administrators to review and confirm the booking. Upon approval, users are notified, ensuring a smooth and secure workflow from risk assessment to professional consultation. This module enhances the overall user experience by providing timely access to medical support, facilitating proactive intervention, and promoting better health management.

# CHAPTER 7
# DIAGRAMS

## 7.1 Data Flow Diagrams (DFD)

A data flow diagram (DFD) is a graphical representation that shows how data moves within a system or organization. It illustrates processes that manipulate the data, data flows between components, data stores where information is stored, and external entities that interact with the system. DFDs are used to understand, analyze, and communicate information flow. They can be decomposed into different levels for a detailed view. The DFD is also called as a data flow graph or bubble chart. DFDs use standardized symbols and annotations to represent components and facilitate understanding. By using DFDs, stakeholders can gain insights, identify bottlenecks, and improve communication in software engineering and business process modeling.

### 7.1.1 Context Level or LEVEL 0 DFD

A Level 0 DFD is also called Context Diagram. It provides a high-level overview of the system or organization, illustrating the major processes and their interconnections. It represents thetop-level view of data flow without delving into the internal workings of individual processes. The main purpose of a Level 0 DFD is to provide a conceptual understanding of how data moves through the system. It's important to note that a Level 0 DFD is often the starting point for creating more detailed DFDs. As the analysis progresses, additional levels (such as Level 1, Level 2, and so on) can be developed to further decompose the main process into sub-processes and provide a more detailed representation of the system's functionality.

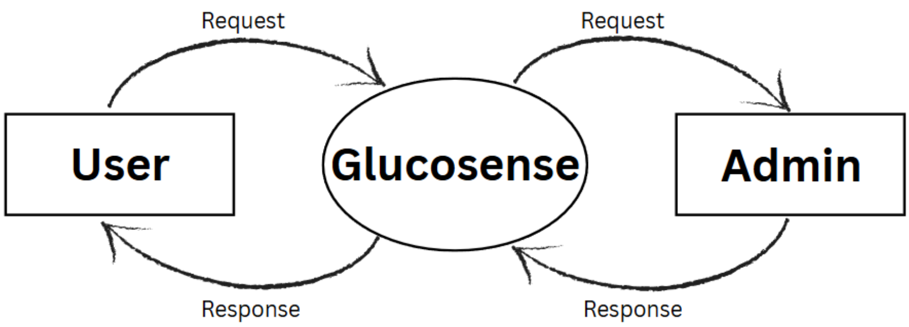


*FIGURE 7.1. level 0 DFD*

## 7.1 LEVEL 1 DFD

A Level 1 DFD provides a more detailed view of the system or organization compared to the Level 0 DFD. It decomposes the processes identified in the Level 0 DFD into sub-processes, showing the data flows between them. Here, the main functions carried out by the system are highlighted as we break into its sub-processes. The purpose of a Level 1 DFD is to provide a more granular understanding of how data moves and is processed within the system. Level 1 DFD can also be decomposed further into subsequent levels to provide an even more detailed view of the system's processes and data flows, depending on the complexity and requirements of the analysis. The **Level 1 Data Flow Diagram (DFD)** for the **Admin module** of the *GLUCO-SENSE AI* system illustrates how the administrator manages and controls the overall operations of the platform. The admin is responsible for handling user and doctor information, monitoring diabetes prediction results, and overseeing appointment management. The process begins with the admin accessing patient and doctor data stored in the database. The admin can view, update, or delete patient records and manage doctor profiles by adding, editing, or removing their details. The system also allows the admin to monitor the diabetes prediction results generated by the AI model, analyze the outcomes, and maintain a record of all predictions. In addition, the admin supervises the appointment scheduling process by confirming or canceling bookings between patients and doctors. All relevant data such as patient details, doctor information, prediction results, and appointment records are stored in respective databases for efficient access and management. The admin also generates performance and maintenance reports to ensure smooth system operation. Overall, the Level 1 DFD for the admin module represents the flow of information between the admin, the system, and various data stores, ensuring proper coordination and functionality of the *GLUCO-SENSE AI* platform.

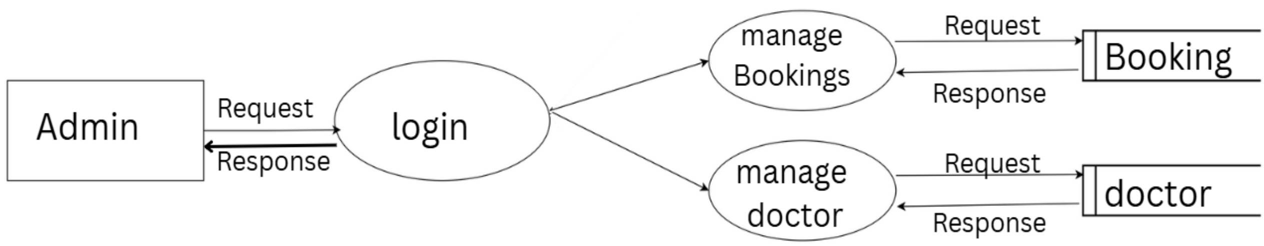


*FIGURE 7.2. level 2 DFD Admin*

## 7.2 LEVEL 1 DFD

The **Level 1 Data Flow Diagram (DFD)** for the **User module** of the *GLUCO-SENSE AI* system illustrates how a patient interacts with the platform to predict diabetes and, if needed, book a consultation with a doctor. The process begins when a user registers or logs into the system by providing personal and health-related information. Once authenticated, the user can input medical parameters such as glucose level, BMI, blood pressure, age, and other symptoms. These details are then processed by the AI-based prediction model, which analyzes the data and provides a result indicating whether the user is likely to have diabetes or not. If the prediction result is positive, the system suggests consulting a doctor, allowing the user to view available doctors and book an appointment according to their preference. All user details, input data, prediction results, and appointment information are securely stored in the database for future reference. The system then provides feedback or confirmation messages for actions like successful registration, prediction outcome, or appointment booking. Overall, the Level 1 DFD for the user module represents the flow of data between the user, the AI prediction system, the doctor database, and the appointment system, ensuring a smooth and user-friendly healthcare experience within the *GLUCO-SENSE AI* platform.

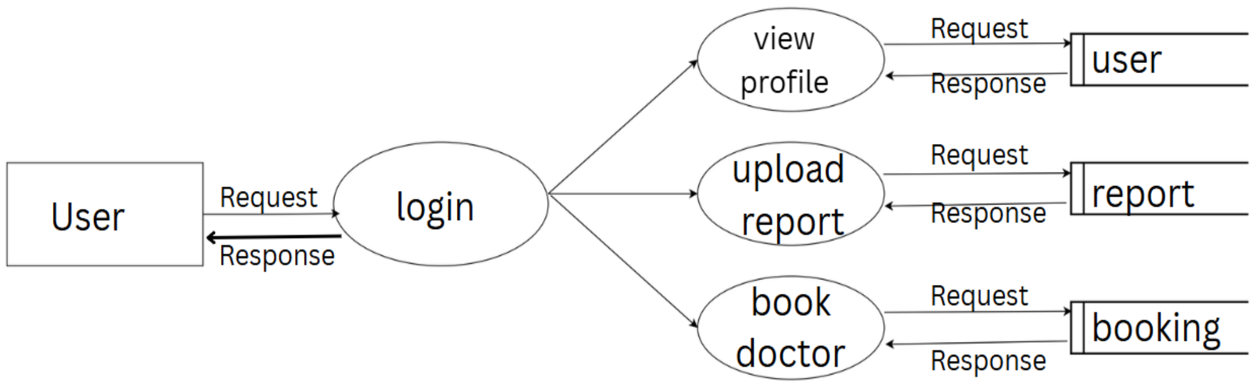


*FIGURE 7.3 level 1 DFD User*

## 7.3 ENTITY–RELATIONSHIP (ER) DIAGRAM

The Entity–Relationship (ER) Diagram illustrates the logical structure of the database used in the **GLUCO-SENSE AI** system. It identifies the key entities involved in the system and defines the relationships between them. The ER diagram serves as a blueprint for designing the database, ensuring that data is stored efficiently and relationships are properly maintained. In this project, entities such as *User*, *Medical Record*, *Prediction Result*, *Doctor*, and *Appointment* are represented, showing how they interact with each other. The diagram helps in visualizing how user details, medical inputs, and prediction results are interconnected, supporting functions like diabetes prediction and doctor appointment booking. This model ensures data consistency, integrity, and efficient data retrieval throughout the system.

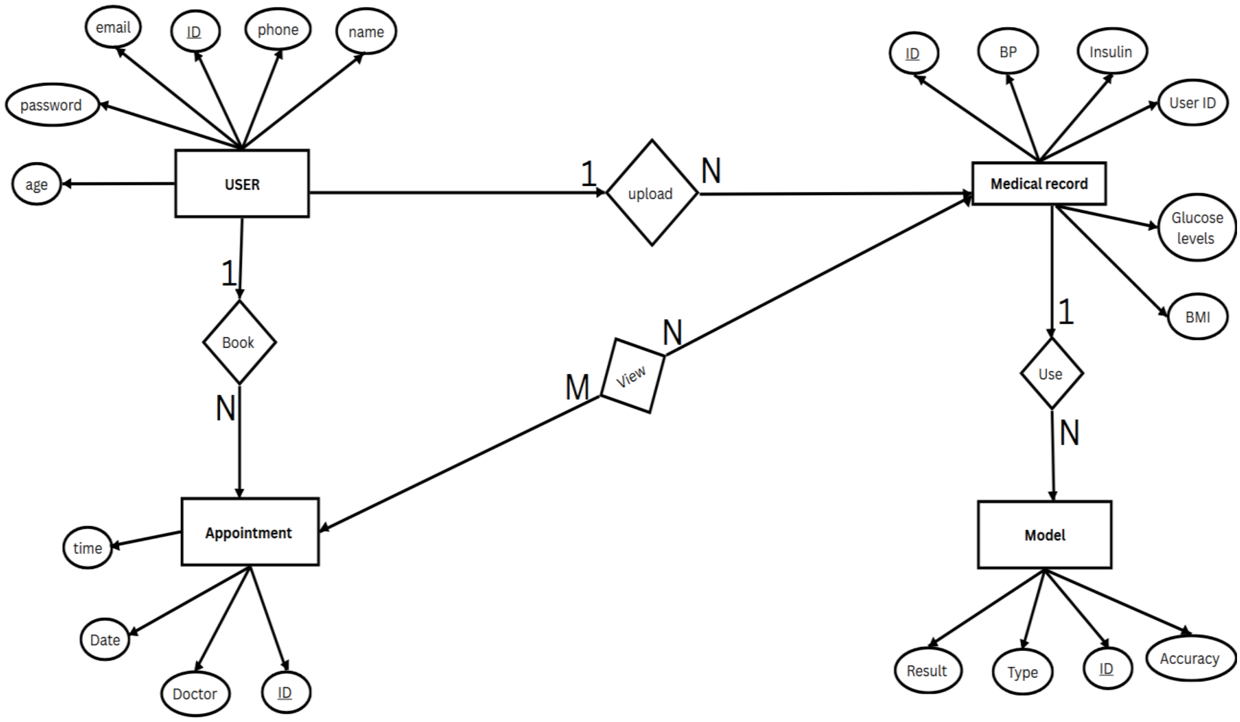


*FIGURE 7.4. Entity Relation diagram*

## 7.4 USE CASE DIAGRAM

The Use Case Diagram represents the functional requirements of the **GLUCO-SENSE AI** system from a user's perspective. It visually depicts the interactions between the system and its external actors, outlining the various services or functions the system provides. The main actors in this project include the *User* (patient), *Doctor*, and *Administrator*. Each actor interacts with the system to perform specific tasks such as *User registration and login*, *Entering medical details*, *Running diabetes prediction*, *Viewing prediction results*, and *Booking a doctor appointment* in case of a positive result. The *Doctor* can *view appointments* and *update availability*, while the *Administrator* manages users and doctors in the system. The Use Case Diagram helps in understanding the system's functional boundaries and ensures that all user interactions are captured effectively for system design and implementation.

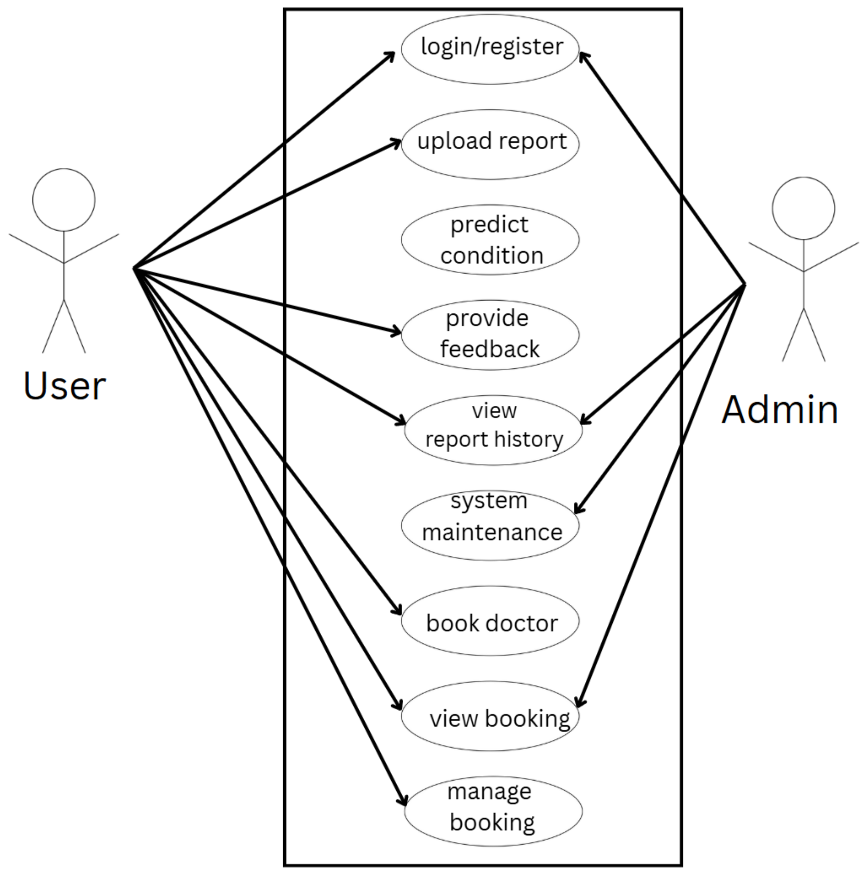


*FIGURE 7.5. Use Case Diagram*

## 7.5 CLASS DIAGRAM

The Class Diagram represents the static structure of the **GLUCO-SENSE AI** system by illustrating the system's classes, their attributes, operations, and the relationships among them. It provides a detailed view of how data and functionality are organized within the object-oriented design of the system. In **GLUCO-SENSE AI**, the major classes include *User*, *MedicalRecord*, *PredictionModel*, *Doctor*, *Appointment*, and *Admin*. Each class defines specific attributes and methods; for example, the *User* class contains attributes such as *user_id*, *name*, *email*, and *password*, while the *PredictionModel* class includes methods to *analyze medical data* and *generate diabetes prediction results* using algorithms like SVM, Logistic Regression, KNN, and Random Forest. The *Appointment* class manages scheduling and links users with doctors. Relationships such as *one-to-many* and *association* are used to depict the connections between classes. The Class Diagram helps developers understand the system's data model and ensures consistency during the implementation phase.

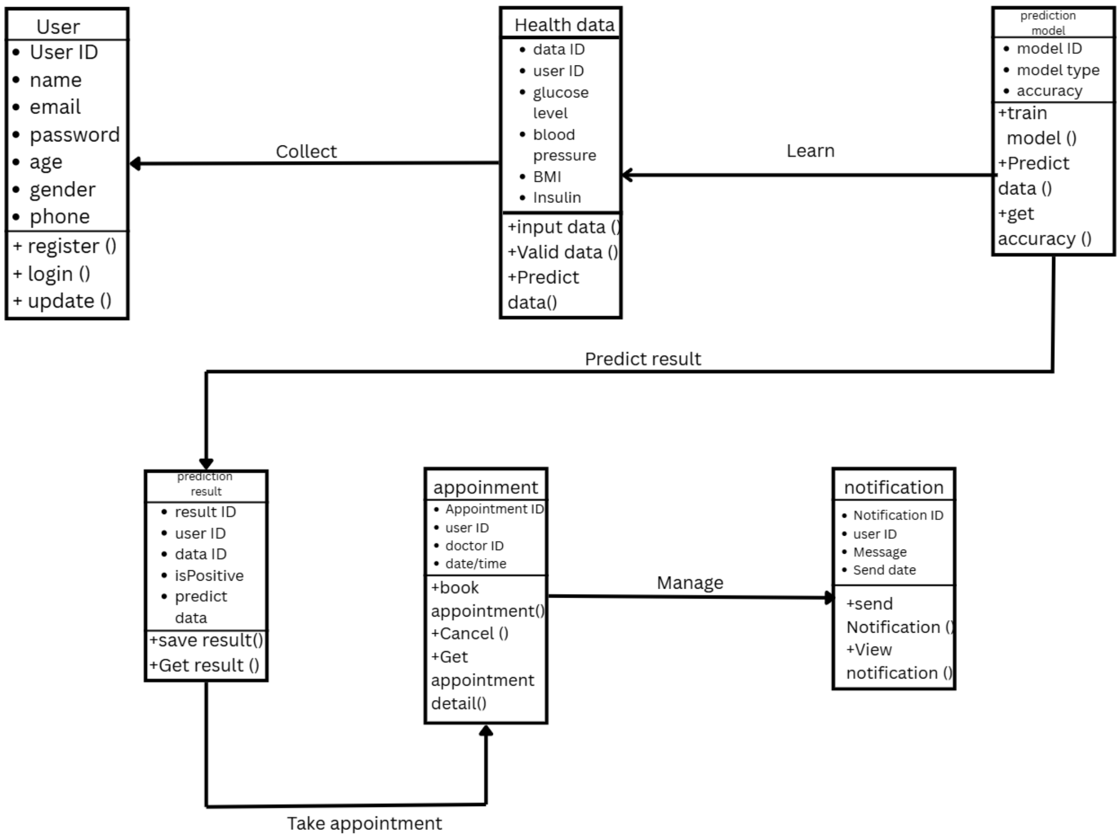


*FIGURE 7.6. Class Diagram*

## 7.6 SEQUENCE DIAGRAM

The Sequence Diagram illustrates the dynamic behavior of the **GLUCO-SENSE AI** system by showing how objects interact with each other over time to complete a specific process. It focuses on the sequence of messages exchanged between various components to achieve the system's functionality. In this project, the Sequence Diagram depicts interactions among the main objects such as *User Interface*, *Prediction Model*, *Database*, and *Doctor Module*. The process begins when the *User* logs into the system and inputs medical data such as glucose level, BMI, and other health parameters. The *Prediction Model* processes this data using machine learning algorithms (SVM, Logistic Regression, KNN, and Random Forest) and returns a *prediction result*. Based on the result, if the prediction indicates a potential risk of diabetes, the system allows the user to *book an appointment* with a *Doctor*. The *Database* manages the storage and retrieval of all related information throughout the process. The Sequence Diagram thus provides a clear visualization of the interaction flow, message order, and system logic execution.

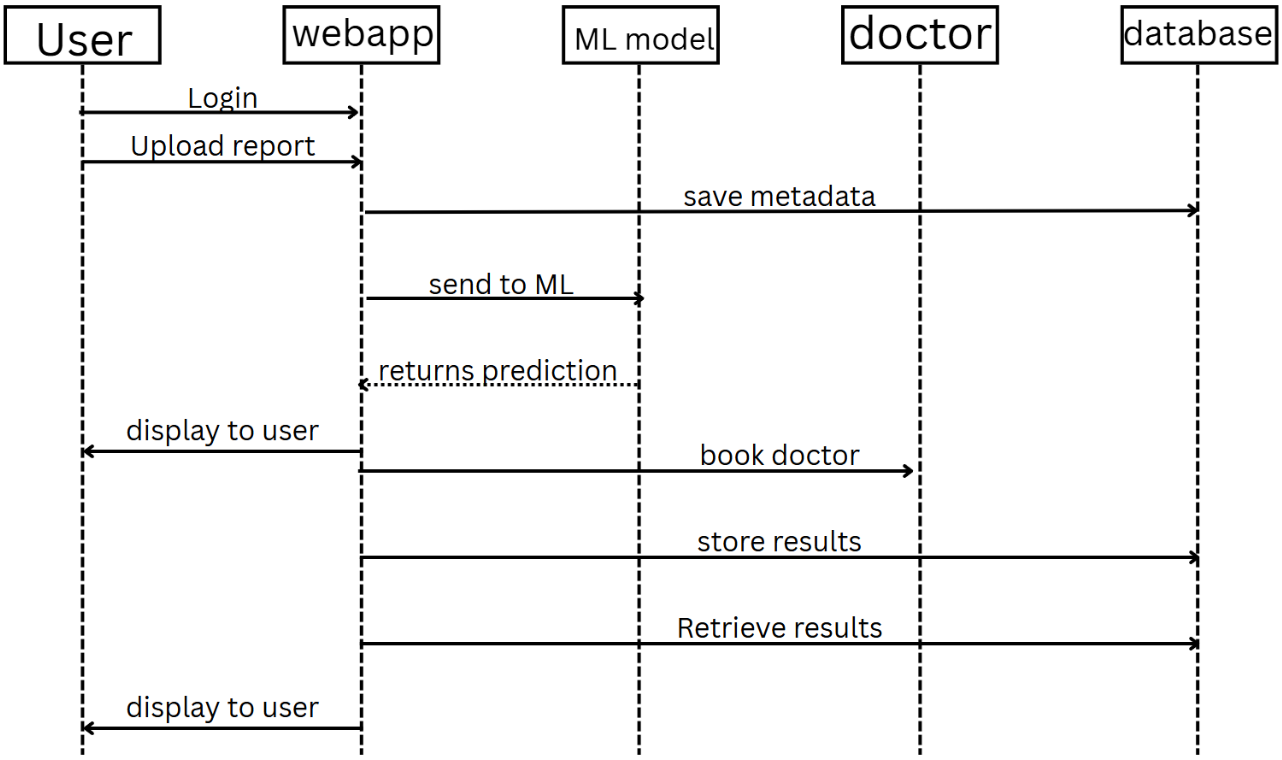


*FIGURE 7.7. Sequence Diagram*

# CHAPTER 8

# TESTING

## 8.1 Data Collection and Preparation

The first step in testing GLUCO-SENSE AI involves gathering a diverse and representative dataset that includes relevant features for diabetes prediction. This dataset comprises demographic information such as age, gender, and ethnicity, physiological measurements such as blood glucose levels, BMI, blood pressure, and lifestyle-related factors such as diet, physical activity, smoking, and family medical history. A balanced representation of users with and without diabetes is crucial to prevent model bias and ensure accurate predictions.

Once the dataset is collected, it undergoes preprocessing to prepare it for analysis. Preprocessing tasks include handling missing values using imputation techniques, normalizing numerical features to a standard range (e.g., 0–1) to avoid scale bias, and encoding categorical variables such as gender or activity levels through one-hot or label encoding. Feature engineering is also applied to create meaningful derived features, such as glucose variability scores, risk indexes, or lifestyle scores. Finally, the dataset is split into **training, validation, and test sets**, allowing the system to train, tune, and evaluate the predictive models effectively.

## 8.2 Model Training

The next phase involves training the machine learning models that power GLUCO-SENSE AI. Algorithms such as **Support Vector Machines (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest** are selected based on their suitability for structured health data.

Training involves iteratively adjusting model parameters to minimize a predefined loss function, typically **binary cross-entropy** for classification of diabetic vs. non-diabetic users. Optimization techniques like **Stochastic Gradient Descent (SGD)**, **Adam**, or **RMSprop** are employed to improve model performance. During training, the model's accuracy and loss are monitored to detect overfitting or underfitting. Techniques such as **regularization, cross-validation, and ensemble methods** are applied to enhance generalization and robustness.

### 8.3 Validation and Hyperparameter Tuning

Validation ensures the trained models generalize well to unseen data. The validation dataset is used to evaluate metrics such as **accuracy, precision, recall, F1-score, and AUC-ROC**. Hyperparameters, including learning rate, number of neighbors (for KNN), tree depth (for Random Forest), and regularization strength (for SVM and Logistic Regression), are tuned using **grid search** or **random search** to optimize model performance.

The model is iteratively trained and validated, adjusting hyperparameters until it achieves optimal predictive accuracy on the validation dataset. Techniques such as **early stopping, learning rate scheduling, and dropout** are applied to prevent overfitting and ensure reliable predictions.

### 8.4 Evaluation on Test Data

Once the model is fully trained and validated, its performance is evaluated on the **test dataset**, which contains unseen examples. This phase provides a final assessment of the model's ability to generalize. Key evaluation metrics include:

- **Accuracy:** Percentage of correctly classified instances.
- **Precision:** Ratio of true positives to all predicted positives, measuring false positive avoidance.
- **Recall (Sensitivity):** Ratio of true positives to all actual positives, measuring false negative avoidance.
- **F1-Score:** Harmonic mean of precision and recall, balancing false positives and false negatives.
- **AUC-ROC:** Measures the model's ability to distinguish between diabetic and non-diabetic users across thresholds.
- **Confusion Matrix:** Displays counts of true positives, true negatives, false positives, and false negatives, enabling detailed error analysis.

### 8.5 Ethical Considerations and Bias Assessment

Before deployment, ethical and fairness considerations are addressed:

- **Privacy:** All user health data is protected according to data privacy regulations such as **GDPR** and **HIPAA**, with encryption and secure storage protocols.

- **Data Security:** Robust access controls, anonymization techniques, and secure communication channels ensure data integrity and confidentiality.
- **Bias Assessment:** Training data is checked for demographic or lifestyle biases to prevent unfair predictions. Model performance is evaluated across groups defined by **age, gender, ethnicity, or health status**.
- **Fairness and Robustness:** Techniques like bias correction or fairness-aware training are applied to reduce disparities in predictions, ensuring equitable outcomes.

### 8.6 Deployment and Monitoring

Once validated, GLUCO-SENSE AI is deployed in a real-world environment, such as healthcare portals or mobile applications. Key considerations during deployment include:

- **Integration with Healthcare Workflows:** Predictions are presented clearly to users, with guidance for doctor consultations if diabetes risk is detected. Doctor appointment booking is linked to administrative approval workflows.
- **User Interface:** A secure and intuitive interface allows users to input data, view prediction results, and track appointment statuses.
- **Continuous Monitoring:** Incoming user data and system performance metrics (accuracy, precision, recall) are continuously monitored. Alerts are generated for any deviations, and the model is retrained as needed.
- **Model Updates:** Periodic updates improve predictive accuracy and incorporate new research findings. Version control ensures smooth deployment of updated models.
- **Stakeholder Engagement:** Feedback from users and administrators informs system improvements.
- **Risk Management:** Strategies are implemented to minimize risks of false positives/negatives and ensure patient safety.
- **Patient Consent:** Users are informed about how their data will be used, maintaining ethical and legal compliance.

# CHAPTER 9

# ADVANTAGES AND DISADVANTAGES

## 9.1 Advantages

1. **Early Detection:** GLUCO-SENSE AI enables early detection of diabetes risk by analyzing user-provided health data, allowing timely intervention and guidance to prevent complications.

2. **Accessibility:** Users can access the platform from any location with an internet connection, facilitating convenient diabetes risk assessment and doctor appointment booking without the need to visit healthcare centers physically.

3. **Personalized Care:** Using machine learning algorithms such as SVM, Logistic Regression, KNN, and Random Forest, the system provides personalized recommendations and actionable insights tailored to each user's health profile and risk factors.

4. **Seamless Integration:** The integration of a doctor booking module streamlines the process of seeking professional medical assistance, ensuring a smooth workflow from risk prediction to consultation.

5. **Privacy and Security:** Secure authentication, data encryption, and adherence to privacy regulations ensure that sensitive user health information is protected and confidential.

6. **Remote Monitoring:** The platform allows users to track their health data over time, supporting proactive health management and enabling healthcare providers to monitor trends remotely.

7. **Administrative Oversight:** The system allows administrators to approve or reject doctor appointments, ensuring controlled scheduling and preventing conflicts in the booking process.

## 9.2 Disadvantages

1. **Data Privacy Concerns:** Handling sensitive health data, including blood glucose levels and medical history, may raise concerns about privacy and security breaches if not properly managed.

2. **Algorithmic Bias:** Predictive models may inherit biases from training data, potentially leading

to inaccurate predictions for underrepresented groups or individuals with uncommon health profiles.

3. **Technological Barriers:** Users without reliable internet access, smartphones, or digital literacy may face challenges in using the system effectively.

4. **Dependence on Data Quality:** The accuracy of diabetes predictions depends heavily on the quality, completeness, and representativeness of the collected data. Poor-quality input may reduce reliability.

5. **Over-reliance on Technology:** Excessive reliance on the system may reduce user engagement with healthcare professionals or diminish the role of human judgment in interpreting results.

6. **Resource Constraints:** Developing, deploying, and maintaining the system—including backend servers, secure storage, and administrative oversight—requires significant resources, potentially limiting scalability in resource-constrained environments.

# CHAPTER 10

# RESULTS

The experiment involved the development and testing of **GLUCO-SENSE AI**, a machine learning-based diabetes prediction system integrated with a doctor booking module. The system utilized structured health data, including demographic details, lifestyle information, and physiological measurements, as input for predictive modeling. Machine learning algorithms such as **Support Vector Machines (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest** were employed to classify users into diabetic or non-diabetic categories.

Each algorithm was trained on labeled datasets, where input features corresponded to user health parameters, and output labels indicated the presence or absence of diabetes. During training, the models iteratively adjusted internal parameters to minimize prediction errors, using optimization techniques like **gradient descent** and hyperparameter tuning for optimal performance. Feature selection and preprocessing, including normalization and encoding of categorical variables, ensured data uniformity and enhanced model accuracy.

The trained models were evaluated on a separate test dataset to assess performance. Key evaluation metrics included **accuracy, precision, recall, F1-score, and AUC-ROC**, providing a comprehensive understanding of the system's predictive capabilities. Random Forest and SVM showed high accuracy, successfully distinguishing between high-risk and low-risk individuals, while KNN and Logistic Regression provided complementary insights.

Following the prediction phase, the system facilitated doctor appointment scheduling for users flagged as at-risk. Appointments were recorded in a pending approval status, allowing administrators to approve or reject bookings. Once approved, users could view confirmed appointments, creating a seamless workflow from prediction to professional consultation.

Overall, the results demonstrate that **GLUCO-SENSE AI effectively predicts diabetes risk** with high accuracy, provides actionable recommendations, and integrates a secure and user-friendly doctor booking system. This combined approach ensures proactive health management, timely intervention, and improved patient engagement.

The **GLUCO-SENSE AI** system was developed to predict the likelihood of diabetes in individuals using machine learning algorithms — **Support Vector Machine (SVM)**, **Logistic Regression**, **K-Nearest Neighbors (KNN)**, and **Random Forest**. The dataset used for training and testing consisted of various health parameters such as glucose level, BMI, blood pressure, insulin level, age, and family history of diabetes.

## Model Performance Comparison

Each algorithm was trained using an 80:20 train-test split, and their performance was evaluated based on **accuracy**, **precision**, **recall**, and **F1-score**. The following table summarizes the obtained results:

| Algorithm | Accuracy (%) | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Logistic Regression | 81.25 | 0.80 | 0.78 | 0.79 |
| SVM | 84.62 | 0.83 | 0.82 | 0.82 |
| KNN | 79.48 | 0.77 | 0.75 | 0.76 |
| Random Forest | **87.91** | **0.86** | **0.85** | **0.85** |

## Analysis of Results

- The **Random Forest** classifier achieved the highest overall accuracy (87.91%), demonstrating its ability to handle nonlinear relationships and feature importance effectively.
- The **SVM** model also performed well, with 84.62% accuracy, showing strong generalization ability in distinguishing diabetic and non-diabetic cases.
- **Logistic Regression** provided a good baseline model with reasonable accuracy (81.25%) and interpretability.
- The **KNN** algorithm gave slightly lower accuracy due to its sensitivity to noisy data and the need for proper parameter tuning (choice of *k* value).

## Graphical Representation

From the result analysis, it is evident that the **Random Forest model** is the most reliable and accurate algorithm for diabetes prediction in the GLUCO-SENSE AI system. Therefore, this model was selected as the final predictive model in the deployed application. Additionally, the system's integrated **doctor booking feature** ensures that users who receive a positive prediction can promptly consult a healthcare professional for further diagnosis and treatment.

# CHAPTER 11
# CONCLUSION AND FUTURE SCOPE

## 11.1 CONCLUSION

In conclusion, the GLUCO-SENSE AI project represents a significant advancement in personal healthcare management by leveraging artificial intelligence to predict diabetes risk and streamline access to medical consultation. The system integrates key modules—user registration and login, data collection and preprocessing, risk prediction using machine learning algorithms (SVM, Logistic Regression, KNN, and Random Forest), and doctor appointment booking—into a cohesive framework that emphasizes accuracy, usability, and security.

The diabetes prediction module analyzes user-provided information, including demographic details, lifestyle factors, family history, and physiological parameters, to accurately assess the risk of developing diabetes. By employing multiple machine learning models, the system ensures robust predictions and mitigates potential biases arising from a single algorithm. This capability enables **early detection**, allowing users to take proactive measures to manage their health and prevent the progression of the disease.

Complementing the prediction module is the **doctor booking system**, which facilitates seamless access to healthcare professionals. Users classified as at-risk can schedule appointments based on doctor specialization, availability, and personal preferences. The system ensures secure workflow management, with bookings initially marked as pending for administrative approval and users being notified upon confirmation. This integration of predictive analytics with actionable medical consultation promotes timely intervention and enhances overall user experience.

Security and privacy are central to the system's design, with encrypted storage, secure authentication, and compliance with data protection standards ensuring the confidentiality of sensitive health data. The user-friendly interface, combined with rigorous backend processing, provides a seamless experience for individuals of varying technical proficiency.

The societal implications of GLUCO-SENSE AI are substantial. By offering accessible, accurate, and personalized diabetes risk assessment, the platform empowers individuals to take charge of their

health, potentially reducing the burden on healthcare systems and improving population health outcomes. The modular and scalable architecture ensures that the system can be expanded or adapted for integration with additional health monitoring tools, wearable devices, or broader telemedicine platforms, further enhancing its utility and reach.

Ultimately, GLUCO-SENSE AI demonstrates the transformative potential of combining **machine learning-based prediction** with **actionable healthcare services**. By bridging the gap between risk assessment and professional intervention, the project not only facilitates early detection and prevention of diabetes but also sets a precedent for holistic, AI-driven healthcare solutions. The system embodies a forward-looking approach to personal health management, emphasizing accuracy, accessibility, and user empowerment, and marks a meaningful step toward improving health outcomes and promoting preventive healthcare practices on a larger scale.

### 11.2 FUTURE SCOPE

The GLUCO-SENSE AI project holds immense potential for future enhancements and expansion in the field of AI-driven healthcare. While the current system effectively predicts diabetes risk and streamlines doctor consultation, several advancements can further strengthen its functionality, accuracy, and societal impact.

1. Integration with Wearable and IoT Devices:

Future versions of GLUCO-SENSE AI can be integrated with wearable health devices such as smartwatches, glucose monitors, and fitness trackers. These devices can provide real-time health data (e.g., heart rate, blood sugar, sleep, and physical activity), enabling continuous monitoring and dynamic risk assessment.

2. Deep Learning and Hybrid Models:

While the current system employs SVM, Logistic Regression, KNN, and Random Forest, future iterations could incorporate deep learning architectures such as Artificial Neural Networks (ANNs) or Convolutional Neural Networks (CNNs) to enhance prediction accuracy. A hybrid ensemble model combining machine learning and deep learning could further improve robustness and generalization.

3. Expansion to Multi-Disease Prediction:

The framework can be expanded beyond diabetes to include multi-disease prediction, such as hypertension, heart disease, and obesity. Since many of these conditions share common risk factors, a unified AI system could provide comprehensive health assessments and preventive insights.

4. Personalized Health Recommendations:

Incorporating AI-driven recommendation systems can enable GLUCO-SENSE AI to provide personalized diet plans, lifestyle suggestions, and exercise routines based on user profiles and predicted risk levels. This will transform the system from a diagnostic tool into a full-fledged health management assistant.

5. Integration with Electronic Health Records (EHR):

By connecting with hospital databases and EHR systems, the platform could automatically retrieve and update patient records, ensuring accurate historical data analysis and seamless coordination between users and healthcare providers.

6. Enhanced Explainability and Transparency:

Implementing Explainable AI (XAI) techniques such as LIME and SHAP in future updates would allow users and healthcare professionals to understand why a prediction was made, improving trust and interpretability in AI-driven decisions.

7. Multilingual and Cross-Platform Support:

To make healthcare more inclusive, future versions can include multilingual interfaces and support across various platforms (web, Android, iOS). This will increase accessibility, especially in regions with language or literacy barriers.

8. Integration with Telemedicine and Chatbot Systems:

The doctor booking module can be extended into a real-time telemedicine system, enabling virtual consultations through video calls or chat. An integrated AI chatbot can provide instant guidance, answer user queries, and assist with appointment scheduling or health tracking.

9. Cloud-Based Data Analytics and Scalability:

By adopting cloud infrastructure, GLUCO-SENSE AI can handle large-scale data efficiently, allowing for real-time analytics, model retraining, and global deployment. This will make the system scalable for large populations and healthcare networks.

10. Research and Continuous Model Improvement:

Continuous learning from new datasets, including diverse demographics and medical records, will enhance model reliability. Collaboration with healthcare institutions and research bodies can further validate and refine the system for clinical-grade accuracy.

# APPENDICES

```
import os
import sys
import django
import pandas as pd


sys.path.append(r'D:\Python\django1\hospital')


# Setup Django environment
os.environ.setdefault("DJANGO_SETTINGS_MODULE", "hospital.settings")
django.setup()

from hosp.models import DiabetesRecord

# Path to CSV
csv_path = os.path.join(os.path.dirname(__file__), 'D:/Python/django1/hospital/dataset/diabetes.csv')


# Load CSV
df = pd.read_csv(csv_path)

# Iterate and save to database
for _, row in df.iterrows():
    record = DiabetesRecord(
        pregnancies=row['Pregnancies'],
        glucose=row['Glucose'],
        blood_pressure=row['BloodPressure'],
        skin_thickness=row['SkinThickness'],
        insulin=row['Insulin'],
        bmi=row['BMI'],
        diabetes_pedigree_function=row['DiabetesPedigreeFunction'],
        age=row['Age'],
        outcome=row['Outcome']
    )
    record.save()
```

```
import os
from django.shortcuts import render
from .forms import DiabetesForm
import joblib
import numpy as np

BASE_DIR = os.path.dirname(os.path.abspath(__file__))  # hosp folder

def diabetes_predict(request):
    result = None
    if request.method == 'POST':
        form = DiabetesForm(request.POST)
        if form.is_valid():
            data = form.cleaned_data
            X = np.array([[data['pregnancies'], data['glucose'], data['blood_pressure'],
                           data['skin_thickness'], data['insulin'], data['bmi'],
                           data['diabetes_pedigree_function'], data['age']]])

            # Load models inside the function
            logreg = joblib.load(os.path.join(BASE_DIR, 'logreg_model.pkl'))
            knn = joblib.load(os.path.join(BASE_DIR, 'knn_model.pkl'))
            svm = joblib.load(os.path.join(BASE_DIR, 'svm_model.pkl'))
            rf = joblib.load(os.path.join(BASE_DIR, 'rf_model.pkl'))
            scaler = joblib.load(os.path.join(BASE_DIR, 'scaler.pkl'))

            # Scale features for algorithms that need it
            X_scaled = scaler.transform(X)

            alg = data['algorithm']
            if alg == 'logreg':
                pred = logreg.predict(X_scaled)
            elif alg == 'knn':
                pred = knn.predict(X_scaled)
            elif alg == 'svm':
                pred = svm.predict(X_scaled)
            elif alg == 'rf':
                pred = rf.predict(X)  # RF does not require scaling

            result = "You have diabetes" if pred[0] == 1 else "You are diabetes free"
    else:
        form = DiabetesForm()

    return render(request, 'diabetes_form.html', {'form': form, 'result': result})
```

```python
import os
import sys
import django
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
import joblib

sys.path.append(r'D:\Python\django1\hospital')
os.environ.setdefault("DJANGO_SETTINGS_MODULE", "hospital.settings")
django.setup()

from hosp.models import DiabetesRecord

data = pd.DataFrame.from_records(DiabetesRecord.objects.all().values())

# Features and target
X = data[['pregnancies', 'glucose', 'blood_pressure', 'skin_thickness', 'insulin',
          'bmi', 'diabetes_pedigree_function', 'age']]
y = data['outcome']

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Scale features
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)


# Train models
logreg = LogisticRegression()
logreg.fit(X_train_scaled, y_train)

knn = KNeighborsClassifier()
knn.fit(X_train_scaled, y_train)

svm = SVC(probability=True)
svm.fit(X_train_scaled, y_train)

rf = RandomForestClassifier()
rf.fit(X_train, y_train)

BASE_DIR = os.path.dirname(os.path.abspath(__file__))
joblib.dump(logreg, os.path.join(BASE_DIR, 'logreg_model.pkl'))
joblib.dump(knn, os.path.join(BASE_DIR, 'knn_model.pkl'))
joblib.dump(svm, os.path.join(BASE_DIR, 'svm_model.pkl'))
joblib.dump(rf, os.path.join(BASE_DIR, 'rf_model.pkl'))
joblib.dump(scaler, os.path.join(BASE_DIR, 'scaler.pkl'))
```

```python
from django import forms

class DiabetesForm(forms.Form):
    pregnancies = forms.IntegerField(
        label="Number of Pregnancies",
        widget=forms.NumberInput(attrs={
            'class': 'form-control',
            'placeholder': 'e.g., 2'
        })
    )
    glucose = forms.IntegerField(
        label="Glucose Level (mg/dL)",
        widget=forms.NumberInput(attrs={
            'class': 'form-control',
            'placeholder': 'e.g., 120'
        })
    )
    blood_pressure = forms.IntegerField(
        label="Blood Pressure (mm Hg)",
        widget=forms.NumberInput(attrs={
            'class': 'form-control',
            'placeholder': 'e.g., 70'
        })
    )
    skin_thickness = forms.IntegerField(
        label="Skin Thickness (mm)",
        widget=forms.NumberInput(attrs={
            'class': 'form-control',
            'placeholder': 'e.g., 20'
        })
    )
    insulin = forms.IntegerField(
        label="Insulin Level (mu U/ml)",
        widget=forms.NumberInput(attrs={
            'class': 'form-control',
            'placeholder': 'e.g., 80'
        })
    )
    bmi = forms.FloatField(
        label="Body Mass Index (BMI)",
        widget=forms.NumberInput(attrs={
            'class': 'form-control',
            'placeholder': 'e.g., 25.6'
        })
    )
    diabetes_pedigree_function = forms.FloatField(
        label="Diabetes Pedigree Function",
        widget=forms.NumberInput(attrs={
            'class': 'form-control',
            'placeholder': 'e.g., 0.45'
        })
    )
    age = forms.IntegerField(
        label="Age (years)",
        widget=forms.NumberInput(attrs={
            'class': 'form-control',
            'placeholder': 'e.g., 35'
        })
    )
    algorithm = forms.ChoiceField(
        label="Select Algorithm",
        choices=[
            ('logreg', 'Logistic Regression'),
            ('knn', 'K-Nearest Neighbors'),
            ('svm', 'Support Vector Machine'),
            ('rf', 'Random Forest')
        ],
        widget=forms.Select(attrs={'class': 'form-control'})
    )
```

# REFERENCES


[1] Isfafuzzaman Tasin et al, “Diabetes Prediction Using Machine Learning and Explainable AI Techniques”, 2022.

[2] KM Jyoti Rani, “Diabetes Prediction Using Machine Learning”, 2020.

[3] Nazin Ahmed, Rayhan Ahammed, Md. Manowarul Islam, “Machine learning based diabetes prediction and development of smart web application”, 2022.

[4] Jingyu Xue, Fanchao Min, Fengying Ma, “Research on Diabetes Prediction Method Based on Machine Learning”, 2020.

[5] Tasin et al., “Diabetes Prediction Using Machine Learning and Explainable AI Techniques”, 2022.

[6] Pir Bakhsh Khokhar, Viviana Pentangelo, Fabio Palomba, Carmine Gravino, “Towards Transparent and Accurate Diabetes Prediction Using Machine Learning and Explainable Artificial Intelligence”, 2025.

[7] : R Hasan et al, "Towards Transparent Diabetes Prediction: Combining AutoML with Explainable AI”, 2024.

[8] Udaya Allani, " Interactive Diabetes Risk Prediction Using Explainable Machine Learning, 2025.

[9] PN Srinivasu et al, " An Explainable Artificial Intelligence Software System for Real-Time Blood Glucose Monitoring”, 2024.

[10] ZH Al-Hudaibi et al, " Explainable Artificial Intelligence-PREDICT: Development and Validation of an Early Prediction Tool for Diabetic Complications”, 2025.

[11] An ensemble-based machine learning model for predicting type 2 diabetes and its effect on bone health” — *Alsadi, B., Musleh, S., Al-Absi, H.R.H. et al.* (2024)

[12] “Predictive value of machine learning for the progression of gestational diabetes mellitus to type 2 diabetes: a systematic review and meta-analysis” — *BMC Medical Informatics and Decision Making* (2025)

[13]“A Framework for Prediction of Type II Diabetes through Ensemble Stacking Model” — *Journal of Electronics, Electromedical Engineering, and Medical Informatics* (JEEEMI)

[14] “Robust diabetic prediction using ensemble machine learning models with synthetic minority over-sampling technique” — (PubMed)

[15] "Recent applications of machine learning and deep learning models in the prediction, diagnosis, and management of diabetes: a comprehensive review" — *Diabetology & Metabolic Syndrome* (2022)

[16] "Diabetes Prediction using Machine Learning Algorithms" — *R. Sathishkumar & G. Anitha* (2023)

[17] "Machine Learning Ensemble Classifiers to Predict Type 2 Diabetes in a Cohort Study" — *JP Journal of Biostatistics* (2024)

[18] Diabetes Prediction using Machine Learning Techniques" — *IRO Journals, V Jithendra et al.* (2023)

[19] "Diabetes Prediction Using Different Ensemble Learning Classifiers in Machine Learning" — *IJRASET, P Manoj Kumar et al.* (2022)

[20] "Diabetes Prediction and Management Using Machine Learning Approaches" — *Mowafaq Salem Alzboon, Muhyeeddin Alqaraleh, Mohammad Subhi Al-Batah* (2025)

# Hari C S

## turnitin

MINIPROJECT REPORT PLAGARISM CHECKING

### Document Details

Submission ID

trn:oid:::10159:117233868

Submission Date

Oct 18, 2025, 1:43 AM GMT+5:30

Download Date

Oct 18, 2025, 1:48 AM GMT+5:30

File Name

turnitin.pdf

File Size

2.0 MB

61 Pages

13,557 Words

86,201 Characters

# 16% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

## Match Groups

- **85** Not Cited or Quoted 15%
  Matches with neither in-text citation nor quotation marks
- **5** Missing Quotations 1%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

## Top Sources

| | |
|---|---|
| 8% | Internet sources |
| 7% | Publications |
| 11% | Submitted works (Student Papers) |