A Project Report on

PREDICTING DIABETES IN PREGNANT WOMAN AND NEONATAL MELLITUS IN NEW BORN CHILD USING MACHINE LEARNING

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

In

Computer Science and Engineering

Submitted by

Gandikota Mahendra
(21H55A0504)
Hotkar Laxman
(21H55A0508)
Munipally Jyoshna
(21H55A0513)
Under the esteemed guidance of
Mr. J. Ranjith
(Assistant professor)

EXPLORE TO INVENT

Department of Computer Science And Engineering

CMR COLLEGE OF ENGINEERING & TECHNOLOGY
(UGC Autonomous)

*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with A^+ Grade

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

2023- 2024

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Major Project Phase I report entitled **"Predicting Diabetes In Pregnant Woman And Neonatal Mellitus In New Born Child Using Machine Learning"** being submitted by G. Mahendra (21H55A0504), H. Laxman (21H55A0508), M. Jyoshna (21H55A0513) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of Bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Mr. J. Ranjith **(Assistant professor)** **Dept. Of CSE**	**HOD/BRANCH**	**EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express our heartfelt gratitude to all the people who helped in making this project a grand success.

We are grateful to **Mr. J. Ranjith, Assistant professor** , Department of Branch Name for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank**, Dr. Siva Skandha Sanagala,** Head of the Department of Dept Name, CMR College of Engineering and Technology, who is the major driving forces to complete our project work successfully.

We are very grateful to **Dr. Ghanta Devadasu**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Dept Name for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary& Correspondent, CMR Group of Institutions, and Shri Ch Abhinav Reddy, CEO, CMR Group of Institutions for their continuous care and support.

Finally, we extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly or indirectly in completion of this project work.

G. Mahendra (21H55A0504)

H. Laxman (21H55A0508)

M. Jyoshna (21H55A0513)

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

One of the main causes of health issues for expectant mothers and their unborn children is diabetes during pregnancy. Machine learning is a crucial technique for estimating the probability of such a development based on the provided data, as gestational diabetes might advance to permanent diabetes. Pregnancy-related diabetes may be predicted by the current study, but neonatal diabetes risk cannot be predicted. Therefore, in order to deliver the most precise results about diabetes persistence in pregnant women and to enhance the forecasting of neonatal mellitus, new characteristics are needed. This may be accomplished with the use of Python scripting and machine learning techniques like K Nearest Neighbors, Support Vector Machines, and Logistic Regression. The preprocessed machine learning dataset on diabetes was gathered via Kaggle and came from the Pima Indian diabetes database. Additionally, the project's dataset now includes two additional attributes. Research suggests that machine learning models using features like SVM and decision trees may be able to accurately predict a pregnant woman's probability of developing diabetes. Numerous variables have been employed to forecast when this illness may manifest during pregnancy.

Keywords: K Nearest Neighbors, Support Vector Machines, and Logistic Regression, Machine learning.

CHAPTER 1
INTRODUCTION

INTRODUCTION

A long-term metabolic disease characterized by high blood glucose levels is called diabetes mellitus. It is a serious public health concern that impacts people of all ages and is regarded as one of the most challenging diseases of the twenty-first century. Diabetes is a serious ailment that has an impact on the developing child as well as the expectant mother. Diabetes during pregnancy increases the risk of complications such as premature labor, cesarean delivery, and infant morbidity and death.

Conversely, neonatal diabetes is a rare form of the disease that appears in the first six months of life. Hyperglycemia sets it apart, and it's commonly confused with type 1 diabetes. Neonatal diabetes can cause convulsions, delayed development, and intellectual impairments if treatment is not received. Diabetes must be identified early in pregnancy in order to successfully treat the condition and prevent problems in newborns and expectant mothers. Machine learning, a subset of artificial intelligence, has become a vital diagnostic and prognostic tool in the medical field.

To find patterns and forecast future data, machine learning algorithms use historical data. When it comes to diabetes prediction, machine learning algorithms may make use of patient demographics, test results, clinical notes, and electronic health record data to pinpoint individuals who are at risk of developing the disease. Numerous studies have examined the application of machine learning algorithms to the prediction of gestational diabetes and neonatal diabetes in infants. For instance, one study used machine learning algorithms to forecast a pregnant woman's risk of gestational diabetes based on clinical and demographic factors. The algorithm's accuracy in predicting gestational diabetes is 89.4%.

1.1.Problem Statement

Diabetes is a chronic condition that affects millions of individuals all over the world and can have major health repercussions, particularly during pregnancy. Gestational diabetes mellitus (GDM) is a frequent condition that can result in neonatal mellitus in neonates. Early identification and treatment of GDM are critical for avoiding negative effects for both the mother and the child. Current GDM detection approaches rely heavily on blood glucose tests and clinical risk factors. However, these strategies may not be precise enough to forecast GDM in all circumstances. Machine learning (ML) has showed potential in predicting GDM by analysing massive datasets and finding risk variables that traditional approaches may miss. The problem statement for utilising machine learning to predict diabetes in pregnant women and neonatal mellitus in newborns entails establishing accurate and reliable ML models that can predict the GDM in pregnant women and neonatal mellitus .

1.2.Research Objective

The project's goal is to use machine learning techniques to forecast the risk of permanent diabetes in pregnant women and Neonatal Mellitus mellitus in newborn children. By developing machine learning models and providing a user interface for on-spot data calculation, the project seeks to facilitate early detection and prevention of diabetes in these populations.

Large datasets will be analysed using machine learning algorithms to uncover patterns and risk variables related with persistent diabetes in pregnant women and Neonatal Mellitus diabetes in newborns. By training these models on comprehensive and diverse datasets, the goal is to ensure their reliability and generalizability in real-world scenarios.

To achieve accurate predictions, the project emphasizes the importance of high accuracy. The models will undergo rigorous evaluation and testing to assess their performance. This evaluation process involves validating the models against large datasets, comparing their predictions with actual outcomes, and measuring their accuracy metrics.

The project's goal is to create a user-friendly interface for healthcare practitioners in addition to generating effective machine learning models. This interface will enable them to input relevant data easily and obtain real-time predictions. The user interface will be designed to be intuitive, allowing healthcare professionals to conveniently access and utilize the predictive models in their clinical practice.

The project's relevance stems from its potential to enhance healthcare outcomes for pregnant women and babies by recognising and managing diabetes risk at an early stage. Diabetes can be reduced in the long run via early identification and prevention, which can lead to prompt treatments and proper care.

Furthermore, by leveraging machine learning techniques, the project seeks to enhance the accuracy of predictions beyond what traditional methods can achieve. The algorithms have the ability to detect subtle patterns and interconnections in the data that human viewers may not see. This can lead to more exact risk assessments and personalised therapies for specific patients.

1.3.Future scope and limitations

The potential of machine learning to predict gestational diabetes and neonatal mellitus is promising and encompasses several potential advancements and areas of exploration. Here are some key aspects that offer significant opportunities for further development.

Accuracy- Researchers can focus on refining existing machine learning models and developing more advanced algorithms to improve the accuracy of predictions. This involves incorporating additional data sources, such as genetic information or nove biomarkers, tenhance the predictive power of the models. Longitudinal Analysis-The integration of longitudinal data, including multiple measurements taken throughout pregnancy, can provide a more comprehensive understanding of the dynamic changes in risk factors and their impact on diabetes development .Personalized Risk Stratification- The future lies in developing personalized risk stratification models that can consider individual characteristics, including genetic predisposition, socio-demographic factors, and lifestyle choices. By tailoring predictions to specific individuals, healthcare providers can offer personalized interventions.

CHAPTER 2
BACKGROUND WORK

2.1 EXISTING METHOD 1:

2.1.1 INTRODUCTION

Diabetes is one of the most hazardous diseases on the planet. Diabetes is caused by obesity, excessive blood glucose levels, and other factors. It alters the hormone insulin, causing aberrant crab metabolism and improving blood sugar levels. When the body does not produce enough insulin, diabetes develops. According to the World Health Organization, 422 million people worldwide suffer from diabetes, with the majority living in low- or middle-income nations. Up until 2030, this figure might be boosted to 490 billion. Diabetes is, nevertheless, prevalent in a number of countries, including Canada, China, and India. With a population of more than 100 million people, India has a total of 40 million diabetes. While we didn't achieve our goal of 100 percent accuracy in diabetes prediction, we did develop a system that can come close to it given enough time and data. As with any project of this nature, there is room for improvement. Because of the nature of this project, multiple algorithms can be combined as modules and their results combined to improve the accuracy of the final result. This research could be expanded to see how likely non-diabetic people are to develop diabetes in the YMER || ISSN : 0044-0477 VOLUME 21 : ISSUE 5 (May) - 2022 http://ymerdigital.com Page No:485 coming years. Thus, for this purpose we apply popular classification and ensemble methods on dataset for prediction

2.1.2 Merits

1. Early detection allows timely interventions and improves outcomes.

2. Personalized care plans based on individual risk factors enhance treatment effectiveness.

3. Risk stratification optimizes healthcare delivery and resource allocation.

4. Cost savings through reduced complications and hospitalizations.

5. Insights gained contribute to further research and understanding of diabetes.

2.1.3 Demerits

1. The assessment is still done manually, resulting in many

2.1.4 Challenges

While the application of ML algorithms for foreseeing diabetes in pregnant women and neonatal diabetes mellitus shows considerable potential, there are various issues that must be solved in order for these models to be accurate and useful. One of the most difficult aspects of forecasting diabetes in pregnant women is the condition's heterogeneity. Gestational diabetes mellitus (GDM) is a complicated illness impacted by a variety of variables such as maternal age, BMI, ethnicity, and diabetes family history. Furthermore, GDM is a changing

2.1.5 Implementation

Gradient Boosting – Gradient Boosting is most powerful ensemble technique used for prediction and it is a classification technique. It combine week learner together to make strong learner models for prediction. It uses Decision Tree model. it classify complex data sets and it is very effective and popular method. In gradient boosting model performance improve over iterations.

Algorithm-

>Consider a sample of target values as P >Estimate

the error in target values.

>Update and adjust the weights to reduce error M.

>P[x] =p[x] +alpha M[x]

>Model Learners are analyzed and calculated by loss function F >Repeat

steps till desired & target result P.

MODEL BUILDING

This is most important phase which includes model building for prediction of diabetes. In this we have implemented various machine learning algorithms which are discussed above for diabetes prediction.

Procedure of Proposed Methodology-

Step1: Import required libraries, Import diabetes dataset.

Step2: Pre-process data to remove missing data.

Step3: Perform percentage split of 80% to divide dataset as Training set and 20% to Test set.
Step4: Select the machine learning algorithm i.e. K- Nearest Neighbour, Support Vector Machine, Decision Tree, Logistic regression, Random Forest algorithm.

2.2 EXISTING METHOD 2

2.2.1 INTRODUCTION

Diabetes is the fast growing disease among the people even among the youngster. In understanding diabetes and how it develops, we need to understand what happens in the body without diabetes. Sugar (glucose) comes from the foods that we eat, specifically carbohydrate foods. Carbohydrate foods provide our body with its main energy source everybody, even those people with diabetes, needs carbohydrate. Carbohydrate foods include bread, cereal, pasta, rice, fruit, dairy product and vegetables (especially starchyvegetables). When we eat these foods, the body breaks them down into glucose The glucose moves around the body in the blood stream. Some of the glucose is taken to our brain to help us think clearly and function. The remainder of the glucose is taken to the cells of our body for energy and also to our liver, where it is stored as energy that is used later by the body. In order for the body to use glucose for energy, insulin is required. Insulin is a hormone that is produced by the beta cells in the pancreas. Insulin works like a key to a door. Insulin attaches itself to doors on the cell, opening the door to allow glucose to move from the blood stream, through the door, and into the cell If the pancreas is not able to produce enough insulin(insulin deficiency) or if the body cannot use the insulin it produces (insulin resistance), glucose builds up in the bloodstream (hyperglycaemia) and diabetes develops. Diabetes Mellitus means high levels of sugar (glucose) in the blood stream and in the urine.

2.2.2 Merits

Higher accuracy can be achieved with less training data.

Use of a Machine Learning Algorithm improves the system's reliability and accuracy.

2.2.3 Demerits

The assessment is still done manually, resulting in many inaccuracies and incorrect treatments

2.2.4 Implementation

The prediction of diabetes in pregnant women and neonatal mellitus using machine learning has several future enhancements that can further improve the accuracy and effectiveness of the predictive models. Here are some key areas for future development.

Incorporating multi-omics data, such as genomes, transcriptomics, proteomics, and metabolomics, can give a more thorough picture of the molecular mechanisms underlying gestational diabetes and neonatal mellitus. Integrating these multi-dimensional datasets with machine learning models can uncover novel biomarkers, identify disease pathways, and enhance the accuracy of predictions.

Real-Time Data Incorporation- The incorporation of real-time data, such as continuous glucose monitoring, maternal vital signs, and foetal monitoring, can enable dynamic and personalized risk assessment. By incorporating real-time data streams into machine learning models, healthcare providers can receive timely alerts and take proactive measures to prevent complications associated with gestational diabetes and neonatal mellitus.

Integration of Electronic Health Records (EHR)- Taking use of the large quantity of information included in electronic health records can provide a comprehensive patient profile for accurate risk assessment. By integrating EHR data with machine learning models, healthcare providers can tap into historical medical records, previous pregnancies, medication history, and comorbidities to improve prediction accuracy and develop personalized interventions

2.3 EXIATING METHOD 3

2.3.1 INTRODUCTION

Diabetes is a common chronic disease that threat to human health. As a result, we use common classification and algorithms on the dataset to make predictions. Diabetes is a prevalent chronic disease that can be extremely dangerous to one's health. Diabetes is diagnosed when blood glucose levels are greater than normal, which is caused by insulin secretion or biological factors. Diabetes can harm our bodies in a variety of ways, including causing tissue, kidney, eye, and blood artery dysfunction. Based on physical examination data and consultation with doctors, machine learning may make a preliminary diagnosis of diabetes mellitus. Many techniques, including machine learning methods like Random Forest, Support Vector Machine, Decision Tree, and others, have recently been utilised to predict diabetes. We can forecast diabetes using machine learning approaches by creating predicting models based on medical datasets.

2.3.2 merits

1. Early detection allows timely interventions and improves outcomes.
2. Personalized care plans based on individual risk factors enhance treatment
3. effectiveness.
4. Risk stratification optimizes healthcare delivery and resource allocation.
5. Cost savings through reduced complications and hospitalizations.
6. Insights gained contribute to further research and understanding of diabetes

2.3.3 demerits

1. Restricted or biased dataset: A lot of current initiatives depend on a biased or constrained dataset that could not be totally representative of the community.
2. This may result in erroneous forecasts or restricted applicability of the findings. Transparency and interpretability issues: Transparency and interpretability issues may arise with some machine learning models utilized in current projects.
3. This makes it difficult to comprehend how the model made its predictions.
4. This may make it more difficult for patients and healthcare professionals to accept and trust the models.

2.3.5 implementation

1. To implement a machine learning model for predicting diabetes in pregnant women and its impact on newborns, you would typically follow these steps:
2. Data Collection: Gather relevant data on pregnant women, including their medical history, lifestyle factors, and any existing conditions, as well as data on newborn health outcomes.
3. Data Preprocessing: Clean the data, handle missing values, and preprocess features (e.g., normalization, encoding categorical variables).
4. Feature Selection/Engineering: Identify important features that may contribute to the prediction task. You may also create new features from existing ones if necessary.

5. Model Selection: Choose appropriate machine learning algorithms for classification tasks. Common choices include logistic regression, decision trees, random forests, support vector machines, or neural networks.

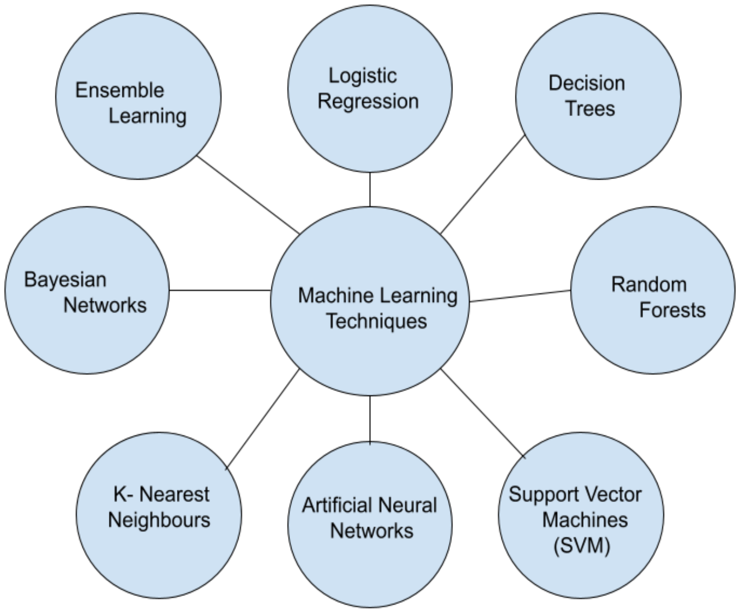
CHAPTER 3
PROPOSED SYSTEM

3.1 PROPOSED SYSTEM

The proposed method of predicting diabetes in pregnant women and neonatal mellitus in newborn children using machine learning holds significant potential in advancing healthcare outcomes in this domain. By leveraging machine learning algorithms and predictive modelling techniques, this approach aims to provide early identification and risk assessment, enabling timely interventions and personalized care. One of the key advantages of this method is its potential to improve maternal and neonatal health outcomes by identifying high-risk cases accurately. Machine learning models can analyze large amounts of data from pregnant women, including medical history, lifestyle factors, and various clinical parameters, to identify patterns and generate predictive insights. This enables healthcare providers to intervene early and implement targeted preventive measures for at-risk individuals, potentially reducing the incidence and severity of gestational diabetes and neonatal mellitus. Additionally, machine learning-based prediction models can assist healthcare professionals in making informed decisions regarding treatment plans and resource allocation.

By accurately identifying pregnant women at risk of developing diabetes or newborns susceptible to neonatal mellitus, medical interventions and support can be focused on those who need them the most, optimizing healthcare resources and improving overall care quality. Moreover, the proposed method has the potential to contribute to medical research by generating new insights and knowledge in the field of gestational diabetes and neonatal health. The analysis of large-scale datasets using Machine Learning approaches, complicated correlations between multiple risk factors and outcomes may be revealed, leading to a better understanding of the underlying mechanisms and supporting the creation of more effective preventative and treatment strategies.

3.2 ALGORITHMS USED FOR PROPOSED MODEL



3.2.1 TYPES OF MACHINE LEARNING TECHNIQUE

Trees of Decisions

A predictive modeling technique used in machine learning is decision trees. Decision trees can be used to forecast the likelihood of developing diabetes in pregnant women and neonatal mellitus in newborn children by analyzing several risk factors and creating a decision tree to assess the possibility of obtaining diabetes.

1.The Forest of Random

Machine learning for predictive modeling employs random forests, an ensemble learning technique. Using random forests to construct many decision trees and combine the results to provide a more accurate prediction, one can predict the likelihood of developing diabetes in expectant mothers and neonatal mellitus in newborns.

2. SVM, or Support Vector Machine

One kind of machine learning technique that can be applied to regression analysis and classification is support vector machines. SVM can be used to predict the likelihood of acquiring diabetes in pregnant women and neonatal mellitus in newborn infants by analyzing several risk indicators and building a model to detect the possibility of having diabetes.

3. Synthetic Neural Nets

A type of machine learning technique called artificial neural networks aims to mimic how the human brain functions. Artificial neural networks can be used to predict the likelihood of developing diabetes in pregnant women and neonatal mellitus in newborn infants by analyzing several risk variables and building a model to identify the possibility of obtaining diabetes.

4. K-Nearest Adjacent

K-Nearest Neighbors is a machine learning technique used for regression analysis and classification. K-Nearest Neighbor’s may be used to predict the likelihood of developing diabetes in pregnant women and neonatal mellitus in newborn children by analyzing a number of risk variables and figuring out the K-nearest neighbour to categorize the chance of having diabetes in pregnant women and neonatal mellitus in newborn children.

5. Bayesian Computation

One kind of probabilistic graphical model used in machine learning for data classification and prediction is the Bayesian network. Bayesian networks can be used to predict the likelihood of acquiring diabetes in pregnant women and neonatal mellitus in newborn infants by analyzing several risk indicators and creating a probabilistic model to determine the possibility of having diabetes.

3.3 DESIGNING

3.3.1 UML DIAGRAM

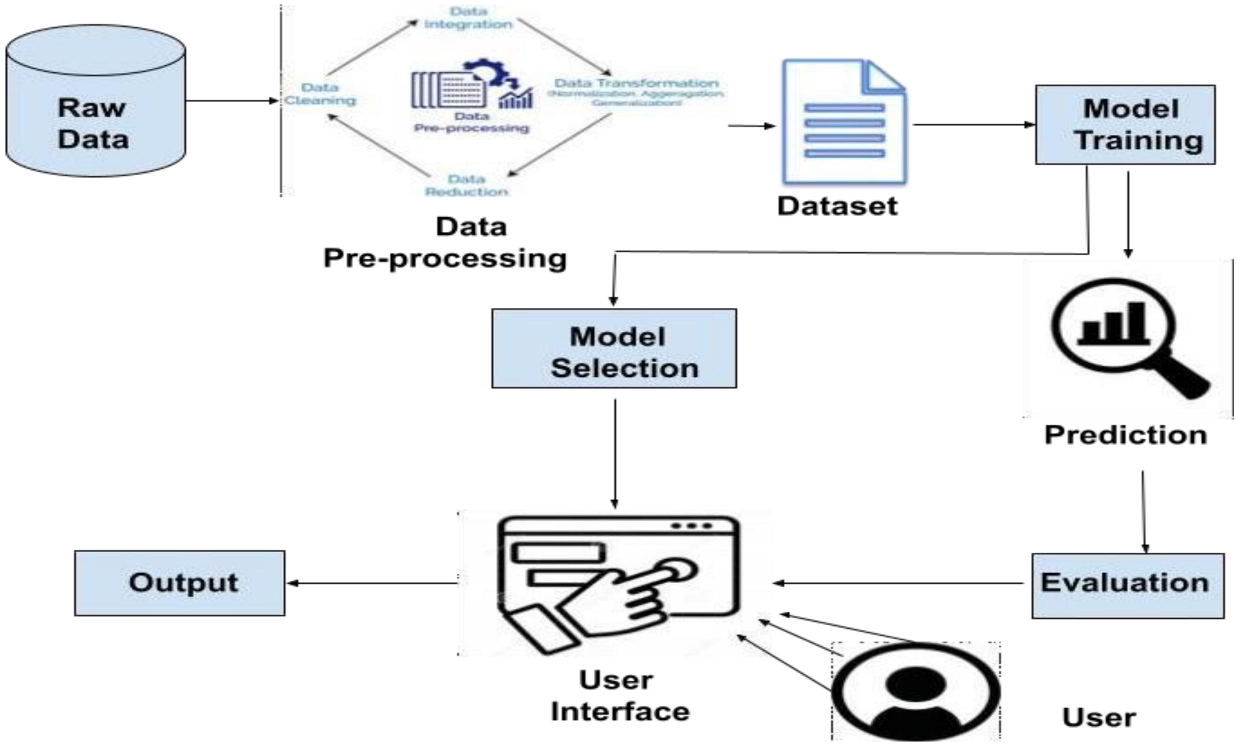


Figure 3.3.1 architecture

1. **Raw Dataset** The Raw Dataset serves as the starting point for the architecture. This dataset is specifically curated to contain a collection of data of various females with symptoms of diabetes. The data is typically sourced from various real-world scenarios or collected under specific conditions.
2. **Data Preprocessing** Data preprocessing is an essential step in preparing the blurred images from the Real Blur dataset for training and evaluation. In this step, several operations are performed to transform the raw image data into a suitable format for feeding into a neural network. The main operations typically include conversion to tensors and resizing.

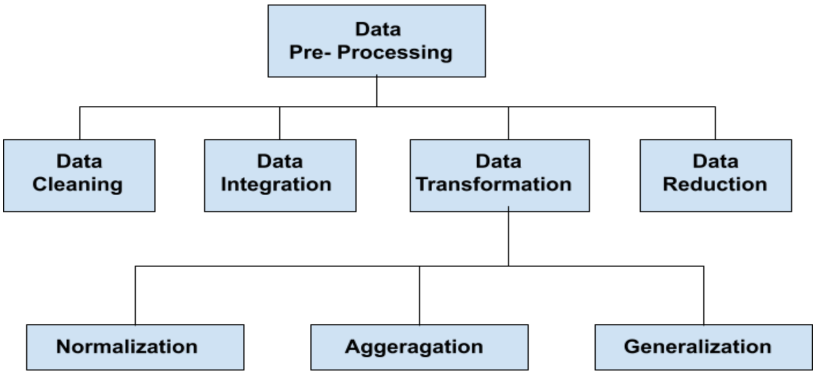


Figure 3.2 Data Preprocessing diagram

3. **Model Training** Model Training is a Crucial step in the project as the various models like Naive Bayes, K-Nearest Neighbour, Random Forest, Logistic Regression, SVM are used to train the dataset and the best of the mentioned is used for final execution.
4. **Prediction** Predictions are drawn from the training process to know that which model is better in the prediction.
5. **Model Evaluation** In this step, the model is evaluated using the preprocessed testing data, which consist of diabetes data of females and hence the best model is drawn out of the five.
6. **User Interface** Here the user can input the data he has and then the algorithm runs in the background to calculate whether the user has diabetes or not.
7. **Output** Here the algorithm generates output and displays prompts either "You are diabetic" or "You are not diabetic". Along with the accuracy of the calculation.

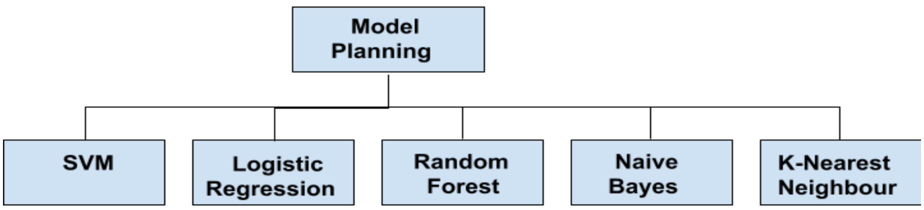


Figure 3.3 Model Planning Diagram

Class Diagram

A popular graphical notation for illustrating the organization and interactions between classes in object-oriented software systems is the UML (Unified Modeling Language) class diagram. Helping with the arrangement, characteristics, and behaviors of classes, they offer a visual blueprint that makes software creation and analysis easier. We shall examine the importance of UML class diagrams and their essential elements in this article.

A UML class diagram's fundamental elements are classes, which are shown as rectangles with lines connecting them to show their relationships. The essential elements of a class diagram are as follows:-

The fundamental components of object-oriented systems are classes. They are abstracted or conceptual representations of real-world phenomena that capture information and actions. Every class in a class diagram is represented by a rectangle with its name, properties, and methods inside. The term of the class is represented by the class name, and the data and activity connected to the class are represented by the attributes and methods, respectively.

The qualities or characteristics of a class are its attributes. They give an explanation of the condition or data components connected to class objects. Typically, attributes are listed inside the class rectangle along with their data types and visibility markers (# for protected, - for private, and + for public). Developers can comprehend the data items that a class has and how they relate to other classes in the system by defining attributes in a class diagram.

Methods: Often referred to as operations, methods specify the actions or behaviors that objects in a class are capable of. They stand for the class's associated functionality. Method names, parameters, return types, and visibility indicators can all be found listed within the class rectangle. Developers can obtain insights into the interactions and collaborations of classes within a system by representing methods in a class diagram.

Relationships: UML class diagrams make it possible to depict the relationships between classes, giving viewers a clear picture of how different classes work together and depend on one another. A class diagram can show many different kinds of relationships, such as.

Software developers, designers, and stakeholders can benefit greatly from the communication and analysis capabilities provided by UML class diagrams. They offer a clear and consistent visual depiction of the classes' relationships and structure, which helps with decision-making, cooperation, and comprehension all the way through the software development lifecycle.

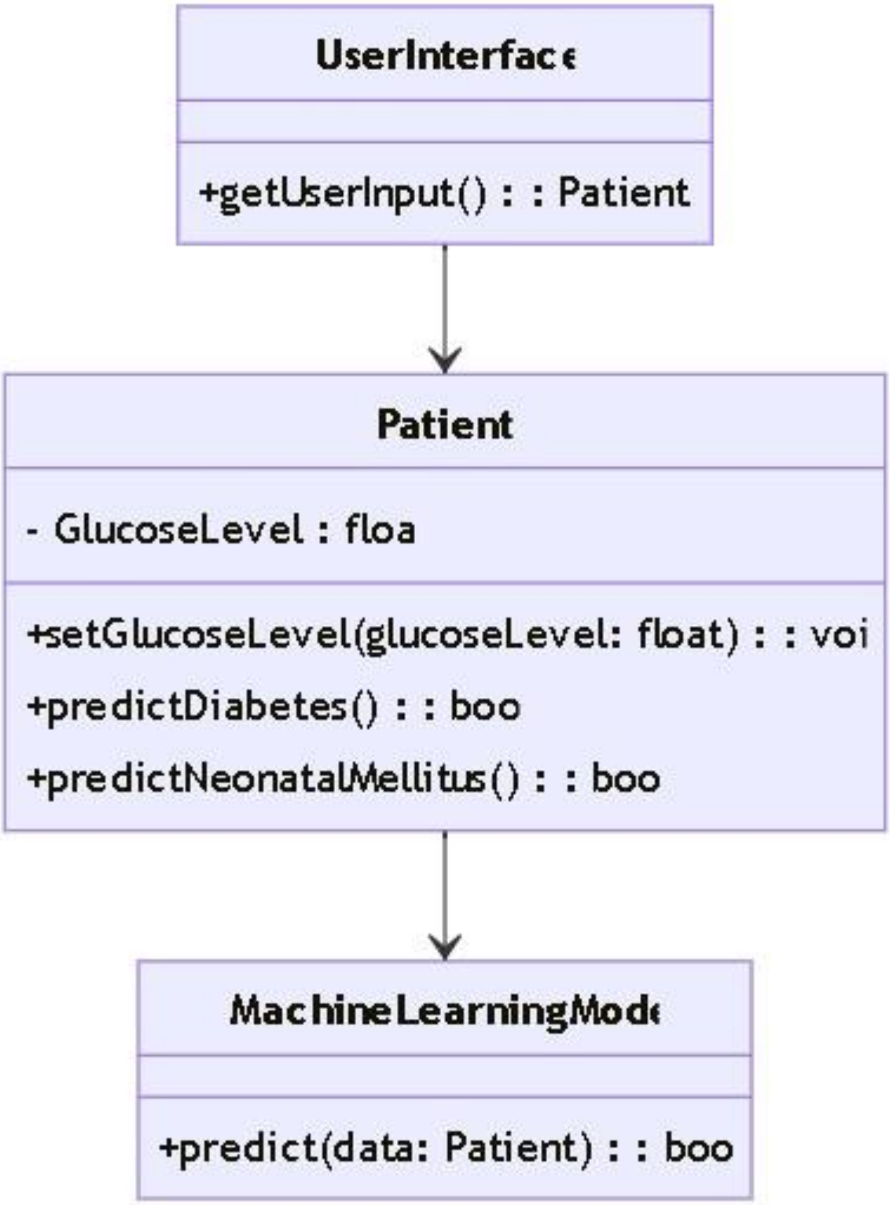


Fig 3.4 Class Diagram

Sequence Diagram

UML sequence diagrams are an effective tool for visualising and comprehending the dynamic interactions between items or components in a system. They give a graphical depiction of the communication flow exchanged between objects, illustrating the order of method calls and the resulting behaviors.

One of the primary purposes of sequence diagrams is to depict the collaboration and communication between different elements in a system. These elements can include objects, actors, and subsystems. By visually representing the sequence of interactions, sequence diagrams enable developers and stakeholders to gain a comprehensive understanding of how the different components of the system work together to accomplish specific tasks. This understanding is crucial for ensuring effective system functionality and identifying any potential issues or bottlenecks.

Sequence diagrams are particularly useful for analyzing system behavior. They allow developers to trace the flow of messages and method calls, enabling them to identify potential errors or inconsistencies in the system's logic. By examining the sequence of interactions, developers can detect issues such as incorrect ordering of operations, missing or redundant messages, or unexpected behavior. This analysis phase is essential for debugging and refining the system's behavior before it is implemented, saving time and effort in the later stages of development.

Moreover, sequence Diagrams are very important in the design and documentation phases of software development. They help developers by providing a visual picture of the system's behaviour and interactions to communicate and discuss the design decisions with other team members or stakeholders. By using sequence diagrams, developers can ensure that the intended behavior and communication patterns are accurately captured and documented, facilitating better collaboration and reducing misunderstandings.

Sequence diagrams in UML are a valuable asset in software development. They offer a visual representation of the dynamic interactions within a system, showcasing the flow of messages and method calls between objects. Sequence diagrams aid in understanding system behavior, identifying potential issues or bottlenecks, and validating the correctness of the system's logic. They are utilized during the analysis, design, and documentation phases of software development, enabling developers and stakeholders to visualize and discuss the behavior and interactions effectively.

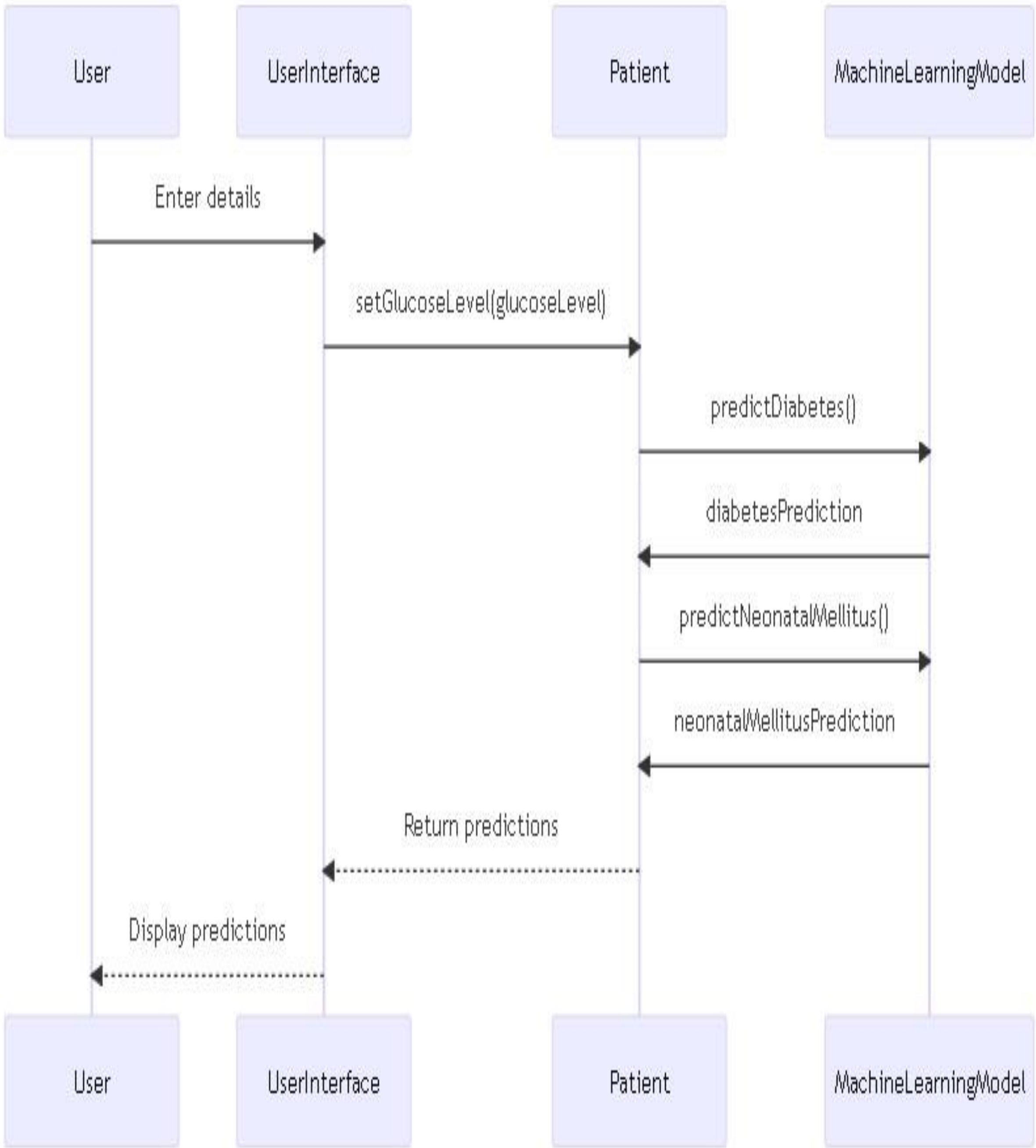


Fig 3.5 Sequence Diagram

State-Transition Diagram

State-transition diagrams, also known as state machines, are a valuable tool in the Unified Modelling Language (UML) is a modelling and representation language ,the behavior of objects or systems. They depict the various states that an object or system can be in and the transitions between these states. State-transition diagrams provide a visual representation of how an object or system responds to external stimuli or events, leading to changes in its state.

State-transition diagrams consist of several key elements, States, transitions, events, and actions are all included. States indicate the many situations or modes in which an item or system can exist. Transitions show the paths or changes between states triggered by events. Events are the stimuli or triggers that cause a transition to occur. Actions represent the specific activities or behaviors that take place when a transition happens.

The primary purpose of state-transition diagrams is to model and analyze the behavior of complex systems with multiple states and transitions. By visually representing the possible states and the conditions for transitioning between them, these diagrams provide a clear understanding of the system's behavior and the various scenarios that can occur. They help stakeholders and developers identify and analyze different behaviors and validate the correctness of the system's logic.

State-transition diagrams are widely used in various domains, including software engineering, control systems, and modeling the behavior of real-world systems. In software engineering, they are instrumental in designing and implementing systems with complex behaviors, such as user interfaces, network protocols, or business processes. State machines aid in the development of robust and reliable systems by ensuring that all possible states and transitions are considered and properly handled. In control systems, state-transition diagrams help in understanding and designing the behavior of physical systems, such as robotic systems or manufacturing processes.

Moreover, state-transition diagrams assist in system testing and validation. By visually depicting the possible states and transitions, developers can create test cases that cover all scenarios, ensuring thorough testing and validation of the system's behavior.

In conclusion, state-transition diagrams provide a powerful means to model and analyze the behavior of objects or systems. They allow for the representation of various states, transitions, events, and actions, facilitating a clear understanding of the system's behavior and the possible scenarios.

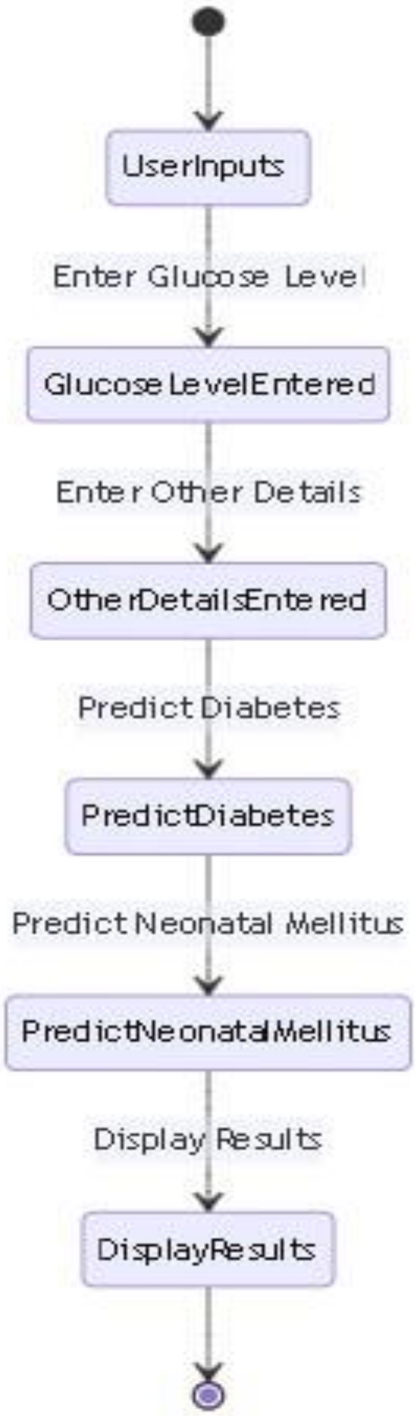


Fig 3.6 State-Transition Diagram

CHAPTER 4

RESULT AND DISCUSSION

Result

The diabetes dataset consists of various attributes that determine a person's risk of being diagnosed by diabetes. It contains data of various patients bearing one or more of the attributes mentioned in the data set. The detailed meanings of each attribute mentioned in the taken dataset is as follows.

Glucose Level- The concentration of glucose (sugar) in the patient's blood is represented by this aspect. Milligrams per deciliter (mg/dL) or millimoles per litre (mmol/L) are the most common units of measurement. Elevated glucose levels might be a sign of poor glucose tolerance or diabetes.

Blood Pressure- Typically, this characteristic comprises two values: systolic and diastolic blood pressure. When the heart contracts, systolic blood pressure indicates the pressure in the arteries, whereas diastolic blood pressure shows the pressure when the heart is at rest. Hypertension (high blood pressure) is a risk factor for gestational diabetes.

BMI (Body Mass Index)- BMI is a measure of body fat that is dependent on a person's height and weight. It is computed by dividing the weight in kilograms by the height in meters squared. BMI indicates if a patient is underweight, normal weight, overweight, or obese.

Obesity is a major risk factor for developing gestational diabetes.

4.1 Performance metrics

Figure 4.1.1 . Input Sliders

	Pregnancies	Glucose	BloodPressure	SkinThickness
count	768	768	768	768
mean	3.8451	120.8945	69.1055	20.5365
std	3.3696	31.9726	19.3558	15.9522
min	0	0	0	0
25%	1	99	62	0
50%	3	117	72	23
75%	6	140.25	80	32
max	17	199	122	99

	se	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
count	68	768	768	768	768	768	768	768
mean	45	69.1055	20.5365	79.7995	31.9926	0.4719	33.2409	0.349
std	26	19.3558	15.9522	115.244	7.8842	0.3313	11.7602	0.477
min	0	0	0	0	0	0.078	21	0
25%	99	62	0	0	27.3	0.2438	24	0
50%	17	72	23	30.5	32	0.3725	29	0
75%	25	80	32	127.25	36.6	0.6263	41	1
max	99	122	99	846	67.1	2.42	81	1

Figure 4.1.2 Training Data Stats-2

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	3	120	70	20	79	20	0.47	33

Figure 4.1.3 Patient Data

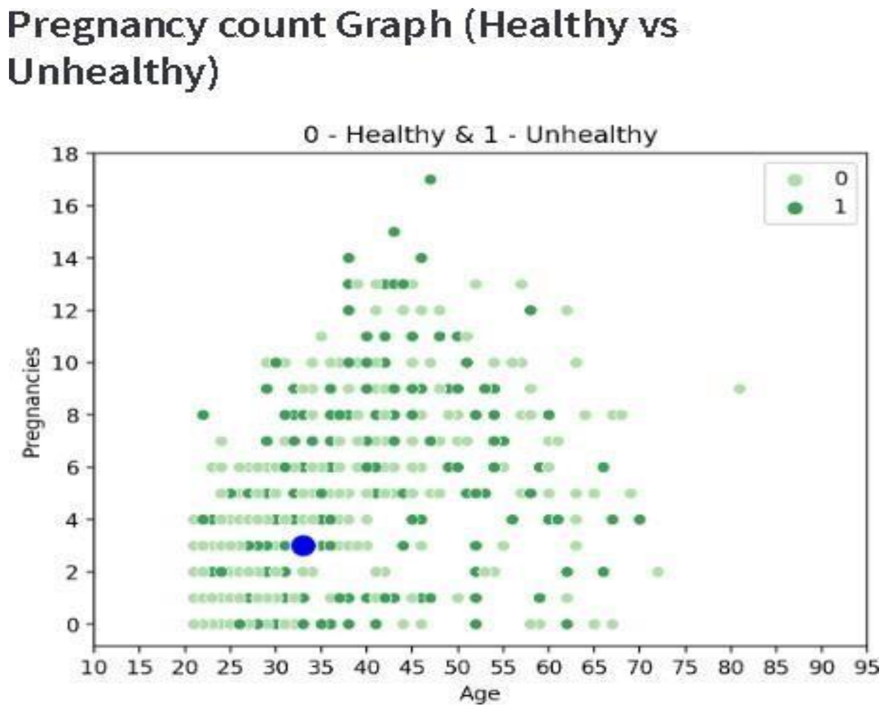


Figure 4.1.4 Pregnancy count graph

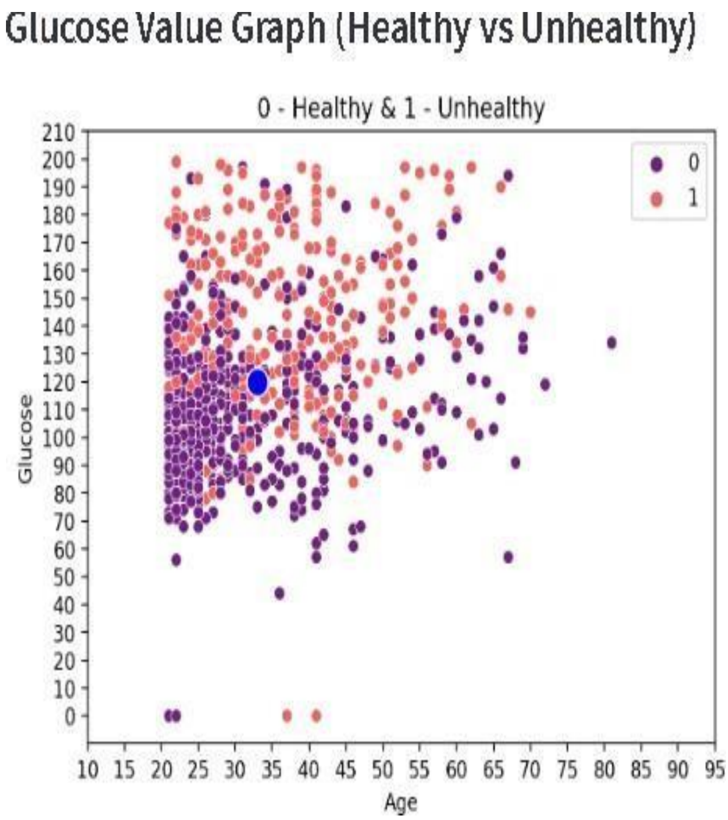


Figure 4.1.5 Glucose Value Graph

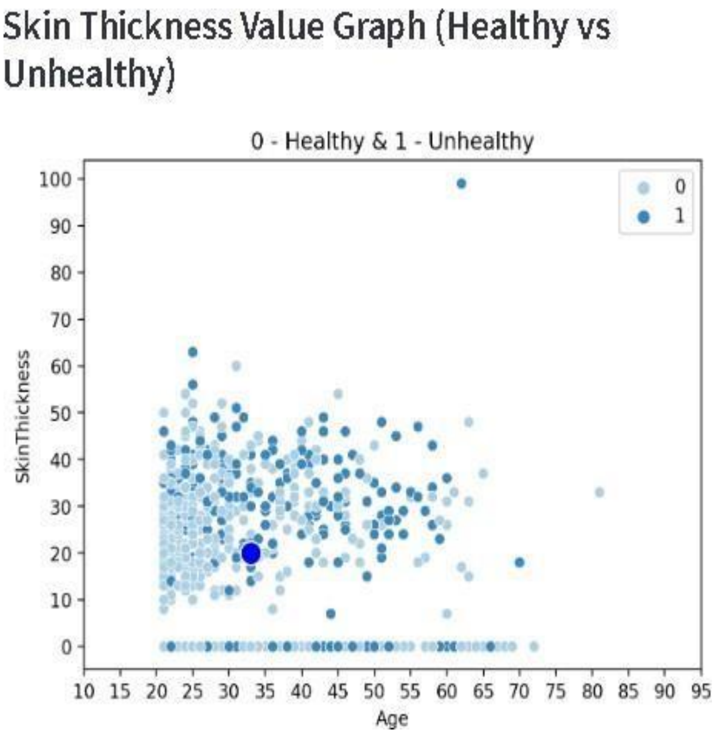


Figure 4.1.6 Skin thickness value count

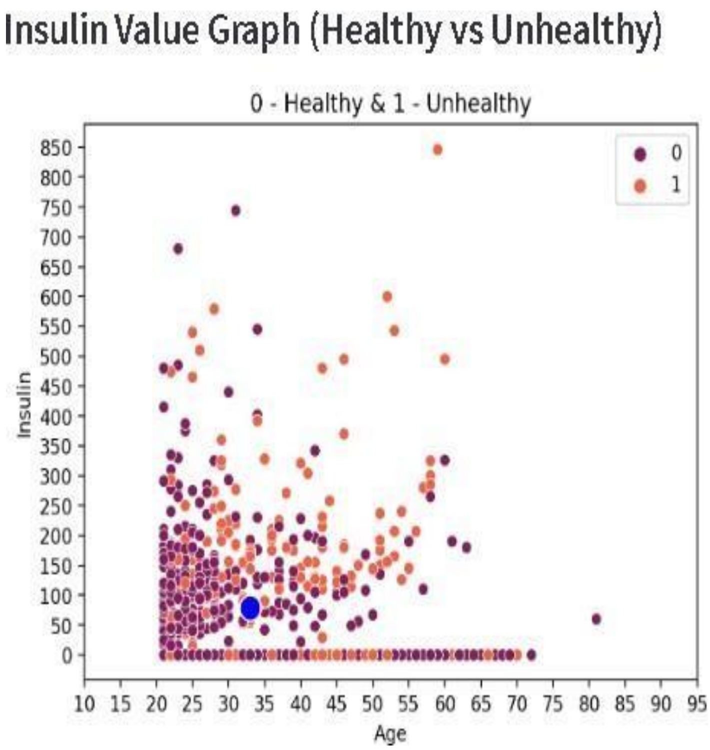


Figure 4.1.7 Insulin value count

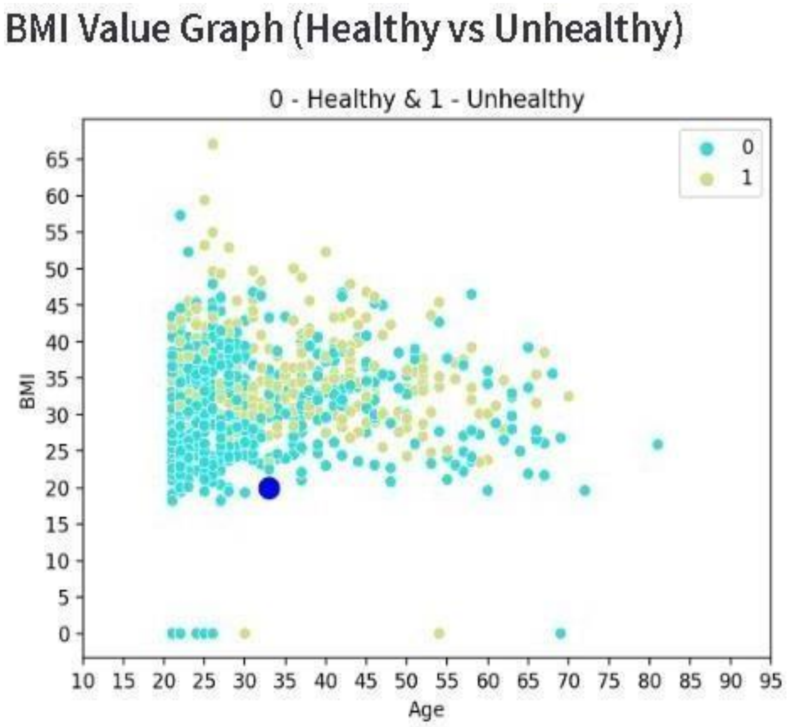


Figure 4.1.8 BMI Value Graph

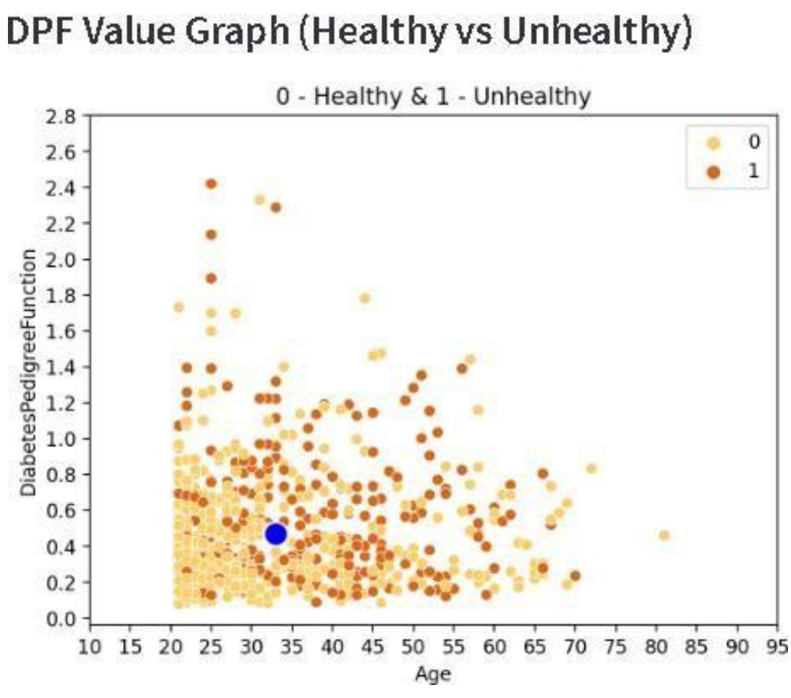


Figure 4.1.9 DPF value count

Fig 4.10 User Output and Accuracy

CHAPTER 5
CONCLUSION

5.1 CONCLUSION AND FUTURE ENHANCEMENT

Prediction of diabetes in pregnant women and neonatal mellitus in newborn children using machine learning revolves around developing predictive models to predict the risk of gestational diabetes in women expectant mothers and the likelihood of neonatal diabetes in newborns. By leveraging machine learning algorithms and analyzing various factors such as medical history, clinical parameters, lifestyle, and genetic predisposition, these models aim to provide early detection and personalized risk assessment. The goal is to enable healthcare professionals to intervene proactively, implement preventive measures. Various machine learning algorithms are applied to the dataset in this study, and classification is done using various algorithms, with Random Forest providing the highest accuracy.

FUTURE ENHANCEMENT

Real-Time Data Incorporation: Dynamic and individualized risk assessment can be made possible by including real-time data, such as continuous glucose monitoring, maternal vital signs, and fetal monitoring. Healthcare professionals can obtain timely alerts and take proactive measures to reduce issues associated with gestational diabetes and neonatal mellitus by integrating real-time data sources into machine learning models.

Electronic Health Record (EHR) Integration:Utilizing the vast amount of data found in electronic health records can yield a thorough patient profile for precise risk assessment. Healthcare professionals can leverage past medical records, past pregnancies, medication history, and comorbidities to enhance prediction accuracy and create customized therapies by merging EHR data with machine learning algorithms.

External Validation and Generalizability: Predictive models ought to be externally validated on a range of patient demographics and healthcare environments in subsequent studies. The practical usability and wider impact of the models will be improved by ensuring their robustness and generalizability across various demographics and healthcare systems.Ethical Issues and Data Privacy: Future improvements should carefully address ethical issues and data privacy, as with any machine learning application in the healthcare industry. This includes highlighting the significance of particular attributes and producing practical suggestions.

CHAPTER 6
REFERENCE

REFERENCE

1 Марко Романович Басараб , Екатерина Олеговна Іванько , Вішвеш Кулкарні, “Prediction of the Development of Gestational Diabetes Mellitus in Pregnant Women Using Machine Learning Methods”.(2021)DOI-10.20535/25234455.mea.228845

2 Zahura Zaman , Md. Ashrak Al Arif Shohas , Mahedi Hasan Bijoy , Meherab Hossain , Shakawat Al Sakib,”Assessing Machine Learning Methods for Predicting Diabetes among PregnantWomen”.(2022)DOI-10.31632/ijalsr.2022.v05i01.005

3 AVSrinivas , Abbireddy Ramya , GT Chandralekha , Bhandaram Vaagdevi , K Anand

4 Quan Zou , Kaiyang Qu , Yamei Luo , Dehui Yin, Ying Ju , Jun Zhang, “Predicting Diabetes Mellitus With Machine Learning Techniques”,(2018)DOI-10.3389/fgene.2018.00515

5 ,“Prediction of gestational diabetes based on explainable deeplearning andfogcomputing”(Lili Wei , Yueshuai Pan , Yan Zhang,“Application of machine learning algorithm for predicting gestational diabetes mellitus in early pregnancy”(2021),DOI10.2478/fon-2021-0022

6 Thach Tran , Jane E.Hirst,“Prognosis Model for Gestational Diabetes Using Machine Learning Techniques”(2021),DOI10.2337/dc12-1418

7 Sara L. White , Debbie A Lawlor , Annette Briley,“Early Antenatal Prediction of Gestational Diabetes in Obese Women”(2016),DOI10.1371/journal.pone.0167846

8 Sumathi Amarnath , Meganathan Selvamani,“Prognosis Model for Gestational Diabetes Using Machine Learning Techniques” (2021),DOI10.18494/sam.2021.3119

9 M Sekulić , D Relić , “The role of machine learning in identification of early gestational diabetes mellitus prediction models”(2020),DOI10.26800/lv-142-supp5- 61

10 Xinxi Lu , Jikai Wang ,“ Prediction of Gestational Diabetes and Hypertension based on Pregnancy Examination Data”(2022), DOI10.1142/s0219519422400012

11 Jie Zhang , Fang Wang,“Prediction of Gestational Diabetes Mellitusunder Cascade and Ensemble learning Algorithm” (2022),DOI10.1155/2022/3212738

12 Nora El-Rashidy , Nesma E. ElSayed2022),DOI10.1007/s00500-022-07420

13 Yang Liu , Hua Sun,“Prediction Method of Gestational Diabetes Based on Electronic Medical RecordData”(2021)DOI10.1155/2021/6672072-

14 Jenny Yang , David Clifton,“Machine Learning-Based Risk Stratification for Gestational Diabetes Management”(2022),DOI10.3390/s22134805-

15 Hang Lai , Huaxiong Huang , Karim Keshavjee , Aziz Guergachi , Xin Gao “Predictive models for diabetes mellitus using machine learningtechniques” (2019)DOI10.1186/s12902-019-0436-

IMPLEMENTATION AND CODE

IMPLEMENTATION AND CODE

```
import streamlit as st
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestClassifier
import pandas as pd
import numpy as np
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split

# Load the diabetes dataset
df = pd.read_csv("diabetes.csv")

# Set the title and sidebar header
st.title('Diabetes Checkup')
st.sidebar.header('Patients Data')
st.subheader('Training Data Stats')

# Display training data statistics
st.write(df.describe())

# Split the data into features (x) and target (y)
x = df.drop(['Outcome'], axis=1)
y = df['Outcome']

# Split the data into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=0)

# Define a function to get user input for patient data
def users_report():
    pregnancies = st.sidebar.slider('Pregnancies', 1, 15, 3)
    glucose = st.sidebar.slider('Glucose', 0, 220, 150)
    bp = st.sidebar.slider('Blood Pressure', 0, 142, 60)
    skin_thickness = st.sidebar.slider('Skin Thickness', 10, 110, 20)
    insulin = st.sidebar.slider('Insulin', 0, 846, 75)
    bmi = st.sidebar.slider('BMI', 0, 69, 25)
    dpf = st.sidebar.slider('Diabetes Pedigree Function', 0.0, 2.9, 0.47)
    age = st.sidebar.slider('Age', 0, 88, 39)

    users_report_data = {
        'Pregnancies': pregnancies,
        'Glucose': glucose,
        'BloodPressure': bp,  # Changed from 'Blood Pressure' to 'BloodPressure'
        'SkinThickness': skin_thickness,
        'Insulin': insulin,
        'BMI': bmi,
```

```
        'DiabetesPedigreeFunction': dpf,  # Changed from 'Diabetes Pedigree Function' to
'DiabetesPedigreeFunction'
        'Age': age
    }

    report_data = pd.DataFrame(users_report_data, index=[0])
    return report_data

# Get user input for patient data
users_data = users_report()

# Display patient data
st.subheader('Patients Data')
st.write(users_data)

# Create and train a RandomForestClassifier model
rfor = RandomForestClassifier()
rfor.fit(x_train, y_train)

# Predict the user's diabetes status
users_result = rfor.predict(users_data)

# Set the color based on the prediction result
if users_result[0] == 0:
    color = 'blue'
else:
    color = 'red'

# Create scatterplots to visualize patient data
st.title('Visualised Patient Report')

# Pregnancy Count Graph
st.header('Pregnancy Count Graph (Healthy v Not Healthy)')
fig_preg = plt.figure()
ax1 = sns.scatterplot(x='Age', y='Pregnancies', data=df, hue='Outcome', palette='Greens')
ax2 = sns.scatterplot(x=users_data['Age'], y=users_data['Pregnancies'], s=150, color=color)
plt.xticks(np.arange(10, 110, 5))
plt.yticks(np.arange(0, 20, 4))
plt.title('0 - Healthy & 1 - Not Healthy')
st.pyplot(fig_preg)

# Glucose Value Graph
st.header('Glucose Value Graph (Healthy v Not Healthy)')
fig_glucose = plt.figure()
ax3 = sns.scatterplot(x='Age', y='Glucose', data=df, hue='Outcome', palette='magma')
ax4 = sns.scatterplot(x=users_data['Age'], y=users_data['Glucose'], s=150, color=color)
```

```
plt.xticks(np.arange(10, 110, 5))
plt.yticks(np.arange(0, 220, 10))
plt.title('0 - Healthy & 1 - Not Healthy')
st.pyplot(fig_glucose)

# Blood Pressure Value Graph
st.header('Blood Pressure Value Graph (Healthy v Not Healthy)')
fig_bp = plt.figure()
ax5 = sns.scatterplot(x='Age', y='BloodPressure', data=df, hue='Outcome', palette='Reds')
ax6 = sns.scatterplot(x=users_data['Age'], y=users_data['BloodPressure'], s=150, color=color)
plt.xticks(np.arange(10, 110, 5))
plt.yticks(np.arange(0, 150, 10))
plt.title('0 - Healthy & 1 - Not Healthy')
st.pyplot(fig_bp)

# Skin Thickness Value Graph
st.header('Skin Thickness Value Graph (Healthy v Not Healthy)')
fig_str = plt.figure()
ax7 = sns.scatterplot(x='Age', y='SkinThickness', data=df, hue='Outcome', palette='Blues')
ax8 = sns.scatterplot(x=users_data['Age'], y=users_data['SkinThickness'], s=110,
color=color)
plt.xticks(np.arange(10, 110, 5))
plt.yticks(np.arange(0, 110, 10))
plt.title('0 - Healthy & 1 - Not Healthy')
st.pyplot(fig_str)

# Insulin Value Graph
st.header('Insulin Value Graph (Healthy v Not Healthy)')
fig_i = plt.figure()
ax9 = sns.scatterplot(x='Age', y='Insulin', data=df, hue='Outcome', palette='rocket')
ax10 = sns.scatterplot(x=users_data['Age'], y=users_data['Insulin'], s=150, color=color)
plt.xticks(np.arange(10, 110, 5))
plt.yticks(np.arange(0, 700, 50))
plt.title('0 - Healthy & 1 - Not Healthy')
st.pyplot(fig_i)

# BMI Value Graph
st.header('BMI Value Graph (Healthy v Not Healthy)')
fig_bmi = plt.figure()
ax11 = sns.scatterplot(x='Age', y='BMI', data=df, hue='Outcome', palette='rainbow')
ax12 = sns.scatterplot(x=users_data['Age'], y=users_data['BMI'], s=150, color=color)
plt.xticks(np.arange(10, 110, 5))
plt.yticks(np.arange(0, 80, 5))
plt.title('0 - Healthy & 1 - Not Healthy')
st.pyplot(fig_bmi)
```

```
# DPF Value Graph
st.header('DPF Value Graph (Healthy v Not Healthy)')
fig_dpf = plt.figure()
ax13 = sns.scatterplot(x='Age', y='DiabetesPedigreeFunction', data=df, hue='Outcome',
palette='YlOrBr')
ax14 = sns.scatterplot(x=users_data['Age'], y=users_data['DiabetesPedigreeFunction'], s=150,
color=color)
plt.xticks(np.arange(10, 150, 5))
plt.yticks(np.arange(0, 3, 0.2))
plt.title('0 - Healthy & 1 - Not Healthy')
st.pyplot(fig_dpf)

# Display the user's report
st.subheader('Your Report')
output = ''
if users_result[0] == 0:
    output = 'You are not Diabetic'
else:
    output = 'You are Diabetic'
st.title(output)

# Display the accuracy of the model
st.subheader('Accuracy')
st.write(str(accuracy_score(y_test, rfor.predict(x_test)) * 100) + '%')
```