**VIVEKANAND EDUCATION SOCIETY’S INSTITUTE OF TECHNOLOGY**

**An Autonomous Institute Affiliated to University of Mumbai**

**Department of Computer Engineering**



Project Report on

MADHUVISTA : YOUR DIABETES ALLY

In partial fulfillment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2023-24

**Submitted by**

Sanjana Asrani (D17B/01)

Karina Karira (D17A/30)

Simran Lahrani (D17A/34)

Roshni Wadhwani (D17A/70)

**Project Mentor**

Mrs. Pallavi Saindane

(2023-24)

**VIVEKANAND EDUCATION SOCIETY’S INSTITUTE OF TECHNOLOGY**

**An Autonomous Institute Affiliated to University of Mumbai**

**Department of Computer Engineering**



**Certificate**

This is to certify that ***Sanjana Asrani, Karina Karira, Simran Lahrani, Roshni Wadhwani*** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on “***MadhuVista : Your Diabetes Ally***” as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor ***Prof. Pallavi Saindane*** in the year 2023-24 .

This thesis/dissertation/project report entitled **MadhuVista: Your Diabetes Ally** by **Sanjana Asrani, Roshni Wadhwani, Karina Karira, Simran Lahrani** is approved for the degree of **B.E. Computer Engineering**.

| Programme Outcomes | Grade |
| --- | --- |
| PO1,PO2,PO3,PO4,PO5,PO6,PO7,  PO8, PO9, PO10, PO11, PO12  PSO1, PSO2 |  |

Date:

Project Guide:

------------------------------------------

**Project Report Approval**

**For**

**B. E (Computer Engineering)**

This Project Report entitled **MadhuVista: Your Diabetes Ally** by **Sanjana Asrani, Roshni Wadhwani, Karina Karira, Simran Lahrani** is approved for the degree of **B.E. Computer Engineering** .

Internal Examiner

---------------------------------------------

External Examiner

---------------------------------------------

Head of the Department

-----------------------------------------------

Principal

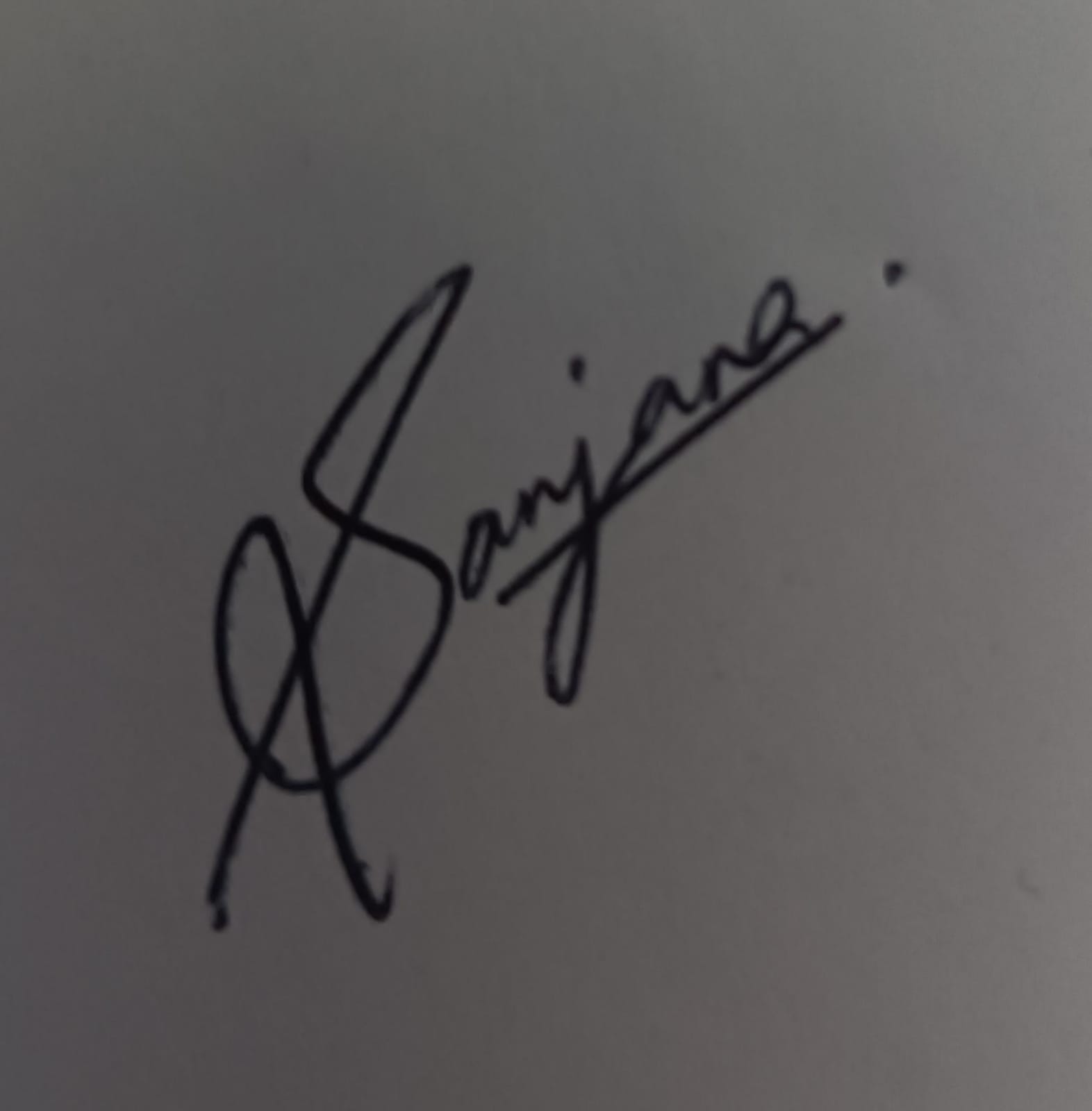
-----------------------------------------------

Date:

Place:

**Declaration**

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

| -----------------------------------------  (Signature)  Sanjana Asrani (01)  -----------------------------------------  (Name of student and Roll No.) | -----------------------------------------  (Signature)  Karina Karira (30)  -----------------------------------------  (Name of student and Roll No.) |
| --- | --- |
| -----------------------------------------  (Signature)  Roshni Wadhwani (70)  -----------------------------------------  (Name of student and Roll No.) | -----------------------------------------  (Signature)  Simran Lahrani (34)  -----------------------------------------  (Name of student and Roll No.) |

Date:

**ACKNOWLEDGEMENT**

We are thankful to our college Vivekanand Education Society’s Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Mrs. Pallavi Saindane** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

**Computer Engineering Department**

**COURSE OUTCOMES FOR B.E PROJECT**

Learners will be to,

| **Course Outcome** | **Description of the Course Outcome** |
| --- | --- |
| CO 1 | Able to apply the relevant engineering concepts, knowledge and skills towards the project. |
| CO2 | Able to identify, formulate and interpret the various relevant research papers and to determine the problem. |
| CO 3 | Able to apply the engineering concepts towards designing solutions for the problem. |
| CO 4 | Able to interpret the data and datasets to be utilized. |
| CO 5 | Able to create, select and apply appropriate technologies, techniques, resources and tools for the project. |
| CO 6 | Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit. |
| CO 7 | Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability. |
| CO 8 | Able to write effective reports, design documents and make effective presentations. |
| CO 9 | Able to apply engineering and management principles to the project as a team member. |
| CO 10 | Able to apply the project domain knowledge to sharpen one’s competency. |
| CO 11 | Able to develop a professional, presentational, balanced and structured approach towards project development. |
| CO 12 | Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project. |

**Index**

**Title page no.**

**Abstract 9**

**Chapter 1: Introduction 10**

1.1 Introduction 10

1.2 Motivation 10

1.3 Problem Definition 11

1.4 Existing Systems 12

1.5 Lacuna of the existing systems 14

1.6 Relevance of the Project 14

**Chapter 2: Literature Survey** **16**

2.1 Research Papers Referred 16

2.2 Patent search 25

2.3 Comparison of existing system and proposed area of work 26

**Chapter 3: Requirement Gathering for the Proposed System 28**

3.1 Introduction to requirement gathering 28

3.2 Functional Requirements 29

3.3 Non-Functional Requirements 30

3.4.Hardware, Software , Technology and tools utilized 31

3.5 Constraints 32

**Chapter** 4**: Proposed Design 34**

4.1 Block diagram of the system 34

4.2 Modular design of the system 36

4.3 Detailed Design 37

4.4 Project Scheduling & Tracking using Timeline / Gantt Chart 39

**Chapter 5: Implementation of the Proposed System 40**

5.1. Methodology employed for development 40

5.2 Algorithms and flowcharts for the respective modules developed 40

5.3 Datasets source and utilization 41

**Chapter 6: Testing of the Proposed System 44**

6.1 . Introduction to testing 44

6.2. Types of tests Considered 44

6.3 Various test case scenarios considered 44

6.4. Inference drawn from the test cases 45

**Chapter 7: Results and Discussion 46**

7.1. Screenshots of User Interface (UI) for the respective module 46

7.2. Performance Evaluation measures 50

7.3. Input Parameters / Features considered 52

7.4. Graphical and statistical output 52

7.5. Comparison of results with existing systems 53

7.6. Inference drawn 55

**Chapter 8: Conclusion 56**

8.1 Limitations 56

8.2 Conclusion 56

8.3 Future Scope 57

**References 58**

**Appendix 61**

List of figures 61

List of tables 62

**1. Paper I & II Details 63**

1. Paper published (Draft) 63
2. Acceptance mail 69
3. Plagiarism report 70
4. Project review sheet 70

**Abstract**

Diabetes Mellitus presents a formidable global health challenge, with India witnessing a concerning surge in its prevalence. According to the International Diabetes Federation, an estimated 77 million individuals in India grapple with this chronic metabolic disorder, a figure that continues to escalate. In response, Madhuvista: Your Diabetes Ally emerges as a groundbreaking solution, leveraging cutting-edge technology to revolutionize diabetes management. At its core, Madhuvista introduces an advanced IR sensor meticulously designed to map glucose and voltage levels in real-time and helps monitor the blood glucose levels continuously without having to use the finger-prick method, thus a non-invasive solution.

**Keywords**: Diabetes Mellitus, Non-invasive, IR sensor, Blood glucose levels

**Chapter 1: Introduction**

**1.1 Introduction**

Diabetes Mellitus, characterized by impaired insulin production or utilization, stands as a pervasive metabolic disorder posing significant global health challenges. With its prevalence surging across the globe, India, in particular, grapples with a burgeoning epidemic, where an estimated 77 million individuals are afflicted, according to the International Diabetes Federation. Over recent decades, both the absolute number of diabetes cases and its prevalence rates have witnessed a troubling upward trend, reflecting the urgent need for effective management strategies.

To address this critical issue, Our aim is to focus on early detection, prediction and post-diagnosis care and management of diabetes. To build the predictive model, we will be leveraging a comprehensive dataset, The PIMA Indian Diabetes Dataset, provided by the National Institute of Diabetes and Digestive and Kidney Diseases that aims to accurately identify individuals at risk of developing diabetes. Early detection can be done using Retinopathy, a medical condition that affects the retina, the light-sensitive tissue located at the back of the eye, resulting in change in retina’s structure and function because of damage to blood vessels caused by high blood sugar levels. This damage disrupts the normal supply of blood and nutrients to the retina, leading to the formation of abnormal blood vessels, leakage, and other changes that affect vision.

Effective monitoring of blood glucose levels, whether through invasive or non-invasive means, constitutes a cornerstone of diabetes management. Regular monitoring empowers individuals to stay vigilant about their glucose levels, facilitating timely interventions such as medication adjustments, dietary modifications, and lifestyle changes. By enabling proactive measures to mitigate dangerous fluctuations, monitoring aids in averting acute complications and mitigating the risk of long-term diabetes-related complications, thereby contributing to improved overall health outcomes for individuals living with diabetes.

**1.2 Motivation**

1. Public Health Impact: Diabetes is a prevalent chronic health condition that affects millions of people worldwide. Working on a diabetes project can have a significant positive impact on public health by developing solutions to prevent, manage, or treat the disease.
2. Patient Well-being: Improving the lives of individuals living with diabetes is a primary motivation. Projects may aim to provide better tools, treatments, or support systems to enhance the well-being and health outcomes of patients.
3. Prevention and Education: Diabetes projects often focus on preventive measures and patient education. Educating individuals about risk factors, lifestyle changes, and early detection can help prevent the onset of the disease.
4. Data-Driven Insights: Leveraging data analytics and machine learning in diabetes projects can lead to valuable insights, early disease detection, and personalized treatment recommendations.
5. Research Opportunities: Diabetes projects offer opportunities for cutting-edge research in areas such as genetics, epidemiology, pharmacology, and behavioral science.

**1.3 Problem Definition**

**Software:**

The aim of this project is to address the challenges associated with diabetes risk prediction and management. Using traditional physical assessments Long queues, limited appointment availability, high costs, and potential geographic barriers hinder timely access to healthcare professionals. This project aims to overcome these challenges by developing an AI-powered application that offers virtual diabetes risk assessment and personalized insights, providing a convenient, cost-effective, and accessible solution for individuals seeking to understand and manage their diabetes risk.

This project endeavors to create ML models that can detect ( using retinopathy) and predict diabetes by analyzing a retinal image dataset and the PIMA dataset involving patient attributes, medical history, and lifestyle choices respectively.

This can be divided into basic steps:

1. Data analysis
2. Exploratory data analysis
3. Model building
4. Saving model

**Hardware:**

1. Non-invasive blood glucose monitoring technologies have long been a goal in the medical field, offering a less intrusive and more convenient alternative to traditional fingerstick methods for diabetic patients. However, achieving accuracy and reliability comparable to invasive methods has been a significant challenge. One major obstacle is the lack of calibration stability, which has limited the widespread adoption of these technologies beyond short-term proof of principle studies.
2. Among various non-invasive technologies explored for blood glucose monitoring, including electrical, thermal, chemical, acoustical, and optical methodologies, optical methods have shown promise as offering the least intrusive means of measurement.
3. Optical methods utilize different aspects of light interaction with biological tissues to infer glucose levels. Here are some common optical techniques investigated for non-invasive blood glucose monitoring:
   * 1. Near-Infrared Spectroscopy (NIRS): Near-infrared light is directed onto the skin, and the light that is scattered back is analyzed. Glucose molecules absorb light in the near-infrared range, and by measuring the amount of absorbed light, glucose levels can be estimated.
     2. Raman Spectroscopy: This technique involves shining laser light onto the skin, which interacts with the molecular vibrations of glucose molecules. The scattered light is then analyzed to determine glucose concentrations.

**1.4 Existing Systems**

**Hardware:**

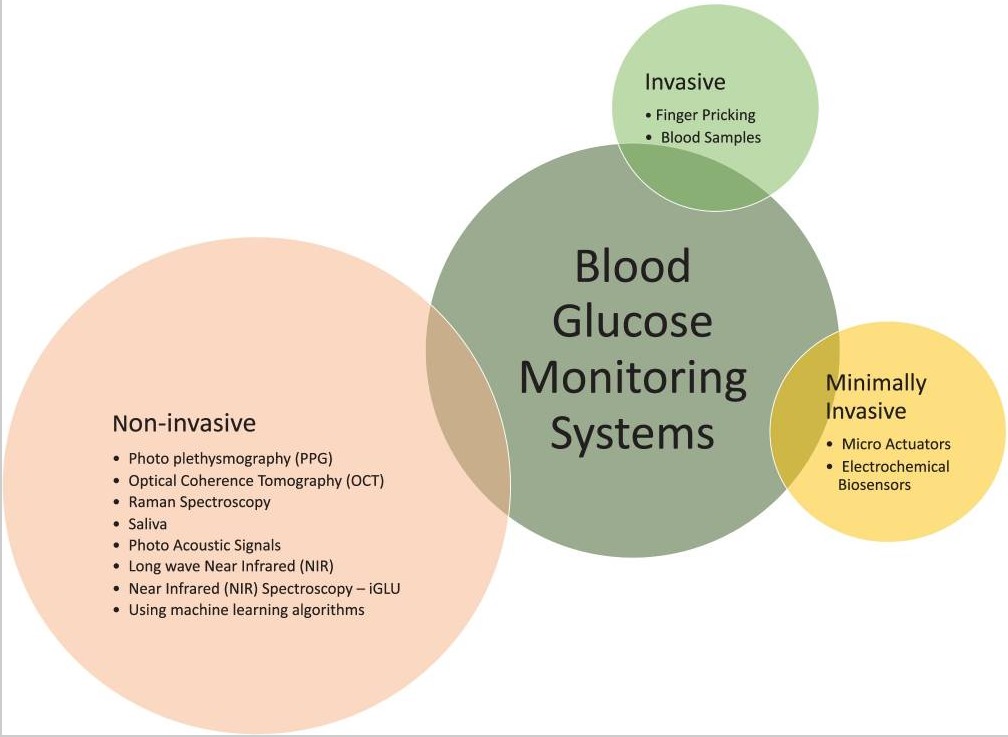


Fig1. Methods of Blood Glucose Measuring system

1. **Invasive Methods:**
   1. Fingerstick Testing: This is the most common method used by diabetic patients to monitor their blood glucose levels. A lancet is used to prick the fingertip, and a small drop of blood is placed on a test strip inserted into a glucose meter. Examples include various glucometers available in the market, such as Accu-Chek, OneTouch, and FreeStyle.
   2. Continuous Glucose Monitoring (CGM): While CGM involves the insertion of a tiny sensor under the skin, it's considered invasive because it requires a minor procedure for placement. The sensor continuously measures glucose levels in the interstitial fluid, providing real-time data. Examples include devices like Dexcom G6 and Medtronic Guardian Sensor.
2. **Non-invasive Methods:**
   1. Optical Methods: As mentioned earlier, optical methods use light to measure glucose levels through the skin without puncturing it.Example: GlucoTrack, which uses a combination of ultrasonic, electromagnetic, and thermal technologies to non-invasively measure glucose levels.
   2. Acoustic Methods: These methods measure glucose levels by analyzing sound waves or vibrations in the body.Example: Echo Therapeutics Symphony tCGM System, which utilizes ultrasound technology to measure glucose levels in the interstitial fluid.
   3. Thermal Methods: These methods measure glucose levels based on changes in skin temperature caused by variations in blood glucose concentration.Example: GlucoWise, which uses radio waves to measure blood glucose levels through the skin's thermal properties.
   4. Electrical Methods: These methods measure glucose levels by analyzing the electrical properties of tissues.Example: Glucowise, which employs electromagnetic waves to detect glucose levels by analyzing the dielectric properties of the skin.
   5. Chemical Methods: These methods involve analyzing bodily fluids other than blood, such as saliva or tears, for glucose levels.Example: GlucoSense, a non-invasive glucose monitoring device that measures glucose levels in tears using a disposable sensor.
3. **Minimally Invasive Methods:**
   1. Microneedle-based Sensors: These sensors penetrate the skin to a minimal depth, allowing for glucose monitoring with minimal discomfort.Example: Symphony tCGM by Echo Therapeutics, which uses a small sensor with a microneedle array to sample interstitial fluid for glucose measurement.
   2. Subcutaneous Sensors: Similar to CGM sensors, these devices are inserted under the skin, but they typically have smaller profiles and cause less tissue trauma. Example: FreeStyle Libre by Abbott, which uses a small, flexible sensor inserted under the skin to measure interstitial glucose levels. While it involves minimal invasiveness, it's categorized as minimally invasive due to the insertion process.

**1.5 Lacuna of the existing systems**

**Software:**

In the realm of diabetes prediction and management, a multitude of research studies have leveraged diverse machine learning techniques, such as Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes, and more. These studies consistently emphasize the significance of data preprocessing, encompassing methods like feature selection through stepwise techniques and dimensionality reduction using Principal Component Analysis (PCA). Furthermore, the utilization of advanced data mining tools like WEKA and the exploration of various datasets, including the Pima Indian diabetes database from the UCI repository, have enabled the development of predictive models for diabetes diagnosis. The adoption of techniques like bootstrapping resampling, SMOTE, and ADASYN to address class imbalances showcases a commitment to improving model accuracy. Comparative evaluations of machine learning algorithms, combined with the integration of explainable AI frameworks, have furthered our understanding of how these models function and make predictions.

**Hardware:**

1. Urine Test Strips**:** Urine test strips detect ketones in urine, indicating potential insulin deficiency, but they don't directly measure blood glucose levels. While simple to use, they're less accurate than blood glucose meters and not ideal for routine monitoring.
2. Laboratory Blood Tests**:** Periodic blood tests measure glycated hemoglobin (HbA1c) levels, reflecting average blood glucose levels over 2-3 months. This method offers convenience and a longer-term perspective but lacks immediate adjustments and requires less frequent monitoring compared to daily testing.
3. Blood Glucose Meters**:** Blood glucose meters provide quick and portable glucose level readings through a small blood sample from a fingertip prick. They offer convenience but require regular fingersticks and testing frequency may vary based on treatment plans.

**1.6 Relevance of the Project**

1. Improving Quality of Life**:** Non-invasive blood glucose monitoring technologies offer the potential to improve the quality of life for individuals with diabetes by providing a more comfortable and convenient alternative to traditional invasive methods. By reducing the need for frequent finger pricks and blood sampling, non-invasive monitoring can enhance patient compliance with glucose monitoring regimens and contribute to better diabetes management.
2. Enhancing Patient Safety**:** Invasive blood glucose monitoring techniques carry inherent risks, such as pain, discomfort, infection, and tissue damage. Non-invasive technologies mitigate these risks by eliminating the need for skin punctures and blood extraction, thereby enhancing patient safety and reducing the likelihood of complications associated with frequent blood sampling.
3. Facilitating Continuous Monitoring**:** Continuous glucose monitoring (CGM) systems based on non-invasive technologies enable real-time monitoring of glucose levels without the need for user intervention. This continuous monitoring capability provides valuable insights into glucose fluctuations, trends, and patterns, allowing for timely intervention and adjustment of treatment regimens to optimize glycemic control.

**Chapter 2: Literature Survey**

**2.1 Research papers referred**

| Sr. No | Title | Summary |
| --- | --- | --- |
| 1 | S. Sunny and S. S. Kumar, "Optical based non invasive glucometer with IoT,"  *International Conference on Power, Signals, Control and Computation (EPSCICON), Thrissur, India, 2018.* [1] | The paper discusses the development of an Optical Based Non-Invasive Glucometer with IoT technology for monitoring glucose levels in diabetic patients. The system aims to provide a painless and convenient method for measuring glucose levels without the need for traditional finger pricking. By utilizing optical methods and IoT integration, the proposed sensor circuit consists of IR LEDs and NIR photodiodes to measure glucose levels through reflected light from the body. The Beer-Lambert law is employed for signal processing, and real-time information transmission is enabled through GSM-based IoT technology. The project is implemented using Arduino IDE for performance evaluation and analytical studies. The system's potential benefits include continuous glucose monitoring, remote monitoring capabilities, and reduced risks associated with traditional invasive methods. |
| 2 | Abith V, Deepika M, Gurumoorthy M, Karpaga Devi V, Nithya R, “Non-Invasive Glucose Estimation Based on Infrared using Finger Plethysmograph ”, *International Research Journal of Modernization in Engineering Technology and Science, 2021.* [2] | The research presented in the International Research Journal of Modernization in Engineering Technology and Science focuses on non-invasive glucose estimation using infrared technology with a finger plethysmograph. The study aims to provide a more convenient and cost-effective method for monitoring blood glucose levels in diabetic patients. By utilizing an Infrared LED, a sensor called Max30100, and an Arduino UNO microcontroller, the system can measure blood glucose levels without the need for invasive procedures like finger pricking. The data obtained can be displayed on an LCD screen and transmitted wirelessly to a smartphone application via a Bluetooth module for real-time monitoring. The results of the study show promising agreement between the non-invasive sensor method and traditional invasive finger prick method in measuring blood glucose levels. This innovative approach has the potential to significantly improve the quality of life for diabetic patients and advance the field of biomedical engineering technology. |
| 3 | Arpitha.B.V, Nithin.G.M, Manoj, Krupan.K.N, Priyanka.R, “ Implementation Of Non-invasive Blood  Glucose Monitoring System  ”, *International Journal of Creative Research Thoughts , 2020.* [3] | The paper discusses the implementation of a Non-Invasive Blood Glucose Monitoring System using Raspberry Pi 3 Model B+ and the MAX30100 sensor. The system utilizes NIR spectroscopy to measure changes in light intensity on the skin tissue for glucose level determination. The sensor transmits light towards the finger, and the Raspberry Pi processes the data to determine blood glucose levels. The system aims for accuracy by conducting experiments on diabetic and non-diabetic individuals and using linear regression analysis to develop an equation for glucose level determination. The system includes a 16x2 LCD display for output. Overall, the project offers a user-friendly and painless method for monitoring blood glucose levels, potentially revolutionizing diabetes management. |
| 4 | Mhd Ayham Darwich, Anas Shahen, Abbas Daoud , Abdullah Lahia, Jomana Diab and Ebrahim Ismaiel, “ Non- invasive IR-based Measurement Of Human Blood Glucose”, *Engineering proceedings, 2023.* [4] | The paper presents a study on the development of a non-invasive blood glucose monitoring device using infrared (IR) light. The device, tested on 30 subjects including individuals with diabetes and healthy controls, utilized a 940 nm IR light source to measure glucose levels based on absorption levels of IR light waves by blood. Results showed varying signal values between normal and diabetic subjects, with the device demonstrating the ability to differentiate between patient groups. The study employed fuzzy logic to calibrate the output voltage of the IR sensor and Arduino controller into reliable glucose concentrations, achieving high accuracy with an error rate of less than 10% . Furthermore, according to the Clarke error grid analysis, the device exhibited high accuracy and reliability, with an error rate of less than 3% . |
| 5 | Nivad Ahmadian, Annamalai Manickavasagan & Amanat Ali, “Comparative assessment of blood glucose monitoring techniques: a review”, *Journal of Medical Engineering & Technology*, *2022.* [5] | The review compares various blood glucose monitoring techniques, including NIR spectroscopy, dielectric spectroscopy, and biofluid-based methods like saliva monitoring. Key points include the importance of glucose monitoring for diabetes management, challenges in NIR spectroscopy due to water interference, and the potential of non-invasive techniques like saliva sampling. The authors discuss the need for processing algorithms, site selection for accurate measurements, and suggest areas for further research to improve these monitoring techniques. |
| 6 | Mojisayo Feyikemi Owoeye & Ayodeji Babatunde Owoeye, “Implementation of a real-time Arduino Based Non-Invasive Blood  Glucose Monitoring System”, *International Journal of Advanced Academic Research, 2024.* [6] | The research by Owoeye and Owoeye focused on implementing a real-time Arduino-based system for non-invasive blood glucose monitoring using near-infrared spectroscopy. The study involved developing an optically-based glucose sensor with components like an Arduino microcontroller, a 940nm LED, a photodiode, a noise filter, an amplifier circuit, and an LED display screen.  The researchers conducted in vitro experiments to calibrate the device by measuring output voltage against infused glucose solution in water. Twelve participants were randomly selected for testing, and their fasting blood glucose levels were monitored twice to determine the mean fasting glucose level. Data analysis using a linear regression model showed a high correlation coefficient of 0.9369 between the device's measurements and actual blood glucose levels. |
| 7 | M. S., R. Selvaraj, G. G, B. S and H. J A, "Development of Non Invasive Blood Glucose Monitoring”, *International Conference on Intelligent Technologies for Sustainable Electric and Communications Systems (iTech SECOM), Coimbatore, India, 2023.* [7] | The research paper discusses the development of a non-invasive blood glucose monitoring system using near-infrared transmittance spectroscopy. The methodology involves the use of IR transmitters and receivers to measure blood glucose levels without invasive procedures like finger pricking. The system integrates a NodeMCU microcontroller and Wi-Fi modules for data processing and remote monitoring. The study compares the accuracy of glucose readings obtained from the non-invasive system with traditional invasive methods, showing promising results. The integration of BMI calculation adds value to the system by offering a comprehensive approach to health monitoring and management. |
| 8 | V. B, N. S, R. R, R. S and M. M, "Investigation and Validation of Non Invasive Blood Glucose Measurement," *International Conference on Recent Advances in Science and Engineering Technology (ICRASET), B G NAGARA, India, 2023*. [8] | The research focuses on investigating and validating non-invasive blood glucose measurement techniques using Galvanic Skin Response (GSR) and various physical parameters like temperature, heart rate, and blood pressure. The study aims to provide an affordable and non-invasive solution for monitoring blood glucose levels, particularly for diabetic individuals. The study aims for 95% accuracy with a 10% error range. Results show promising feasibility and accuracy in predicting blood glucose levels compared to conventional methods, highlighting the potential for effective diabetes management. |
| 9 | Jaya Rubi , Thella Shalem Rahul, G.Srividhya , A.Keerthana, “Non-Invasive Blood Glucose Monitoring Device”, *International Journal of Recent Technology and Engineering (IJRTE), 2019.* [9] | The Paper introduces a non-invasive blood glucose monitoring device that employs red laser light and spectroscopy techniques to measure glucose levels based on the refractive index of transmitted light. The device's methodology involves emitting red laser light through a finger hose, detecting the light with a laser detector, and using the Beer-Lambert law to calculate glucose levels based on the intensity difference in the laser light. This approach aims to provide continuous and accurate monitoring of blood glucose levels without the need for invasive procedures, offering improved linearity and accuracy compared to traditional methods. |
| 10 | A. Kassem, M. Hamad, G. G. Harbieh and C. El Moucary, "A Non-Invasive Blood Glucose Monitoring Device," *IEEE 5th Middle East and Africa Conference on Biomedical Engineering (MECBME), Amman, Jordan, 2020*. [10] | The paper introduces Glucotect, a non-invasive blood glucose monitoring device developed and tested using Near Infrared (NIR) technology. Integrated with an Arduino Uno board and a mobile application, Glucotect demonstrates an approximately 8% improvement in accuracy compared to invasive glucose monitoring devices. The study includes preliminary tests validating its reliability and establishes a manual calibration method based on correlating voltage values with invasive blood glucose readings, forming a linear calibration equation. |
| 11 | H. D. Bader and M. S. Jarjees, "Infrared-Based Non-Invasive Blood Glucose Measurement and Monitoring System," *International Conference on Engineering, Science and Advanced Technology (ICESAT), Mosul, Iraq, 2023.* [11] | The paper presents a non-invasive method for measuring blood glucose levels using near-infrared (NIR) technology, employing a PT333 sensor and an IR333 LED. By analyzing how glucose molecules absorb NIR light, the system enables glucose level assessment without invasive procedures like finger pricks. Although promising, further research is required to optimize performance, particularly for patients with diverse skin colors and conditions. Integration with mobile devices enhances usability and potential for diabetes management applications. |
| 12 | Sivaranjani S, Ananya S, Aravinth J, Karthika R, “ Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction”,  *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS),* March 2021.[12] | The research paper investigates the prediction of diabetes-related diseases utilizing machine learning techniques, namely Support Vector Machine (SVM) and Random Forest (RF) algorithms. Using the Pima Indian diabetes dataset from the UCI repository, the study focuses on data preprocessing, including handling missing values and ensuring data quality through imputation. Feature selection methods such as step forward and step backward feature selection are employed to identify the most relevant features for prediction, while Principal Component Analysis (PCA) is utilized for dimensionality reduction. The findings reveal that the RF model achieves superior accuracy (83%) compared to SVM (81.4%) after feature selection and dimensionality reduction, suggesting RF's effectiveness in predicting diabetes-related diseases in this context. |
| 13 | S.Saru, S.Subashree, “ Analysis and Prediction of Diabetes Using Machine Learning”, *International Journal of Emerging Technology and Innovative Engineering Volume 5, Issue 4,* April 2019.[13] | The research employs WEKA software to diagnose diabetes using the Pima Indian diabetes dataset. Various machine learning algorithms including Decision Trees, Naïve Bayes, and k-Nearest Neighbors (k-NN) are compared for predictive accuracy. Decision Trees achieve 94.4%, k-NN (k=1) achieves 93.79%, and Naïve Bayes achieves 79.84% accuracy. Bootstrapping resampling enhances accuracy, notably improving k-NN (k=1) from 69.93% to 93.79%. Ensemble methods are found effective, yielding a 90.36% accuracy rate for diabetes prediction. Future research may extend the method to other diseases and explore different classifiers. |
| 14 | Muhammad Azeem Sarwar, Nasir Kamal,Wajeeha Hamid; “Prediction of Diabetes Using Machine Learning Algorithms in Healthcare”; 24th International Conference on Automation and Computing (ICAC): 06-07 September 2018. [14] | The research paper delves into predictive analytics in healthcare, applying six different machine learning algorithms (KNN, Naive Bayes, SVM, Decision Tree, Logistic Regression, Random Forest) to the PIMA dataset. The performance and accuracy of these algorithms are evaluated and discussed. Notably, SVM and KNN achieve the highest accuracy of 77% compared to other algorithms. |
| 15 | Ahamed BS, Arya MS and Nancy V AO , “Prediction of Type-2 Diabetes Mellitus Disease Using Machine Learning Classifiers and Techniques”, Front. Computer. Sci. 4:835242. doi: 10.3389/fcomp.2022.835242, 2022. [15] | The research paper explores predictive analytics, comparing the performance of several classifiers including logistic regression, XGBoost, gradient boosting, decision trees, ExtraTrees, random forest, and LGBM. Utilizing the PIMA Indian Dataset from UC Irvine Repository, the study evaluates the accuracy of these algorithms. Results reveal that the LGBM classifier achieves the highest accuracy of 95.20% compared to other algorithms. The researchers suggest further exploration with different datasets and parameters to enhance prediction accuracy, and also hint at the potential for utilizing advanced versions of LGBM in future research endeavors. |
| 16 | Tasin, I., Nabil, T.U., Islam, S., Khan, R., “Diabetes prediction using machine learning and explainable AI techniques”, Healthc. Technol. Lett. 10, 1–10 , 2023.[16] | The research paper introduces an automatic diabetes prediction system developed using a private dataset of female patients in Bangladesh and the PIMA dataset. Employing extreme gradient boosting and techniques like SMOTE and ADASYN for class imbalance, the study compares various machine learning algorithms. The XGBoost classifier with ADASYN achieves the highest accuracy of 81%, accompanied by an F1 score of 0.81 and an AUC of 0.84. Furthermore, explainable AI techniques using SHAP and LIME libraries enhance model interpretability. The system includes a website framework and an Android app for real-time diabetes prediction. |
| 17 | Chang, V., Bailey, J., Xu, Q.A. *et al.* ,“Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms”, *Neural Comput & Applic* 35, 16157–16173 (2023).[17] | The research paper evaluates three interpretable supervised ML models—Naïve Bayes, random forest, and J48 decision tree—using the Pima Indians diabetes dataset in R programming. Naïve Bayes performs well with refined feature selection for binary classification, while random forest excels with more features. The study develops classification models for electronic diagnostic systems in the Internet of Medical Things (IoMT), where the random forest model outperforms Naïve Bayes and J48 decision tree models in accuracy, precision, specificity, F-score, and AUC. The paper proposes an IoMT-based e-diagnosis system for diabetes prediction and management, with future work aiming to enhance accuracy through preprocessing techniques and extend the approach to other diseases. |
| 18 | Kirti Kangra, Jaswinder Singh, “Comparative analysis of predictive machine learning algorithms for diabetes mellitus “, Bulletin of Electrical Engineering and Informatics Vol. 12, No. 3, June 2023.[18] | The research paper evaluates various machine learning algorithms, including SVM, Naïve Bayes, KNN, RF, LR, and DT, using the Pima Indian diabetic (PID) and Germany diabetes datasets. Experimentation conducted in WEKA 3.8.6 assesses performance metrics and error rates. For PID, SVM achieves 74% accuracy, while KNN and RF perform better for the Germany dataset with 98.7% accuracy. The study suggests LR as the preferable choice for both datasets, with higher accuracy and better performance in terms of ROC area. Future work may explore hybrid models and their performance on these datasets and real-time data. |
| 19 | Muhammad Exell Febriana, Fransiskus Xaverius Ferdinana , Gustian Paul Sendani, Kristien Margi Suryanigrum, Rezki Yunanda “ Diabetes prediction using supervised machine learning ” , 7th International Conference on Computer Science and Computational Intelligence, 2020.[19] | The research paper compares the effectiveness of K-Nearest Neighbor (KNN) and Naive Bayes algorithms in building an intelligent predictive model for diabetes prediction based on health attributes in the PIMA dataset. Through experiments and Confusion Matrix evaluation, Naive Bayes is found to outperform KNN in terms of accuracy, precision, and recall. Naive Bayes achieves an average accuracy of 76.07%, precision of 73.37%, and recall of 71.37%, while KNN achieves an average accuracy of 73.33%, precision of 70.25%, and recall of 69.37%. The study suggests future research could explore integrating other algorithms like neural networks and employing techniques such as Particle Swarm Optimization to further enhance accuracy and precision, along with developing practical application programs. |
| 20 | Md Shahin Ali , Md Khairul Islam , A. Arjan Das , D. U. S. Duranta , Mst. Farija Haque , and Md Habibur Rahman , “A Novel Approach for Best Parameters Selection and Feature Engineering to Analyze and Detect Diabetes: Machine Learning Insights”, Hindawi BioMed Research International Volume , 2023.[20] | The paper presents the development of a finely-tuned Random Forest algorithm with optimized parameters (RFWBP) for early detection of diabetes patients. Various data processing techniques were employed, including normalization and feature enhancement through data mining methods. RFWBP's performance was compared with other conventional methods using the PIMA dataset, employing 5-fold cross-validation to improve accuracy. Results indicate RFWBP achieves an accuracy of 95.83% with cross-validation and 90.68% without it, outperforming conventional machine learning methods in diabetes detection. Future research aims to expand analysis by incorporating more subjects and diverse datasets to provide deeper insights for more precise diabetes patient identification. |
| 21 | Alain Hennebelle , Huned Materwala, Leila Ismail, “HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System”, The 14th International Conference on Ambient Systems, Networks and Technologies (ANT) , March 15-17, 2023. [21] | The research paper presents HealthEdge, a smart healthcare framework utilizing machine learning for predicting type 2 diabetes within an integrated IoT-edge-cloud computing system. Comparative analysis using Random Forest (RF) and Logistic Regression (LR) algorithms with real-life diabetes datasets shows RF achieving a 6% higher accuracy on average compared to LR. HealthEdge is evaluated on PIMA Indian and Sylhet datasets, where RF outperforms LR with an accuracy of 78.27% for PIMA Indian and 97.23% for Sylhet. The study highlights the importance of diverse datasets and provides a foundation for future IoT-based healthcare applications. |
| 22 | Umair Muneer Butt, Sukumar Letchmunan, Mubashir Ali, Fadratul Hafinaz Hassan, Anees Baqir, and Hafiz Husnain Raza Sherazi, “Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications”, Journal of Healthcare Engineering-Hindawi, Volume 2021. [22] | The article presents a machine learning-based approach for diabetes classification, early-stage identification, and prediction, alongside an IoT-based diabetes monitoring system. Three classifiers—Random Forest (RF), Multilayer Perceptron (MLP), and Logistic Regression (LR)—are employed for diabetes classification, while Long Short-Term Memory (LSTM), Moving Averages (MA), and Linear Regression (LR) are used for predictive analysis. Utilizing the PIMA Indian Diabetes dataset, MLP achieves the highest accuracy at 86.08%, while LSTM significantly improves diabetes prediction with 87.26% accuracy. Comparative analysis with existing techniques is also demonstrated. Additionally, the paper introduces an IoT-based system for real-time diabetes monitoring using BLE sensors and smartphones, aiming to monitor vital signs, predict diabetes, with accuracy rates of 86.083% for MLP and 87.26% for LSTM. Future work includes developing an Android application and exploring genetic algorithms for improved monitoring. |
| 23 | Kopitar, L., Kocbek, P., Cilar, L. *et al.*, “Early detection of type 2 diabetes mellitus using machine learning-based prediction models”. *Sci Rep* 10, 11981 (2020). [23] | The study compares machine learning-based prediction models (Glmnet, RF, XGBoost, LightGBM) with traditional regression models for predicting undiagnosed Type 2 Diabetes Mellitus (T2DM) using fasting plasma glucose levels. Assessing prediction performance with 100 bootstrap iterations simulating new incoming data in 6-month batches, the simple regression model achieved the lowest average Root Mean Square Error (RMSE) of 0.838, followed by RF, LightGBM, Glmnet, and XGBoost. Glmnet showed the highest improvement rate (+3.4%) with additional data, while LightGBM demonstrated the highest stability in variable selection over time. The study highlights the advantages of simpler models in terms of visualization and stability, with LightGBM notably stable. Future research should explore ensemble methods while considering challenges in result interpretation for healthcare decisions. |
| 24 | Ashwini Tuppad , Shantala Devi Patil, “Machine learning for diabetes clinical decision support: a review”, *Advances in Computational Intelligence , 2022.* [24] | The review paper examines the application of machine learning (ML) in clinical decision support for type 2 diabetes, addressing gaps in existing medical knowledge and highlighting ML's potential contributions. It explores ML research in risk assessment, diagnosis methods, and prognosis modeling for predicting diabetes incidence and complications. The paper emphasizes ML's potential for diabetes prevention and management while identifying research gaps that need attention to develop clinically reliable decision support models for diabetes care. |
| 25 | Soumya K N, Vigneshwaran P, “Prediction on Type-2 Diabetes Mellitus Using Machine Learning Methods”, Volume 41: Advances in Parallel Computing Algorithms, Tools and Paradigms, 2022. [25] | The text discusses the increasing prevalence of Type II diabetes due to lifestyle changes and its association with complications such as insulin resistance, kidney, eye, and heart issues. It also explores ongoing research into the potential link between diabetes and cancer. Using dimensionality reduction, classification, and clustering techniques, the study compares existing classifiers with the PIMA Indian diabetes dataset and Stanford AIM-94 dataset. The text underscores the extensive research efforts worldwide in this field, providing motivation for further exploration. While many tests haven't conclusively linked Type II diabetes and cancer, the focus remains on understanding factors contributing to malignant growth and differentiating between harmful and benign tumors. The current work aims to establish potential connections between Type II diabetes and various cases through data-driven information design and machine learning techniques for accurate classification into normal and abnormal categories. |

**Table 1**. Literature Survey

**2.2 Patent Search**

1. **Non-Invasive Blood Glucose Monitoring System**

The present disclosure introduces a compact and portable spectroscopy module integrated with advanced image processing and machine learning algorithms for non-invasive blood glucose monitoring. Positioned on a user's finger or ear, the device captures spectroscopic images using a laser and camera, enabling glucose estimation based on tissue light absorption. Image processing techniques extract relevant features fed into machine learning models trained on a large dataset, including linear models and Artificial Neural Networks. Real-time glucose readings are displayed on a user-friendly interface, offering convenient and reliable monitoring. Evaluation shows promising accuracies, with about 79% for finger readings and 62% for ear readings. AdaBoost trained with KNeighbors emerged as an advantageous model, with color intensity data from the red channel proving beneficial. Additionally, accuracy testing across demographics showed consistent performance, with minor impacts from factors like nail polish color and skin pigmentation. [26]

1. **Smart-Tooth Blood Glucose Measurement Device**

The smart-tooth glucose monitoring device revolutionizes blood glucose measurement by offering a non-invasive or partially-invasive alternative to traditional glucometer methods. By embedding sensors and an optic source within the pulp chamber of a molar tooth, this innovative device facilitates direct blood measurement, eliminating the need for lancets and strips. Through pulsed photoacoustic spectroscopy, it accurately correlates glucose levels with sensor readings, providing real-time data for effective diabetes management. Beyond glucose monitoring, the device demonstrates potential for broader healthcare applications, such as measuring hematocrit, body temperature, and more. This breakthrough technology represents a significant advancement in personalized healthcare, offering a convenient and efficient solution for continuous monitoring of vital health parameters. [27]

1. **Diabetes prediction using glucose measurements and machine learning**

The method described involves predicting diabetes using machine learning and glucose measurements. It utilizes historical glucose measurements and outcome data from a user population to train a machine learning model. This model predicts an individual user's diabetes classification based on their glucose measurements collected by a wearable monitoring device over several days. The prediction, which can categorize the user's diabetes state or predict adverse effects, can be communicated to the user through notifications or user interfaces. The historical data includes measurements from wearable devices and diagnostic measurements like HbAlc and FPG, indicating clinical diabetes diagnosis. The platform provides insights into the user's diabetes status and prompts necessary actions, such as contacting a healthcare provider.[28]

**2.3 Comparison of Existing system and proposed area of work**

In the realm of diabetes prediction and management, existing research has extensively utilized diverse machine learning techniques such as Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes. These studies underscore the importance of data preprocessing, including feature selection and dimensionality reduction, to develop predictive models for diabetes diagnosis. Additionally, various data mining tools like WEKA have been employed alongside datasets like the Pima Indian diabetes database to enhance model accuracy. Techniques such as bootstrapping resampling and class imbalance addressing methods like SMOTE and ADASYN have been pivotal in improving model performance. Comparative evaluations of machine learning algorithms, coupled with explainable AI frameworks, have furthered our understanding of model functionality and prediction mechanisms.

On the other hand, existing research on non-invasive blood glucose monitoring has explored techniques such as Near-Infrared (NIR) Spectroscopy, red laser light, Raman spectroscopy, and analysis of biofluids and physiological parameters. While promising results have been reported, challenges persist in ensuring accuracy across diverse skin tones, calibration against finger-prick measurements, and developing advanced algorithms for real-time glucose estimation.

**Proposed solution:**

MadhuVista presents a novel approach to blood glucose monitoring by combining advanced IR sensor technology, Attiny85 microcontroller integration, and Decision Tree Regressor models implemented via TinyML. Unlike invasive finger pricking methods, MadhuVista offers continuous, non-invasive monitoring of blood glucose levels in real-time. By utilizing Attiny85 for analog voltage output and mapping it to glucose levels, MadhuVista ensures streamlined operation and reduced dependency on external hardware. Furthermore, the integration of Decision Tree Regressor models via TinyML enables accurate glucose level estimation, even with real-time data. The solution exhibits a mean absolute error (MAE) of 3.00, mean squared error (MSE) of 36, and root mean squared error (RMSE) of 6, thus promising a more effective and user-friendly approach to diabetes management.

Additionally, our solution stands out by implementing five different algorithms for comparison and statistical analysis, addressing a gap in existing systems. Visualizations such as correlation heatmaps, feature importance plots, and AUC-ROC curves have been integrated, enhancing interpretability and insights into model performance. Furthermore, the implementation of Optical Character Recognition (OCR) technique facilitates data extraction and user input, streamlining the process. To enhance user experience, we have improved the user interface by incorporating sliders in input fields, making the system more intuitive and user-friendly.

**Chapter 3: Requirement Gathering for the Proposed System**

**3.1 Introduction to requirement gathering**

For the software part of the project, the PIMA datasets from National Institute of Diabetes and Digestive and Kidney Diseases, Sylhet dataset, and a custom dataset made using pathology lab reports were used.

For the hardware part of the project, data from various research papers for univariate relationship between blood glucose levels and analog voltage levels was collected, using which implementation of the work on the existing research papers was done. Along with that, a custom dataset was created by making use of the following components:

1. Arduino: A microcontroller platform used for building digital devices and interactive objects that can sense and control physical devices.
2. IR Sensor: An infrared sensor that detects infrared radiation emitted by objects to measure distance, proximity, or object detection.
3. ATtiny85: A low-power, high-performance 8-bit AVR microcontroller from Atmel, often used in projects where space and power consumption are critical.
4. Glucometer: A medical device used to measure the concentration of glucose in the blood, commonly used by individuals with diabetes to monitor their blood sugar levels.
5. LED (Light-Emitting Diode): A semiconductor light source that emits light when an electric current passes through it, commonly used for visual indicators or illumination in electronic devices.
6. OLED (Organic Light-Emitting Diodes) enable emissive displays - which means that each pixel is controlled individually and emits its own light (unlike LCDs in which the light comes from a backlighting unit)
7. Breadboard: A solderless breadboard is a fundamental component used for prototyping and testing electronic circuits. It allows you to quickly connect and disconnect components without soldering.
8. Zero PCB: For testing out circuits before soldering the actual PCB.
9. Jumper Wires: These wires are used to create connections between components on a breadboard. They come in different lengths and colors to help organize and troubleshoot your circuit.
10. Resistors: Resistors are passive electronic components that limit the flow of electric current in a circuit. They are commonly used to control the voltage and current in various parts of the circuit.
11. Coin Battery: To power up the circuit.

**3.2 Functional Requirements**

**Software:**

1. User Registration and Authentication: Users should be able to register for an account. The system should support secure authentication methods (e.g., username/password, biometrics) to ensure data privacy.
2. Medical Records Management: Users (patients and healthcare professionals) should be able to input, edit, and view medical records, including diagnosis, treatment plans, lab results, and medications.
3. Accuracy and reports: Patients should be able to check the accuracy and correct prediction for the data that he entered.

**Hardware:**

Data Acquisition:

1. For acquiring data from sensors to create a relationship between the voltage levels and the corresponding blood glucose levels, you would need appropriate sensor modules like HW201 IR Sensor which is then interfaced with the microcontroller platform (e.g., Arduino or ATTiny85).
2. Glucose sensors typically provide analog voltage outputs proportional to glucose levels. You can use ADC (Analog-to-Digital Converter) pins available on the Arduino or ATTiny85 to convert these analog signals into digital values.
3. Ensure that the sensor readings are stable and accurate by implementing constant environment conditions and calibration procedures during collection and processing.
4. Data Processing:
5. Once the sensor data is acquired, the microcontroller (Arduino or ATTiny85) should perform necessary calculations or conversions to extract glucose level information.
6. For glucose sensors, you may need to apply calibration factors or equations provided by the sensor manufacturer to convert raw sensor readings into glucose concentration values.
7. Additionally, if temperature compensation is required for glucose measurements, you can incorporate temperature sensor readings into the calculation algorithm.
8. Implement error-checking mechanisms to ensure the integrity of acquired data and processed results.
9. Display:
10. To display the glucose concentration values, you can use an output device such as an LCD or an OLED display or a serial monitor for plotting and visualization. We have used an OLED display.
11. For the OLED, Connection to the microcontroller platform (Arduino or ATTiny85) using appropriate pins and libraries has been done. Display the glucose concentration values in a clear and readable format on the screen.
12. Alternatively, A HC-05 Bluetooth module or a wifi module can be interfaced.

**3.3 Non-Functional Requirements**

**Software:**

1. Performance: Response Time: The system should respond to user interactions within a reasonable timeframe (e.g., sub-second response for most actions).
2. Scalability: The system should be scalable to accommodate an increasing number of users and data volume.
3. Throughput: It should handle a certain number of simultaneous users or requests without performance degradation.
4. Reliability: The system should be highly available and reliable, minimizing downtime or service interruptions. It should have mechanisms for disaster recovery and data backup.
5. Security: Patient data should be encrypted both in transit and at rest. Access to patient records should be role-based and follow strict access controls.
6. Usability: The user interface should be intuitive, user-friendly, and designed with user experience (UX) principles in mind.

**Hardware:**

1. Performance: Use efficient algorithms and coding practices to minimize processing overhead. Optimize loops, minimize unnecessary operations, and use hardware resources effectively. Employ interrupt-based programming techniques where applicable to ensure timely response to sensor inputs without blocking the main execution flow.Utilize microcontroller platforms with sufficient processing power and memory to handle the required computations and data storage efficiently. Consider implementing parallel processing or multitasking strategies if the system complexity demands concurrent execution of multiple tasks.Optimize sensor sampling rates and data transmission rates to achieve real-time or near-real-time glucose level measurements.
2. Reliability: Design the hardware with robustness in mind, using high-quality components and ensuring proper electrical connections and shielding to minimize susceptibility to environmental factors and electromagnetic interference.Implement error detection and handling mechanisms to detect and recover from unexpected inputs, sensor malfunctions, or communication errors.Include built-in self-test routines to verify the integrity of critical system components and perform periodic calibration checks to maintain measurement accuracy. Employ watchdog timers or other hardware-based watchdog mechanisms to monitor the system's operation and reset it if it becomes unresponsive or enters an error state. Conduct thorough testing and validation under various environmental conditions to ensure the reliability and durability of the hardware implementation.
3. Accuracy: Calibrate sensors regularly according to manufacturer guidelines and compensate for any drift or variation in sensor output over time.Account for sensor accuracy limitations and potential sources of measurement error, such as environmental conditions, interferences, and physiological factors.Implement error correction algorithms or filtering techniques to mitigate noise and improve the accuracy of glucose level measurements. Validate the accuracy of the hardware implementation through comparison with reference measurements or clinical data and refine the calibration algorithms as needed.Provide feedback mechanisms or indicators to alert users of potential measurement errors or out-of-range readings and guide them in taking corrective actions.

**3.4. Hardware, Software , Technology and tools utilized**

**Software:**

1. Operating System: Compatible operating systems include Windows, macOS, or Linux for Android application development.
2. Integrated Development Environments (IDEs): Jupyter Notebook Google Colab Visual Studio Code
3. Streamlitto build the webapp.

**Hardware:**

1. Languages: Arduino Sketch. (based on C/C++)
2. IDE: Arduino IDE

**3.5 Constraints**

1. Budget Constraints: Limited financial resources may necessitate prioritization of essential hardware and software components over more advanced or optional features.Consider open-source or low-cost alternatives for hardware platforms, sensors, and development tools to minimize expenses. Optimize resource utilization and minimize operational costs by selecting energy-efficient components, streamlining processes, and leveraging cloud services for data storage and processing.
2. Time Constraints: Tight deadlines require efficient project planning and execution, including clear task prioritization, resource allocation, and milestone tracking.Adopt agile development methodologies to facilitate iterative development and accommodate changes in requirements or priorities. Allocate sufficient time for testing, validation, and regulatory compliance activities to ensure the reliability and safety of the final product. Collaborate closely with stakeholders and team members to identify potential bottlenecks or delays early and implement mitigation strategies proactively.
3. Regulatory and Compliance Constraints: Compliance with healthcare regulations such as HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation) is essential to protect patient privacy and ensure data security. Allocate resources for legal and regulatory compliance expertise or consulting services to navigate complex regulatory requirements effectively. Implement robust data encryption, access controls, and audit trails to safeguard sensitive health information and demonstrate compliance with regulatory standards. Stay updated on evolving regulatory requirements and industry best practices to adapt the system architecture and policies accordingly.
4. Data Availability Constraints: Limited access to comprehensive and accurate patient health data may require collaboration with healthcare providers, research institutions, or data repositories to access relevant datasets. Explore partnerships or collaborations with organizations that have access to large-scale health data repositories or patient cohorts to facilitate data collection and analysis.Implement data anonymization and aggregation techniques to protect patient privacy while still deriving meaningful insights from available data sources. Consider leveraging synthetic or simulated data generation techniques to augment limited datasets and train machine learning models effectively.
5. Technological Constraints: Outdated or limited technology infrastructure may require investment in upgrading hardware, software, or networking infrastructure to support the system's capabilities and scalability. Conduct a thorough assessment of existing technological constraints and identify areas for improvement or modernization to enhance system performance and reliability. Prioritize interoperability and compatibility when selecting hardware and software components to ensure seamless integration and future scalability. Explore cloud-based solutions or platform-as-a-service (PaaS) offerings to offload infrastructure management and leverage scalable computing resources on-demand.

**Chapter 4: Proposed Design**

**4.1 Block diagram of the system**

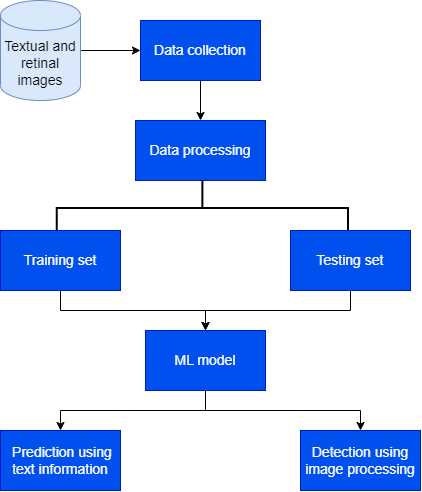


Fig2. Block diagram for the software part

The above block diagram gives a brief overview of the project. Being a Machine Learning project, the foremost step is data collection from the most relevant sources in the required format.We required text-based data for prediction of diabetes while image-based data for detection of diabetes using Retinopathy.

In the next step, the data collected is in the raw format and hence it needs to be pre-processed wherein we fill the missing values(if there are any in the dataset), perform normalization, etc. After preprocessing, the data is divided into training and testing sets. Usually, 75% or more data is used as training samples and the rest as testing samples. The ML model is then using a specific algorithm such as Random Forest, Support Vector Machine(SVM), etc.

Once the model is trained, we can analyze its performance using various accuracy measures like precision, recall, f-score, confusion-matrix, etc.

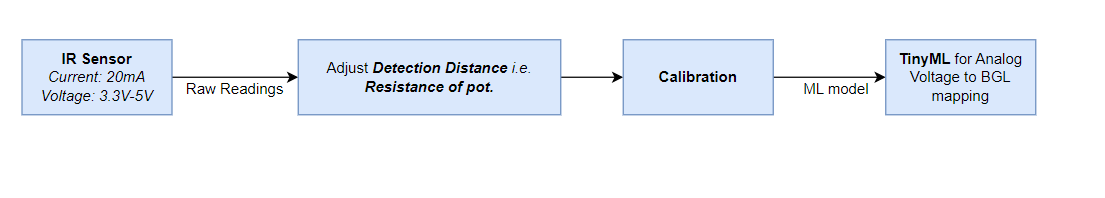


Fig 3.Block diagram for the hardware part

The block diagram shows a representation of how the hardware works in general. We have used an IR sensor. An infrared (IR) sensor is an electronic device that measures and detects infrared radiation in its surrounding environment. An IR receiver LED and an IR transmitter LED are both types of light-emitting diodes (LEDs) that are used in infrared (IR) communication.

An IR receiver LED is a device that detects infrared signals from remote controls and other IR sources.An IR transmitter LED, on the other hand, is a device that emits infrared light in order to send signals to other devices. The device has a trimmer pot to adjust the distance of objects. This distance can be adjusted according to the requirements. Once this distance is fixed, it is calibrated. Using TinyML, the analog voltages from the sensor are used for mapping with the Blood Glucose Level.

IR sensors can be sensitive to changes in temperature, leading to drift in readings over time if not properly compensated. Ambient light, especially sunlight or artificial lighting, can interfere with IR sensor readings, causing inaccuracies.Improper calibration or drift in calibration settings can result in inaccurate readings. Regular recalibration may be necessary to maintain accuracy. Some IR sensors may be sensitive to the distance and angle between the sensor and the target object. Environmental conditions such as humidity, dust, and air quality can affect IR sensor performance, especially for outdoor applications. Some IR sensors may have limited detection ranges, leading to errors if the target object is beyond the sensor's range.

**4.2 Modular design of the system**

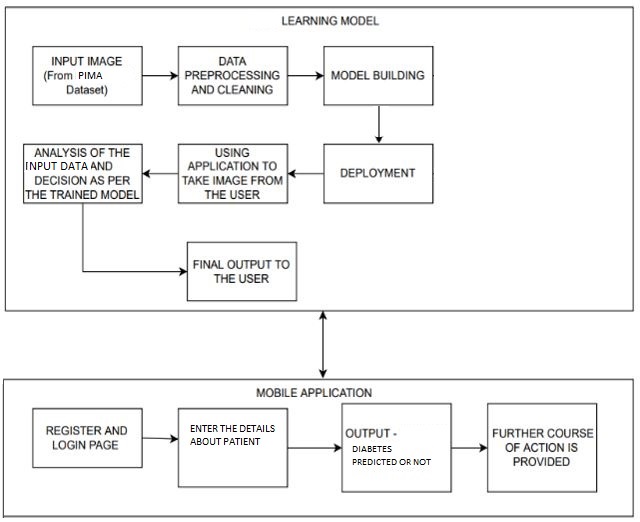


Fig 4. Modular design for software part

The modular diagram explains the model as two sections. The first section i.e.the learning model explains the steps required for training the model and the input given to the model and output given as a classification result based on the model.

The application section explains how the model can be used in real-life scenarios. A user-interface will collect all the required inputs from the user that will be given to the model. Based on the input, the model predicts whether the user is diabetic or not.

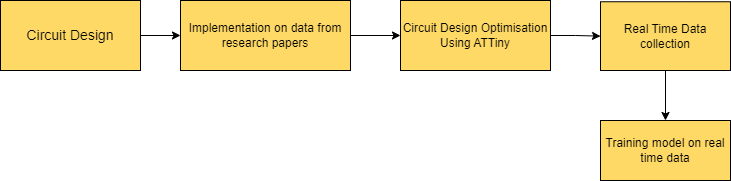


Fig 5. Modular design for hardware part

The modular diagram explains the models employed for achieving the task. It begins with designing the circuit to measure the voltage with the help of an IR sensor. The circuit consists of an arduino working as the microcontroller, IR sensor and OLED to get the analog values of voltage. The voltage is mapped to the Blood Glucose Level measured through the traditional methods and this data is used for training the model.

The model is first trained with data collected from various research papers. We measure the accuracy in each case. Furthermore, the circuit design can be optimized by using ATTiny.

With the help of ATTiny, we could achieve the following optimizations:

* circuit simplification
* easier to carry around for more data collection to enhance accuracy measure

The next module is real time data collection. The Blood Glucose Level is measured with the help of Glucometer using appropriate precautions. The data collected is mapped to the voltage values measured using the designed hardware. A classification model is trained on this data and analyzed for testing. In this way, we could determine whether the person is diabetic or not with the help of the voltage measurement recorded on our instrument non-invasively.

**4.3 Detailed Design**

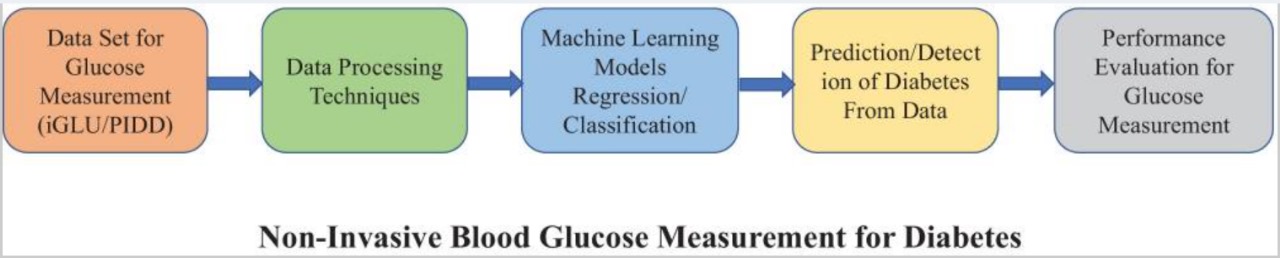


Fig 6. Detailed design of the system

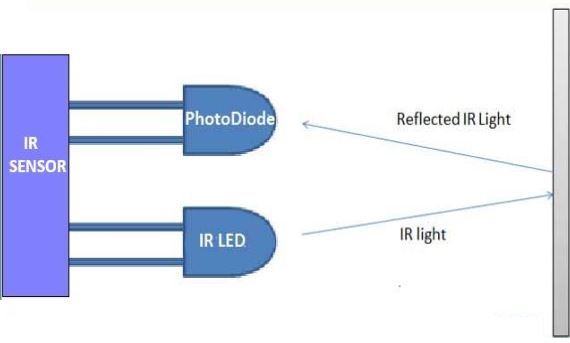


Fig 7. IR sensor

The working of IR sensor can be explained as follows:

When an IR signal is detected, the IR receiver LED will emit a small amount of visible light, which can be used to confirm that a signal has been received.

An IR transmitter LED, on the other hand, is a device that emits infrared light in order to send signals to other devices. It is typically a small, clear, or translucent device that emits IR light in a specific frequency range. With the help of an appropriate circuit consisting of Arduino as the microcontroller, we can get the voltage value on an OLED device.

This reading of the sensor is processed and fed to the Machine Learning model. A regression model is used to find the relationship between glucose level and voltage value. After a relation is established, the model is evaluated using various accuracy measures. It can then be used to predict the value of glucose level provided a voltage value.

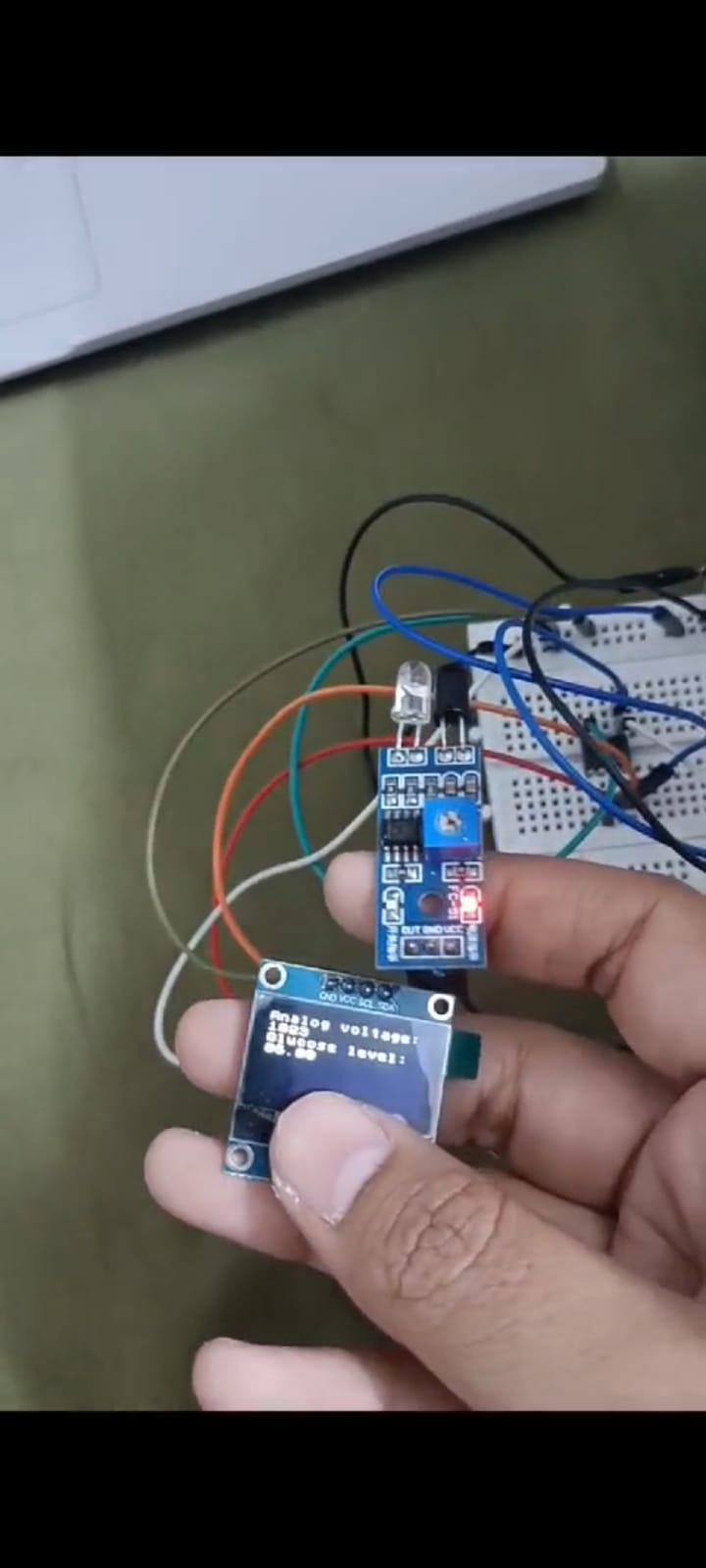


Fig 8. Testing the ATtiny85 Circuit on Breadboard before soldering it to zero PCB

**4.4 Project Scheduling & Tracking using Timeline / Gantt Chart**

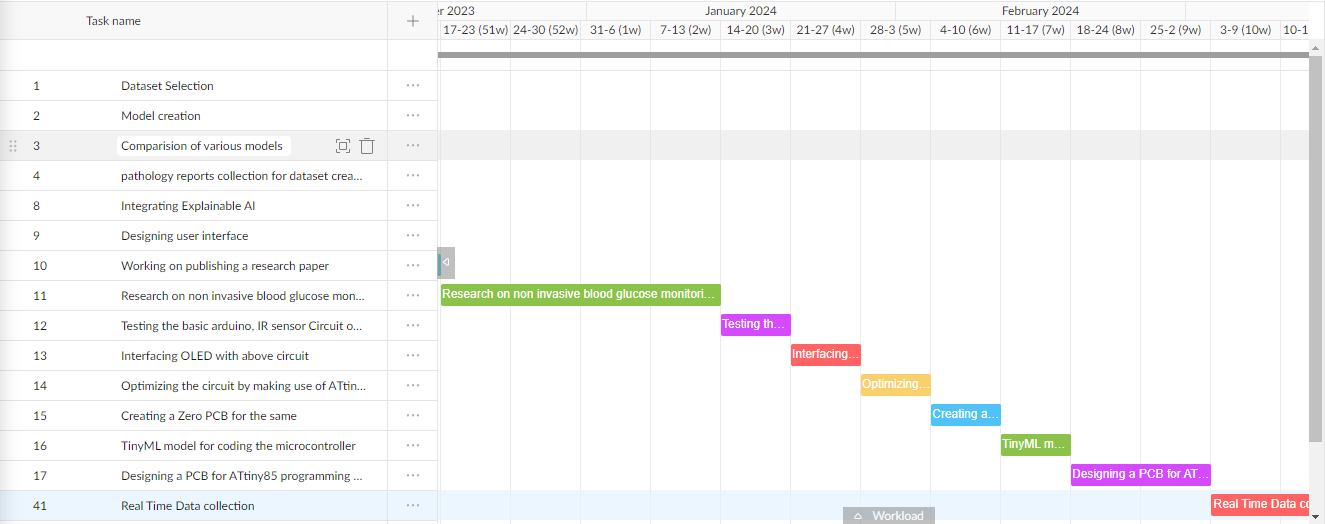


Fig 9. Gantt chart

**Chapter 5: Implementation of the Proposed System**

**5.1. Methodology employed for development**

**Software:**

Calculating the Evaluation Measures for following Algorithms

1. Decision Tree
2. Random Forest
3. SVM
4. XGBoost
5. AdaBoost

Using the

1. PIMA Dataset
2. Sylhet dataset.
3. The first step is for the sensor to take raw readings of its environment. This could be anything from the distance of an object to the temperature of a room. The raw readings are then sent to the microcontroller for Calibration and Processing.
4. The collected sensor readings of analog voltage and blood glucose levels are then put through a ML model to create a mathematical relationship between the two. This model is then used to predict glucose levels of people whose analog voltage of IR sensor is known. The ML model is made using **TinyML** and **m2cgen** library is used to convert the ML model code to C language which is compatible with Arduino Uno and the Attiny85.
5. Also, the voltage of the sensor is 3.3-5V. This information is likely important for the person who is calibrating the sensor, as it can help them to ensure that the sensor is functioning properly.
6. Features like thickness of skin, age, gender, blood pressure also have an effect on the blood glucose levels, so a **multivariate regression model** has to be made.

**5.2 Algorithms and flowcharts for the respective modules developed**

For establishing relationship between glucose and voltage level, we used the following algorithms:

1. Linear Regression

Linear regression is a statistical method used to model the relationship between two variables by fitting a linear equation to observed data points. One variable is considered the independent or explanatory variable, while the other is the dependent variable. The goal is to find the best-fitting line that minimizes the differences between the observed and predicted values. Linear regression is widely used in various fields, including economics, finance, and social sciences, to analyze and predict the behavior of dependent variables based on one or more independent variables.

1. Decision Tree Regression

Decision tree regression is a supervised learning technique used for predicting continuous outcomes. It involves constructing a decision tree model where each internal node represents a decision based on the value of a feature, and each leaf node represents the prediction for the target variable. Decision trees are built recursively by splitting the data into subsets based on the values of features, with the goal of minimizing the variance of the target variable within each subset. Decision tree regression is commonly used in fields such as finance, healthcare, and engineering for tasks such as risk assessment and predictive modeling.

1. Random Forest Regression

Random forest regression is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and robustness. It works by constructing a multitude of decision trees during training and outputting the average prediction of the individual trees for regression tasks. Each tree is trained on a random subset of the training data and a random subset of features, which helps to reduce overfitting and improve generalization performance. Random forest regression is widely used in fields such as finance, bioinformatics, and marketing for tasks such as stock price prediction, gene expression analysis, and customer churn prediction.

**5.3 Datasets source and utilization**

For our project, we used datasets that are available on the internet along with the data collected realtime by us. The data sources can be listed as:

1. <https://iopscience.iop.org/article/10.1088/1742-6596/2325/1/012021/meta>
2. <https://www.semanticscholar.org/paper/Non-Invasive-Monitoring-of-Glucose-Level-in-Blood-Bobade-Patil/0b8e694b2308097513da3f7cefc8c5462e1de67a>
3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10669386/>

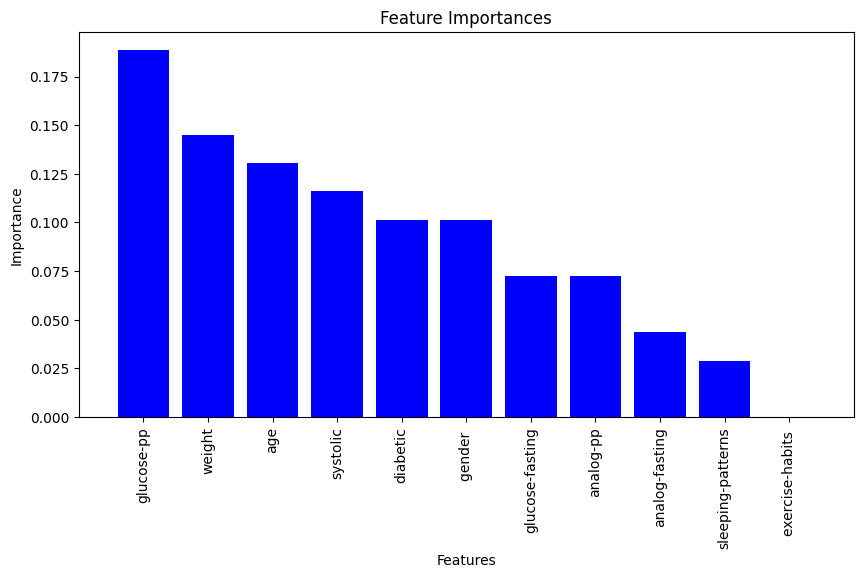
Real time data collection:

* For our project, we collected data from individuals whom we knew so that the analysis gets real.
* It very well proved useful for the project.



Fig 10. Real-time data collection

Since, the model's accuracy was limited having considered only voltage to determine the BGL, we tried creating a multi variate model where we used variables such as age, gender weight, exercise habits, sleep patterns, and blood pressure apart from voltage. We trained three models namely Decision tree and random forest on the data collected with these parameters. The feature importance plot for Random Forest classifier can be seen as below:



Here's the interpretation of each column in the provided dataset:

1. age: Represents the age of the individual in years.
2. gender: Indicates the gender of the individual (e.g., "F" for female).
3. weight: Denotes the weight of the individual in kilograms.
4. sleeping-patterns: Describes the sleeping patterns of the individual (e.g., "regular").
5. exercise-habits: Specifies the exercise habits of the individual (e.g., "active").
6. blood-pressure: Represents the blood pressure of the individual, typically measured as systolic pressure over diastolic pressure (e.g., "115/75").
7. analog-fasting: Indicates the analog fasting glucose level of the individual.
8. analog-pp: Denotes the analog postprandial (after meal) glucose level of the individual.
9. glucose-fasting: Represents the fasting glucose level of the individual.
10. glucose-pp: Specifies the postprandial glucose level of the individual.
11. diabetic: Indicates whether the individual is diabetic or not (e.g., "yes" or "no").
12. Unnamed: 11: Additional information or comments associated with the individual.

**Chapter 6: Testing of the Proposed System**

**6.1 . Introduction to testing**

Testing is integral to ensuring the functionality and reliability of any system. In this chapter, we detail the testing procedures undertaken for the system, encompassing both the user interface (UI) and the hardware component for non-invasive blood glucose monitoring.

**6.2. Types of tests Considered**

1. Unit Testing : Modules were tested individually, including UI components and hardware modules, initially on breadboard and later on zero PCB.
2. Integration Testing : Interaction between UI modules and backend components was validated, alongside the integration of hardware components with the software interface.
3. Functional Testing : Core functionalities, such as accurate diabetes prediction and blood glucose level estimation, were thoroughly tested.
4. Performance Testing : System response time, prediction accuracy, and stability were evaluated under various conditions.
5. Usability Testing : The UI's ease of use and functionality were assessed, gathering user feedback to refine the interface.

**6.3 Various Test Case Scenarios Considered**

Test cases included:

1. User Interface Testing : Ensuring proper functionality of input fields, buttons, and form validation, along with testing different input parameter combinations for accurate prediction.
2. Model Testing : Utilizing TinyML, Python ML models were converted to Code compatible for Arduino (C), and their accuracy was verified using diverse datasets.
3. Hardware Testing : Validation of IR sensor accuracy for capturing voltage values, and testing the mapping model for accurate blood glucose level estimation.
4. Data Validation : Verification of data integrity and accuracy, ensuring consistency with expected relationships between sensor readings and blood glucose levels.
5. Integration Testing : Verifying communication between UI and backend services, and ensuring data consistency between hardware and software components.

**6.4. Inference drawn from the test cases**

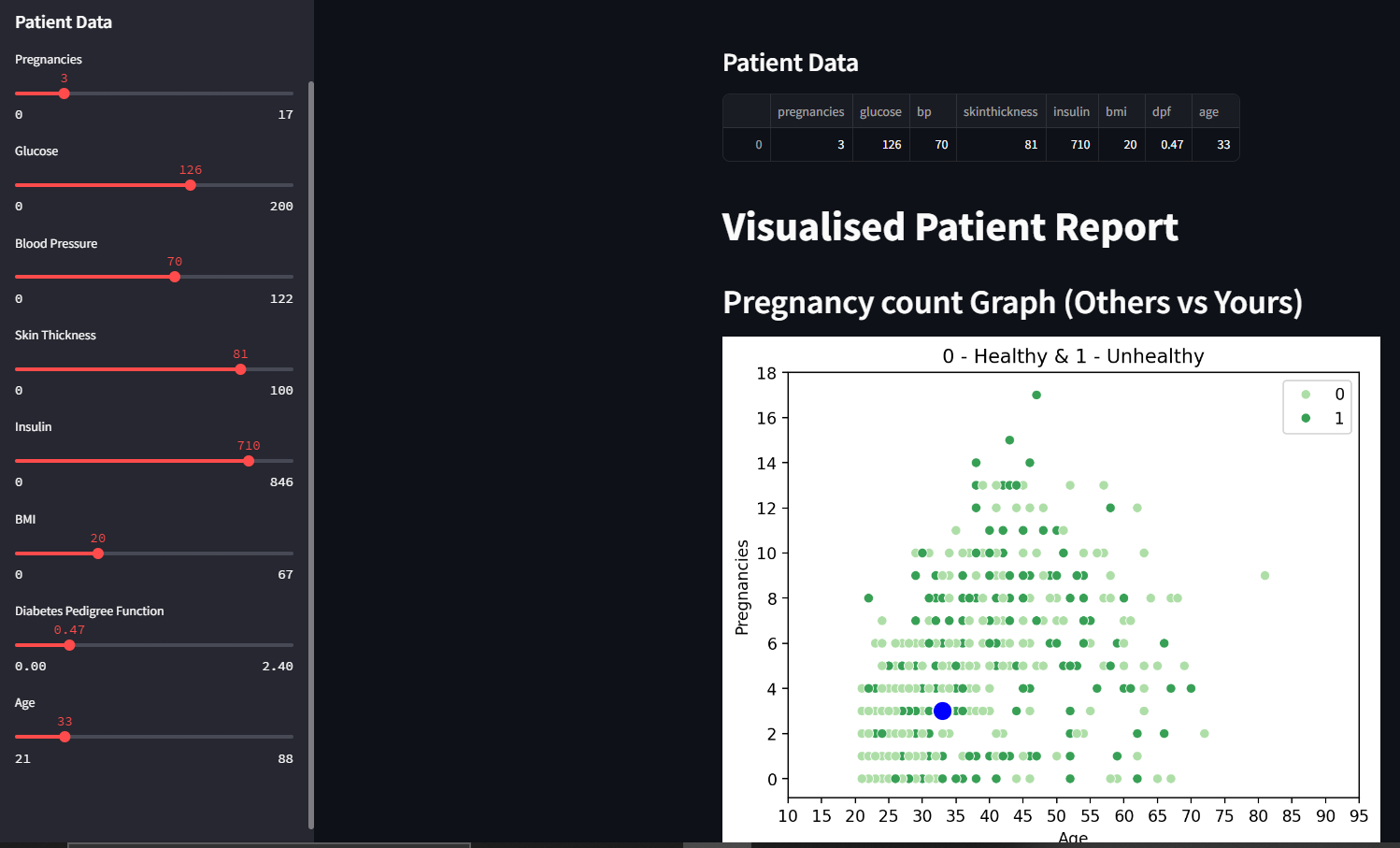
Based on conducted tests:

1. Accuracy of Prediction : ML models demonstrated high accuracy in diabetes prediction, validated through rigorous testing.
2. Functionality of Hardware : Hardware components, from breadboard to zero PCB, provided reliable blood glucose estimates under varying conditions.
3. Future Integration : The possibility of integrating WiFi and Bluetooth modes was identified for seamless integration of the system.

**Chapter 7: Results and Discussion**

**7.1. Screenshots of User Interface (UI) for the respective module (Implementation)**

**Software:**

****

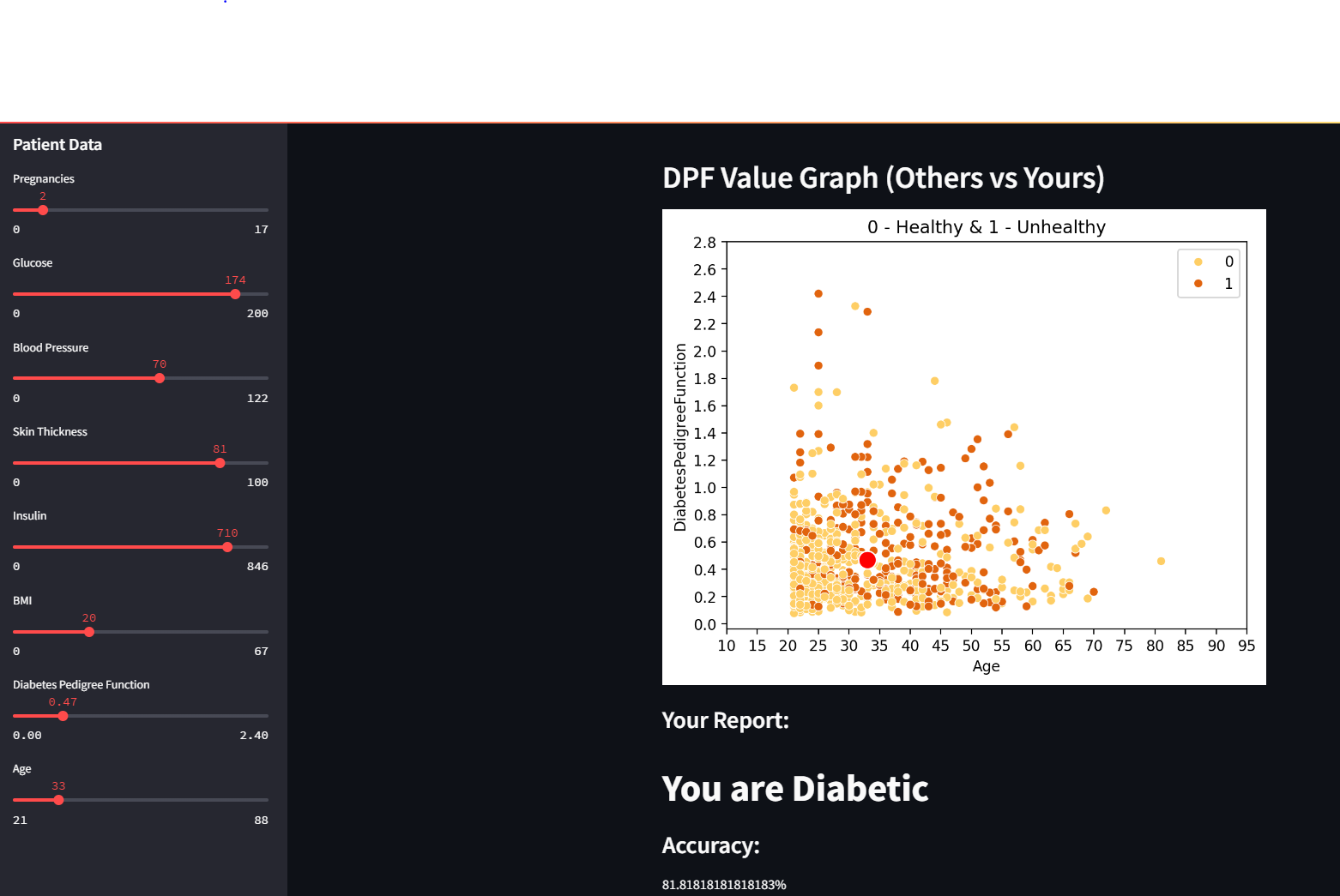
****

Fig 11. UI pages

**Hardware:**

1. **Arduino and IR sensor**

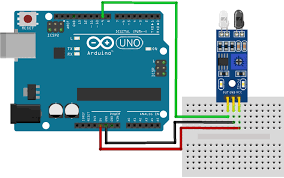


Fig 12. Arduino and IR sensor

(Readings displayed on Serial Monitor).

1. **Interfacing OLED with Arduino and IR sensor**

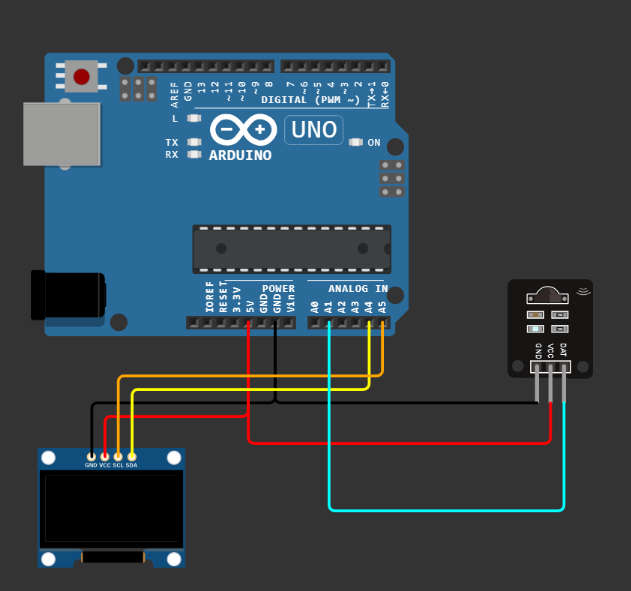


Fig 13. Interfacing OLED with Arduino and IR sensor

1. **Interfacing of Attiny85 with Oled & IR Sensor (Realization of simplified version of above circuit, alternative to using arduino)**

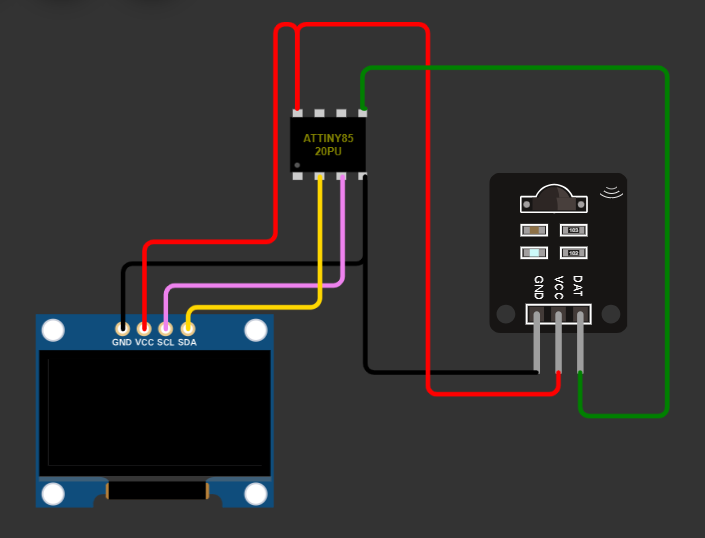


Fig 14. Interfacing of Attiny85 with Oled & IR Sensor

Working circuit on Zero PCB:



Fig 15. Working circuit on Zero PCB

1. Since use of a smaller Microcontroller ATtiny85 can be made, the programming of Attiny85 still requires an Arduino as an ISP (In-Circuit Serial Programmer), Using the following connections after burning the bootloader on ATtiny85:

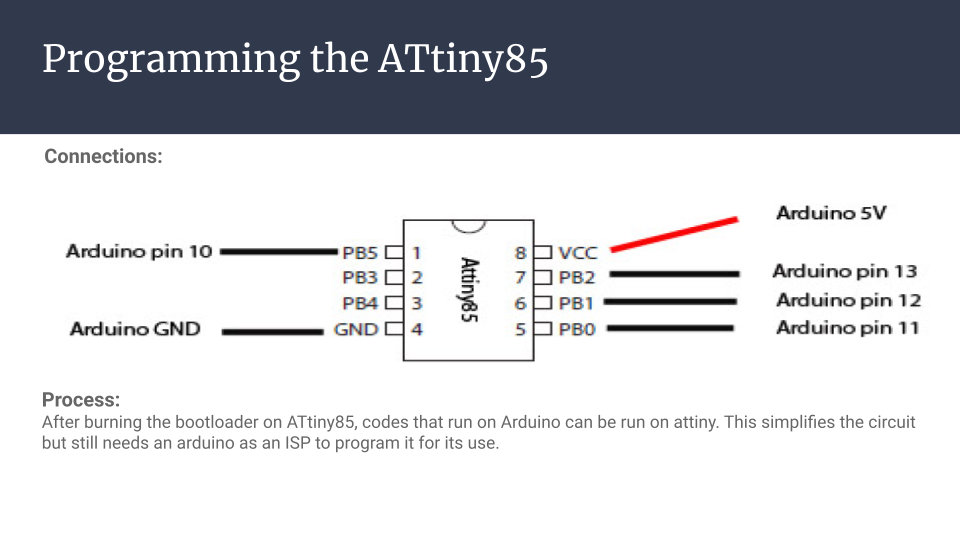


Fig 16. Pin diagram of Attiny85

Thus, a PCB (Printed Circuit Board) has been made, so the ATtiny85 can be programmed by itself as an ISP.

Schematic for the same:

****

Fig 17. Schematic of PCB

**The PCB:**

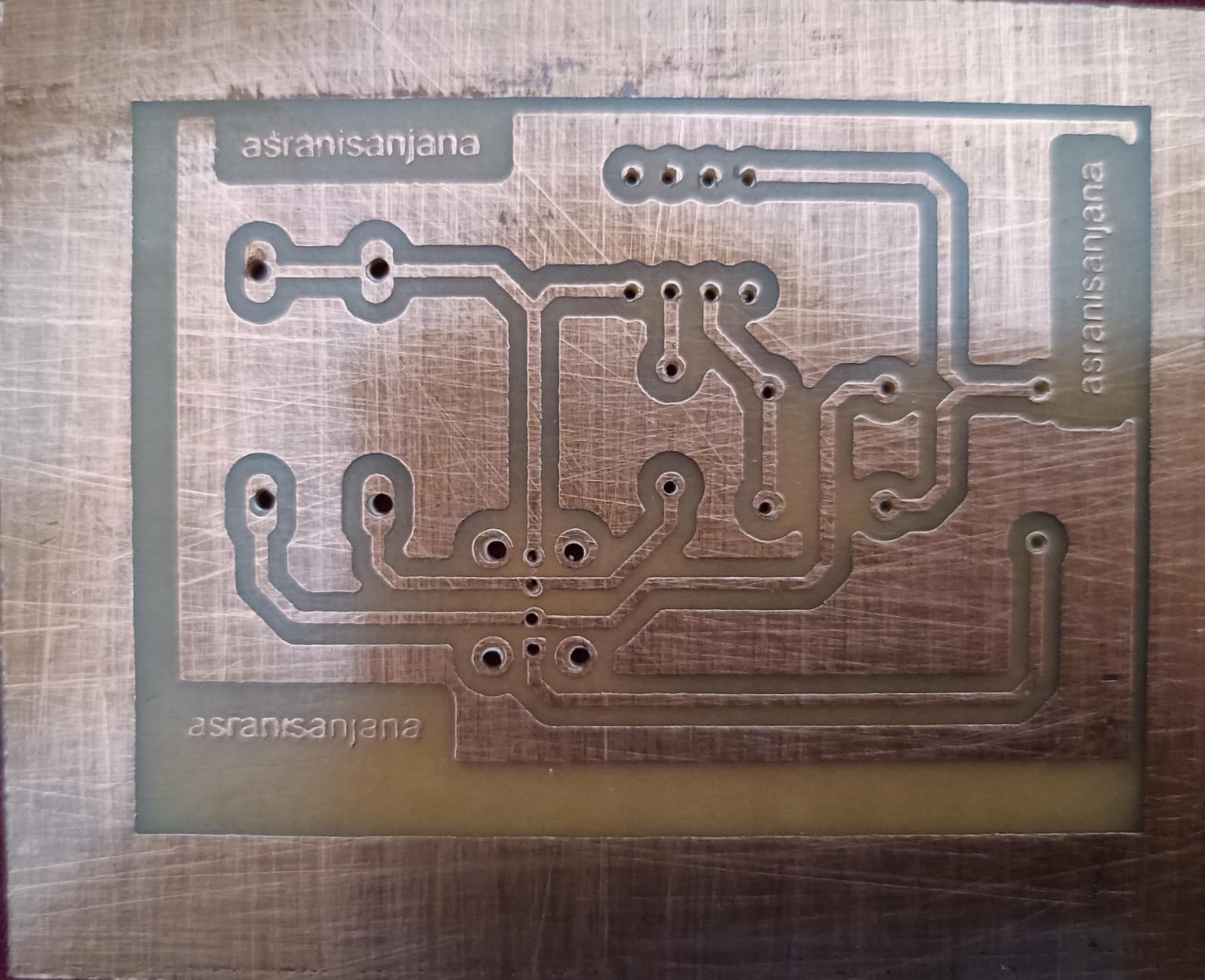
****

Fig 18. Printed Circuit Board (PCB)

**ML Model for Analog voltage to glucose level mapping:**

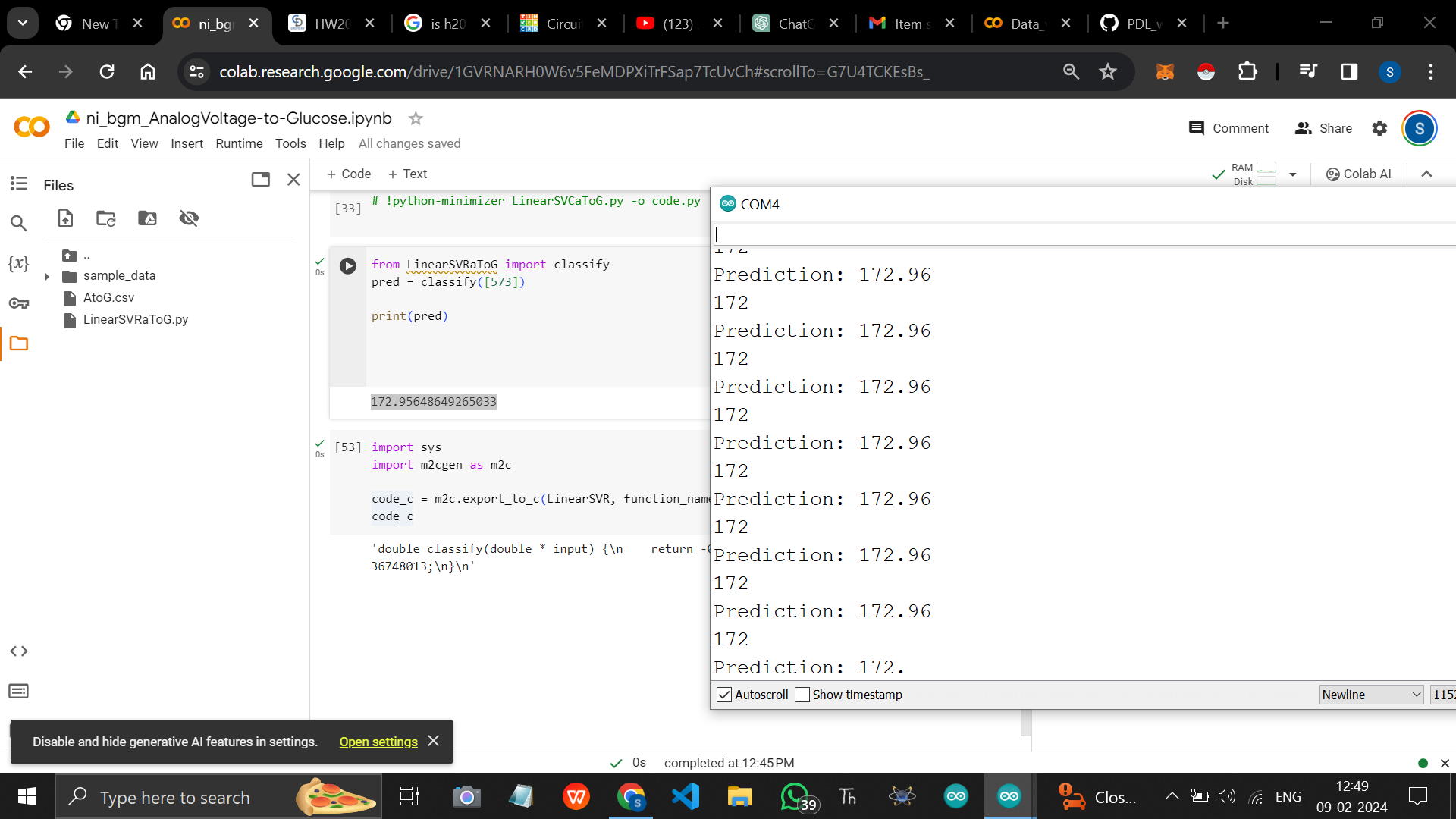
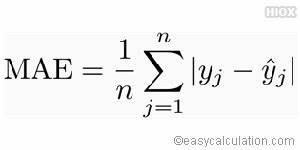
****

Fig 19. Analog voltage values as output on the screen

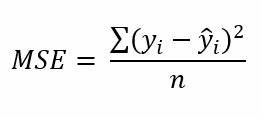
**7.2. Performance Evaluation measures**

Performance evaluation measures are essential metrics used to assess the effectiveness and accuracy of predictive models in machine learning and data analysis tasks.

1. Mean Absolute Error (MAE):
   1. Mean Absolute Error (MAE) is a metric used to evaluate the accuracy of a predictive model by measuring the average absolute difference between the predicted values and the actual values. It provides a straightforward indication of how close the predictions are to the true values, regardless of their direction. MAE is calculated by taking the average of the absolute differences between each predicted value and its corresponding actual value. A lower MAE indicates better prediction accuracy, with zero representing a perfect prediction.
   2. 

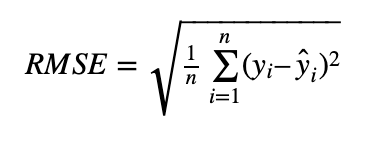
where:

* + - 1. n is the number of observations.
      2. yj is the actual value of the target variable for observation i.
      3. ŷj is the predicted value of the target variable for observation i.
      4. | · | denotes the absolute value.

1. Mean Squared Error (MSE):
   1. Mean Squared Error (MSE) is a commonly used metric for assessing the performance of regression models. It calculates the average of the squared differences between the predicted values and the actual values. MSE penalizes larger errors more heavily than smaller ones due to the squaring operation. It is calculated by summing the squared errors and dividing by the number of observations. MSE provides insight into the average magnitude of errors but does not offer a direct interpretation of the error scale since it is in squared units.
   2. 

where:

* + - 1. n is the number of observations.
      2. yi is the actual value of the target variable for observation i.
      3. ŷi is the predicted value of the target variable for observation i.

1. Root Mean Squared Error (RMSE):
   1. Root Mean Squared Error (RMSE) is the square root of the MSE and is often used as a more interpretable measure of prediction error. RMSE shares the same units as the target variable, making it easier to understand the scale of prediction errors. By taking the square root of the MSE, RMSE provides a measure of the average magnitude of prediction errors in the original units of the data. Like MSE, lower RMSE values indicate better model performance, with zero representing perfect predictions. RMSE is particularly useful for comparing models and understanding the practical significance of prediction errors.
   2. 

where:

1. n is the number of observations.
2. yi is the actual value of the target variable for observation i.
3. ŷi is the predicted value of the target variable for observation i.
4. √ denotes the square root.

**7.3. Input Parameters / Features considered**

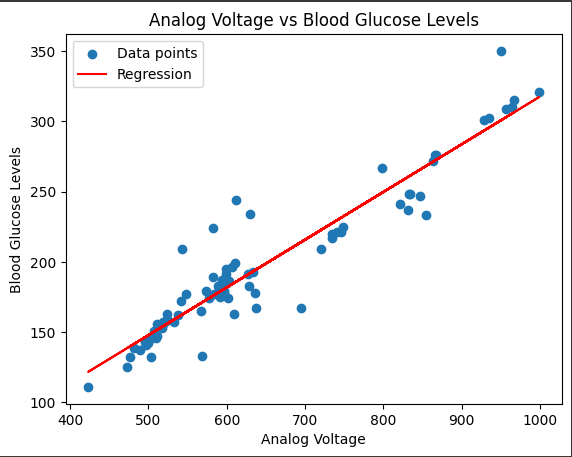
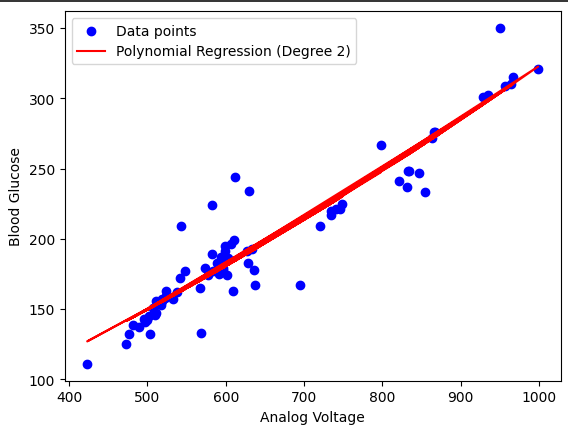
In the context of blood glucose monitoring, two key parameters are considered: voltage and glucose concentration. Voltage, a measurable electrical signal, is typically captured using an infrared (IR) sensor, which detects and quantifies the intensity of infrared radiation emitted or reflected by the skin. This voltage reading acts as a vital input in the monitoring process. Furthermore, Age, Gender, Skin Thickness and Blood Pressure have to be considered as well for input parameters.

To establish a correlation between voltage and glucose concentration, a mapping table is constructed based on empirical data or calibration experiments. This table correlates different voltage readings obtained from the IR sensor with corresponding glucose concentrations measured through traditional invasive methods, Here, blood sampling. Through this correlation, a mapping function or algorithm is developed, enabling the estimation of glucose concentrations based solely on voltage measurements.

In this mapping framework, voltage serves as the independent variable, representing the input provided by the IR sensor. Conversely, glucose concentration acts as the dependent variable, which depends on the aforesaid demographic parameters like age, gender, etc. Using which, we can accurately estimate blood glucose concentrations non-invasively, offering valuable insights into an individual's metabolic status without the need for frequent blood sampling.

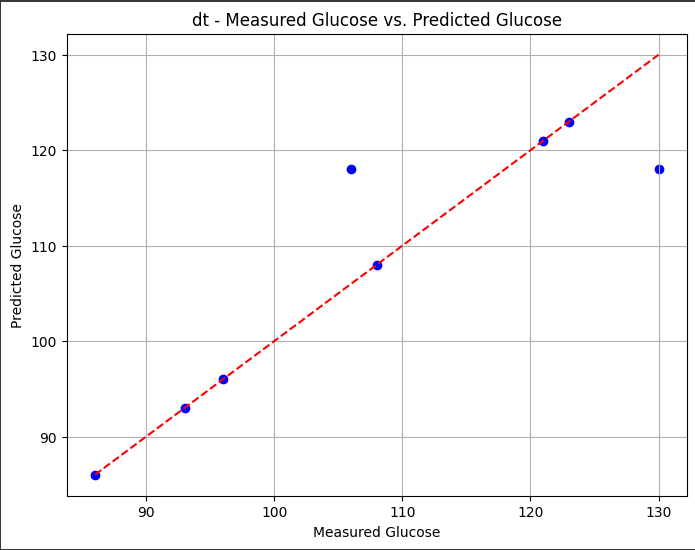
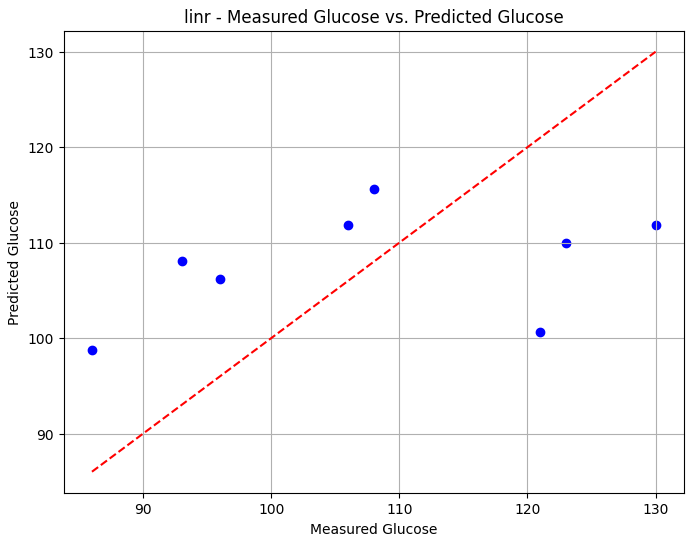
**7.4. Graphical and statistical output**

1. **Graphs below are based on already available dataset:**
2. Polynomial regression 2. Linear regression

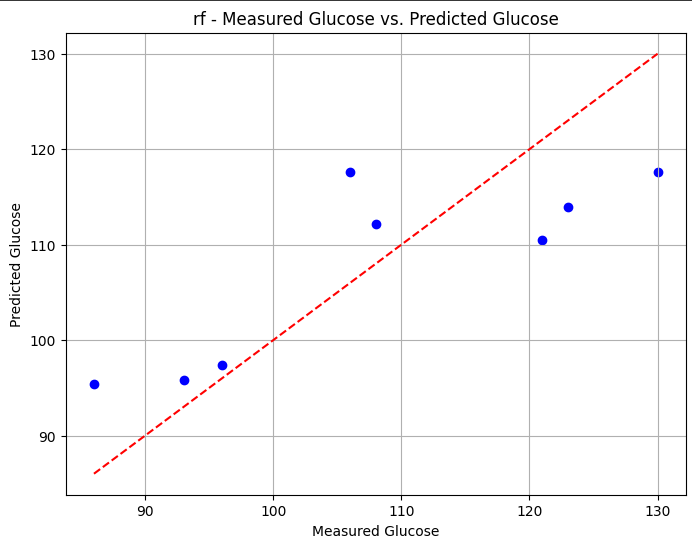


**2) Graphs below are based on real data that we collected:**

1. Linear regression 2. Decision tree



3. Random forest



**7.5. Comparison of results with existing systems**

**Software:**

We employed decision trees, Support Vector Machines (SVM), Random Forest classifiers, XGBoost and AdaBoost and evaluated their accuracy using different data split ratios (80:20 and 70:30) and cross-validation techniques (5-fold and 10-fold) for both of the datasets.

Accuracy Comparison for all algorithms

PIMA:

|  | SVM ( non-linear: rbf kernel ) | Decision tree | Random Forest | Adaboost | XGBoost |
| --- | --- | --- | --- | --- | --- |
| 80:20 | 75% | 75% | 76.6% | 73.37 | 73.37% |
| 70:30 | 74% | 70% | 75.3% | 74.45 | 73.59% |
| 5-fold cross | 76.17% | 73% | 85.7% | 80.52 | 76.43% |
| 10-fold cross | 76.43% | 71% | 80.2% | 84.21 | 76.17% |

**Table 2**. All Algorithms accuracy comparison wrt PIMA dataset

The other dataset:

|  | SVM | Decision tree | Random Forest | Adaboost | XGBoost |
| --- | --- | --- | --- | --- | --- |
| 80:20 | 89% | 96% | 100% | 93.26 | 99.03% |
| 70:30 | 94% | 97% | 96.7% | 92.94 | 98.71% |
| 5-fold cross | 91.15% | 96% | 100% | 90.38 | 96.35% |
| 10-fold cross | 92.31% | 97% | 100% | 94.23 | 96.92% |

**Table 3**. All Algorithms accuracy comparison wrt the other dataset

**Hardware:**

In the realm of comparing results with existing systems, a notable gap emerges in the lack of comprehensive comparison charts provided by the prior systems. However, this lacuna spurred our endeavor to implement a thorough comparative analysis within our proposed solution. Through meticulous execution, we curated a detailed comparison chart, juxtaposing the performance metrics of the algorithms integrated into our system with those of existing methodologies.

1. Comparison of algorithms implemented on data collected from research papers-I

|  | Decision Tree | Random Forest | Linear Regression |
| --- | --- | --- | --- |
| Mean Absolute Error | 11.78 | 7.19 | 7.62 |
| Mean Squared Error | 333.76 | 141.71 | 79.93 |
| Root Mean Squared Error | 18.26 | 11.9 | 8.94 |

**Table 4 :** Comparison of implemented algorithms wrt research paper - 1

1. Comparison of algorithms implemented on data collected from research papers-II

|  | Decision Tree | Random Forest | Linear Regression |
| --- | --- | --- | --- |
| Mean Absolute Error | 14.8 | 12.31 | 8.57 |
| Mean Squared Error | 353.2 | 225.06 | 107.05 |
| Root Mean Squared Error | 18.79 | 15 | 10.35 |

**Table 5 :** Comparison of implemented algorithms wrt research paper - 2

1. Comparison of algorithms implemented on real data

|  | Decision Tree | Random Forest | Linear Regression |
| --- | --- | --- | --- |
| Mean Absolute Error | 3.00 | 7.72 | 12.89 |
| Mean Squared Error | 36.00 | 74.64 | 187.71 |
| Root Mean Squared Error | 6.00 | 8.64 | 13.70 |

**Table 6 :** Comparison of implemented algorithms wrt real data

**7.6. Inference drawn**

After all the comparisons,it is observed that on real data, the decision tree works the best as MAE is 3 percent followed by random forest with MAE 7.72 and finally Linear regression with 12.89 percent.Thus in hardware implementation we integrated decision tree as our prime algorithm.

**Chapter 8: Conclusion**

**8.1 Limitations**

1. Limited Dataset:

The accuracy of the model could be hindered by the availability of a limited dataset. Access to larger and more diverse datasets could enhance the model's performance and reliability.

1. Performance Variability:

Madhuvista's performance may vary based on environmental factors, individual differences among users, and other external influences such as ambient light or skin tone. These variations could impact the accuracy and reliability of glucose level measurements.

1. Disparate Report Formats

One limitation we encountered was the disparate formats of reports collected from various labs, necessitating manual processing. Moreover, these reports risked becoming obsolete over time, posing challenges for maintaining the dataset's relevance and usability.

**8.2 Conclusion**

In our study, we implemented decision trees, Support Vector Machines (SVM), Random Forest classifiers, XGBoost, and AdaBoost, evaluating their accuracy using different data split ratios and cross-validation techniques on two datasets (Sylhet and PIMA). Additionally, we collected pathological reports to build our own dataset. However, we encountered limitations such as different reports collected from different labs having different formats, requiring manual processing and becoming obsolete over time.

Transitioning to non-invasive blood glucose monitoring technology, our proposed solution offers a promising approach to tackle the increasing prevalence of diabetes mellitus. Leveraging **state-of-the-art** IR sensor technology and innovative microcontroller integration, the proposed solution provides a user-friendly and continuous glucose monitoring system without the discomfort of traditional finger-prick methods. However, it's essential to acknowledge the limitations of the proposed solution, including potential challenges related to environmental factors, user adherence, maintenance demands, and privacy considerations. Overcoming these hurdles will be crucial to ensuring widespread acceptance and sustained effectiveness. Moving forward, continued research and development efforts are necessary to enhance the solution's accuracy, reliability, and compatibility with existing healthcare infrastructure, ultimately empowering individuals with diabetes to better manage their condition and enhance their overall well-being.

The critical conclusions emerged from our analysis were:

1. Addressing accuracy and overfitting necessiates acquiring more data.
2. Employing multiple sensors and refining their calibration processes are pivotal for enhanced performance.

**8.3 Future Scope**

1. Clinical Validation:

Future advancements in clinical validation will likely focus on expanding the scope and scale of large-scale trials to encompass diverse populations and environments. Additionally, the integration of real-world data and long-term monitoring studies may provide further insights into the performance and reliability of non-invasive blood glucose monitoring systems.

1. Advanced Machine Learning:

The future of machine learning in blood glucose monitoring may involve the development of personalized models tailored to individual users, leveraging continuous feedback and adaptive learning techniques. Exploring novel architectures and algorithms, such as recurrent neural networks and generative adversarial networks, could further enhance prediction accuracy and robustness.

1. Sensor Technology Optimization:

Continued research and development efforts may lead to advancements in NIR light sources, including the exploration of novel materials and fabrication techniques to improve efficiency and spectral characteristics. Integration of additional physiological sensors, such as sweat sensors and temperature sensors, could enable comprehensive health monitoring and provide valuable contextual information for glucose measurements.

1. User Experience and Integration:

Develop a user-friendly app and integrate with smartwatches and healthcare platforms for seamless data management.

**References**

1. S. Sunny and S. S. Kumar, "Optical based non invasive glucometer with IoT,"  *International Conference on Power, Signals, Control and Computation (EPSCICON), Thrissur, India, 2018.*Available: [https://ieeexplore.ieee.org/document/8379597](https://ieeexplore.ieee.org/document/8379597/)

1. Abith V, Deepika M, Gurumoorthy M, Karpaga Devi V, Nithya R, “Non-Invasive Glucose Estimation Based on Infrared using Finger Plethysmograph ”, *International Research Journal of Modernization in Engineering Technology and Science, 2021.* Available: <https://www.academia.edu/51124895/NON_INVASIVE_GLUCOSE_ESTIMATION_BASED_ON_INFRARED_USING_FINGER_PLETHYSMOGRAPH>

1. Arpitha.B.V, Nithin.G.M, Manoj, Krupan.K.N, Priyanka.R, “ Implementation Of Non-invasive Blood Glucose Monitoring System”, *International Journal of Creative Research Thoughts , 2020.*

Available*:* <https://ijcrt.org/papers/IJCRT2006301.pdf>

1. Mhd Ayham Darwich, Anas Shahen, Abbas Daoud , Abdullah Lahia, Jomana Diab and Ebrahim Ismaiel, “ Non- invasive IR-based Measurement Of Human Blood Glucose”, *Engineering proceedings, 2023.*

Available: <https://www.mdpi.com/2673-4591/35/1/27>

1. Nivad Ahmadian, Annamalai Manickavasagan & Amanat Ali, “Comparative assessment of blood glucose monitoring techniques: a review”, *Journal of Medical Engineering & Technology*, *2022.*

Available:<https://pubmed.ncbi.nlm.nih.gov/35895023/>

1. Mojisayo Feyikemi Owoeye & Ayodeji Babatunde Owoeye, “Implementation of a real-time Arduino Based Non-Invasive Blood Glucose Monitoring System”, *International Journal of Advanced Academic Research, 2024.*

Available: <https://www.openjournals.ijaar.org/index.php/ijaar/article/view/409>

1. M. S., R. Selvaraj, G. G, B. S and H. J A, "Development of Non Invasive Blood Glucose Monitoring”, *International Conference on Intelligent Technologies for Sustainable Electric and Communications Systems (iTech SECOM), Coimbatore, India, 2023.*

Available: <https://ieeexplore.ieee.org/abstract/document/10435168>

1. V. B, N. S, R. R, R. S and M. M, "Investigation and Validation of Non Invasive Blood Glucose Measurement," *International Conference on Recent Advances in Science and Engineering Technology (ICRASET), B G NAGARA, India, 2023*.

Available: <https://ieeexplore.ieee.org/abstract/document/10420263>

1. Jaya Rubi , Thella Shalem Rahul, G.Srividhya , A.Keerthana, “Non-Invasive Blood Glucose Monitoring Device”, *International Journal of Recent Technology and Engineering (IJRTE), 2019.*

Available: <https://www.researchgate.net/publication/364087495_Non-Invasive_Blood_Glucose_Monitoring_Device>

1. A. Kassem, M. Hamad, G. G. Harbieh and C. El Moucary, "A Non-Invasive Blood Glucose Monitoring Device," *IEEE 5th Middle East and Africa Conference on Biomedical Engineering (MECBME), Amman, Jordan, 2020*.

Available: <https://ieeexplore.ieee.org/document/9265170>

1. H. D. Bader and M. S. Jarjees, "Infrared-Based Non-Invasive Blood Glucose Measurement and Monitoring System," *International Conference on Engineering, Science and Advanced Technology (ICESAT), Mosul, Iraq, 2023.*

Available: <https://ieeexplore.ieee.org/document/10347308>

1. Sivaranjani S, Ananya S, Aravinth J, Karthika R, “ Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction”,  *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS),* March 2021.

Available: <https://ieeexplore.ieee.org/document/9441935>

1. S.Saru, S.Subashree, “ Analysis and Prediction of Diabetes Using Machine Learning”, *International Journal of Emerging Technology and Innovative Engineering Volume 5, Issue 4,* April 2019.

Available: <https://ssrn.com/abstract=3368308>

1. Muhammad Azeem Sarwar, Nasir Kamal,Wajeeha Hamid; “Prediction of Diabetes Using Machine Learning Algorithms in Healthcare”; *24th International Conference on Automation and Computing (ICAC)*: 06-07 September 2018.

Available: <https://ieeexplore.ieee.org/abstract/document/8748992/authors#authors>

1. Ahamed BS, Arya MS and Nancy V AO , “Prediction of Type-2 Diabetes Mellitus Disease Using Machine Learning Classifiers and Techniques”, *Front. Computer. Sci. 4:835242. doi: 10.3389/fcomp.2022.835242,* 2022.

Available: <https://www.frontiersin.org/articles/10.3389/fcomp.2022.835242/full>

1. Tasin, I., Nabil, T.U., Islam, S., Khan, R., “Diabetes prediction using machine learning and explainable AI techniques”, *Healthc. Technol. Lett*. 10, 1–10 , 2023.

Available: <https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/htl2.12039>

1. Chang, V., Bailey, J., Xu, Q.A. *et al.* ,“Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms”, *Neural Comput & Applic* 35, 16157–16173 (2023).

Available: <https://link.springer.com/article/10.1007/s00521-022-07049-z>

1. Kirti Kangra, Jaswinder Singh, “Comparative analysis of predictive machine learning algorithms for diabetes mellitus “, *Bulletin of Electrical Engineering and Informatics* *Vol. 12, No. 3, June 2023.*

Available: <https://www.beei.org/index.php/EEI/article/view/4412>

1. Muhammad Exell Febriana, Fransiskus Xaverius Ferdinana , Gustian Paul Sendani, Kristien Margi Suryanigrum, Rezki Yunanda “ Diabetes prediction using supervised machine learning ” , *7th International Conference on Computer Science and Computational Intelligence*, *2020*.

Available: [//pdf.sciencedirectassets.com/280203/](https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050922X00197/1-s2.0-S1877050922021858/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEK3%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJHMEUCIQDiR%2BvVna7sWtMNVyBSY7lHy7e3bwHIPatBjgrpNkhxdAIgQX%2B2F7n81t4rqqYlfU1hUwV06iMZZiPDjXLTPCOk7moqswUIdhAFGgwwNTkwMDM1NDY4NjUiDPD7ulw8cAiBCHrBfCqQBTt1koFd7wvINXRTfB9%2BiOaHHtjaxzG12ljURdYH%2F1r2lD6YyEyI1lsUSOmLDyeAcCN%2FHOwRnoW1dA8P5NZuea39HyPgRIZAPOgJ8NW%2BhK%2FiG%2FGRKkxVnrAgkoJ19QtlGf1xU1AerXCHoce83JMAE5yZCafO7Qb6qNtI0sAB%2BpDCBUZfaVbj2aQaOwbCVEczSBEJvhvZnQYI5c2BfLDeINYPQcJSmMk8SUJaEOfQdWwSKAjQ8wEMXxyexnHH1mCcpdWT2jKnQaMu%2BK91B65JkyO9E0k7%2BowW%2FiN21CypeY7uKwZvcNF0BC5kSDyuOtmMbf92JN3Uaf9Db3tM0RbNKDuKznKdJdiu7DUf2kalWEV0WxZVPPJqGM3U5gwLm7lVXyXGMqEFMXGt7Fi7zqQqi3w3SJLVzOxfjVmKEnTdoiPoWCGFX8pu8yDuB4rV9%2BvRD1y9Lkc6TALMLFZGrOwRPKIypD785MxtpQiXqDEkQ3Upij7EBeHjc1ukhrW8KLhvEeusLf2oXgNGMzWefnc3NrqJNhrPicXueZfwbnlTXRbVLr5rJwMItboclEwt789L6Nmq%2FHHifn7sBl3rY6Yy5B9ZpYJNaCuHQtuvcoT5fauTX%2F%2FZoN1Z2adeOAlOWUC4IlwovsMCx2rsaAzrJqbyyWuU%2Fi5SVqd8mWkOK0cQ%2BRD7P%2BxTFkVpKWKoqYvSo8DsFYkZydg%2FIGh2sKkMWO4zPqSXj1CuAlavoZeNCX%2B2AWaYQWA8SjqNEFom8U6ArusiALNn4SKQD67c%2FqO2zW6QZMoCpTmx5uwBE%2B4fjSQreezXXK6ETtWkfERvFpldsdHGKQMg7mhkSFgYzBTWeutaMXcwwDZzXWXx8SFaP%2FHo%2B9ZcMMmJracGOrEBVQVfZa4rCr5dHLOuopOs5R2M3zKNjDPSU5ENzfunljXiYvzbUGoSWHHSX8l3ho%2BicCRjiSWQsiPyRf46mkmNl1PcTqe6Ckml%2Fvt1gW7omffH327m7yxrr9adqpcpIpeH15FtJo7hlGUG%2B%2Fp0WCvatLTiv%2FXTb86I9WEeUDgjaqnTEfqUqH%2BuE2q8O0t%2FUaFwBf6kxlG00BnPvIUjritp%2FarWFXx0VIYwatKJiOFDKkx4&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20230827T125908Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTYSJIHJVLY%2F20230827%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=be7762148e054ecd7f68d3aab17301487e30d723dc88caa784b912ad653a108e&hash=557dc406a605886f99ea61e24017c2e4725acd355388fa1615135bc8f498be87&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S1877050922021858&tid=spdf-6bd91c46-17b3-4a94-aaf8-ef8e66e7808d&sid=cd59b35d4cf74348b31b04e739a946edf7a6gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=13085705005a500a525603&rr=7fd47d71af916ec2&cc=in)

1. Md Shahin Ali , Md Khairul Islam , A. Arjan Das , D. U. S. Duranta , Mst. Farija Haque , and Md Habibur Rahman , “A Novel Approach for Best Parameters Selection and Feature Engineering to Analyze and Detect Diabetes: Machine Learning Insights”, *Hindawi BioMed Research International Volume , 2023*.

Available: <https://www.hindawi.com/journals/bmri/2023/8583210/>

1. Alain Hennebelle , Huned Materwala, Leila Ismail, “HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System”, *The 14th International Conference on Ambient Systems, Networks and Technologies (ANT) , March 15-17, 2023*.

Available: [pdf.sciencedirectassets.com](https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050923X00040/1-s2.0-S1877050923005781/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEK7%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJGMEQCIGg5u7m6W0zDwU70IW3KWbneSatpIYawP%2BS%2BxU9cLMkLAiALE4%2BWROrBy7lq6TgOqAqlToJBLwwZ%2FNAhy2rWFRI7XCqzBQh3EAUaDDA1OTAwMzU0Njg2NSIM6%2BaRHw7J7kNrs5snKpAFUs1wPq8MWV%2BjYKXP1p0%2FKd7%2F58nZnsYnx0Dj2tgNNbHylcIRu%2Ba4b13KkSzOUFVaJCob%2BaicL4W3Uv%2F1X6nPejnV9sHHhPWa%2BH%2FuxgpLn1g9aNuPMvkNk2rbCUIVGK8m4XLJFCMhMTmUkpXgIRsrKmwrg9aqiYkur5M0GrZfJdj0CRiR7GOHR%2Fx274qGHINkk%2FshHa7GA%2BsWk8hPoCqyszt8PGs9Ubz%2F0tpX7w6T6xfr9mK7LvtB6qbIa1Tjy4SLJdSy3%2FeWf5r6Icm4qxdpYfG0OKZhVvixjL0Qd5mCZWS0C0mDTAG51XX6615jVvuZhwmS0l%2B1o6xSdsXZHVKdDTL5UfUaj6NMaZ6PIcyml8a3fomWfNLQPRWKmbFCJTz9Eszz1l%2BJNyZuV59tNwsPUzfTcCIiumizBGkBtNAQnVwO0zLUTbq%2BC93%2FMNsIQJxN61vMVHbE7dY5a8JoW8BddFaMwV49qHu%2BYsv6Q%2B0oQgvxjDI3NiMTM0czrvVWnI3aSfzHlwTlRWEXKCTcvYJRtUr1LgJ%2FDZxgVPjrOLQIvXRyMgBRc%2FOlnF%2FmRAYgaW5Xxli7hDM5DF0MuTd2nEVa1Yt57tEjgPkBr3XFgYI1Bc%2FWRDJZfKETlg%2BCOpxVoHPKYGyjp9wAs%2BOYgVB3ohi71GIxuF4dEGyHRjaGD125X9fnEE5gbA5Ptyl3YxZ%2BMTvS0rWxk5DM7%2BLQ1M9%2FMKevkj1S3GFoIpsp%2B2VHoYOAnn8%2BYO6nX7Z9UygV4qn48rLIeC1b8EnTM9%2BVb%2Fh2Gl8y3K3rnqujo63BzjgOgi4xuAqpnUOUfXhCJg56Ri2d3gikh%2FtW9uwRLZhUMfXwV8oL2fdneIti71uS24Gv2D%2BAcBAw7aStpwY6sgEYsqS94eGvEVsngZfYdx3%2F%2B4nfRHul0pm79EBcYbtlf%2F6PFegRi78ZEnj%2BVKPQnP%2B7H9OguykbKZlJ6nfrOyS0KVBJ9Rr6KN7RLJIggCw4%2FgTwzjLZgeyGEVxXyi1Wm%2B%2FuEmNs9TeuB8Iug5vPGy3Jww6jlcLi8mpPkIbMFVXNvGqJrPZxtYnEYnTxYaZCBBid4ZnANt3l5aKhk5bryB3bLQo6XER3mzAR7CM8vq3AEkdu&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20230827T140226Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTY63JA427P%2F20230827%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=ea950e7140bc63802a8ebcd50c75b4659eeb0cad5a852dded3a9e8fd843a29f7&hash=74396542bbb37620a495a5d81c15aa19e5fd1fc7b748155bebfb7bfaa16844c6&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S1877050923005781&tid=spdf-e700d030-f49c-4c17-b14d-02a79c1c3fbc&sid=cd59b35d4cf74348b31b04e739a946edf7a6gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=13085705005a030f575e04&rr=7fd4da29f9e885f7&cc=in)

1. Umair Muneer Butt, Sukumar Letchmunan, Mubashir Ali, Fadratul Hafinaz Hassan, Anees Baqir, and Hafiz Husnain Raza Sherazi, “Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications”, *Journal of Healthcare Engineering-Hindawi, Volume 2021.*

Available: <https://www.hindawi.com/journals/jhe/2021/9930985/>

1. Kopitar, L., Kocbek, P., Cilar, L. *et al.*, “Early detection of type 2 diabetes mellitus using machine learning-based prediction models”. *Sci Rep* 10, 11981 (2020).

Available: <https://www.nature.com/articles/s41598-020-68771-z>

1. Ashwini Tuppad , Shantala Devi Patil, “Machine learning for diabetes clinical decision support: a review”, *Advances in Computational Intelligence , 2022.*

Available: <https://link.springer.com/article/10.1007/s43674-022-00034-y>

1. Soumya K N, Vigneshwaran P, “Prediction on Type-2 Diabetes Mellitus Using Machine Learning Methods”, *Volume 41: Advances in Parallel Computing Algorithms, Tools and Paradigms, 2022.*

Available: <https://ebooks.iospress.nl/volumearticle/61506>

1. Valero de Clemente, Maria [US]. (2024, January 25). Non-Invasive Blood Glucose Monitoring System [US Patent No. US2024023838 A1]. Kennasaw State Univ Research and Service Foundation Inc. [US].

Available: <https://worldwide.espacenet.com/publicationDetails/biblio?II=0&ND=4&adjacent=true&locale=en_EP&FT=D&date=20240125&CC=US&NR=2024023838A1&KC=A1>

1. Parsi Joseph John. (2023, September 21). Smart-Tooth Blood Glucose Measurement Device (US Patent Application No. US 2023293055 A1) [Applicant/Inventor].

Available: <https://worldwide.espacenet.com/publicationDetails/inpadocPatentFamily?CC=US&NR=2023293055A1&KC=A1&FT=D&ND=3&date=20230921&DB=&locale=en_EP>

1. Frank, Spencer, Price, David, Stroyeck, Chuck, Hames, Kazanna Calais, DexCom, Inc. (2022, August 30). Diabetes Prediction Using Glucose Measurements and Machine Learning (United States Patent No. US 11,426,102 B2)..

Available: <https://ebooks.iospress.nl/volumearticle/61506>

**Appendix**

**List of Figures:**

| **Number** | **Heading** | **Page no.** |
| --- | --- | --- |
| Fig.1 | Methods of Blood Glucose Measuring system | 12 |
| Fig 2. | Block diagram for the software part | 34 |
| Fig 3. | Block diagram for the hardware part | 35 |
| Fig 4. | Modular design for software part | 36 |
| Fig 5. | Modular design for hardware part | 36 |
| Fig 6. | Detailed design of the system | 37 |
| Fig 7. | IR sensor | 38 |
| Fig 8. | Testing the ATtiny85 Circuit on Breadboard before soldering it to zero PCB | 39 |
| Fig 9. | Gantt chart | 39 |
| Fig 10. | Real-time data collection | 42 |
| Fig 11. | UI pages | 45 |
| Fig 12. | Arduino and IR sensor | 46 |
| Fig 13. | Interfacing OLED with Arduino and IR sensor | 46 |
| Fig 14. | Interfacing of Attiny85 with Oled & IR Sensor | 47 |
| Fig 15. | Working circuit on Zero PCB | 47 |
| Fig 16. | Pin diagram of Attiny85 | 47 |
| Fig 17. | Schematic of PCB | 48 |
| Fig 18. | Printed Circuit Board (PCB) | 48 |
| Fig 19. | Analog voltage values as output on the screen | 49 |

**List of Tables:**

| **Number** | **Heading** | **Page no** |
| --- | --- | --- |
| Table. 1 | Literature Survey | 16 |
| Table. 2 | All Algorithms accuracy comparison wrt PIMA dataset | 53 |
| Table. 3 | All Algorithms accuracy comparison wrt the other dataset | 53 |
| Table. 4 | Comparison of implemented algorithms wrt research paper - 1 | 54 |
| Table. 5 | Comparison of implemented algorithms wrt research paper - 2 | 54 |
| Table. 6 | Comparison of implemented algorithms wrt real data | 54 |

**1. Paper I & II Details**

1. Paper published

***CC: Draft***

**Comparative Analysis of varied ML algorithms for Diabetes datasets**

Sanjana Vashdev Asrani1

[2020.sanjana.asrani@ves.ac.in](mailto:2020.sanjana.asrani@ves.ac.in)

Karina Karira1

[2020.karina.karira@ves.ac.in](mailto:2020.karina.karira@ves.ac.in)

Roshni Wadhwani1

[2020.roshni.wadhwani@ves.ac.in](mailto:2020.roshni.wadhwani@ves.ac.in)

Simran Lahrani1

[2020.simran.lahrani@ves.ac.in](mailto:2020.simran.lahrani@ves.ac.in)

and Mrs. Pallavi Saindane1

[pallavi.saindane@ves.ac.in](mailto:pallavi.saindane@ves.ac.in)

1 Vivekanand Education Society’s Institute of Technology Chembur, Maharashtra, India.

***Abstract*—Diabetes Mellitus, a chronic metabolic disorder, poses a significant global health challenge impacting millions of people worldwide. In India, diabetes has reached alarming proportions, affecting an estimated 77 million people, as reported by the International Diabetes Federation. Over recent decades, there has been a consistent rise in both the incidence and prevalence of diabetes. To address this critical issue, early detection, prediction and post-diagnosis care and management of diabetes are important. This research focuses on developing a predictive model for diabetes risk identification that fits the PIMA Indian Diabetes Dataset from the National Institute of Diabetes and Digestive and Kidney Diseases accurately by exploring multiple algorithms with varied parameters.**

***Keywords***—Diabetes, PIMA dataset, ML, XAI

# 1. INTRODUCTION

Diabetes, a multifaceted metabolic disorder, arises from a combination of genetic predisposition, lifestyle choices, and environmental factors. This chronic condition disrupts the body's ability to regulate blood sugar levels effectively. Insufficient insulin production or decreased insulin sensitivity leads to elevated blood sugar levels, a condition known as hyperglycemia. Long-term hyperglycemia can lead to cardiovascular disease, kidney damage, nerve dysfunction, and vision problems. Thus making Early detection, prediction and post-diagnosis care and management of diabetes a very crucial task. Approach:

1. Prediction into categories like Diabetic, Non Diabetic, Pre Diabetic using clinical Datasets.
2. Probabilistic Range of Risk/Health.
3. Type1, Type2 Identification using the Pathological Report Datasets

2. OVERVIEW

## 2.1. Motivation

In the face of surging diabetes cases, particularly in India, a diabetes prediction and detection project fueled by machine learning emerges as a critical intervention. Rapid urbanization, sedentary lifestyles, and genetic predispositions contribute to escalating risks. Early diagnosis through our ML model becomes a beacon of hope, offering timely interventions to curb the rising tide of diabetes-related complications. By addressing the unique challenges prevalent in India's healthcare landscape, we strive to empower individuals, alleviate strain on healthcare infrastructure, and foster a paradigm shift towards proactive diabetes management.

## 2.2. Problem Statement

The aim of this project is to address the challenges associated with diabetes risk prediction and management. Using traditional physical assessments Long queues, limited appointment availability, high costs, and potential geographic barriers hinder timely access to healthcare professionals. This project aims to overcome these challenges by developing an AI-powered application that offers virtual diabetes risk assessment and personalized insights, providing a convenient, cost-effective, and accessible solution for individuals seeking to understand and manage their diabetes risk.

## 2.3. Objectives

* Diabetes risk prediction and management
* Overcome disadvantages of traditional physical assessments: Long queues, limited appointment availability, high costs, and potential geographic barriers hinder timely access to healthcare professionals.
* virtual diabetes risk assessment and personalized insights.

# 3. LITERATURE REVIEW

## 3.1. Survey of Existing Systems

Authors Sivaranjani S et al. presented a study at the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), focusing on predicting diabetes-related diseases. They utilized Support Vector Machine (SVM) and Random Forest (RF) on the Pima Indian diabetes dataset from the UCI repository. Results showed RF outperformed SVM, achieving an 83% prediction accuracy after feature selection and dimensionality reduction, contributing to the literature on diabetes prediction and demonstrating the effectiveness of machine learning techniques [[1]](#bookmark=id.lh0j3kb7v4ja). S. Saru et al. used WEKA software and the Pima Indian diabetes dataset. Employing Decision Trees, Naïve Bayes, and k-Nearest Neighbors, the study achieves high accuracies (e.g., 94.4% for Decision Trees). Bootstrapping improves accuracy, particularly for k-NN (k=1), from 69.93% to 93.79%. The research suggests that ensemble methods enhance prediction performance, achieving an overall accuracy of 90.36% [[2]](#bookmark=id.qjerple4kfzn). In their paper presented at the 24th International Conference on Automation and Computing, Muhammad Azeem Sarwar et al. discuss predictive analytics in healthcare. They apply six machine learning algorithms, including KNN and SVM, to the PIMA dataset for diabetes prediction. SVM and KNN stand out with the highest accuracy of 77%, highlighting their effectiveness in healthcare applications [[3]](#bookmark=id.vfc5hnyw5t16). In their paper on predicting Type-2 Diabetes Mellitus, Ahamed BS et al. employ various machine learning classifiers, including logistic regression, XGBoost, and LGBM. Using the PIMA Indian Dataset, LGBM stands out with the highest accuracy of 95.20%, outperforming other algorithms. The study suggests avenues for future research, including exploring different datasets and considering advanced versions of LGBM for enhanced prediction accuracy [[4]](#bookmark=id.en5qjwcj1ugg). In their 2023 paper titled "Diabetes prediction using machine learning and explainable AI techniques," Tasin et al. developed an automatic diabetes prediction system using a private dataset of female patients in Bangladesh and the PIMA dataset. They employed a semi-supervised model with extreme gradient boosting to predict insulin features, addressing class imbalance with SMOTE and ADASYN. Various machine learning classification methods were explored, with XGBoost and ADASYN yielding the best results, achieving 81% accuracy, 0.81 F1 coefficient, and an AUC of 0.84. The authors implemented explainable AI techniques, specifically LIME and SHAP frameworks, to understand the XGBoost model's predictions. Additionally, they created a website and an Android application for instantaneous diabetes prediction [[5]](#bookmark=id.jdi91y1c5zab) .In the study by Kirti Kangra et al. on predictive machine learning algorithms for diabetes mellitus, various algorithms, including SVM, Naïve Bayes, KNN, RF, LR, and DT, were evaluated using the Pima Indian diabetic and Germany diabetes datasets. Using WEKA 3.8.6, SVM achieved 74% accuracy for PID, while KNN and RF performed better with 98.7% accuracy for the Germany dataset. The study suggests Logistic Regression (LR) as a preferable choice for both datasets based on accuracy and ROC area. Future work may involve exploring hybrid models and assessing their performance on real-time data [[7]](#bookmark=id.79v6wmu65wpp). Muhammad Exell Febriana et al. (2020) compared K-Nearest Neighbor (KNN) and Naive Bayes algorithms for diabetes prediction using supervised machine learning. Evaluating the PIMA dataset, they found Naive Bayes outperformed KNN with an average accuracy of 76.07%, precision of 73.37%, and recall of 71.37%, while KNN achieved 73.33%, 70.25%, and 69.37% respectively. The study suggests future research incorporating neural networks, Particle Swarm Optimization, and practical application development to enhance accuracy and precision [[8]](#bookmark=id.7shiv898ipll). Md Shahin Ali et. al developed the RFWBP algorithm for early diabetes detection using the PIMA dataset. Achieving a remarkable 95.83% accuracy with cross-validation and 90.68% without, RFWBP outperformed conventional methods (AdaBoost, SVM, logistic regression, naive Bayes, multilayer perceptron, and regular Random Forest). The study utilized data processing and mining techniques to enhance the dataset. Future research aims to broaden the analysis by incorporating more subjects and diverse datasets for precise diabetes identification [[9]](#bookmark=id.1k0aeze6ku6f).Umair Muneer Butt et al. present a machine learning approach for diabetes classification and prediction using Random Forest, Multilayer Perceptron, and Logistic Regression. The study, based on the PIMA Indian Diabetes dataset, shows MLP achieving 86.08% accuracy for classification, while LSTM significantly improves prediction with 87.26% accuracy. Kopitar et al. compared machine learning (Glmnet, RF, XGBoost, LightGBM) and traditional regression models for early detection of Type 2 Diabetes Mellitus. Using 6 months of data, a simple regression model had the lowest RMSE at 0.838. Glmnet showed the highest improvement rate (+3.4%) with more data. LightGBM demonstrated stability in variable selection. While machine learning didn't significantly outperform regression models, all methods improved with additional data. The study suggests exploring ensemble methods in future research but warns about challenges in interpreting results for healthcare decisions [[12]](#bookmark=id.pb36utsie4e3).

## 3.2. Limitations and Research Gap

The limitations highlight the need for further research in several key areas. One critical area is the diversification of datasets to enhance the generalizability of predictive models across various populations. Additionally, the exploration of emerging algorithms and hybrid models, as well as the systematic evaluation of ensemble methods and alternative classifiers, holds promise for improving predictive accuracy. Furthermore, the integration of explainable AI techniques and the development of real-time monitoring and IoT applications represent promising avenues for enhancing the interpretability and practicality of diabetes prediction models. Finally, cross-disease evaluation, long-term predictive analytics, and considerations for ethnic and gender differences should be prioritized to broaden the applicability and effectiveness of machine learning-based diabetes prediction systems.

## 3.3. Project Contribution

Our project makes a significant contribution to diabetes risk prediction by incorporating a multifaceted approach to data collection. Inputs are sourced from users, capturing vital information like pregnancies, glucose levels, blood pressure, and other pertinent features, as well as potential user interface data. The inclusion of model input, derived from a diverse dataset such as PIMA and Sylhet, enhances the generalizability of our predictive models, addressing a key limitation in existing research. We employ a variety of machine learning algorithms, including Decision Tree, Support Vector Machine, Random Forest, Adaboost, and XGBoost, to ensure a comprehensive evaluation of diabetes risk. The transparency of our predictions, coupled with detailed explanations provided to the user, facilitates informed decision-making regarding diabetes risk management. By utilizing the PIMA dataset, and the Sylhet dataset from Bangladesh, our system caters to diverse demographic groups, contributing to the inclusivity of diabetes prediction models. Overall, our project not only advances the accuracy of diabetes risk prediction but also addresses the limitations of existing research by incorporating diverse datasets and fostering user understanding through detailed explanations.

## 3.4. Dataset Description

Following two datasets have been used:

**PIMA Dataset**:

The dataset originates from the National Institute of Diabetes and Digestive and Kidney Diseases and serves the purpose of predicting whether a patient has diabetes based on specific diagnostic measurements. Stringent criteria were applied when selecting instances from a broader database, focusing exclusively on female patients aged at least 21 years with Pima Indian ancestry. These datasets encompass various medical predictor variables alongside a single target variable, Outcome. Predictor variables encompass factors such as the patient's number of pregnancies, BMI, insulin levels, age, and others.

**Table 1.** Pima Dataset Features

| **Features** | **Min value** | **Max Value** | **Avg Value** |
| --- | --- | --- | --- |
| Pregnancies | 0.0 | 17.0 | 3.85 |
| Glucose | 0.0 | 199.0 | 120.89 |
| BloodPressure | 0.0 | 122.0 | 69.11 |
| SkinThickness | 0.0 | 99.0 | 20.54 |
| Insulin | 0.0 | 846.0 | 79.8 |
| BMI | 0.0 | 67.1 | 31.99 |
| DiabetesPedigreeFunction | 0.078 | 2.42 | 0.47 |
| Age | 21.0 | 81.0 | 33.24 |

**Sylhet dataset**: Diabetes stands as one of the most rapidly spreading chronic, life-threatening conditions, having impacted an estimated 422 million individuals globally, as reported by the World Health Organization (WHO) in 2018. Given its lengthy asymptomatic period, detecting diabetes at an early stage is crucial for achieving meaningful clinical outcomes. Alarmingly, approximately 50% of diabetes cases remain undiagnosed due to this prolonged asymptomatic phase. The dataset under discussion comprises 520 observations featuring 17 distinct characteristics. These data points were gathered through direct questionnaires and diagnostic findings from patients treated at the Sylhet Diabetes Hospital in Sylhet, Bangladesh.

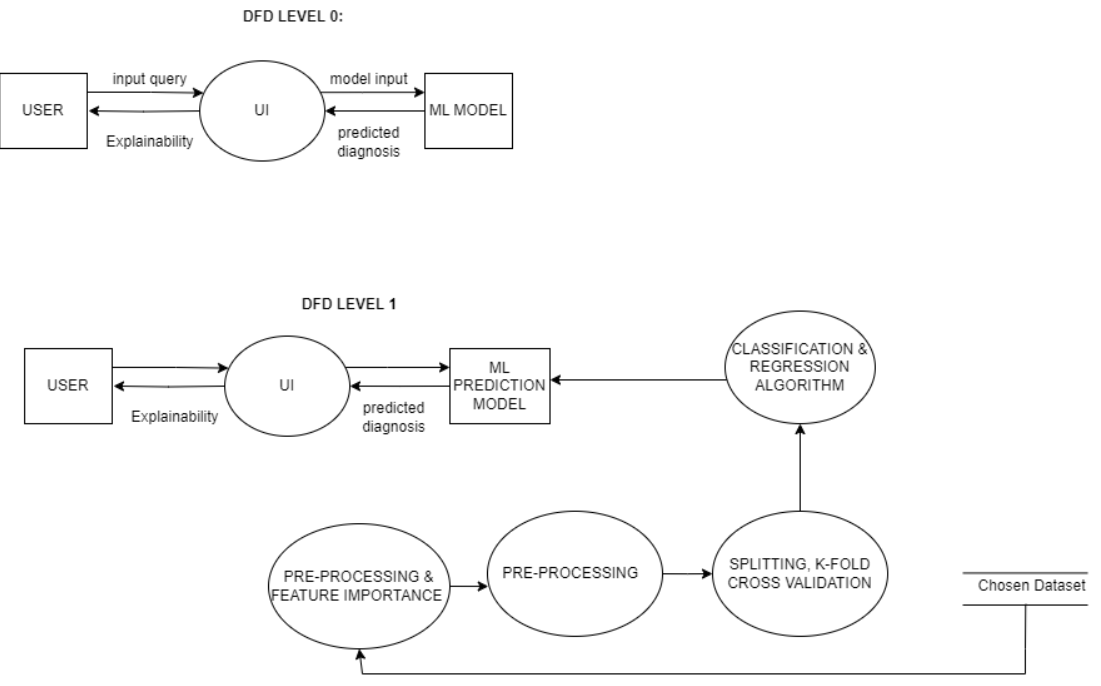
**Table 2.** Sylhet Dataset Features

| **Features** | **Meaning** |
| --- | --- |
| age | Age of the patient |
| gender | Gender of the patient (Male/Female) |
| polyuria | Excessive urination |
| polydipsia | Excessive thirst |
| sudden\_weight\_loss | Rapid, unintentional weight loss |
| weakness | Generalized weakness |
| polyphagia | Excessive hunger |
| genital\_thrush | Fungal infection in genital area |
| visual\_blurring | Unclear vision |
| itching | Skin itching |
| irritability | Feeling easily annoyed or frustrated |
| delayed\_healing | Slow healing of wounds |
| partial\_paresis | Partial loss of voluntary muscle movement |
| muscle\_stiffness | Stiffness or discomfort in muscles |
| alopecia | Loss of hair |
| obesity | Excessive body weight |

# 4. IMPLEMENTATION

## 4.1. Detailed Design

User provides information which is input to the model. It requires all the features of pima dataset to be inputted.The Algorithm used here finally is *Random Fores*t for deployment while *Decision Tree*, *SVM* (for non-linear data), *XGBoost* and *ADABoost* were also implemented along with the *XAI* techniques i.e. *LIME* and *SHAP*. It predicts whether the user is diabetic or not using ML techniques. The explanation of the prediction is given in XAI techniques to help users make informed decisions on how to manage their risk of developing diabetes.

****

**Fig 1.** DFD

Algorithms executed:

**1. Decision Tree:**

Decision trees iteratively partition the data based on provided features to form a hierarchical structure comprising nodes and leaves. At each node, the algorithm selects the feature that best separates the data into pure classes (maximizes information gain or minimizes impurity) depending upon this. The process persists until a stopping criterion is satisfied, such as reaching a predefined tree depth or attaining a minimum number of samples in a leaf node, as predetermined.

**2. Random Forest:**

Random Forest builds multiple decision trees by using random subsets of both the data and the features given. The final prediction is an average (classification) or a mean (regression) of the predictions of individual trees depending upon the data.Random Forest introduces diversity and reduces overfitting by combining the predictions of multiple trees.

**3. Support Vector Machine (SVM):**

SVM finds the hyperplane that best separates classes in a high-dimensional space.It maximizes the margin between classes by transforming input data using different kernel functions.

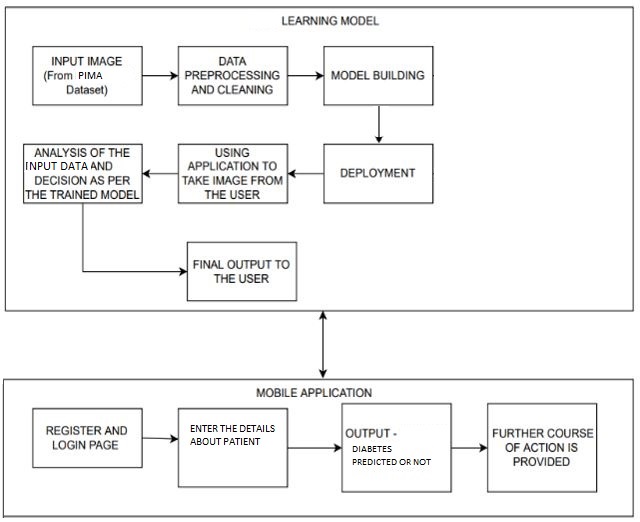
**4. XGBoost (Extreme Gradient Boosting):**

XGBoost, a form of gradient boosting, iteratively builds decision trees to rectify errors made by preceding trees. It uses a gradient descent algorithm to minimize the loss function.

5**. AdaBoost (Adaptive Boosting):**

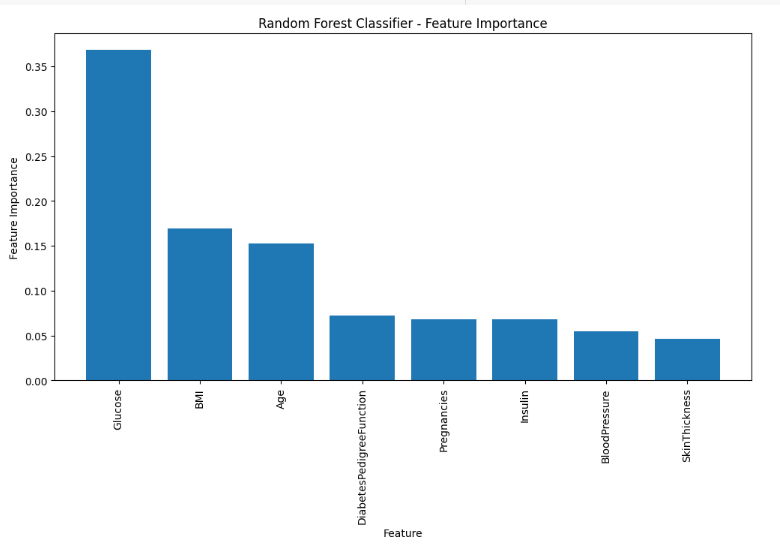
AdaBoost combines multiple weak learners (simple models that perform slightly better than random guessing). It allocates weights to instances, prioritizing those incorrectly classified by preceding models. The final prediction is a combination of these weighted weak learners.

4.2. Modular Diagram

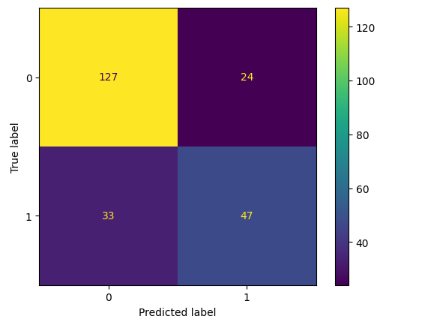


**Fig 2.** Modular Diagram

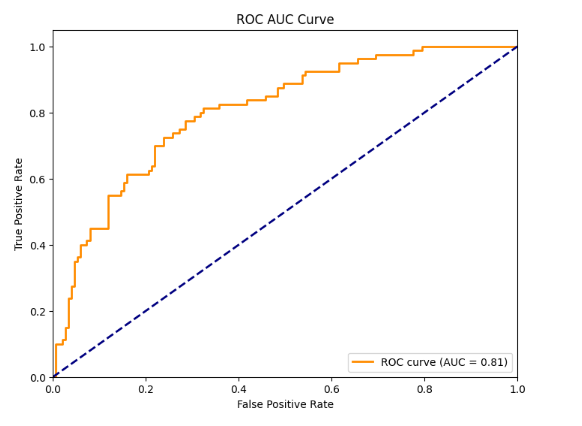
The modular diagram explains the model as two sections. A user-interface will collect all the required inputs from the user that will be given to the model. Based on the input, the model predicts whether the user is diabetic or not.Both the sections are combined together for proper functioning of diabetes prediction and detection.



**Fig** 3. Feature Importance Plot for PIMA Dataset



**Fig 4.** Confusion Matrix for RF, PIMA Dataset



**Fig.** 5. ROC AUC Curve for RF, PIMA Dataset

# 5. RESULTS AND EVALUATION

There are various evaluation measures that are used for determining the performance of a model. Some of them are: Accuracy, precision, recall, F1-score, etc.

**Accuracy:** It is a metric representing the proportion of correct predictions made by a model in relation to the total number of input samples.

Accuracy=(TP+TN)/(TP+FP+TN+FN)

**Confusion matrix:** A confusion matrix offers a summary of a machine learning model's performance on a test dataset, particularly for classification tasks. It displays key metrics like true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

**F1 score:** The F1 score is a balanced metric for assessing the accuracy of a test, calculated as the harmonic mean of precision and recall in classification tasks.

F1 score= 2\*precision\*recall/precision+recall

**Precision:** This quantity represents the ratio of correct positive predictions made by the classifier to the total number of positive instances predicted by the classifier.

Precision=TP/(TP+FP)

**Recall:** Itis the ratio of correct positive predictions to all relevant samples, which include all instances that should have been identified as positive.

Recall=TP/(TP+FN)

There are a number of classification algorithms that help determine the appropriate class. While developing the model, we implemented various algorithms to determine the accuracy of each algorithm.The Random Forest algorithm gave the best accuracy for the dataset chosen.

Accuracy Comparison for all 5 algorithms:

PIMA:

|  | SVM( non-linear: rbf kernel ) | Decision tree | Random Forest | Adaboost | XGBoost |
| --- | --- | --- | --- | --- | --- |
| 80:20 | 75% | 75% | 76.6% | 73.37% | 73.37% |
| 70:30 | 74% | 70% | 75.3% | 74.45% | 73.59% |
| 5-fold | 76.17% | 73% | 85.7% | 80.52% | 76.43% |
| 10-fold | 76.43% | 71% | 80.2% | 84.21% | 76.17% |

**Table 3**. All Algorithms accuracy comparison wrt PIMA dataset

The sylhet dataset:

|  | SVM | Decision tree | Random Forest | Adaboost | XGBoost |
| --- | --- | --- | --- | --- | --- |
| 80:20 | 89% | 96% | 100% | 93.26% | 99.03% |
| 70:30 | 94% | 97% | 96.7% | 92.94% | 98.71% |
| 5-fold | 91.15% | 96% | 100% | 90.38% | 96.35% |
| 10-fold | 92.31% | 97% | 100% | 94.23% | 96.92% |

**Table 4**. All Algorithms accuracy comparison wrt the sylhet dataset

As per our evaluation, Random Forest Algorithm works best on both the datasets. However we observe the accuracy in the second dataset is 100% which shows that the second dataset is overfitting.

# 6. CONCLUSION

This study compares machine learning algorithms (Decision Tree, Random Forest, SVM, XGBoost, AdaBoost) using PIMA and Sylhet datasets. It assesses accuracy metrics with varied data splits and cross-validation. It also employs XAI techniques (LIME, SHAP) for model interpretability. However, limitations are encountered, such as the inability to create a robust model from the collected reports of diabetic patients due to variations in patient reports. Moreover, the Sylhet dataset demonstrates 100% accuracy, indicating potential overfitting issues. Future work includes integrating hardware for voltage measurement to estimate glucose levels accurately and enhancing voltage accuracy by adjusting parameters.

# 7. REFERENCES

1.Book Referred

[1].Soumya K N, Vigneshwaran P, “Prediction on Type-2 Diabetes Mellitus Using Machine Learning Methods”, *Volume 41: Advances in Parallel Computing Algorithms, Tools and Paradigms, 2022.*

Available:<https://ebooks.iospress.nl/volumearticle/61506>

2.Research Papers referred/ Journals/ Articles referred

[1]Sivaranjani S, Ananya S, Aravinth J, Karthika R, “Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction”,  *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS),* March 2021.

Available:<https://ieeexplore.ieee.org/document/9441935>

[2]S.Saru, S.Subashree, “ Analysis and Prediction of Diabetes Using Machine Learning”, *International Journal of Emerging Technology and Innovative Engineering Volume 5, Issue 4,* April 2019.

Available: <https://ssrn.com/abstract=3368308>

[3]Muhammad Azeem Sarwar, Nasir Kamal,Wajeeha Hamid; “Prediction of Diabetes Using Machine Learning Algorithms in Healthcare”; *24th International Conference on Automation and Computing (ICAC)*: 06-07 September 2018.

Available:<https://ieeexplore.ieee.org/abstract/document/8748992/authors#authors>

[4]Ahamed BS, Arya MS and Nancy V AO , “Prediction of Type-2 Diabetes Mellitus Disease Using Machine Learning Classifiers and Techniques”, *Front. Computer. Sci. 4:835242. doi: 10.3389/fcomp.2022.835242,* 2022.

Available:<https://www.frontiersin.org/articles/10.3389/fcomp.2022.835242/full>

[5]Tasin, I., Nabil, T.U., Islam, S., Khan, R., “Diabetes prediction using machine learning and explainable AI techniques”, *Healthc. Technol. Lett*. 10, 1–10 , 2023.

Available:<https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/htl2.12039>

[6]Chang, V., Bailey, J., Xu, Q.A. *et al.* ,“Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms”, *Neural Comput & Applic* 35, 16157–16173 (2023).

Available:<https://link.springer.com/article/10.1007/s00521-022-07049-z>

[7]Kirti Kangra, Jaswinder Singh, “Comparative analysis of predictive machine learning algorithms for diabetes mellitus “, *Bulletin of Electrical Engineering and Informatics* *Vol. 12, No. 3, June 2023.*

Available:<https://www.beei.org/index.php/EEI/article/view/4412>

[8]Muhammad Exell Febriana, Fransiskus Xaverius Ferdinana , Gustian Paul Sendani, Kristien Margi Suryanigrum, Rezki Yunanda “ Diabetes prediction using supervised machine learning ” , *7th International Conference on Computer Science and Computational Intelligence*, *2020*.

Available: [//pdf.sciencedirectassets.com/280203/](https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050922X00197/1-s2.0-S1877050922021858/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEK3%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJHMEUCIQDiR%2BvVna7sWtMNVyBSY7lHy7e3bwHIPatBjgrpNkhxdAIgQX%2B2F7n81t4rqqYlfU1hUwV06iMZZiPDjXLTPCOk7moqswUIdhAFGgwwNTkwMDM1NDY4NjUiDPD7ulw8cAiBCHrBfCqQBTt1koFd7wvINXRTfB9%2BiOaHHtjaxzG12ljURdYH%2F1r2lD6YyEyI1lsUSOmLDyeAcCN%2FHOwRnoW1dA8P5NZuea39HyPgRIZAPOgJ8NW%2BhK%2FiG%2FGRKkxVnrAgkoJ19QtlGf1xU1AerXCHoce83JMAE5yZCafO7Qb6qNtI0sAB%2BpDCBUZfaVbj2aQaOwbCVEczSBEJvhvZnQYI5c2BfLDeINYPQcJSmMk8SUJaEOfQdWwSKAjQ8wEMXxyexnHH1mCcpdWT2jKnQaMu%2BK91B65JkyO9E0k7%2BowW%2FiN21CypeY7uKwZvcNF0BC5kSDyuOtmMbf92JN3Uaf9Db3tM0RbNKDuKznKdJdiu7DUf2kalWEV0WxZVPPJqGM3U5gwLm7lVXyXGMqEFMXGt7Fi7zqQqi3w3SJLVzOxfjVmKEnTdoiPoWCGFX8pu8yDuB4rV9%2BvRD1y9Lkc6TALMLFZGrOwRPKIypD785MxtpQiXqDEkQ3Upij7EBeHjc1ukhrW8KLhvEeusLf2oXgNGMzWefnc3NrqJNhrPicXueZfwbnlTXRbVLr5rJwMItboclEwt789L6Nmq%2FHHifn7sBl3rY6Yy5B9ZpYJNaCuHQtuvcoT5fauTX%2F%2FZoN1Z2adeOAlOWUC4IlwovsMCx2rsaAzrJqbyyWuU%2Fi5SVqd8mWkOK0cQ%2BRD7P%2BxTFkVpKWKoqYvSo8DsFYkZydg%2FIGh2sKkMWO4zPqSXj1CuAlavoZeNCX%2B2AWaYQWA8SjqNEFom8U6ArusiALNn4SKQD67c%2FqO2zW6QZMoCpTmx5uwBE%2B4fjSQreezXXK6ETtWkfERvFpldsdHGKQMg7mhkSFgYzBTWeutaMXcwwDZzXWXx8SFaP%2FHo%2B9ZcMMmJracGOrEBVQVfZa4rCr5dHLOuopOs5R2M3zKNjDPSU5ENzfunljXiYvzbUGoSWHHSX8l3ho%2BicCRjiSWQsiPyRf46mkmNl1PcTqe6Ckml%2Fvt1gW7omffH327m7yxrr9adqpcpIpeH15FtJo7hlGUG%2B%2Fp0WCvatLTiv%2FXTb86I9WEeUDgjaqnTEfqUqH%2BuE2q8O0t%2FUaFwBf6kxlG00BnPvIUjritp%2FarWFXx0VIYwatKJiOFDKkx4&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20230827T125908Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTYSJIHJVLY%2F20230827%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=be7762148e054ecd7f68d3aab17301487e30d723dc88caa784b912ad653a108e&hash=557dc406a605886f99ea61e24017c2e4725acd355388fa1615135bc8f498be87&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S1877050922021858&tid=spdf-6bd91c46-17b3-4a94-aaf8-ef8e66e7808d&sid=cd59b35d4cf74348b31b04e739a946edf7a6gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=13085705005a500a525603&rr=7fd47d71af916ec2&cc=in)

[9]Md Shahin Ali , Md Khairul Islam , A. Arjan Das , D. U. S. Duranta , Mst. Farija Haque , and Md Habibur Rahman , “A Novel Approach for Best Parameters Selection and Feature Engineering to Analyze and Detect Diabetes: Machine Learning Insights”, *Hindawi BioMed Research International Volume , 2023*.

Available:<https://www.hindawi.com/journals/bmri/2023/8583210/>

[10]Alain Hennebelle , Huned Materwala, Leila Ismail, “HealthEdge: A Machine Learning-Based Smart Healthcare Framework for Prediction of Type 2 Diabetes in an Integrated IoT, Edge, and Cloud Computing System”, *The 14th International Conference on Ambient Systems, Networks and Technologies (ANT) , March 15-17, 2023*.

Available: [pdf.sciencedirectassets.com](https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050923X00040/1-s2.0-S1877050923005781/main.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEK7%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCXVzLWVhc3QtMSJGMEQCIGg5u7m6W0zDwU70IW3KWbneSatpIYawP%2BS%2BxU9cLMkLAiALE4%2BWROrBy7lq6TgOqAqlToJBLwwZ%2FNAhy2rWFRI7XCqzBQh3EAUaDDA1OTAwMzU0Njg2NSIM6%2BaRHw7J7kNrs5snKpAFUs1wPq8MWV%2BjYKXP1p0%2FKd7%2F58nZnsYnx0Dj2tgNNbHylcIRu%2Ba4b13KkSzOUFVaJCob%2BaicL4W3Uv%2F1X6nPejnV9sHHhPWa%2BH%2FuxgpLn1g9aNuPMvkNk2rbCUIVGK8m4XLJFCMhMTmUkpXgIRsrKmwrg9aqiYkur5M0GrZfJdj0CRiR7GOHR%2Fx274qGHINkk%2FshHa7GA%2BsWk8hPoCqyszt8PGs9Ubz%2F0tpX7w6T6xfr9mK7LvtB6qbIa1Tjy4SLJdSy3%2FeWf5r6Icm4qxdpYfG0OKZhVvixjL0Qd5mCZWS0C0mDTAG51XX6615jVvuZhwmS0l%2B1o6xSdsXZHVKdDTL5UfUaj6NMaZ6PIcyml8a3fomWfNLQPRWKmbFCJTz9Eszz1l%2BJNyZuV59tNwsPUzfTcCIiumizBGkBtNAQnVwO0zLUTbq%2BC93%2FMNsIQJxN61vMVHbE7dY5a8JoW8BddFaMwV49qHu%2BYsv6Q%2B0oQgvxjDI3NiMTM0czrvVWnI3aSfzHlwTlRWEXKCTcvYJRtUr1LgJ%2FDZxgVPjrOLQIvXRyMgBRc%2FOlnF%2FmRAYgaW5Xxli7hDM5DF0MuTd2nEVa1Yt57tEjgPkBr3XFgYI1Bc%2FWRDJZfKETlg%2BCOpxVoHPKYGyjp9wAs%2BOYgVB3ohi71GIxuF4dEGyHRjaGD125X9fnEE5gbA5Ptyl3YxZ%2BMTvS0rWxk5DM7%2BLQ1M9%2FMKevkj1S3GFoIpsp%2B2VHoYOAnn8%2BYO6nX7Z9UygV4qn48rLIeC1b8EnTM9%2BVb%2Fh2Gl8y3K3rnqujo63BzjgOgi4xuAqpnUOUfXhCJg56Ri2d3gikh%2FtW9uwRLZhUMfXwV8oL2fdneIti71uS24Gv2D%2BAcBAw7aStpwY6sgEYsqS94eGvEVsngZfYdx3%2F%2B4nfRHul0pm79EBcYbtlf%2F6PFegRi78ZEnj%2BVKPQnP%2B7H9OguykbKZlJ6nfrOyS0KVBJ9Rr6KN7RLJIggCw4%2FgTwzjLZgeyGEVxXyi1Wm%2B%2FuEmNs9TeuB8Iug5vPGy3Jww6jlcLi8mpPkIbMFVXNvGqJrPZxtYnEYnTxYaZCBBid4ZnANt3l5aKhk5bryB3bLQo6XER3mzAR7CM8vq3AEkdu&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20230827T140226Z&X-Amz-SignedHeaders=host&X-Amz-Expires=300&X-Amz-Credential=ASIAQ3PHCVTY63JA427P%2F20230827%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=ea950e7140bc63802a8ebcd50c75b4659eeb0cad5a852dded3a9e8fd843a29f7&hash=74396542bbb37620a495a5d81c15aa19e5fd1fc7b748155bebfb7bfaa16844c6&host=68042c943591013ac2b2430a89b270f6af2c76d8dfd086a07176afe7c76c2c61&pii=S1877050923005781&tid=spdf-e700d030-f49c-4c17-b14d-02a79c1c3fbc&sid=cd59b35d4cf74348b31b04e739a946edf7a6gxrqb&type=client&tsoh=d3d3LnNjaWVuY2VkaXJlY3QuY29t&ua=13085705005a030f575e04&rr=7fd4da29f9e885f7&cc=in)

[11]Umair Muneer Butt, Sukumar Letchmunan, Mubashir Ali, Fadratul Hafinaz Hassan, Anees Baqir, and Hafiz Husnain Raza Sherazi, “Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications”, *Journal of Healthcare Engineering-Hindawi, Volume 2021.*

Available:<https://www.hindawi.com/journals/jhe/2021/9930985/>

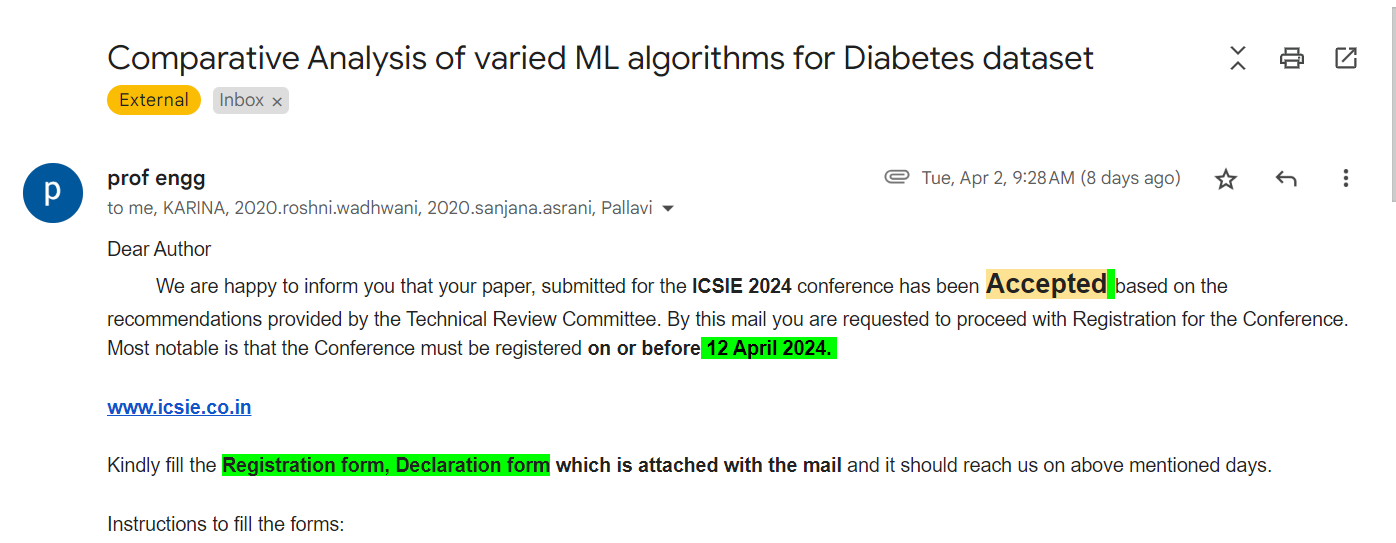
[12]Kopitar, L., Kocbek, P., Cilar, L. *et al.*, “Early detection of type 2 diabetes mellitus using machine learning-based prediction models”. *Sci Rep* 10, 11981 (2020).

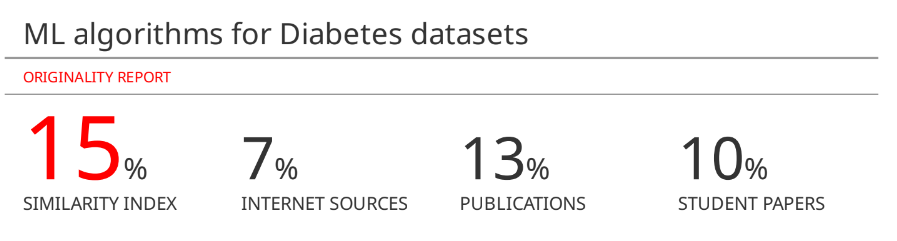
Available:<https://www.nature.com/articles/s41598-020-68771-z>

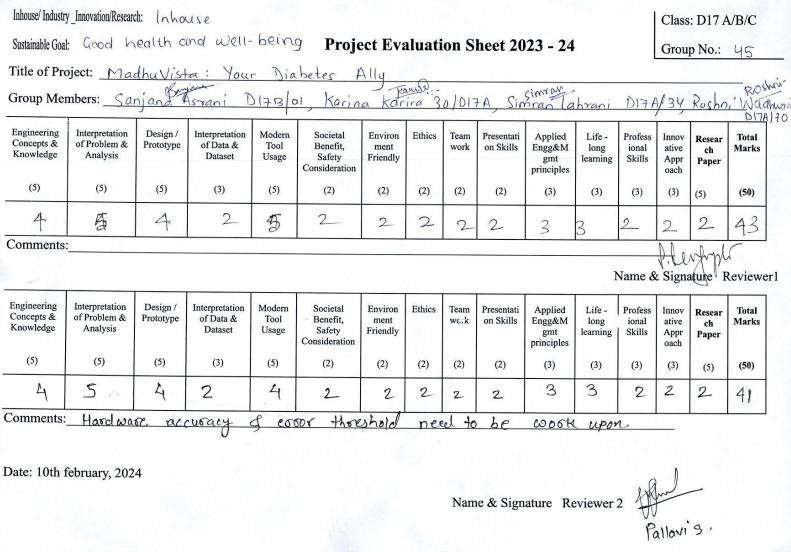
[13]Ashwini Tuppad , Shantala Devi Patil, “Machine learning for diabetes clinical decision support: a review”, *Advances in Computational Intelligence , 2022.*

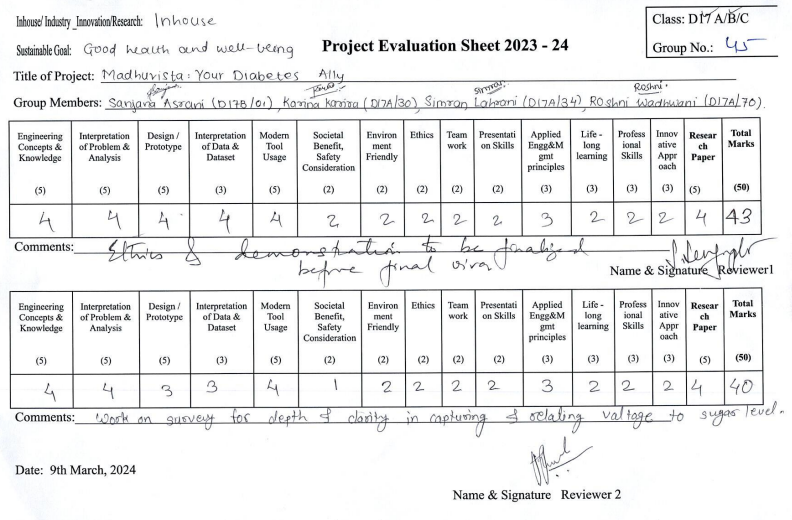
Available:<https://link.springer.com/article/10.1007/s43674-022-00034-y>

1. Acceptance mail:



1. Plagiarism report
2. Project review sheet

****

****