**VIVEKANAND EDUCATION SOCIETY’S**

**INSTITUTE OF TECHNOLOGY**

**Department of Computer Engineering**



Project Report on

MadhuVista : Your Diabetes Ally

In partial fulfillment of the Fourth Year (Semester–VII), Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2023-2024

**Prof. Pallavi Saindane**

**Submitted by**

Sanjana Asrani, D17B/01,

Karina Karira, D17A/30,

Simran Lahrani, D17A/34,

Roshni Wadhwani, D17A/70

(2023-24)

**VIVEKANAND EDUCATION SOCIETY’S**

**INSTITUTE OF TECHNOLOGY**

**Department of Computer Engineering**

**CERTIFICATE of Approval**

This is to certify that *Sanjana Asrani,Karina Karira,Simran Lahrani,Roshni Wadhwani* of Fourth Year Computer Engineering studying under the University of Mumbai has satisfactorily presented the project on “*MadhuVista: Your Diabetes Ally*” as a part of the coursework of PROJECT-I for Semester-VII under the guidance of *Mrs.Pallavi Saindane* in the year 2023-2024.

**\_\_\_\_\_\_\_\_\_\_\_\_**

Date

**\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_**

Internal Examiner External examiner

\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_

Project Mentor Head of the Department Principal

Mrs.Pallavi Saindane Dr.Mrs.Nupur Giri Dr.J.M.Nair

**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** | **Description of the Course Outcome** |
| --- | --- |
| **Outcome** |  |
|  |  |
| CO 1 | Do literature survey/industrial visit and identify the problem of |
|  | the selected project topic. |
| CO2 | Apply basic engineering fundamental in the domain of practical |
|  | applications for problem identification, formulation and solution |
|  |  |
| CO 3 | Attempt & Design a problem solution in a right approach to |
|  | complex problems |
|  |  |
| CO 4 | Cultivate the habit of working in a team |
|  |  |
| CO 5 | Correlate the theoretical and experimental/simulations results |
|  | and draw the proper inferences |
| CO 6 | Demonstrate the knowledge, skills and attitudes of a |
|  | professional engineer & Prepare report as per the standard |
|  | guidelines. |
|  |  |

**ABSTRACT of the project**

Diabetes Mellitus, a chronic metabolic disorder, is a global health concern affecting millions of individuals worldwide. In India, diabetes has reached alarming proportions, affecting an estimated 77 million people, as reported by the International Diabetes Federation. Both the number of cases and the prevalence of diabetes have been steadily increasing over the past few decades.

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** will 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. It is a common complication of diabetes, thus making retinopathy a way to detect it in its early stages, thus allowing for timely intervention that can slow the progression of the condition and preserve vision by early diagnosis and management of this vision-threatening complication.

**Index**

Chapter 1 : Introduction

1.1 Introduction

1.2 Motivation

1.3 Problem Definition

1.4 Relevance of the Project

1.5 Methodology used

Chapter 2: Literature Survey

2.1 Research Papers referred

2.2.Books/ Journals/ Articles referred

2.3 Patent search

2.4 Exiting Systems

2.5. Lacuna in the existing systems

2.6 Comparison of existing systems and proposed area of work

2.7. Focus Area

Chapter 3: Requirements

3.1 Proposed model

3.2 Functional Requirements

3.3.Non-Functional Requirements

3.4.Hardware & Software Requirements

3.5.Technology and Tools utilized

3.6.Constraints of working

Chapter 4: Proposed Design

4.1 Block Diagram of the proposed system

4.2 Modular diagram

4.3 Detailed Design

4.4. Proposed algorithms

4.3. Project Scheduling & Tracking using Timeline / Gnatt Chart

5. Results and Discussions

6. Plan of action for the next semester

7. Conclusions

8. References

9. Appendix

**Chapter 1 : Introduction**

1.1 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. Prolonged hyperglycemia can result in a range of complications, including cardiovascular disease, kidney damage, nerve dysfunction, and vision problems.

Diabetic retinopathy, a diabetes-related complication affecting the eyes, often starts with minimal or no symptoms, possibly causing slight vision issues. However, if left untreated, it can progress to blindness. Therefore, early detection is vital for preventing such outcomes.

1.2 Motivation

* 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.
* 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.
* 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.
* Data-Driven Insights: Leveraging data analytics and machine learning in diabetes projects can lead to valuable insights, early disease detection, and personalized treatment recommendations.
* Research Opportunities: Diabetes projects offer opportunities for cutting-edge research in areas such as genetics, epidemiology, pharmacology, and behavioral science.

1.3 Problem Definition

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

Nearly all [patients with diabetes](https://diabetesjournals.org/care/article/27/suppl_1/s84/24669/Retinopathy-in-Diabetes#:~:text=Diabetic%20retinopathy%20is%20the%20most,type%202%20diabetes%20have%20retinopathy.) will eventually develop some form of diabetic retinopathy within 15-20 years. The disease has four main stages and if it’s detected in the earlier stages, it is more easily treatable.

* Stage 1: Mild nonproliferative diabetic retinopathy.
* Stage 2: Moderate nonproliferative diabetic retinopathy.
* Stage 3: Severe nonproliferative diabetic retinopathy.
* Stage 4: Proliferative diabetic retinopathy.

1.4 Relevance of the Project

* Prevalence of Diabetes: Diabetes is a widespread chronic health condition that affects millions of people worldwide. Its prevalence continues to rise, making it a pressing public health concern.
* Healthcare Burden: Diabetes places a substantial burden on healthcare systems, leading to increased healthcare costs and resource utilization. Effective diabetes management and prevention are critical for alleviating this burden.
* Quality of Life: Diabetes significantly affects the quality of life for individuals living with the condition. A diabetes project can improve the daily lives of patients by offering better treatment options, monitoring tools, and support systems.
* Complications and Comorbidities: Diabetes can lead to severe complications, such as heart disease, kidney disease, vision problems, and nerve damage. Managing diabetes effectively can reduce the risk of these complications.
* Early Detection and Prevention: Projects focused on diabetes often emphasize early detection and prevention strategies. Identifying individuals at risk and implementing preventive measures can reduce the incidence of the disease.

1.5 Methodology used

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.

**Chapter 2: Literature Survey**

2.1 Research Papers referred

**Table 1**. Study of Reference papers and their inferences

| Sr. No | Title | Abstract | Inference |
| --- | --- | --- | --- |
| [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.[1] | * Different Machine Learning techniques, Support Vector Machine (SVM) and Random Forest (RF) algorithms are used for predicting Diabetes Related Diseases. * Data preprocessing, feature selection through step forward and backward methods, and dimensionality reduction using Principal Component Analysis (PCA) were performed. * The RF model achieved an 83% prediction accuracy, surpassing SVM's 81.4% accuracy. | * The Pima Indian diabetes dataset from the UCI repository is used for analysis and prediction. * The research focuses on predicting the potential chances of individuals getting affected by diabetes-related diseases. Machine learning algorithms, specifically Support Vector Machine (SVM) and Random Forest (RF), are employed for this purpose. * The data used for prediction is pre-processed to handle missing values and ensure data quality. Imputation is used to replace missing values with the mean values of existing data. * Feature selection techniques, such as step forward and step backward feature selection, are employed to identify the most relevant features for diabetes prediction. These selected features are used to enhance the accuracy of the predictive model. * Principal Component Analysis (PCA) is used for dimensionality reduction, which helps in reducing the complexity of the data while preserving its essential information. * Random Forest (RF) outperforms Support Vector Machine (SVM) in terms of accuracy for diabetes prediction. After feature selection and dimensionality reduction, RF achieves an accuracy of 83%, while SVM achieves an accuracy of 81.4%. |
| [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.[2] | * The WEKA software was employed as a mining tool for diagnosing diabetes. * The Pima Indian diabetes database from UCI repository was used for analysis. * The model built was used to predict and diagnose diabetes disease. * The bootstrapping resampling technique was used to enhance the accuracy and then Naïve Bayes, Decision Trees and (KNN) algorithms were applied and their performance was compared. | * The Pima Indian diabetes dataset from the UCI repository is used for analysis and prediction. * The study employs various machine learning algorithms, including Decision Trees, Naïve Bayes, and k-Nearest Neighbors (k-NN), to build predictive models. * The accuracy of different algorithms is compared. For instance, Decision Trees achieve an accuracy of 94.4%, k-NN with k=1 achieves an accuracy of 93.79%, and Naïve Bayes achieves an accuracy of 79.84%. * Bootstrapping resampling techniques are applied to enhance the accuracy of the models. For instance, the accuracy of k-NN with k=1 improves from 69.93% to 93.79% after bootstrapping. * The research suggests that ensemble methods, which combine multiple classifiers, can provide better prediction performance than individual algorithms. * Future work may involve extending the application of this method to other diseases, using larger datasets, and exploring different base classifiers such as Artificial Neural Networks (ANN) and Random Forest. * The proposed method in this study provides a high accuracy rate of 90.36% for diabetes prediction. Ensemble methods are found to be effective in improving prediction performance. |
| [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. [3] | * Predictive analytics in healthcare is discussed. * Six different machine learning algorithms (KNN, Naive Bayes, SVM, Decision Tree, Logistic Regression, Random forest) are applied on the dataset and their performance and the accuracy is discussed. | * PIMA dataset is used and 6 different machine learning algorithms are applied on the dataset and their performance is discussed. * SVM and KNN provide 77% accuracy, which is highest as compared to others. |
| [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. [4] | * Predictive analytics is discussed. * The classifiers taken are logistic regression, XGBoost, gradient boosting, decision trees, ExtraTrees, random forest, and light gradient boosting machine (LGBM) and their performance is compared. * The dataset used is PIMA Indian Dataset sourced from UC Irvine Repository. The performance of these algorithms is compared. * The results obtained from these classifiers show that the LGBM classifier has the highest accuracy of 95.20% in comparison with the other algorithms. | * The researchers experimented with several machine learning algorithms, including logistic regression, XGBoost, gradient boosting, decision trees, ExtraTrees, random forest, and LGBM. * The dataset used in this study was the PIMA Indian Dataset from UC Irvine Repository. * The researchers compared the performance of the different algorithms in terms of accuracy. The LGBM classifier achieved the highest accuracy at 95.20%, outperforming the other algorithms considered in the study. * The study suggests that further research can involve different datasets and explore additional parameters to fine-tune algorithms for even better prediction accuracy. It also hints at the potential for using advanced versions of LGBM. |
| [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.[5] | * An automatic diabetes prediction system has been developed using a private dataset of female patients in Bangladesh and PIMA dataset. * A semi-supervised model with extreme gradient boosting has been utilized to predict the insulin features of the private dataset. * SMOTE and ADASYN approaches have been employed to manage the class imbalance problem. * The authors used machine learning classification methods, i.e., decision tree, SVM, Random Forest, Logistic Regression, KNN, and various ensemble techniques, to determine which algorithm produces the best prediction results.The best result is provided by the XGBoost classifier with the ADASYN approach with 81% accuracy, 0.81 F1 coefficient and AUC of 0.84. * The domain adaptation method has been implemented to demonstrate the versatility of the proposed system. * The explainable AI approach with LIME and SHAP frameworks is implemented to understand how the model predicts the final results. * A website framework and an Android smartphone application have been developed to input various features and predict diabetes instantaneously. | * In this paper, the authors developed a machine learning model to predict diabetes using explainable AI techniques. * Explainable AI techniques were used to understand how the XGBoost model was making its predictions. This was done by using the SHAP and LIME libraries. * They used a private dataset of female patients in Bangladesh and various machine learning techniques. * They found that the XGBoost classifier with ADASYN outperformed other methods, with an accuracy of 81%, F1 score of 0.81, and AUC of 0.84. * Even a website framework and an Android smartphone application have been developed to input various features and predict diabetes instantaneously. |
| [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).[6] | * This research explores three interpretable supervised ML models: Naïve Bayes, random forest, and J48 decision tree, trained and tested using the Pima Indians diabetes dataset in R programming. * Performance analysis assessed accuracy, precision, sensitivity, and specificity, concluding that Naïve Bayes excels with a refined feature selection for binary classification, whereas random forest performs better with more features. | * This paper developed classification models for electronic diagnostic systems in the IoMT. * Using the Pima Indian Diabetes dataset, they trained three machine learning algorithms to predict diabetes presence. * The random forest model outperformed Naïve Bayes and J48 decision tree models in terms of accuracy, precision, specificity, F-score, and AUC. The J48 model had the best sensitivity. * They proposed an IoMT-based e-diagnosis system for diabetes prediction and management. * Future work includes improving accuracy with preprocessing techniques and extending the approach to other diseases. |
| [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.*[7] | * In this study, various algorithms, including SVM, Naïve Bayes, KNN, RF, LR, and DT, were assessed using two datasets: Pima Indian diabetic (PID) and Germany diabetes. * The WEKA 3.8.6 tool was used for experimentation, evaluating performance metrics and error rates. * For PID, SVM achieved 74% accuracy, while for the Germany dataset, KNN and RF performed better with 98.7% accuracy. | * This study explored various machine learning predictive algorithms for diabetes prediction, including Naïve Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), and Logistic Regression (LR). * They conducted experiments using two datasets: PID and Germany diabetes datasets in WEKA software. * The results showed that SVM and LR were the most accurate for the PID dataset, with LR performing better in terms of ROC area. For the Germany dataset, KNN and RF demonstrated higher accuracy, and RF excelled in terms of the ROC area. The research also discussed error rates for these classifiers. * The overall analysis suggested that LR is a preferable choice for both datasets. * Future work may involve using hybrid models and evaluating their performance on these datasets and real-time data. |
| [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*.[8] | * This study compares K-Nearest Neighbor (KNN) and Naive Bayes algorithms to build an intelligent predictive model. * After conducting experiments and assessing the algorithms using the Confusion Matrix, the study concludes that Naive Bayes is more effective than KNN in predicting diabetes based on health attributes in the dataset. | * In this study, the performance of two k-Nearest Neighbor (KNN) algorithms and the Naive Bayes algorithm in predicting diabetes based on health attributes in PIMA dataset using supervised machine learning was compared. * Based on the experiment results and the evaluation using a Confusion Matrix, the Naive Bayes algorithm outperformed KNN. Naive Bayes achieved an average accuracy of 76.07 percent, precision of 73.37 percent, and recall of 71.37 percent, while KNN had an average accuracy of 73.33 percent, precision of 70.25 percent, and recall of 69.37 percent. * Future research directions could involve incorporating other algorithms such as neural networks and employing techniques like Particle Swarm Optimization to enhance accuracy and precision, as well as developing application programs for practical use. |
| [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*.[9] | * In this study, a finely-tuned Random Forest algorithm with optimized parameters (RFWBP) was developed to detect diabetes patients at an early stage. The research involved various data processing techniques, such as normalization and converting raw data into numerical format during preprocessing. Additional data mining techniques were applied to enhance the dataset with relevant features. * The study compared the performance of RFWBP with other conventional methods including AdaBoost, support vector machine, logistic regression, naive Bayes, multilayer perceptron, and a regular Random Forest. 5-fold cross-validation was used to improve RFWBP's performance. | * PIMA dataset is used. * The results showed that RFWBP achieved an accuracy of 95.83% with cross-validation and 90.68% without it. The experimental results demonstrated that RFWBP outperformed the conventional machine learning methods in diabetes detection. * Future research aims to expand the analysis by including more subjects and diverse datasets, both balanced and imbalanced. This approach intends to offer deeper insights for more precise diabetes patient identification. |
| [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*. [10] | * This paper introduces HealthEdge, a smart healthcare framework based on machine learning, designed for predicting type 2 diabetes within an integrated IoT-edge-cloud computing system. * The study conducted numerical experiments and comparative analysis using two commonly used machine learning algorithms, Random Forest (RF) and Logistic Regression (LR), with real-life diabetes datasets. * The findings indicate that, on average, RF achieves a 6% higher accuracy in diabetes prediction compared to LR. | * The study introduced HealthEdge, a smart healthcare framework for predicting type 2 diabetes using IoT devices in an integrated edge-cloud system. * It compared the Random Forest and Logistic Regression algorithms on PIMA Indian and Sylhet datasets. Random Forest outperformed Logistic Regression, with an accuracy of 78.27% for PIMA Indian and 97.23% for Sylhet. * Feature selection improved or maintained accuracy, while dataset balancing had mixed results. * The research underscores the importance of diverse datasets and serves as a foundation for future IoT-based healthcare applications. |

2.2.Books/ Journals/ Articles referred

| Sr.No | Title | Abstract | Inference |
| --- | --- | --- | --- |
| [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.* [11] | * This article introduces 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)—were employed for diabetes classification. For predictive analysis, Long Short-Term Memory (LSTM), Moving Averages (MA), and Linear Regression (LR) were utilized. * The study used the PIMA Indian Diabetes dataset and found that MLP achieved the highest accuracy at 86.08%, while LSTM significantly improved diabetes prediction with 87.26% accuracy. * Comparative analysis with existing techniques is also demonstrated. | * This paper introduces two algorithms, one based on Multilayer Perceptrons (MLP) for diabetes classification and another using Long Short-Term Memory (LSTM) for diabetes prediction. * It also proposes an IoT-based system for real-time diabetes monitoring, using BLE sensors and smartphones. * The system aims to help users monitor vital signs, predict diabetes, and offers accuracy rates of 86.083% for MLP and 87.26% for LSTM when tested with the PIMA Indian Diabetes dataset. * Future work includes developing an Android application and exploring genetic algorithms for improved monitoring. |
| [2] | 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). [12] | * This study compared machine learning-based prediction models (Glmnet, RF, XGBoost, LightGBM) with traditional regression models for predicting undiagnosed Type 2 Diabetes Mellitus (T2DM). * The prediction performance for fasting plasma glucose levels was assessed using 100 bootstrap iterations with data subsets simulating new incoming data in 6-month batches. * With only 6 months of available data, a simple regression model achieved the lowest average Root Mean Square Error (RMSE) of 0.838, followed by RF (0.842), LightGBM (0.846), Glmnet (0.859), and XGBoost (0.881). * When more data was added, Glmnet showed the highest improvement rate (+3.4%). LightGBM models demonstrated the highest stability in variable selection over time. | * This study compared machine learning-based prediction models with multivariable regression models using electronic health record (EHR) data. * It found that rebuilding models with new EHR data improved performance and variable importance stability, but machine learning models did not significantly outperform regression models. Simpler models were advantageous in terms of visualization and stability. LightGBM was notably stable. As more data became available, all methods showed improved predictive performance. * Future research should explore ensemble methods but consider challenges in result interpretation for healthcare decisions. |
| [3] | Ashwini Tuppad , Shantala Devi Patil, “Machine learning for diabetes clinical decision support: a review”, *Advances in Computational Intelligence , 2022.* [13] | * This review begins by identifying gaps in our medical knowledge about diabetes from existing articles, highlighting areas where machine learning (ML) can help. * It then explores ML research in three key areas: (1) assessing the risk of diabetes, (2) diagnosing diabetes using various methods, and (3) predicting outcomes, including complications. * The review also discusses the limitations and gaps in current ML approaches for diabetes, suggesting areas for future research. * Overall, it offers a comprehensive overview of how ML can be applied in diabetes-related decision support, while pointing out both medical and technological areas that need further attention. | * This paper offers a review of machine learning (ML) research in the context of clinical decision support for type 2 diabetes. It begins by identifying gaps in diabetes knowledge, guidelines, and practice that ML can address. * The review covers recent research in three key areas: (1) statistical and ML-based risk assessment, (2) non-invasive and invasive diagnosis methods, and (3) prognosis modeling for predicting type 2 diabetes incidence and related complications. * The paper emphasizes the potential of ML applications for the prevention and management of type 2 diabetes while pointing out research gaps that must be addressed to create clinically reliable ML decision support models for diabetes care. |
| [4] | 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.* [I] | * The text discusses the increasing prevalence of diabetes due to lifestyle changes and poor habits, with a focus on Type II diabetes and its link to insulin resistance and complications like kidney, eye, and heart issues. * There's also mention of research into a potential connection between diabetes and cancer. The text describes an overview of these findings and related cancer research efforts. * The work involves using dimensionality reduction, classification, and clustering techniques to compare with existing classifiers, utilizing the PIMA Indian diabetes dataset and Stanford AIM-94 dataset as benchmark datasets for experimentation. | * The text highlights the extensive research and efforts by scientists worldwide, which provide significant motivation and inspiration for various scopes in the field. * It mentions that many tests have not definitively shown a direct link between preexisting Type II diabetes and cancer. The focus is on understanding the factors contributing to malignant growth and distinguishing between harmful and benign tumors. * The current work involves developing a new methodology to establish potential connections between Type II diabetes and various types of cases. This is achieved through data-driven information design and machine learning techniques to classify data accurately into normal and abnormal categories. |

2.3 Patent search

| Sr.No | Title | Abstract | Inference |
| --- | --- | --- | --- |
| [1] | Spencer Frank, David Price, Chuck Stroyeck , Kazanna Calais Hames , DexCom , Inc. Diabetes prediction using glucose measurements and machine learning, United States patent US 11,426,102 B2, Aug. 30 , 2022 .[14] | * This describes a method for predicting diabetes using machine learning and glucose measurements. * It involves training a machine learning model using historical glucose measurements and outcome data from a user population. * The model is used to predict a diabetes classification for an individual user based on their glucose measurements collected by a wearable glucose monitoring device over several days. * The prediction can be communicated to the user through notifications or user interfaces. * The historical data includes information from both glucose monitoring devices and other diagnostic measurements from independent sources. | * It describes an observation analysis platform that employs a machine learning model trained with historical glucose measurements and outcome data from a user population to predict an individual user's diabetes classification. * The glucose measurements come from wearable monitoring devices used by the user population, while the outcome data includes diagnostic measurements like HbAlc, FPG, and 2 Hr - PG, which can indicate clinical diabetes diagnosis. * Once trained, the machine learning model predicts the user's diabetes classification based on glucose measurements collected over several days. This classification can categorize the user's diabetes state (e.g., diabetes, prediabetes, or no diabetes) or predict adverse effects of diabetes. * The prediction can be shared through notifications or user interfaces, such as a healthcare provider report indicating the diabetes classification or a notification advising the user to contact their healthcare provider. |

2.4 Exiting Systems

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.

2.5. Lacuna in the existing systems

1.Data Availability and Quality**:** Many prediction systems rely on patient data such as blood glucose levels, medical history, and lifestyle factors. Data may not always be readily available, and the quality of data can vary, which can affect the accuracy of predictions.

2. Lack of Personalization: Some prediction systems may not account for individual variations in diabetes management. Personalized medicine and treatment plans are essential, as diabetes varies greatly from person to person.

3. Limited Predictive Power: Some systems might not have highly accurate predictive capabilities, leading to false positives and negatives. This can cause unnecessary stress or complacency for patients.

4. Incomplete Data: Many prediction models may not incorporate all relevant variables that can influence diabetes risk, such as genetic factors, environmental factors, and lifestyle changes.

5. Overfitting: Overfitting occurs when a model is trained too closely on a specific dataset and may not generalize well to new data. It can lead to unreliable predictions.

6.. Interpretability: Some advanced machine learning models, like deep learning, may lack interpretability. Understanding how and why a prediction is made is crucial for both healthcare professionals and patients.(Explainable AI is not integrated).

2.6 Comparison of existing systems and proposed area of work

-The existing systems have not employed all the algorithms while we have implemented 5 different algorithms to compare the accuracy and check the statistics.

-Visualisations are very rarely found in the existing systems.we have implemented the visuals such as correlation heatmap, Feature importance plot,AUC-ROC curve.

-We have tried to implement OCR technique which will take input as parameters from the user and get the data collected to extract the values.

-The interface provided in the existing system is not so user friendly.Thus,we tried to improve the UI by giving sliders in the input fields.

2.7. Focus Area

-Accurate free prediction of diabetes:

The main focus area is the correct prediction of diabetes based on the input provided by the users.For this,we implemented various classification algorithms like Decision Tree,SVM,Random Forest,XGBoost and Adaboost.

-To provide easily accessible interface to the user

It is essential that the model that we employed provides the results whenever the user needs.Thus, we provided a website where the user can easily select the input parameters and get the results.

-Predict the results with appropriate reasons

With using explainable-AI tools like Lime and Shap we are able to provide the reasons why the predictions are positive or negative.

**Chapter 3: Requirements**

3.1 Proposed model

The core functionality of the application involves users providing input such as patient attributes (age, gender, family history), medical history (glucose levels, blood pressure), and lifestyle factors (diet, physical activity). This input data will be processed by the ML model, which has been trained on extensive datasets to recognize patterns and correlations indicative of diabetes risk. The model's output will provide a prediction regarding the likelihood of the user developing diabetes in the future.

Our application will also be able to detect whether an individual is diabetic with the help of retinal images. This is called Retinopathy. Diabetes, in severe cases,can be determined from the retina.Thus, it aims to determine the same from the retinal images by using the technique of image processing.

Users input their data, enabling the machine learning model to predict their risk of diabetes and offer personalized insights. The application offers tailored health recommendations, lifestyle modifications, and potential referrals to healthcare professionals based on the user's risk level. By integrating continuous progress tracking and updates, the application empowers users to take informed actions towards better health, bridging the gap between risk assessment and medical interventions.

3.2 Functional Requirements

* 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.
* 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.
* Accuracy and reports: Patients should be able to check the accuracy and correct prediction for the data that he entered.

3.3.Non-Functional Requirements

* Performance: Response Time: The system should respond to user interactions within a reasonable timeframe (e.g., sub-second response for most actions).
* Scalability: The system should be scalable to accommodate an increasing number of users and data volume.
* Throughput: It should handle a certain number of simultaneous users or requests without performance degradation.
* Reliability: The system should be highly available and reliable, minimizing downtime or service interruptions. It should have mechanisms for disaster recovery and data backup.
* 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.
* Usability: The user interface should be intuitive, user-friendly, and designed with user experience (UX) principles in mind.

3.4.Hardware and Software Requirements

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

3.5.Technology and Tools utilized

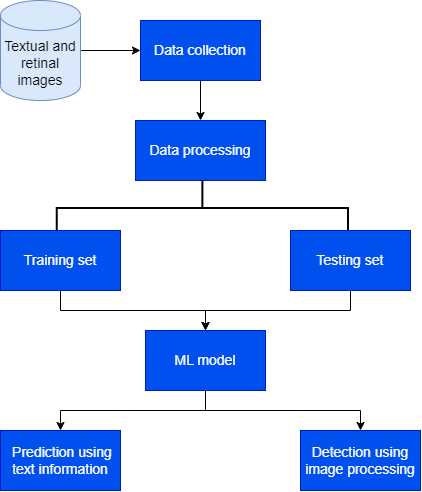
* **Languages**: Python for the ML model
* **Python Stack**:Scikit Learn, Numpy, Matplotlib, Pandas, Keras, Kivy, LIME, SHAP, TensorFlow
* **To build the Application**:Streamlit library

3.6.Constraints of working

* Budget Constraints: Limited financial resources may restrict the availability of funds for hardware, software, personnel, and ongoing operational costs.
* Time Constraints: The project may have tight deadlines, especially if it is related to healthcare services or research with critical timelines. Regulatory and Compliance Constraints: Compliance with healthcare regulations (e.g., HIPAA, GDPR) and data privacy laws can impose constraints on data handling and storage.
* Data Availability Constraints: Access to comprehensive and accurate patient health data may be limited, affecting the development of machine learning models and data-driven insights.
* Technological Constraints: Outdated or limited technology infrastructure may constrain the system's capabilities and scalability.

**Chapter 4: Proposed Design**

4.1 Block Diagram of the proposed system



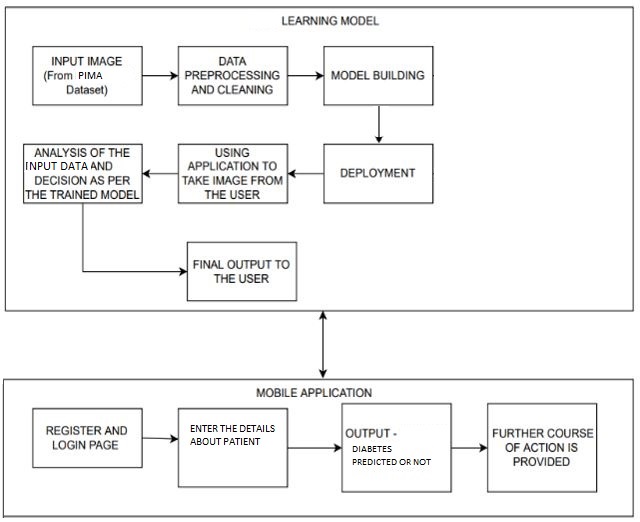
**Fig 1. Block Diagram**

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.

4.2 Modular diagram

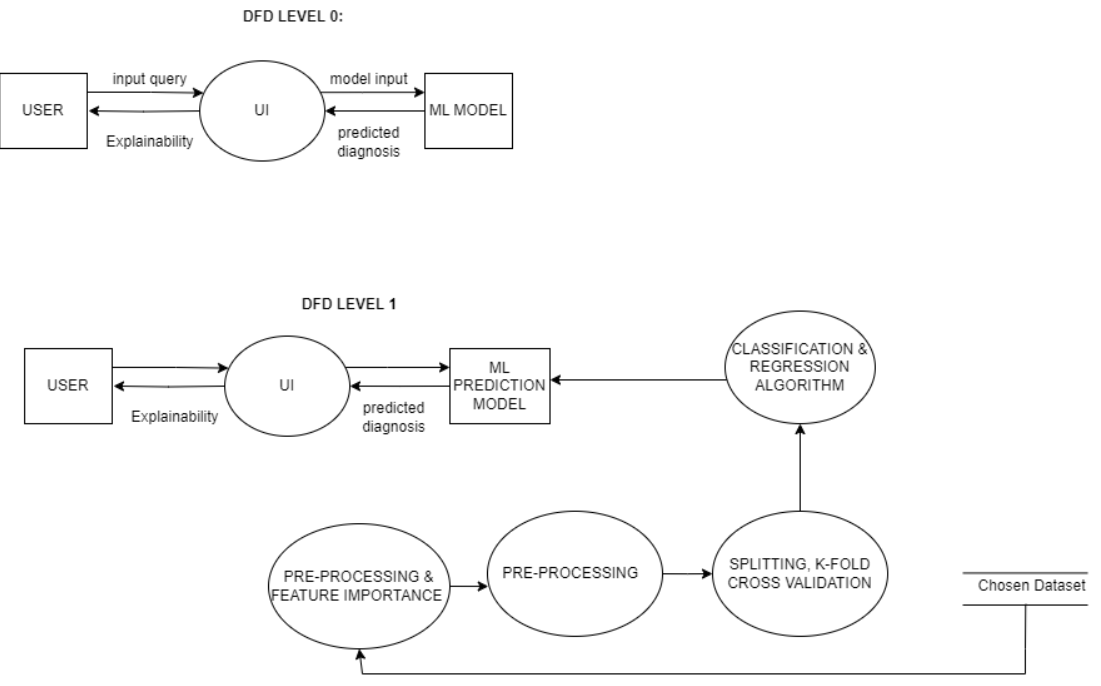


**Fig 2. Modular Diagram**

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.

4.3 Detailed Design

****

The system takes input from a variety of sources, including:

User: This includes information, such as Pregnancies(Number of times pregnant), Glucose(Plasma glucose concentration a 2 hours in an oral glucose tolerance test), BloodPressure(Diastolic blood pressure (mm Hg)), SkinThickness(Triceps skin fold thickness (mm)), Insulin(2-Hour serum insulin (mu U/ml)), BMI(Body mass index (weight in kg/(height in m)^2)), DiabetesPedigreeFunction, Age(years).

UI: This could include data from the user interface.

Model input: This is the data that will be used to train the machine learning model.

The system then uses a variety of machine learning techniques to predict the user's risk of developing diabetes. The predicted diagnosis is then output to the user, along with an explanation of the prediction.

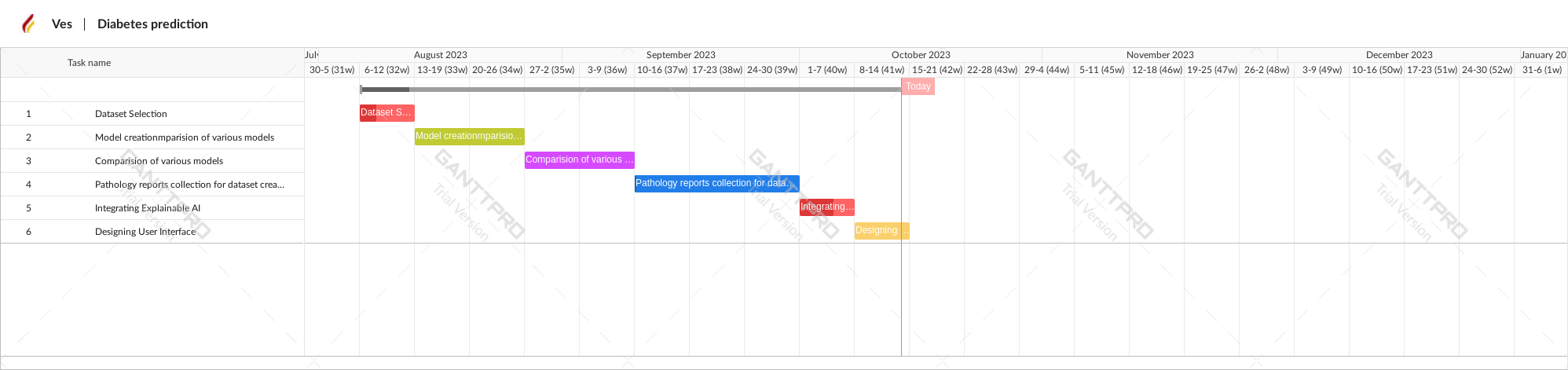
The explanation of the prediction is important because it helps the user to understand the factors that contributed to the prediction. This can help the user to make informed decisions about how to manage their risk of developing diabetes.

4.4. Proposed algorithms

We have implemented the following classification algorithms to predict whether a person is diabetic or not:

* Decision Tree (DT)
* Support Vector Machine (SVM)
* Random Forest (RF)
* Adaboost
* XGBoost

4.5. Project Scheduling and; Tracking using Timeline / Gantt Chart

****

**Chapter 5. Results and Discussions**

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:

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

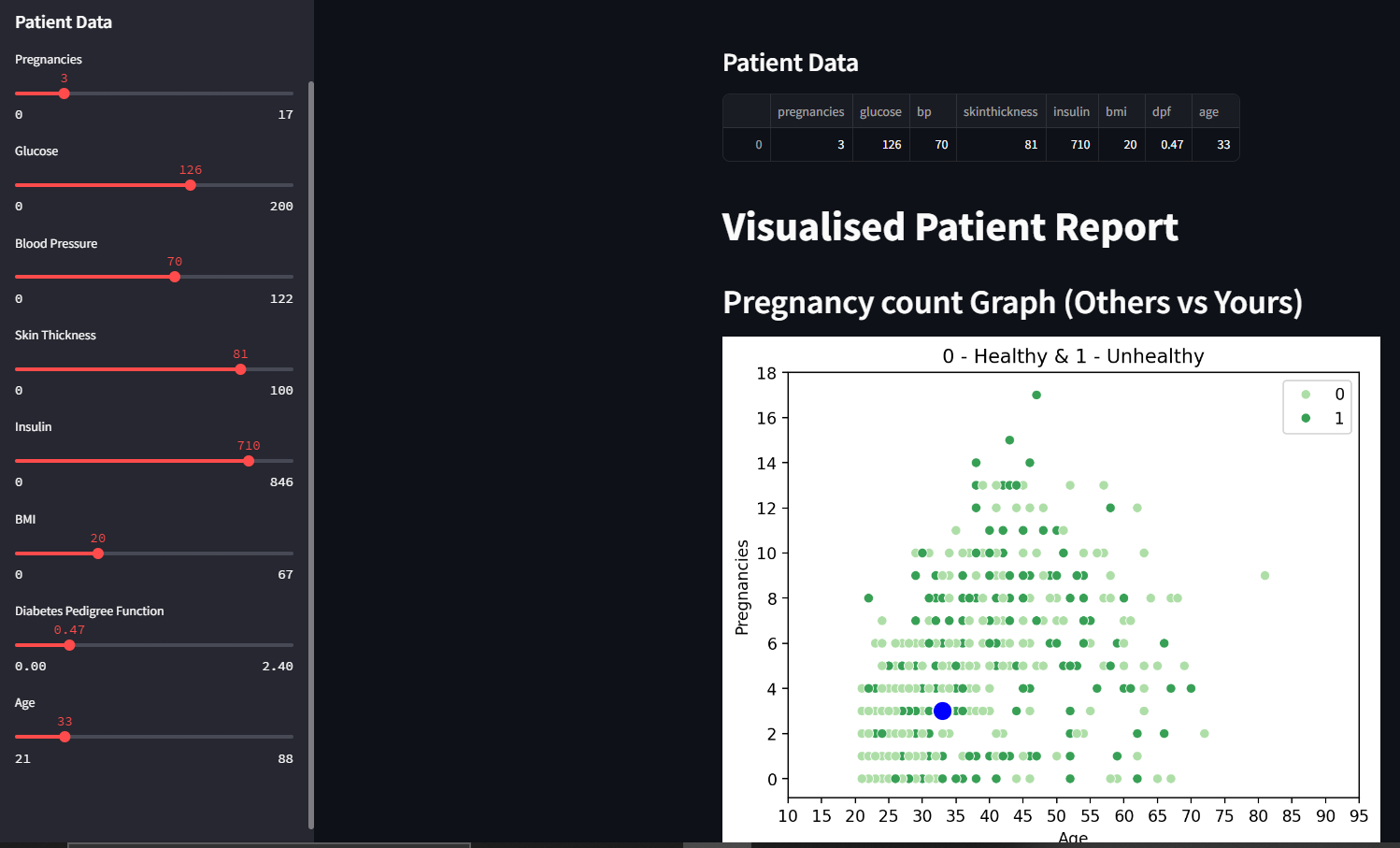
|  | SVM ( non-linear: rbf kernel ) | Decision tree | Random Forest | Adaboost | XGBoost |
| --- | --- | --- | --- | --- | --- |
| 80:20 | 75% | 75% | 76.6% | 73.37 | 73.37% |
| 70:30 | 74% | 70% | 75.3% | 74.45 | 73.59% |
| 5-fold cross | 76.17% | 73% | 85.7% | 80.52 | 76.43% |
| 10-fold cross | 76.43%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

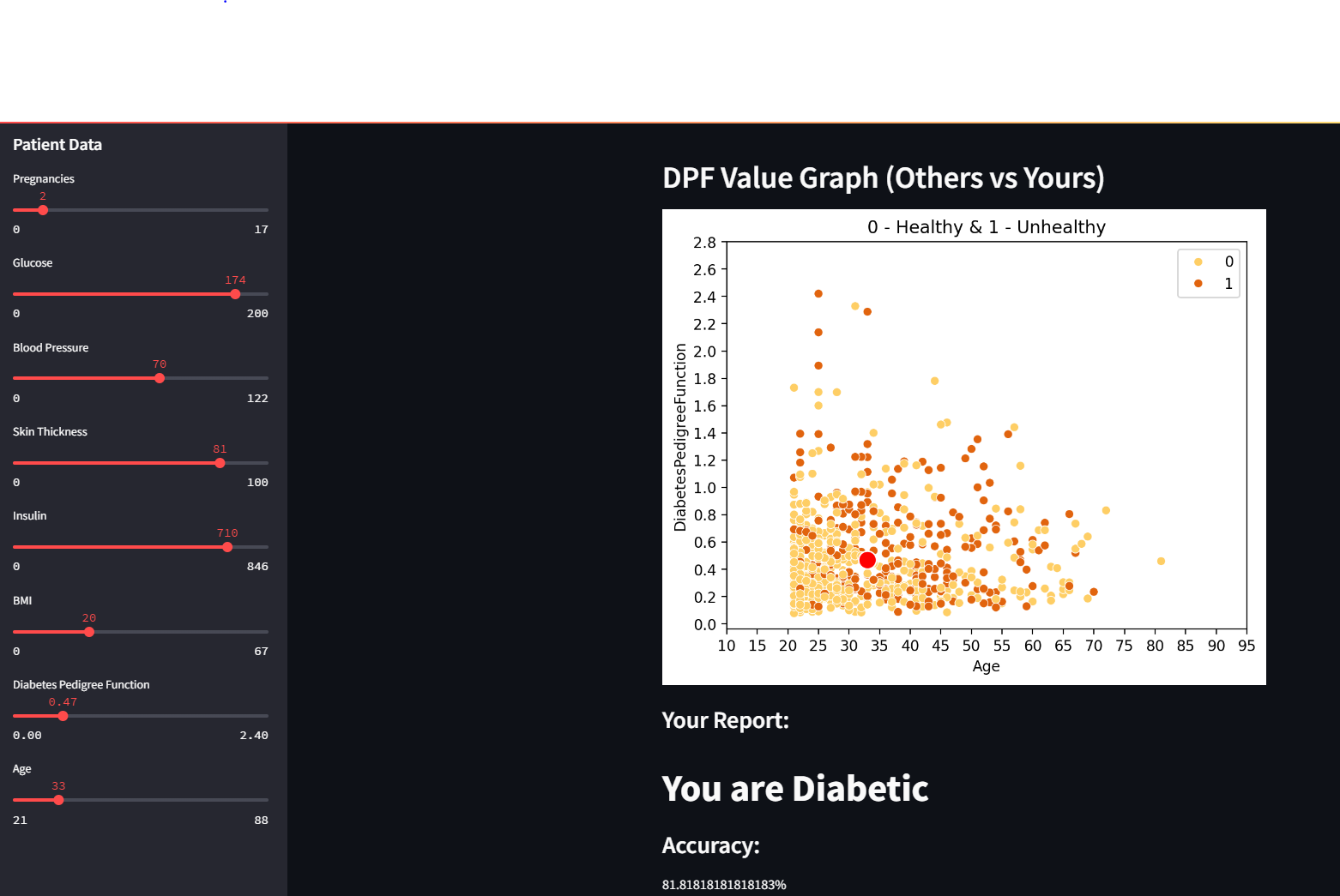
The other dataset:

**Table 3**. All Algorithms accuracy comparison wrt 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% |

**UI for the website:**

****

****

**6. Plan of action for the next semester**

1. Build an ensemble model using the models implemented along with ADAboost, XGBoost and a few more datasets if found.
2. Implement Explainable AI to validate the predictions and make users trust the system.
3. Enhancing the UI
4. Retinopathy dataset for Detection Model.

**7. Conclusions**

Thus 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. Furthermore, We collected pathological Reports to build our own dataset. Additionally, we have built the user interface.

**8. References**

i. 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>

ii.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>

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. Spencer Frank, David Price, Chuck Stroyeck , Kazanna Calais Hames , DexCom , Inc. Diabetes prediction using glucose measurements and machine learning, United States patent US 11,426,102 B2, Aug. 30 , 2022 .

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

**9. Appendix**

a. List of Figures

| **Figure Number** | **Heading** | **Page no.** |
| --- | --- | --- |
| Fig.1 | Block Diagram | 40 |
| Fig 2. | Modular Diagram | 41 |

b. List of tables

| **Table number** | **Heading** | **Page no.** |
| --- | --- | --- |
| Table 1 | Study of Reference papers and their inferences | 11 |
| Table 2 | All Algorithms accuracy comparison wrt PIMA dataset | 44 |
| Table 3 | All Algorithms accuracy comparison wrt the other dataset | 44 |

c. Paper Publications :-

Review-1 marksheet:

