A
Mini-Project Report on

DIABETES PREDICTION

Submitted in partial fulfillment of the requirements for
the degree of
BACHELOR OF ENGINEERING
IN
Computer Science & Engineering
(Artificial Intelligence & Machine
Learning)

by

Himanshu Rajput (21106055)
Vinayak Kokare (21106052)
Priyesh Mangela (21106049)
Harsh Kokitkar (21106029)

Under the guidance of

Prof. Ranjita Asati

Department of Computer Science & Engineering
(Artificial Intelligence & Machine Learning)
A. P. Shah Institute of Technology
G. B. Road, Kasarvadavali, Thane (W)-400615
University of Mumbai

2023-2024

A. P. SHAH INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled "**Diabetes Prediction"** is a bonafide work of Himanshu Rajput(21106055), Harsh Kokitkar (21106029), Vinayak Kokare (21106052), Priyesh Mangela (21106049) submitted to the University of Mumbai in partial fulfillment of the requirement for the award of **Bachelor of Engineering** in **Computer Science & Engineering (Artificial Intelligence & Machine Learning).**

Prof. Ranjita Asati	Dr. Jaya Gupta
Mini Project Guide	Head of Department

A.P. SHAH INSTITUTE OF TECHNOLOGY

Project Report Approval

This Mini project report entitled "**Diabetes Prediction"** by **Himanshu Rajput, Vinayak Kokare, Priyesh Mangela, Harsh Kokitkar** *is* approved for the degree of ***Bachelor of Engineering*** in ***Computer Science &Engineering***, (AI&ML) ***2023-24***.

External Examiner: _____

Internal Examiner: _____

Place: APSIT, ThaneDate:

Declaration

We declare that this written submission represents our ideas in our own words and where other's 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.

Himanshu Rajput	Vinayak Kokare	Priyesh Mangela	Harsh Kokitkar
(21106055)	(21106052)	(21106049)	(21106029)

ABSTRACT

Diabetes mellitus remains a significant global health concern, with a rising prevalence that places immense burdens on healthcare systems and individual well-being. The development and evaluation of a comprehensive diabetes prediction model, which harnesses the power of machine learning and data analytics to predict the risk of diabetes onset. The model is designed to assist healthcare professionals in identifying high-risk individuals and implementing preventive measures to mitigate the impact of diabetes.

A diverse and extensive dataset comprising clinical, lifestyle, and genetic factors was used to train and validate the predictive model. A wide range of machine learning algorithms, including logistic regression, decision trees, random forests, and deep neural networks, were explored to identify the most effective approach for diabetes risk prediction. Feature engineering and selection techniques were employed to optimize the model's performance and interpretability.

Diabetes Mellitus is among critical diseases and lots of people are suffering from this disease. People having diabetes have high risk of diseases like heart disease, kidney disease, stroke, eye problem, nerve damage, etc. Current practice in hospital is to collect required information for diabetes diagnosis through various tests and appropriate treatment is provided based on diagnosis.

Index

List of Figures

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

Diabetes is a chronic metabolic disorder that affects millions of people worldwide. It occurs when the body either cannot produce enough insulin or cannot effectively utilize the insulin it produces, leading to abnormal blood sugar levels. Diabetes can result in serious health complications, including heart disease, kidney problems, vision impairment, and nerve damage. Early diagnosis and management of diabetes are crucial to prevent these complications and improve the quality of life for individuals with the condition.[1]

Predictive modeling has emerged as a valuable tool in the field of healthcare, with the potential to revolutionize diabetes care. By leveraging data and advanced analytics, healthcare professionals can predict the risk of diabetes in individuals, enabling early intervention and personalized treatment strategies. Diabetes prediction models utilize various factors, such as genetics, lifestyle, medical history, and biomarkers, to assess an individual's susceptibility to developing the disease.[2]

This predictive approach has the potential to benefit both healthcare providers and patients. Healthcare providers can identify high-risk individuals, offer preventive measures, and allocate resources more efficiently. Patients can take proactive steps to modify their lifestyle, monitor their health, and reduce their risk of diabetes.[3]

This introduction provides a glimpse into the world of diabetes prediction, where data-driven approaches hold the promise of a healthier and more informed future for those at risk of this prevalent and life-altering condition. In the following sections, we will delve deeper into the methodologies, data sources, and the significance of diabetes prediction in the context of modern healthcare.[4]

CHAPTER 2

LITERATURE SURVEY

1. LITERATURE SURVEY

2.1-HISTORY

The history of diabetes dates back thousands of years, with references to a condition similar to diabetes appearing in ancient medical texts. Here are some key milestones in the history of diabetes:

Ancient References: The earliest known reference to a condition resembling diabetes comes from an Egyptian papyrus dating back to around 1550 BCE. It describes a condition called "too great emptying of the urine."[1]

"Diabetes" Coining: The word "diabetes" is derived from the Greek word meaning "siphon" or "to pass through," which reflects the excessive urination characteristic of the condition. The term was coined by Aretaeus of Cappadocia, a Greek physician, around the first century CE.[2]

Early Treatment Attempts: Throughout history, various treatments were attempted for diabetes, including dietary restrictions and herbal remedies. These early attempts had limited success in managing the condition.[2]

Discovery of Insulin: One of the most significant breakthroughs in diabetes history came in the early 20th century. In 1921, Canadian scientists Frederick Banting and Charles Best discovered and isolated insulin, a hormone produced by the pancreas. This discovery revolutionized the treatment of type 1 diabetes, which was previously a fatal condition[3].

Classification of Diabetes: In the 20th century, researchers further classified diabetes into different types. Type 1 diabetes, previously known as juvenile diabetes, is characterized by the immune system attacking the insulin-producing cells in the pancreas. Type 2 diabetes, which is more common, is associated with insulin resistance and often linked to lifestyle factors.

Advancements in Treatment: Over the years, there have been significant advancements in diabetes treatment and management. This includes the development of various medications, glucose monitoring devices, and insulin delivery methods.[4]

Diabetes Epidemic: In recent decades, diabetes has reached epidemic proportions, with a significant increase in the number of people diagnosed with the condition, particularly type 2 diabetes. This increase is largely attributed to sedentary lifestyles, poor dietary choices, and rising obesity rates.

Research and Prevention: Ongoing research is focused on understanding the genetic and environmental factors contributing to diabetes. Public health initiatives and educational programs aim

to prevent and manage diabetes through lifestyle modifications, early detection, and effective treatment.

The history of diabetes is marked by significant milestones, from the ancient descriptions of the condition to the discovery of insulin and the advancements in treatment and prevention. Diabetes continues to be a major public health challenge, and research and healthcare efforts are ongoing to address this complex and widespread disease.[5]

2.2-LITERATURE REVIEW

[1] Hasan, Md Kamrul, Md Ashraful Alam, Dola Das, Eklas Hossain, and Mahmudul Hasan. "Diabetes prediction using ensembling of different machine learning classifiers." ***IEEE Access*** **8 (2020): 76516-76531.**

Diabetes, also known as chronic illness, is a group of metabolic diseases due to a high level of sugar in the blood over a long period. The risk factor and severity of diabetes can be reduced significantly if the precise early prediction is possible. The robust and accurate prediction of diabetes is highly challenging due to the limited number of labeled data and also the presence of outliers (or missing values) in the diabetes datasets. In this literature, we are proposing a robust framework for diabetes prediction where the outlier rejection, filling the missing values, data standardization, feature selection, K-fold cross-validation, and different Machine Learning (ML) classifiers (k-nearest Neighbour, Decision Trees, Random Forest, AdaBoost, Naive Bayes, and XGBoost) and Multilayer Perceptron (MLP) were employed. The weighted ensembling of different ML models is also proposed, in this literature, to improve the prediction of diabetes where the weights are estimated from the corresponding Area Under ROC Curve (AUC) of the ML model. AUC is chosen as the performance metric, which is then maximized during hyperparameter tuning using the grid search technique. All the experiments, in this literature, were conducted under the same experimental conditions using the Pima Indian Diabetes Dataset. From all the extensive experiments, our proposed ensembling classifier is the best performing classifier with the sensitivity, specificity, false omission rate, diagnostic odds ratio, and AUC as 0.789, 0.934, 0.092, 66.234, and 0.950 respectively which outperforms the state-of-the-art results by 2.00 % in AUC. Our proposed framework for the diabetes prediction outperforms the other methods discussed in the article.

[2] Jayanthi, Neelampalli, B. Vijaya Babu, and N. Sambasiva Rao. "Survey on clinical prediction models for diabetes prediction." ***Journal of Big Data*** **4 (2017): 1-15.**

Predictive analytics has gained a lot of reputation in the emerging technology Big data. Predictive analytics is an advanced form of analytics. Predictive analytics goes beyond data mining. A huge amount of medical data is available today regarding the disease, their symptoms, reasons for illness, and their effects on health. But this data is not analysed properly to predict or to study a disease. The aim of this paper is to give a detailed version of predictive models from base to state-of-art, describing various types of predictive models, steps to develop a predictive model, their applications in health care in a broader way and particularly in diabetes.

[3] El_Jerjawi, Nesreen Samer, and Samy S. Abu-Naser. "Diabetes prediction using artificial neural network." (2018).

Diabetes is considered as one of the deadliest and chronic diseases which causes an increase in blood sugar. Many complications occur if diabetes remains untreated and unidentified. The tedious identifying process results in visiting of a patient to a diagnostic center and consulting doctor. But the rise in machine learning approaches solves this critical problem. The motive of this study is to design a model which can prognosticate the likelihood of diabetes in patients with maximum accuracy. Therefore, three machine learning classification algorithms namely Decision Tree, SVM and Naive Bayes are used in this experiment to detect diabetes at an early stage. Experiments are performed on Pima Indians Diabetes Database (PIDD) which is sourced from UCI machine learning repository. The performances of all the three algorithms are evaluated on various measures like Precision, Accuracy, F-Measure, and Recall. Accuracy is measured over correctly and incorrectly classified instances. Results obtained show Naive Bayes outperforms with the highest accuracy of 76.30% comparatively other algorithms.

[4] Khan, Farrukh Aslam, Khan Zeb, Mabrook Al-Rakhami, Abdelouahid Derhab, and Syed Ahmad Chan Bukhari. "Detection and prediction of diabetes using data mining: a comprehensive review." *IEEE Access* 9 (2021): 43711-43735.

Diabetes is one of the most rapidly growing chronic diseases, which has affected millions of people around the globe. Its diagnosis, prediction, proper cure, and management are crucial. Data mining based forecasting techniques for data analysis of diabetes can help in the early detection and prediction of the disease and the related critical events such as hypo/hyperglycemia. Numerous techniques have been developed in this domain for diabetes detection, prediction, and classification. In this paper, we present a comprehensive review of the state-of-the-art in the area of diabetes diagnosis and prediction using data mining. The aim of this paper is twofold; firstly, we explore and

investigate the data mining-based diagnosis and prediction solutions in the field of glycemic control for diabetes. Secondly, in the light of this investigation, we provide a comprehensive classification and comparison of the techniques that have been frequently used for diagnosis and prediction of diabetes based on important key metrics. Moreover, we highlight the challenges and future research directions in this area that can be considered in order to develop optimized solutions for diabetes detection and prediction.

[5] Jayanthi, Neelampalli, B. Vijaya Babu, and N. Sambasiva Rao. "Survey on clinical prediction models for diabetes prediction." ***Journal of Big Data*** **4 (2017): 1-15.**

Predictive analytics has gained a lot of reputation in the emerging technology Big data. Predictive analytics is an advanced form of analytics. Predictive analytics goes beyond data mining. A huge amount of medical data is available today regarding the disease, their symptoms, reasons for illness, and their effects on health. But this data is not analysed properly to predict or to study a disease. The aim of this paper is to give a detailed version of predictive models from base to state-of-art, describing various types of predictive models, steps to develop a predictive model, their applications in health care in a broader way and particularly in diabetes.

CHAPTER 3

Problem Statement

2.Problem Statement

Diabetes is a global health crisis, with millions of individuals affected by this chronic metabolic disorder. The problem for diabetes encompasses a range of issues related to its prevention, diagnosis, management, and the associated healthcare burden.

It is to create a user-friendly tool that helps individuals assess their risk of developing diabetes by providing a simple, non-invasive way to determine susceptibility. Users will input their Body Mass Index (BMI), insulin levels, glucose levels, and age into the tool, which will then analyze this information and provide a clear indication of whether they are at risk for diabetes. The primary objective is to offer an accessible and intuitive solution for early diabetes risk assessment, empowering individuals to take proactive steps towards a healthier lifestyle.

CHAPTER 4

Experimental Setup

4. Experimental Setup

4.1 Hardware Setup

- Operating System – Windows version 10/11

4.2 Software Setup

- Python - version (3.10)
- VS Code
- Google Colab
- GUI - Flask Python

CHAPTER 5

Proposed System & Implementation

3. Proposed system & Implementation

5.1 Block diagram of proposed system

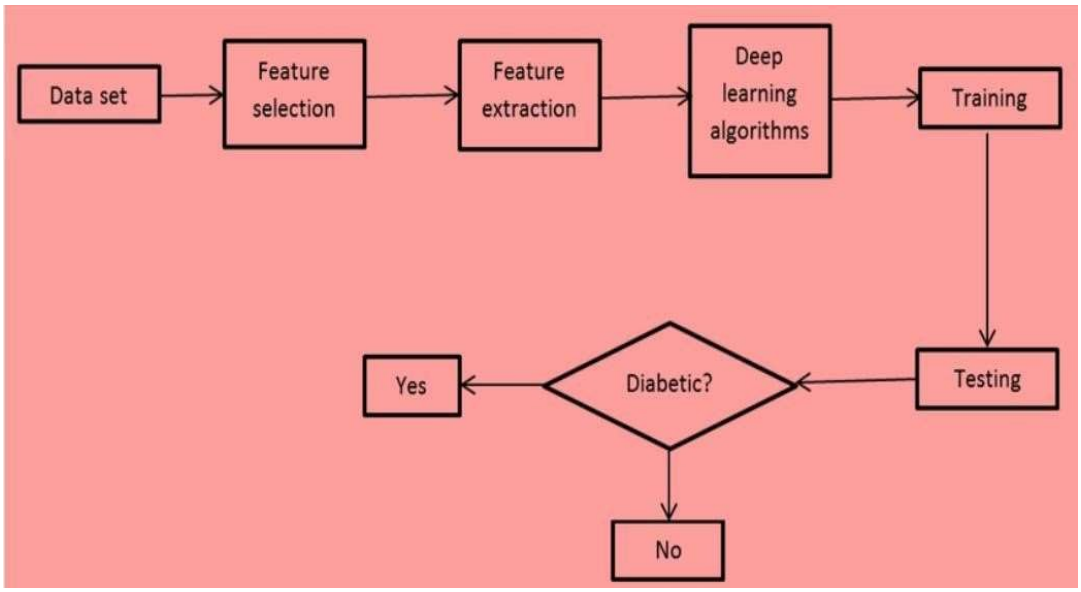

Fig.5.1 Block Diagram

5.2 Implementation

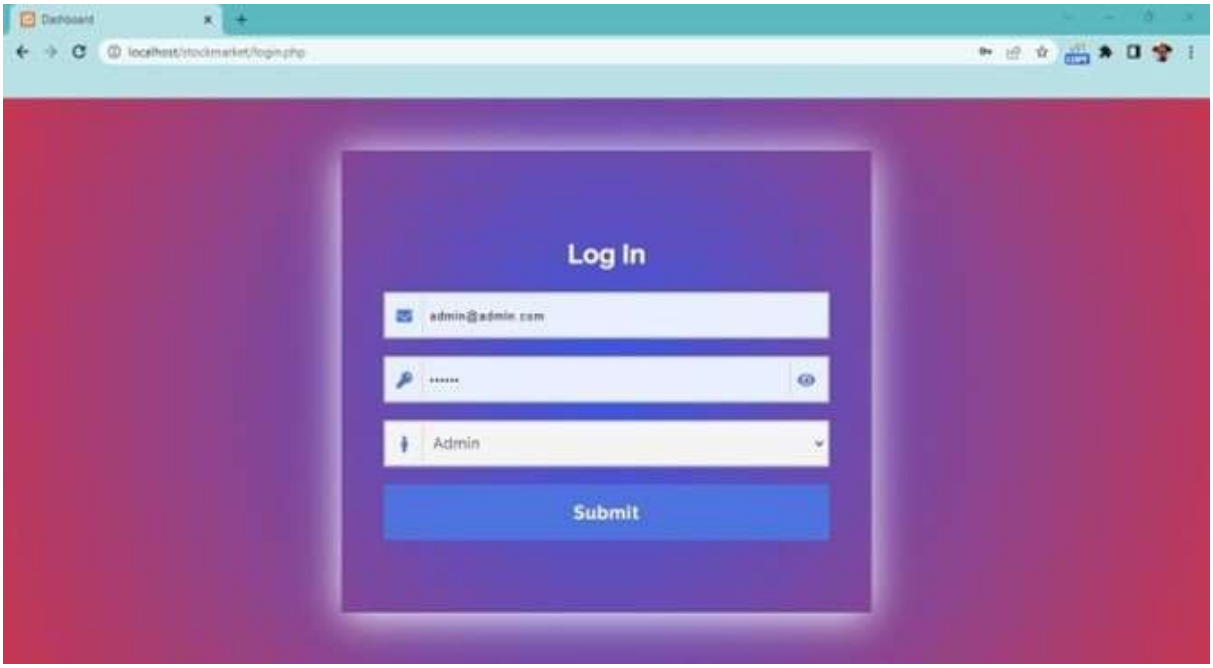

Fig.5.2 GUI Panel

5.3 Application

- The field of diabetes management and prevention has seen significant advancements in recent years, driven by innovative applications and technologies. These applications, often in the form of mobile apps, web platforms, and wearable devices, play a critical role in helping individuals with diabetes, healthcare providers, and researchers. Here are some key applications of technology in diabetes:

- Blood Glucose Monitoring Apps: Mobile apps and wearable devices allow individuals with diabetes to easily monitor their blood glucose levels. These applications often provide real-time data, trend analysis, and reminders for medication and testing. Some can even transmit data directly to healthcare providers for remote monitoring and intervention.
- Diet and Nutrition Apps: Nutrition plays a crucial role in managing diabetes. Various apps help users track their food intake, count carbohydrates, and make healthier dietary choices. These apps can provide personalized meal plans and recipes, making it easier for individuals to manage their blood sugar levels.
- Physical Activity and Fitness Apps: Regular physical activity is essential for diabetes management. Fitness apps provide exercise routines, track daily activity, and encourage users to stay active. They can also monitor heart rate and other relevant metrics.

- Insulin Dose Calculators: For individuals with type 1 diabetes or those on insulin therapy, insulin dose calculators can help determine the correct dosage based on factors such as blood glucose levels, carbohydrate intake, and activity levels.
- Medication Adherence Apps: Medication non-adherence is a common issue in diabetes management. These apps send reminders for medication schedules and help individuals keep track of their medications, reducing the risk of missed doses.

- Telemedicine and Virtual Healthcare: Telehealth platforms and apps allow individuals with diabetes to consult with healthcare professionals remotely. They can receive expert advice, discuss their condition, and get prescriptions without the need for in-person visits, improving access to care, especially in rural or underserved areas.

CHAPTER 6

Conclusion

4.Conclusion

Our simple diabetes prediction project has shown promising results in helping individuals assess their risk of developing diabetes. By inputting basic information such as BMI, insulin levels, glucose levels, and age, users can gain valuable insights into their susceptibility to diabetes. While this project is not a substitute for professional medical advice, it can serve as a useful initial screening tool to raise awareness about potential risk factors.

By making this tool accessible and user-friendly, we hope to encourage people to take proactive steps towards a healthier lifestyle and seek medical advice when necessary. Early detection and lifestyle modifications can play a crucial role in preventing or managing diabetes.

References

[1] Zelinsky, Alex. "Learning OpenCV---Computer vision with the OpenCV library (Bradski, GR et al.; 2008)[On the Shelf]." IEEE Robotics & Automation Magazine 16, no. 3 (2009): 100100.

[2]Haralick, Robert M., Hyonam Joo, Chung-Nan Lee, Xinhua Zhuang, Vinay G. Vaidya, and ManBae Kim. "Pose estimation from corresponding point data." IEEE Transactions on Systems, Man, and Cybernetics 19, no. 6 (1989): 1426-1446 [3] https://mpeg.chiariglione.org/who-we-are

[3] Yuan, Zhaocong, Adam W. Hall, Siqi Zhou, Lukas Brunke, Melissa Greeff, Jacopo Panerati, and Angela P. Schoellig. "Safe-control-Gym: A unified benchmark suite for safe learning-based control and reinforcement learning in robotics." IEEE Robotics and Automation Letters 7, no. 4 (2022): 11142-11149.

[4] Saravanan, Renuka, and Pothula Sujatha. "A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification." In 2018 Second international conference on intelligent computing and control systems (ICICCS), pp. 945-949. IEEE, 2018.

[5] Taware, Gourangi, Reena Kharat, Pratik Dhende, Prathamesh Jondhalekar, and Rohit Agrawal. "AI-based Workout Assistant and Fitness guide." In 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA, pp. 14. IEEE, 2022.