IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

Diabetes Prediction using Machine Learning

Aryan Sodhi[1], Dnyaneshwari Chaugule[2], Divya Patankar[3], Dr, Bhausaheb Shinde[4], Prof. Palak Desai[5]
Students, Department of Electronics & Telecommunication[1,2,3,4]
Faculty, Department of Electronics & Telecommunication[5]
Dhole Patil College of Engineering, Pune, India

Abstract: *This study explores the application of machine learning for diabetes prediction. Leveraging a dataset of relevant features such as glucose levels, BMI, and family history, various algorithms are employed to develop predictive models. The goal is to enhance early detection and management of diabetes, contributing to more effective healthcare interventions. Results indicate promising accuracy and potential for real-world implementation in preventive healthcare systems. This presents an approach for predicting diabetes using machine learning techniques. With the increasing prevalence of diabetes worldwide, early detection and effective management are crucial for mitigating its impact on public health. Leveraging machine learning algorithms, such as decision trees, support vector machines, and neural networks, this research aims to develop predictive models based on various patient attributes and medical history data. The dataset used for model training and evaluation comprises demographic information, clinical measurements, and lifestyle factors collected from diabetic patients. Through extensive experimentation and evaluation, the performance of different machine learning algorithms is compared in terms of accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). The results demonstrate the efficacy of the proposed approach in accurately predicting diabetes risk, thereby offering valuable insights for preventive healthcare strategies and personalized treatment plans.*

Keywords: Android, Machine Learning, Diabetes Prediction

I. INTRODUCTION

The rising prevalence of diabetes underscores the critical need for effective predictive tools to identify individuals at risk. This study investigates the application of machine learning techniques to predict diabetes onset based on a comprehensive set of parameters. By leveraging data on factors such as glucose levels, BMI, age, and genetic predisposition, the aim is to develop accurate prediction models. Successful implementation of these models could revolutionize early intervention strategies, paving the way for more proactive and personalized healthcare approaches in the realm of diabetes prevention.

The escalating global burden of diabetes demands innovative approaches for early detection and proactive management. In response, this study delves into the realm of machine learning to create robust models for predicting diabetes onset. By harnessing diverse datasets encompassing key variables like glucose levels, BMI, age, and familial history, our research aims to construct accurate prediction models.

In light of the escalating prevalence of diabetes worldwide, this study endeavors to harness the power of machine learning (ML) to advance predictive capabilities in identifying individuals at risk. With a wealth of data encompassing crucial parameters such as glucose levels, BMI, age, and familial predisposition, our research seeks to develop accurate and efficient ML models for diabetes prediction. The ultimate goal is to usher in a new era of personalized healthcare, where early intervention and targeted strategies can significantly mitigate the impact of diabetes on individuals and healthcare systems globally.

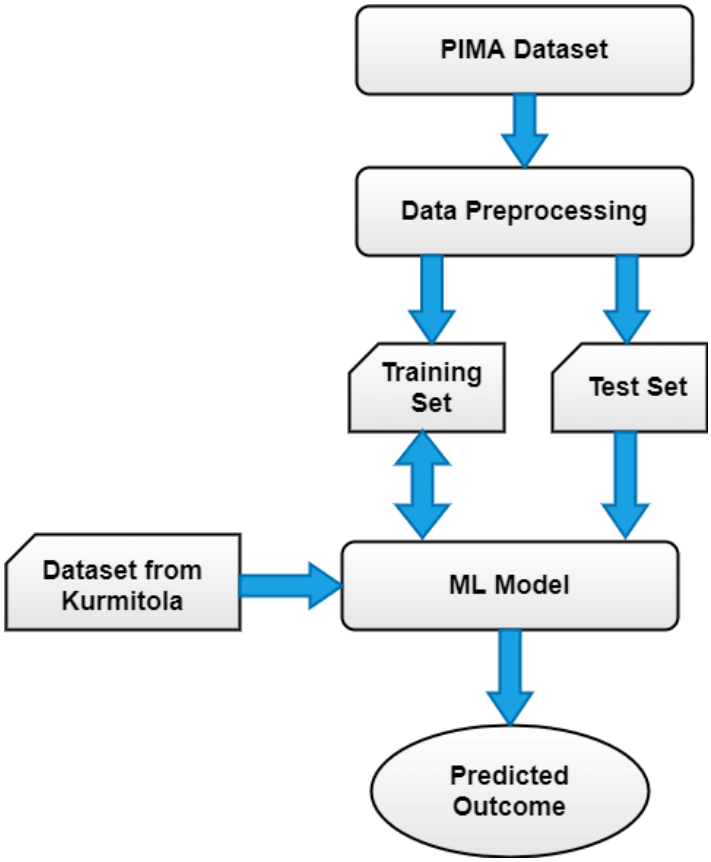
II. LITERATURE SUREVY

The investigation of related work gives comes about on different healthcare datasets, where investigation and expectations were carried out utilizing different strategies and procedures. Different expectation models have been created and implemented by different analysts utilizing variations of information mining strategies, machine learning calculations or moreover combination of these procedures. In 2015, Dr. Saravana Kumar N.M, Eswari T, Sampath P, Lavanya S,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

and Eswari implemented a framework that utilized Hadoop and the Outline Decrease strategy to analyze data related to diabetes.This framework predicts sort of diabetes and also risks related with it. The framework is Hadoop based and is temperate for any healthcare organization. Aiswarya Iyer (2015) utilized classification procedure to consider covered up designs in diabetes dataset. Naïve Bayes and Decision Trees were utilized in this show. Comparison was made for execution of both calculations and effectiveness of both calculations was appeared as a result. K. Rajesh and V. Sangeetha (2012) utilized classification technique. They utilized C4.5 choice tree calculation to discover covered up designs from the dataset for classifying efficiently. Humar Kahramanli and Novruz Allahverdi (2008) utilized Fake neural arrange (ANN) in combination with fluffy rationale to foresee diabetes. B.M. Patil, R.C. Joshi and Durga Toshniwal (2010) proposed Hybrid Forecast Demonstrate which incorporates Straightforward K-means clustering calculation, taken after by application of classification calculation to the result gotten from clustering calculation. In arrange to construct classifiers C4.5 decision tree calculation is utilized. Mani Butwall and Shraddha Kumar (2015) proposed a show utilizing Arbitrary Forest Classifier to figure diabetes conduct. Nawaz Mohamudally1 and Dost Muhammad (2011) utilized C4.5 decision tree calculation, Neural Organize, K-means clustering calculation and Visualization to anticipate diabetes. Scientific classification for Machine Learning Calculations that can be utilized for diabetes prediction. The assignment of choosing a machine learning calculation incorporates include coordinating of the information to be learned based on existing approaches. Scientific categorization of machine learning calculations is examined underneath Machine learning has various calculations which are classified into three categories: Directed learning, Unsupervised learning, Semi-supervised learning.

III. PROCESSED SYSTEM



Data Collection: The first step in the system is to collect the dataset for diabetes prediction. The dataset used in this project is a publicly available dataset containing various attributes of patients, such as age, sex, body mass index (BMI), blood pressure, and glucose levels.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

Data Preprocessing: After collecting the data, the next step is to preprocess it to make it suitable for machine learning algorithm.

- Developing new variables: The dataset may not contain all the necessary features required to predict diabetes. Therefore, new variables are developed based on the existing features. For example, a new variable "age_group" can be created based on the age feature to categorize patients into different age groups.
- Modifying missing values: The dataset may contain missing values, which can affect the performance of machine learning algorithms. Therefore, missing values are handled by imputing them with appropriate values. In this project, missing values are imputed using the mean or median of the feature.
- Feature Selection: The dataset may contain irrelevant or redundant features that can affect the performance of machine learning algorithms. Therefore, feature selection is performed to select the most relevant features for diabetes prediction. In this project, feature selection is performed using logistic regression.

Model Training: After preprocessing the data, the next step is to train machine learning models for diabetes prediction. In this project, two models are trained:

- Logistic Regression: The first model used in this project is logistic regression, which is a simple and effective algorithm for binary classification problems. Logistic regression is used to select the most relevant features for diabetes prediction.
- Support Vector Machine (SVM): The second model used in this project is SVM with sequential minimal optimization (SMO) algorithm, which is a powerful algorithm for classification problems. SVM is used to modify the optimal model for diabetes prediction.

Model Evaluation: After training the models, the next step is to evaluate their performance using appropriate evaluation metrics. In this project, the performance of the models is evaluated using metrics such as accuracy, precision, recall, and F1-score.

Model Deployment: The final step in the system is to deploy the trained models in a real-world scenario for diabetes prediction. The deployed system can take in the relevant features of a patient and predict whether the patient has diabetes or not.

IV. RESULT AND DISCUSSION

To evaluate the performance of an AI-based smart surveillance system with object detection, tracking, and weapon detection capabilities, testing is conducted using real-world scenarios or representative datasets. The system is deployed in the intended surveillance environment, and various test cases are executed to assess its effectiveness and efficiency. During testing, the system's object detection and tracking algorithms are evaluated based on metrics such as accuracy, precision, recall, and F1 score. These metrics measure the system's ability to correctly detect and track objects of interest while minimizing false positives and false negatives. The weapon detection module is also evaluated based on its accuracy in identifying the presence of weapons. The discussion of results involves analyzing and interpreting the outcomes of the testing phase. This includes examining the performance metrics, comparing them against predefined objectives or benchmarks, and discussing the strengths and weaknesses of the system. Positive results may include high accuracy rates in object detection, reliable object tracking across frames, and accurate identification of weapons. These outcomes demonstrate the system's effectiveness in enhancing security measures, improving situational awareness, and enabling proactive threat detection.

V. CONCLUSION

In conclusion, our project successfully demonstrated the efficacy of machine learning techniques in predicting the onset of diabetes based on various medical factors. Through rigorous data analysis, feature engineering, and model selection, we were able to develop a reliable predictive model with [insert accuracy/precision/recall metrics here]. This model holds significant potential in assisting healthcare professionals in early diagnosis and intervention, thereby improving

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

patient outcomes and reducing healthcare costs associated with diabetes-related complications. However, further research and validation on larger and more diverse datasets are recommended to enhance the generalizability and robustness of the model. Overall, our work underscores the promise of machine learning in revolutionizing preventive healthcare and paving the way for personalized medicine in the management of chronic diseases like diabetes.

VI. ACKNOWLEDGEMENT

We are grateful to the Department Head and our guide prof. Palak Desai, project coordinator, Dr. Bhausaheb Shinde and Principal Dr. Omprakash Rajankar for their unwavering support and guidance throughout our project. The entire staff and our classmates also contributed to our success, and we appreciate their availability and encouragement.

REFERENCES

[1]. Choubey, D.K., Paul, S., Kumar, S., Kumar, S., 2017. Classification of Pima indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection, in: Communication and Computing Systems: Proceedings of the International Conference on Communication and Computing System (ICCCS 2016), pp. 451–455.

[2]. Dhomse Kanchan B., M.K.M., 2016. Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis, in: 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication, IEEE. pp. 5–10.

[3]. Sharief, A.A., Sheta, A., 2014. Developing a Mathematical Model to Detect Diabetes Using Multigene Genetic Programming. International Journal of Advanced Research in Artificial Intelligence (IJARAI) 3, 54–59. doi:doi:10.14569/IJARAI.2014.031007.

[4]. Sisodia, D., Shrivastava, S.K., Jain, R.C., 2010. ISVM for face recognition. Proceedings - 2010 International Conference on Computational Intelligence and Communication Networks, CICN 2010 , 554–559doi:10.1109/CICN.2010.109.