



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

⊕ www.ijirset.com 🖂 ijirset@gmail.com 🖄 +91-9940572462 🕓 +91 63819 07438



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)



www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404720|

Diabetic Prediction using Machine Learning

Nishigandha Charhate, Maithili Burghate, Rahasya Gawande Prof. P. R. Nerkar

U.G. Student, Department of Information Technology Prof. Ram Meghe Institute of Technology & Research,

Badnera, India

U.G. Student, Department of Information Technology Prof. Ram Meghe Institute of Technology & Research,

Badnera, India

U.G. Student, Department of Information Technology Prof. Ram Meghe Institute of Technology & Research,

Badnera, India

U.G. Student, Department of Information Technology Prof. Ram Meghe Institute of Technology & Research,

Badnera, India

Associate Professor, Department of Information Technology Prof. Ram Meghe Institute of Technology & Research,

Badnera, India

ABSTRACT: Diabetes Mellitus is a chronic disease affecting millions worldwide. Early detection can significantly reduce complications and improve patient outcomes. This research leverages machine learning algorithms to predict diabetes using patient diagnostic data. Techniques such as K-Nearest Neighbors (KNN), Logistic Regression, Random Forest, and Support Vector Machines (SVM) are evaluated. Data preprocessing, feature reduction through Principal Component Analysis (PCA), and model evaluation using accuracy, precision, recall, and F1-score metrics are employed. Results show that Logistic Regression delivers the highest predictive accuracy, offering a robust tool for early diagnosis. Machine Learning and Data mining has the ability to extract hidden knowledge from a huge amount of diabetes-related data. This paper reviewed and analyzed the current studies on classification of Diabetes. Furthermore, the study has developed a classification model for diabetes using decision tree, Naïve Bayes, Support vector machine and k nearest neighbor Algorithm. The classification model is based on a dataset of 15000 cases collected from different National Institute of Diabetes and Digestive and Kidney Diseases. The results of the Naïve Bayes can be used by medical specialist to classify and diagnose diabetic patients. These results help the medical doctors in the classification process of diabetes. This study follows different machine learning algorithms to predict diabetes disease at an early stage. Such as, KNN, Naïve Bayes, Decision Tree, and Support Vector machine to predict this chronic disease at an early stage for safe human life.

KEYWORDS: Diabetes Prediction, Machine Learning, Supervised Learning, , Logistic Regression, Data Preprocessing, Hyperglycemia, Blood Glucose Monitoring, PIMA Indian Diabetes Dataset, Model Evaluation (Precision, Recall, F1-Score),System Architecture, Web Application Development, Flask Framework, Python Programming, Classification Algorithms, Prediction Accuracy, Cross-validation, Data Visualization.

I. INTRODUCTION

In this day and age, one of the most notorious diseases to have taken the world by storm is Diabetes, which is a disease which causes an increase in blood glucose levels as a result of the absence or low levels of insulin. Due to the many criteria to be taken into consideration for an individual to harbor this disease, it's detection and prediction might be tedious or sometimes inconclusive. Nevertheless, it isn't impossible to detect it, even at an early stage . Performance and accuracy of the applied algorithms is discussed and compared to predict the Diabetes Disease using Machine learning and study the proposed hypothesis that supervised ML algorithm can improve health care by the accurate and early detection of diseases. It affects the organs of the human body. It can be controlled by predicting this disease earlier. If diabetics patient is untreated for a long time, it may lead to increase blood sugar. Objectives of the Unified Teacher Guardian Platform:



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)



www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404720|

This research work aims to analyze the Diabetes dataset, design, and implement a Diabetes prediction and recommendation system utilizing machine learning classification techniques. The specific objectives of this project work are:

- i. To review existing literature along the area of diabetes diagnosis and prediction.
- ii. Design and develop a model using machine learning techniques.
- iii. To analyze the Diabetes dataset and use Support Vector Machine and Random Forest algorithms to develop a prediction engine.

II. LITERATURE SURVEY

Diabetes prediction using machine learning (ML) techniques has been a significant focus in recent biomedical research. Various studies have explored different algorithms, feature engineering techniques, and performance evaluations to enhance prediction accuracy.

Arwatki Chen Lyngdoh et al. (2020) in the IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES) proposed diabetes prediction models utilizing multiple ML algorithms. Their work emphasized algorithm comparison for improving disease prediction accuracy [1]. Mitushi Soni and Dr. Sunita Varma (2020) published in IJERT applied several machine learning techniques, showcasing their effectiveness in early diabetes detection. Their study highlighted that the choice of classifier significantly impacts the prediction results [2].

Sivaranjani S et al. (2021) during the **7th ICACCS** conference presented a model incorporating **feature selection** and **dimensionality reduction**. Their work demonstrated that reducing irrelevant features enhances both the efficiency and performance of prediction models [3]. Shejal Kale et al. (2022) in **IJSDR** explored various ML approaches for diabetes prediction. They compared algorithms like Decision Trees, Random Forests, and SVMs, identifying the models best suited for clinical data [4].

Ashwini R and colleagues (2022) in **IJRTI** focused on building a diabetes prediction system using ML models and stressed the importance of data preprocessing to boost model accuracy [5]. G. Tripathi and R. Kumar (2020) at the **8th ICRITO** conference discussed early diabetes prediction using ML, particularly focusing on enhancing reliability and optimization of the prediction models [6][10].

Debadri Dutta et al. (2018) in an IEEE paper analyzed feature importance for diabetes prediction. Their research showed that understanding which features most affect model outcomes can lead to better interpretability and predictive performance [7]. Vijayakumar et al. (2019) explored the use of the Random Forest algorithm for diabetes prediction during the International Conference on Systems Computation Automation and Networking. Their findings reinforced the robustness and accuracy of ensemble learning methods in healthcare datasets [8]. Md. Faisal Faruque et al. (2019) at the International Conference on Electrical, Computer and Communication Engineering (ECCE) performed a comparative performance analysis of various ML techniques for predicting diabetes mellitus. Their work systematically evaluated models based on different metrics such as accuracy, precision, and recall [9].

III. METHODOLOGY

This study aims to propose a new model for diabetic's classification. Numerous algorithms and different approaches have been applied, such as traditional machine learning algorithms, ensemble learning approaches and association rule learning in order to achieve the best classification accuracy.

The methods employed in this research are split by the four main phases of the research work, which are the problem formulation phase, the dataset collection phase, and the experimentation phase and the results summarizing. This research started with formulating the research problem that is reviewing of the literature and formulating of the research problem. After the research problem formulation, this research identified the scope of the research, the objectives, and limitations of the research procedure. The second phase of the study is the dataset collection. The



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)



www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404720|

dataset items were collected from National Institute of Diabetes and Digestive and Kidney Diseases. The third phase of the study was the data preparation which included:

- Converting Data to Appropriate format
- Data Preprocessing

• Use Machine Learning to manipulate Data. In the experimentation phase several experiments were conducted and results were collected.

IV. EXPERIMENTAL

The project predicts the onset of diabetes in a person based on the relevant medical details collected. When the person enters all the relevant medical data required in the online Web portal, this data is then passed on to the trained model for it to make predictions whether the person is diabetic or non-diabetic the model then makes the prediction with an accuracy of **91.48** %, which is fairly good and reliable. Following figure shows the basic UI form which requires the user to enter the specific medical data fields. These parameters help determine if the person is prone to develop diabetes Our research has the added benefit of an associated Web app, which makes the model more user friendly and easily understandable



Fig : Website Include : Login Page, State Diagram of System.

V. CONCLUSION

The aim of this project is to assess the effectiveness or efficiency of Logistic Regression with other linear classifiers, Examples of such classifiers include SVM, KNN, and Random Forest(RF). The results of the comparison revealed that Logistic Regression outperformed all the other classifiers. The accuracy of Logistic Regression was found to be the highest, at 91.48%. The proposed approach utilized ensemble learning and classification methods, which resulted in high accuracy levels. These experimental results can assist healthcare professionals by enabling early predictions and informed decisions, these classifiers can aid in the treatment of diabetes and potentially save human lives.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404720|

The main aim of this project was to design and implement Diabetes Prediction Using Machine Learning Methods and Performance Analysis of that methods and it has been achieved successfully. The proposed approach uses various classification and ensemble learning method in which SVM, KNN, Random Forest, Decision Tree, Logistic Regression and Gradient Boosting classifiers are used. and 91.48 % classification accuracy has been achieved. The Experimental results can be asst health care to take early prediction and make early decision to cure diabetes and save humans life.

REFERENCES

- 1. Arwatki Chen Lyngdoh, Nurul Amin Choudhury, Soumen Moulik, "Diabetes Disease Prediction Using Machine Learning Algorithms", IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES) 2020.
- 2. Mitushi Soni, Dr. Sunita Varma " Diabetes Prediction using Machine Learning Techniques", Journal of Engineering Research & Technology (IJERT) 2020,
- 3. Sivaranjani S, Ananya S, Aravinth J, Karthika R, " Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction", 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021.
- 4. Shejal Kale, Priti Rahane, Mansi Ghumare, Snehal Patil B "Diabetes Prediction Using Different Machine Learning Approaches" IJSDR | Volume 7 Issue 5, 2022.
- 5. Ashwini r, s m aiesha Afshin, kavya v, Deepthi raj "diabetes prediction using machine learning" ijrti | volume 7, issue 7 | issn: 2456-3315, 2022 .
- G. Tripathi and R. Kumar, "Early Prediction of Diabetes Mellitus Using Machine Learning", 2020 8th International Conference on Reliability Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020.
- 7. Debadri Dutta, Debpriyo Paul, Parthajeet Ghosh, "Analyzing Feature Importance's for Diabetes Prediction using Machine Learning". IEEE, pp 942-928, 2018.
- 8. Vijayakumar, B.Lavanya, I.Nirmala, S.Sofia Caroline, "Random Forest Algorithm for the Prediction of Diabetes ".Proceeding of International Conference on Systems Computation Automation and Networking, 2019.
- 9. Md. Faisal Faruque, Asaduzzaman, Iqbal H. Sarker, "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus". International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.
- G. Tripathi and R. Kumar, "Early Prediction of Diabetes Mellitus Using Machine Learning", 2020 8th International Conference on Reliability Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1009-1014,









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com