International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 6-11, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402002
ISSN (Online): 2583-7257

Automated Diabetes Risk Assessment Using Machine Learning Algorithms

[1]G. Swathi, [2]A. Swathi, [3]M. Tharun Kumar Reddy, [4]M. Sravanthi, [5]S. Yatheesha, [6]P. Vikram Simha Reddy, [7]A. Prashanth

[1,2,3,4,5,6,7]Department of Computer Science Engineering (Artificial Intelligence), Gates Institute of Technology, Gooty, Andhra Pradesh, India
E-mail: [1]swathi.ganta20@gmail.com, [2]arleswathi2610@gmail.com, [3]tharunkumarreddmylarapu@gmail.com, [4]msravanthireddy31@gmail.com, [5]sunkarayatheesha@gmail.com, [6]vvikramsimha451@gmail.com, [7]amidalaprashanth@gmail.com

***Abstract* - Diabetes Mellitus is among critical diseases and lots of people are suffering from this disease. Age, obesity, lack of exercise, hereditary diabetes, living style, bad diet, high blood pressure, etc. can cause Diabetes Mellitus. 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. Big Data Analytics plays an significant role in healthcare industries. Healthcare industries have large volume databases. Using big data analytics one can study huge datasets and find hidden information, hidden patterns to discover knowledge from the data and predict outcomes accordingly. In existing method, the classification and prediction accuracy is not so high. In this Project, we have proposed a diabetes prediction model for better classification of diabetes which includes few external factors responsible for diabetes along with regular factors like Glucose, BMI, Age, Insulin, etc. Classification accuracy is boosted with new dataset compared to existing dataset. Further with imposed a pipeline model for diabetes prediction intended towards improving the accuracy of classification.**

Keywords: Diabetes Prediction, Machine Learning, Classification Algorithms, Artificial Intelligence, Healthcare Analytics.

I. INTRODUCTION

Healthcare sectors have large volume databases. Such databases may contain structured, semi-structured or unstructured data. Big data analytics is the process which analyses huge data sets and reveals hidden information, hidden patterns to discover knowledge from the given data. Considering the current scenario, in developing countries like India, Diabetic Mellitus (DM) has become a very severe disease. Diabetic Mellitus (DM) is classified as Non-Communicable Disease (NCB) and many people are suffering from it. Around 425 million people suffer from diabetes according to 2017 statistics. Approximately 2-5 million patients every year lose their lives due to diabetes. It is said that by 2045 this will rise to 629 million.[1] Diabetes Mellitus (DM) is classified asType-1 known as Insulin-Dependent Diabetes Mellitus (IDDM). Inability of human's body to generate sufficient insulin is the reason behind this type of DM and hence it is required to inject insulin to a patient. Type-2 also known as Non-Insulin-Dependent Diabetes Mellitus (NIDDM). This type of Diabetes is seen when body cells are not able to use insulin properly. Type-3 Gestational Diabetes, increase in blood sugar level in pregnant woman where diabetes is not detected earlier results in this type of diabetes. DM has long term complications associated with it. Also, there are high risks of various health problems for a diabetic person. A technique called, Predictive Analysis, incorporates a variety of machine learning algorithms, data mining techniques and statistical methods that uses current and past data to find knowledge and predict future events. By applying predictive analysis on healthcare data, significant decisions can be taken and predictions can be made. Predictive analytics can be done using machine learning and regression technique. Predictive analytics aims at diagnosing the disease with best possible accuracy, enhancing patient care, optimizing resources along with improving clinical outcomes.[1] Machine learning is considered to be one of the most important artificial intelligence features supports development of computer systems having the ability to acquire knowledge from past experiences with no need of programming for every case. Machine learning is considered to be a dire need of today's situation in order to eliminate human efforts by supporting automation with minimum flaws. Existing method for diabetes detection is uses lab tests such as fasting blood glucose and oral glucose tolerance. However, this method is time consuming. This paper focuses on building predictive model using machine learning algorithms and data mining techniques for diabetes prediction.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 6-11, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402002
ISSN (Online): 2583-7257

II. RELATED WORK

Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop: Now days from health care industries large volume of data is generating. It is necessary to collect, store and process this data to discover knowledge from it and utilize it to take significant decisions. Diabetic Mellitus (DM) is from the Non-

Communicable Diseases (NCD), and lots of people are suffering from it. Now days, for developing countries such as India, DM has become a big health issue. The DM is one of the critical diseases which has long term complications associated with it and also follows with various health problems. With the help of technology, it is necessary to build a system that store and analyze the diabetic data and predict possible risks accordingly. Predictive analysis is a method that integrates various data mining techniques, machine learning algorithms and statistics that use current and past data sets to gain insight and predict future risks. In this work machine learning algorithm in Hadoop MapReduce environment are implemented for Pima Indian diabetes data set to find out missing values in it and to discover patterns from it. This work will be able to predict types of diabetes are widespread, related future risks and according to the risk level of patient the type of treatment can be provided.

Predictive Analytics in Health Care Using Machine Learning Tools and Techniques: When we have a huge data set on which we would like to perform predictive analysis or pattern recognition, machine learning is the way to go. Machine Learning (ML) is the fastest rising arena in computer science, and health informatics is of extreme challenge. The aim of Machine Learning is to develop algorithms which can learn and progress over time and can be used for predictions. Machine Learning practices are widely used in various fields and primarily health care industry has been benefitted a lot through machine learning prediction techniques. It offers a variety of alerting and risk management decision support tools, targeted at improving patients' safety and healthcare quality. With the need to reduce healthcare costs and the movement towards personalized healthcare, the healthcare industry faces challenges in the essential areas like, electronic record management, data integration, and computer aided diagnoses and disease predictions. Machine Learning offers a wide range of tools, techniques, and frameworks to address these challenges. This paper depicts the study on various prediction techniques and tools for Machine Learning in practice. A glimpse on the applications of Machine Learning in various domains are also discussed here by highlighting on its prominence role in health care industry.

III. PROPOSED SYSTEM

Diabetes can be controlled if it is predicted earlier. To achieve this goal this project work we will do early prediction of Diabetes in a human body or a patient for a higher accuracy through applying, Various Machine Learning Techniques. Machine learning techniques Provide better result for prediction by con- structing models from datasets collected from patients. In this work we will use Machine Learning Classification and ensemble techniques on a dataset to predict diabetes. Which are Artificial Neural Networks (ANN), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), and Ensemble Classifier. The accuracy is different for every model when compared to other models. The Project work gives the accurate or higher accuracy model shows that the model is capable of predicting diabetes effectively. Our Result shows that Random Forest achieved higher accuracy compared to other machine learning techniques.

IV. GOALS OF THE SYSTEM

The goals of an Automated Diabetes Risk Assessment System using Machine Learning are like the stars guiding a ship across the uncertain sea of modern healthcare—they give direction, purpose, and power to the project. Here's a well-rounded list of the system's key goals:

1. Early Detection of Diabetes Risk: To identify individuals at risk of developing diabetes before symptoms appear, allowing for timely intervention, lifestyle changes, and medical support.

2. Use of Predictive Machine Learning Models: To build and train accurate, reliable, and interpretable machine learning models (e.g., Logistic Regression, Random Forest, SVM, etc.) that can analyze health data and predict diabetes risk with high performance.

3. Minimize Dependence on Costly Tests: To reduce the need for lab-based tests by using easily obtainable features like age, BMI, blood pressure, family history, and lifestyle habits—making the system affordable and non-invasive.

4. Improve Accessibility in Remote and Underserved Areas: To create a scalable and deployable system (e.g., mobile or web-based app) that can be used in rural, low-resource, or medically underserved regions, helping bridge the healthcare gap.

5. Ensure Data Privacy and Ethical Use: To implement robust security measures, respect patient confidentiality, and

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 6-11, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402002
ISSN (Online): 2583-7257

ensure compliance with data protection laws like HIPAA or GDPR.

6. Raise Awareness and Promote Preventive Healthcare: To encourage users to take preventive action—like better nutrition, exercise, and regular check-ups—through accessible, data-driven health insights.

V. ADVANTAGES OF PROPOSED SYSTEM

1. Early Detection, Timely Intervention: Your system enables early identification of individuals at risk for diabetes, giving them the chance to take preventive measures before symptoms or complications arise. This reduces long-term health issues and healthcare costs.

2. Fast, Automated, and Scalable: The system offers real-time predictions with minimal user input, thanks to automation. It can be deployed at scale across clinics, mobile apps, or remote villages—making it a true ally in public health.

3.Cost-Effective and Non-Invasive No need for expensive lab tests or invasive diagnostics. Your model uses easily available health parameters like age, BMI, and blood pressure to assess risk, making it ideal for low-resource settings.

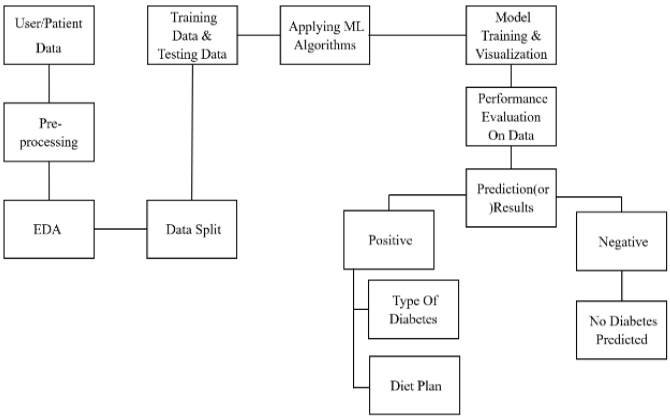
4. High Accuracy with Smart Models: By using advanced machine learning algorithms, your system can achieve higher accuracy than traditional rule-based systems, reducing the chances of false negatives or false positives.

5. Secure and Privacy-Compliant: Designed with data protection in mind, your system can implement encryption, anonymization, and compliance with HIPAA, GDPR, or national data policies earning user trust.

6. Promotes Preventive Healthcare Culture: Most importantly, your system shifts the healthcare model from reactive treatment to proactive prevention, encouraging healthier lifestyles and spreading awareness about diabetes risk.

VI. ARCHITECTURE

The process starts with gathering user or patient data, which includes relevant health parameters such as glucose levels, BMI, age, insulin levels, blood pressure, and other medical information. This raw data is essential for training and testing the machine learning model.



Once the data is collected, it undergoes pre-processing to ensure it is clean and suitable for analysis. Pre-processing involves steps such as handling missing values, removing duplicates, normalizing data, and converting categorical variables into numerical format if necessary. After pre-processing, Exploratory Data Analysis (EDA) is conducted to gain insights into the data. EDA helps in understanding the distribution of variables, identifying trends, and detecting any outliers or anomalies that may impact model performance.

Following EDA, the data is split into training and testing datasets. This step is crucial to evaluate the model's performance accurately and prevent overfitting. The training data is used to train the machine Learning algorithms, while the testing data helps assess how well the model generalizes to unseen data. Various machine learning algorithms are then applied to the training data. The choice of algorithm depends on the nature of the dataset and the desired accuracy and performance.

After applying the algorithms, the model undergoes training and visualization. During this phase, the model learns from the data and performance metrics such as accuracy, precision, recall, and F1-score are visualized to better understand the model's effectiveness. This is followed by a performance evaluation, where the model is tested against the test data to evaluate its predictive capabilities.

The next phase involves generating predictions or results. Based on the analysis, the model predicts whether the user is at risk of diabetes. If the result is positive, the system proceeds to determine the type of diabetes, such as Type 1, Type 2, or gestational diabetes. Additionally, a suitable diet plan is recommended to help the user manage or mitigate the disease. If the result is negative, the system confirms that there is no disease predicted, giving the user a clear indication of their health status.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 6-11, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402002
ISSN (Online): 2583-7257

VII. DATA FLOW DIAGRAM

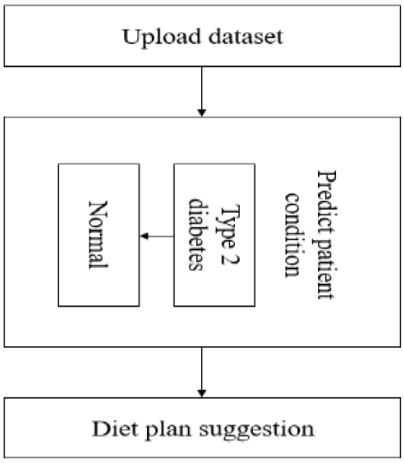


Fig: Level-1 DFD of user Application

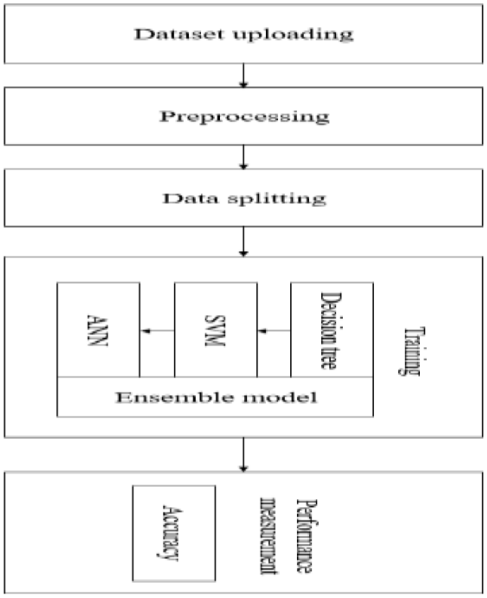


Fig: Level-2 DFD of cloud Application

The basic workflow of the diabetes prediction system using machine learning algorithms.

1. Upload Dataset: The process begins with uploading a dataset containing patient health attributes such as glucose levels, BMI, age, etc.

2. Predict Patient Condition: Once the data is uploaded, the system uses trained machine learning models to analyse the input and predict the patient's condition. Based on the analysis, the patient is classified as either Normal or having Type 2 Diabetes.

3. Diet Plan Suggestion: If the patient is predicted to have Type 2 Diabetes, the system proceeds to recommend a suitable diet plan. This ensures that along with diagnosis, the patient receives guidance for managing their condition through lifestyle changes.

The workflow of the machine learning model development process for automated diabetes risk assessment.

1. Dataset Uploading: The process starts with uploading the dataset that contains medical records and health attributes of patients relevant to diabetes prediction.

2. Preprocessing: The data undergoes cleaning and preprocessing steps such as handling missing values, normalization, and feature selection to prepare it for training.

3. Data Splitting: The dataset is split into training and testing sets to allow model training and performance evaluation.

4. Model Training: Multiple machine learning algorithms including ANN (Artificial Neural Network), SVM (Support Vector Machine), and Decision Tree are individually trained on the training data. These models are then combined using an Ensemble Model to enhance prediction accuracy and robustness.

5. Performance Measurement: The trained ensemble model is tested, and its performance is evaluated based on metrics like accuracy, helping determine its effectiveness in predicting diabetes.

VIII. RESULTS

The user interface includes input fields for entering relevant data (such as medical information), buttons to trigger actions like data prediction and diet plan recommendation, and visualizations for displaying results. This application serves as the front-end interaction point for users to input their data, receive predictions, and get diet recommendations. Involves a systematic and user-friendly application designed to assess diabetes risk and suggest diet plans accordingly. The main interface provides multiple functionalities such as uploading datasets, preprocessing data, and applying machine learning algorithms including Decision Tree, Support Vector Machine (SVM), Artificial Neural Network (ANN), and an Ensemble Classifier Model. Once the dataset is uploaded, the system displays the total number of records for reference, which in this case is 768. The user can then apply various algorithms, and the system computes and displays the accuracy of each. Among the models, the Ensemble Classifier achieves the highest accuracy, making it the most effective in predicting diabetes.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 6-11, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402002
ISSN (Online): 2583-7257

Once all models have been applied, this screen displays the accuracy of each algorithm:

1.Decision Tree: 74.6%

2.SVM: 78.5%

3.ANN: 68.6%

4.Ensemble Model:83.7%

This helps in identifying the most effective model for diabetes prediction.

A visual bar chart is generated to compare the performance of each model, providing a clear understanding of their efficiency. The system also includes a personalized user interface that allows individual users to upload their own health data. Upon submission, the system evaluates the user's health parameters using the trained model and predicts whether the user is at risk of Type 2 Diabetes. If the result is positive, the system goes a step further by suggesting a detailed diet plan tailored to the user's condition. This plan includes meal suggestions for breakfast, lunch, snacks, and dinner along with estimated calorie counts, helping users manage or prevent diabetes through proper nutrition. This end-to-end process not only ensures accurate diabetes prediction but also promotes health awareness through actionable dietary guidance.

IX. CONCLUSION

In this project, various machine learning algorithms are applied on the dataset and the classification has been done using various algorithms of which Ensemble Classifier model gives highest accuracy of 88.7%. Second highest accuracy gave the Support Vector Machine (SVM) with accuracy of 84.4%. We have seen comparison of machine learning algorithm accuracies with two different datasets. It is clear that the model improves accuracy and precision of diabetes prediction with this dataset compared to existing dataset. Further this work can be extended to find how likely non-diabetic people can have diabetes in next few years.

This study demonstrates the effectiveness of machine learning (ML) algorithms in automating the risk assessment of diabetes, offering a promising tool for early detection and intervention. By applying algorithms such as Support Vector Machine (SVM), Logistic Regression (LR), and an Ensemble Classifier Model and Artificial Neural Network (ANN), we were able to predict the likelihood of diabetes in individuals based on various health parameters. The results indicate that these algorithms can achieve a high degree of accuracy in predicting diabetes risk, with ensemble models providing superior performance by combining the strengths of multiple classifiers.

The outcomes of this project highlight the potential of ML-based systems in healthcare applications, particularly for preventing and managing diabetes, which remains a significant global health challenge. The model can be further refined by incorporating additional features, improving training data, or exploring other advanced ML techniques.

In conclusion, the automated diabetes risk assessment system not only contributes to better prediction accuracy but also presents an accessible tool for clinicians, allowing for quicker and more efficient risk identification. By integrating this system into clinical practices, we can facilitate timely interventions and enhance overall patient care.

X. FUTURE SCOPE

The project focuses on predicting diabetes based on key health attributes and applying machine learning algorithms to achieve faster and more accurate results. In the future, this system can be enhanced by integrating it with real-time health monitoring devices, allowing for instant data input and immediate risk assessment. Additionally, the model can be expanded to provide personalized diet plans for individuals who test positive, offering actionable health guidance. Developing the system into a mobile or web-based application would improve accessibility for both users and healthcare professionals. Regular updates to the model using new data can further improve its accuracy and relevance over time. Moreover, adding support for regional languages can help make the system more inclusive and beneficial for a wider audience, particularly in rural or underserved areas.

REFERENCES

[1] Sharma, A., & Parmar, D. (2020). Diabetes prediction using machine learning techniques. Materials Today: Proceedings, 33, 3738–3743.

[2] Muhammad, L. J., Algehyne, E. A., Usman, S. S., Yusuf, A., & Ahmad, A. (2020). Supervised machine learning models for prediction of diabetes mellitus. SN Computer Science, 1(6), 1–10.

[3] Vakil, V. et al. (2021). Explainable predictions of different machine learning algorithms used to predict Early Stage diabetes. arXiv preprint arXiv:2111.09939.

[4] Rana, R., & Sharma, D. (2021). Early diagnosis and prediction of diabetes using machine learning techniques. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7(2), 2456–3307.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 6-11, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402002
ISSN (Online): 2583-7257

[5] Dritsas, E., & Trigka, M. (2022). Data-Driven Machine-Learning Methods for Diabetes Risk Prediction. Sensors, 22(14), 5304.
[6] Imrie, F. et al. (2022). Auto Prognosis 2.0: Democratizing Diagnostic and Prognostic Modeling in Healthcare with Automated Machine Learning. arXiv preprint arXiv:2210.12090.
[7] Khan, M. Y. et al. (2022). An AI-based approach for early diabetes prediction using machine learning techniques. Computers in Biology and Medicine, 141, 105004.
[8] Abegaz, T. M. et al. (2023). Application of Machine Learning Algorithms to Predict Uncontrolled Diabetes Using the All of Us Research Program Data. Healthcare, 11(8), 1138.
[9] Kakoly, I. J., Hoque, M. R., & Hasan, N. (2023). Data-Driven Diabetes Risk Factor Prediction Using Machine Learning Algorithms with Feature Selection Technique. Sustainability, 15(6), 4930.
[10] Kumar, A., & Rajasekaran, C. (2023). A hybrid ensemble approach for diabetes prediction using ML algorithms. Informatics in Medicine Unlocked, 38, 101138.
[11] Gupta, P., & Sindhu, R. (2024). Diabetes Prediction Using Machine Learning. Journal of Electrical Systems, 20(7s).
[12] Karmand, M. et al. (2024). Machine-learning algorithms in screening for type 2 diabetes mellitus: Data from Fasa Adults Cohort Study. Endocrinology, Diabetes & Metabolism.

Citation of this Article:

G. Swathi, A. Swathi, M. Tharun Kumar Reddy, M. Sravanthi, S. Yatheesha, P. Vikram Simha Reddy, A. Prashanth. (2025). Automated Diabetes Risk Assessment Using Machine Learning Algorithms. *International Current Journal of Engineering and Science - ICJES*, 4(2), 6-11. Article DOI: https://doi.org/10.47001/ICJES/2025.402002

****

