JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Diabetes Detection Using Logistic Regression

[1] Ms.B.Sujani Reddy, [2] Dr. P. Shruthi,[3] Vegiraju Varun Sai NagaRaju, [4] Vavilala Ashwin Reddy,

[5] Thatipally Satwika

[1]Asst. Professor, Department of CSE (AI&ML), CMR College of Engineering & Technology, Hyderabad, Telangana
[2]HOD, Department of CSE (AI&ML), CMR College of Engineering &Technology, Hyderabad, Telangana
[3,4,5]UG Student, Department of CSE (AI&ML), CMR College of Engineering & Technology, Hyderabad,Telangana

Abstract: Diabetes, a persistentailment with the potential to precipitate aglobal healthcare crisis, affects a substantial portion of the global populace. According to the International Diabetes Federation, a staggering 382 million individuals currently grapple with diabetes worldwide. The underlying pathology of diabetes lies in the escalation of blood glucose levels, resulting in distressing symptoms like frequent urination, and insatiable hunger. Tragically, diabetes serves as the principal catalyst for an array of grave health complications, such as renal failure, heart failure, blindness, and so on, and strokes. Themalady manifests in diverse forms, with diabetes type 1 and type 2 being the most prevalent, while gestational diabetes emerges during pregnancy, among other variations. This entails amalgamating the outcomes including several machine learning methods, such as Random Forest, Support Vector Machine, Logistic Regression, and Decision Tree algorithms.

Keywords: Diabetes,blood-glucose, pregnancy, Random forest, Logistic Regression

I. INTRODUCTION

Diabetes is a Long-term illness that impacts how your body processes blood sugar (glucose). There are primarily two kinds of diabetes: type 1 and type 2. Type 1 diabetes is an autoimmune illness that manifests as your immune system attacks and destroys the cells in your pancreas that produce insulin. Type 2 diabetes occurs when your body becomes resistant to insulin or doesn't produce enough insulin to maintain normal blood sugar levels. Diabetes is a dangerous illness that may cause a range of health issues, such as heart disease, stroke, kidney disease, nerve damage, and blindness. However, with proper management and treatment, people with diabetes can lead a long healthy life. Doctors rely on common knowledge for treatment. When common knowledge is lacking, studies are summarized after some number of cases have been studied. However, this procedure takes time. whereas if machine learning is used, the patterns can be identified earlier. For using machine learning, a huge amount of data is required. There is a very limited amount of data available depending on the disease. Also, the number of samples having no diseases is very high compared to the number of samples actually having the disease. We have some existing solutions of thus diabetes detection using the adap algorithm[4], SVC, and Random forest[2]. The primary aim of this project is to analyze the Diabetes Dataset and use Logistic Regression, Random forest[2], and Support vector machine algorithms[3][7] for prediction and to develop a prediction engine. The secondary aim is to develop a web application with the following features. Allow users to predict diabetes utilizing the prediction engine. The objective is to achieve the aims of the project through Research on statistical models in machine learning and to understand how the algorithms work.

II. RELATED WORK

When conducting related work for diabetes detection using logistic regression, you'd typically want to explore existing research and studies in several areas:

Diabetes Detection Techniques: Look into various methods and methods that have been employed for diabetes detection, encompassing, but not restricted to logistic regression. This could include algorithms for machine learning, statistical models, and traditional medical diagnostic approaches.

Logistic Regression in Medical Diagnosis: Investigate how logistic regression has been applied in medical diagnosis, especially in light of diabetes detection. Examine studies that have employed logistic regression for predicting diabetes risk factors or diagnosing diabetes based on certain criteria.

Comparison with Other Methods: Compare the performance of logistic regression with other algorithms or methods used for diabetes detection. This could include decision trees, ensemble techniques, neural networks, and support vector machines. Highlight the advantages and limitations of logistic regression compared to these approaches.

Feature Selection and Engineering: Explore how feature selection and feature engineering work have been done in previous studies related to diabetes detection. Identify relevant features and variables used in models of logistic regression for the prediction of diabetes.

Datasets Used: Discuss the datasets that have been utilized in previous research for training and testing diabetes detection models based on logistic regression. Evaluate the characteristics of these datasets, including size, diversity, and representativeness.

Performance Evaluation Metrics: Review the performance evaluation metrics utilized to evaluate the efficacy of logistic regression models in the identification of diabetes. This could consist of sensitivity, specificity, accuracy, F1 score, area under the ROC curve, and other relevant metrics. Clinical Relevance and Interpretability: Consider the clinical relevance and interpretability of various models for logistic regression within the framework of diabetes detection. Discuss how Healthcare practitioners can evaluate the findings of logistic regression analysis and use them to guide clinical decision-making.

Recent Advances and Future Directions: Highlight any recent advances or emerging trends in utilizing logistic regression for diabetes detection. Identify gaps in the existing literature and suggest some directions for further study in this field.

By thoroughly examining these aspects in your related work section, you can provide a comprehensive overview of the existing literature and establish the context for your research on diabetes detection using logistic regression.

III.METHODOLOGY

1. Data Collection

Gather data from various sources including medical records, surveys, or research studies. The dataset should contain relevant parameters such as age, weight, family history of diabetes, blood pressure, glucose levels, etc., along with a label indicating whether the individual has diabetes or not.

2. Data Preprocessing

Clean the data by handling missing values, outliers, and inconsistencies. Normalize or standardize numerical features to guarantee that every feature adds the same amount to the model. Encode categorical variables if necessary.

3. Feature Selection

Identify the most important elements influencing the likelihood of having diabetes. Feature selection techniques like correlation analysis, feature importance, or domain knowledge can help in selecting the right set of features. Train-Test Split: Divided the dataset into training and testing sets. Typically, around 70-80% of the data is used for training and the remaining for testing.

4. Model Training

Apply the logistic regression algorithm to the data used for training. One type of regression is called logistic regression binary classification algorithm that calculates the likelihood of a binary outcome (in this case, whether an individual has diabetes or not) based on one or more predictor variables.

5. Model Evaluation

Utilizing the testing data, assess the performance of the trained model. Typical metrics for evaluating binary classification include accuracy, precision, recall, F1-score, and ROC-AUC score.

6. Hyperparameter Tuning

Fine-tune the hyperparameters within the logistic regression model to optimize its performance. Techniques like grid search or random search can be Suitable for hyperparameter tuning**.**

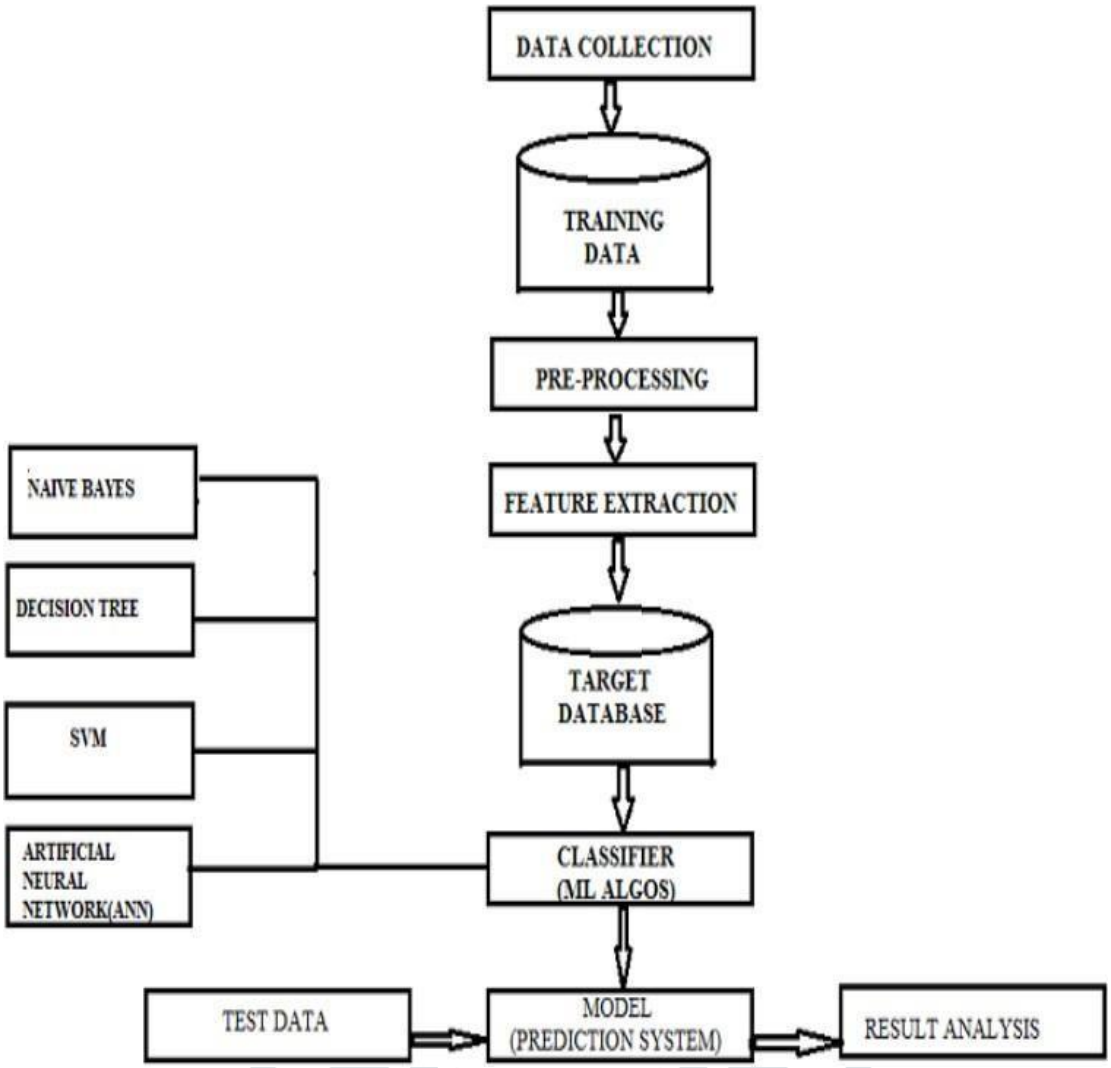
7. Validation

Validate the model's performance on unseen data using techniques like cross-validation to ensure its generalizability.

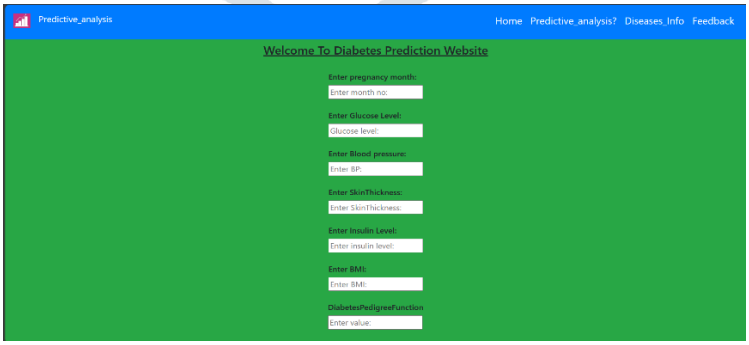
8. Interpretation

Interpret the coefficients of the logistic regression model, one can comprehend the effects of each feature on the likelihood of having diabetes. Positive coefficients indicate a positive association with diabetes, while negative coefficients indicate a negative association.

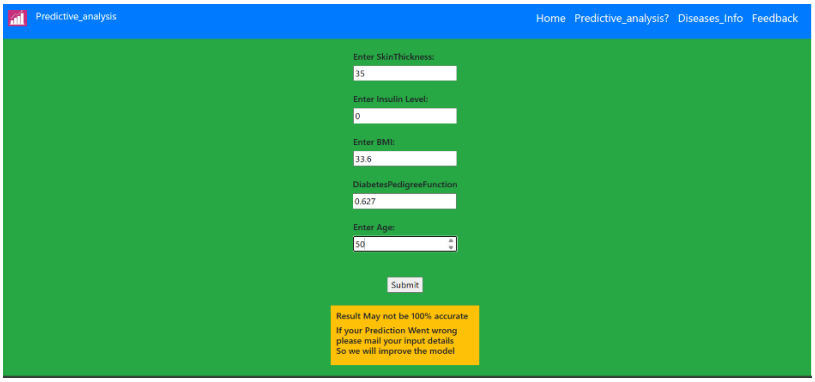
IV.FLOW CHART



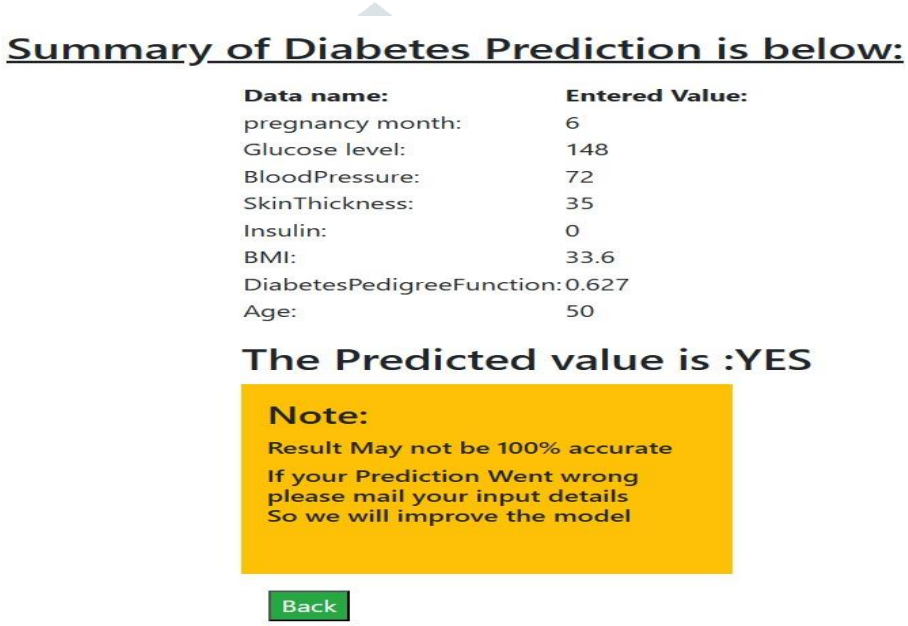
V.RESULT AND DISCUSSION



Upon opening the website, you'll see the initial screen. There after, you can enter values for the specified parameters.



In the above screen, we can see once you've entered the values for the specified parameters, proceed to click the submit button.



After clicking the submit button, you'll be able to view all the entered values along with the predicted outcome indicating whether the person is diabetic or not.

VI. CONCLUSION

Using logistic regression to detect diabetes has several benefits, such as its simplicity, interpretability, and ability to handle binary classification tasks effectively. Through the review of related work, it becomes evident that the use of logistic regression has been widely utilized in medical diagnosis, including the detection of diabetes. Existing Research has indicated the effectiveness of logistic regression models in accurately predicting diabetes based on various risk factors and clinical indicators. While logistic regression may not always outperform more complex machine learning algorithms, it remains a valuable tool, particularly in scenarios where transparency and interpretability are crucial. However, it's essential to acknowledge the limitations of logistic regression, such as its assumption of linear relationships between predictors and the log odds of the outcome. Additionally, feature selection and engineering play a crucial part in the execution of logistic regression models, requiring careful consideration and domain expertise. Looking ahead, further research in diabetes detection using logistic regression could focus on refining feature selection techniques, exploring novel datasets, and integrating additional clinical variables to enhance predictive accuracy. Moreover, the integration of logistic regression using other machine learning approaches, such as ensemble methods or deep learning may enhance performance while maintaining interpretability. Overall, while logistic regression serves as a valuable tool in diabetes detection, ongoing research efforts are needed to advance its capabilities and address the evolving challenges in medical diagnosis. Through continued innovation and cooperation between scholars and healthcare professionals, logistic regression can continue to contribute significantly to the early identification and management of diabetes.

VII. ACKNOWLEDGMENT

The Author is grateful to the CMR College of Engineering & Technology for providing better facilities and practical requirements.

References

[1] B. Wu, W. Zhu, F. Shi, S. Zhu, and X. Chen, "Automatic detection of microaneurysms in retinal fundus images," Computerized Medical Imaging and Graphics, vol. 55, pp. 106–112, 2017.

[2] VijiyaKumar, K., Lavanya, B., Nirmala, I., Caroline, S.S.: Random forest algorithm for the prediction of diabetes. In: International Conference on System, Computation, Automation and Networking, pp. 1–5 (2019).

[3] Mohan, N., Jain, V.: Performance analysis of support vector machine in diabetes prediction. In: International Conference on Electronics, Communication and Aerospace Technology, pp. 1–3 (2020) [Google Scholar]

[4] Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., Johannes, R.S.: Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In: Annual Symposium on Computer Applications in Medical Care pp. 261–265 (1998) [Google Scholar]

[5] Aurélien, G.: Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc., Sebastopol, CA [Google Scholar]

[6] Hasan, M.K. , Alam, M.A. , Das, D. , Hossain,

E., Hasan, M.: Diabetes prediction using ensembling of different machine learning classifiers. IEEE Access 8, 76516–76531, (2020) [Google Scholar]

[7] M. Gandhi and R. Dhanasekaran, "Diagnosis of diabetic retinopathy using morphological process and SVM classifier," in Communications and Signal Processing (ICCSP), 2013 International Conference on. IEEE, 2013, pp. 873–877.