Sugar Sense: Machine Learning for Predicting Diabetes Risks

[1]Ch. Sai Siva Durga, [2]S V V Lakshman Kumar, [3]J Prasanna Anjaneyulu, [4]Y Chetan, [5]Dwarakalal
[1] *Assistant professor,*
[2,3,4,5] *Student*

Abstract- Diabetes is a prevalent health concern affecting individuals worldwide. Identifying the risk factors associated with diabetes and predicting its onset can significantly improve preventive interventions and management strategies. In this paper, we explore the application of machine learning techniques to predict diabetes risk using everyday language and accessible concepts.

In today's world diabetes is the major health challenges in India. It is a group of a syndrome that results in too much sugar in the blood. It is a protracted condition that affects the way the body mechanizes the blood sugar. Prevention and prediction of diabetes mellitus is increasingly gaining interest in medical sciences.

When we eat, our body turns food into sugars, or glucose. At that point, our pancreas is supposed to release insulin. Insulin serves as a key to open our cells, to allow the glucose to enter and allow us to use the glucose for energy. But with diabetes, this system does not work. Type 1 and type 2 diabetes are the most common forms of the disease, but there are also other kinds, such as gestational diabetes, which occurs during pregnancy, as well as other forms. Machine learning

is an emerging scientific field in data science dealing with the ways in which machines learn from experience.

The aim of this project is to develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by combining the results of different machine learning and deep learning techniques. The algorithms like K nearest neighbour, Decision tree, Multi layer perceptron are used. The accuracy of the model using each of the algorithms is calculated. Then the one with a good accuracy is taken as the model for predicting the diabetes.

INTRODUCTION

Diabetes Mellitus (DM), commonly referred to as diabetes, is a chronic metabolic disorder characterized by elevated blood sugar levels over a prolonged period. It poses a significant global health challenge, affecting millions of individuals worldwide and causing severe complications if left untreated. Accordingto the World Health Organization (WHO), the prevalence of diabetes has been steadily increasing over the past few decades, with an estimated 422 million adults living with diabetes in 2014, and this number is projected to rise to 642 million by 2040 if effective interventions are not implemented.

Diabetes is a long lasting disease and its affects people worldwide. It happens when the body is not capable of producing enough insulin.

Insulin which is secreted by pancreas, one of the most important hormones in the body, which is needed to maintain the level of glucose.

It may produce the symptoms of frequent urination, increased thirst and hunger. Diabetes can be controlled with the help of insulin injections, a healthy diet and regular exercise. Diabetes can also leads to other disease such as blindness, blood pressure, heart disease, and kidney disease etc. There are four types of diabetes.

Type 1 Diabetes: - Insulin isn't delivered sufficiently by pancrers, then type-1 diabetes may take place in body. It may arise any stage of life. e.g.: kids, youngsters [2].

Type 2 Diabetes: - Insulin doesn't in adequate amount and it is not sufficient for body need, then type-2 diabetes arises. Because of parent's inheritance, seniority, corpulence expands the danger of getting type diabetes. For the most part happens at 40 years old.

Gestational Diabetes: -It is the third principle shape, significantly happens with the pregnant ladies because of abundance glucose equal level in the body.

Pregestational Diabetes: It is another form of diabetes and it occurs when insulin-subordinate diabetes earlier getting to be pregnant [2].

Advancements in machine learning (ML) and deep learning (DL) present promising opportunities to revolutionize diabetes management by enabling the development of predictive models that can forecast blood sugar levels based on various input parameters. These models, when integrated into wearable devices or mobile applications, have the potential to provide real-time feedback and personalized recommendations to individuals with diabetes, empowering them to make informed decisions about their lifestyle, diet, and medication regimen.

Sugar Sense is a project aimed at harnessing the power of ML and DL techniques to create an innovative solution for predicting blood sugar levels in individuals with diabetes. By leveraging large datasets of blood glucose measurements, along with relevant physiological and contextual factors such as diet, physical activity, sleep patterns, and stress levels, Sugar Sense aims to develop accurate and reliable predictive models that can anticipate blood sugar fluctuations with high accuracy.The primary objective of Sugar Sense is to improve the quality of life for individuals with diabetes by offering them a non-invasive, convenient, and proactive approach to managing their condition. By providing timely insights and actionable recommendations, Sugar Sense strives to empower individuals to take control of their health, minimize the risk of complications, and enhance overall well-being.

LITERATURE SURVEY

Utilizing Machine Learning for Comprehensive Analysis and Predictive Modelling of Diabetes Risks

1)Analysis of a Population of Diabetic Patients Databases in WeakTool

AUTHORS: Yashoda and M. Kannan

The classification on diverse types of datasets that can be accomplished to decide if a person is diabetic or not. The diabetic patient's data set is established by gathering data from hospital warehouse which contains two hundred and forty-nine instances with seven attributes. These instances of this dataset are referring to two groups i.e. blood tests and urine tests. In this study the implementation can be done by using WEKA to classify the data and the data is assessed by means of 10-fold cross validation approach, as it performs very well on small datasets, and the outcomes are compared. The naïve Bayes, J48, REP Tree and Random Tree are used. It was concluded that J48 works best showing an accuracy of 60.2% among others.

2)Accuracy, Sensitivity and Specificity Measurement of Various Classification Techniques on Healthcare Data

AUTHORS: N. Niyati Gupta, A. Rawal, and V. Narasimhan

This aims to find and calculate the accuracy, sensitivity and specificity percentage of numerous classification methods and also tried to compare and analyze the results of several classification methods in WEKA, the study compares the performance of same classifiers when implemented on some other tools which includes RapidMiner and Matla busing the same parameters (i.e. accuracy, sensitivity and specificity). They applied JRIP, Jgraft and Bayes Net algorithms. The result shows that Jgraft shows highest accuracy i.e. 81.3%, sensitivity is 59.7% and specificity is 81.4%. It was also concluded that WEKA works best than MATLAB and RapidMiner.

3)Big Data Analytics Predicting Risk of Re-admissions of Diabetic Patients

AUTHORS: Salian and G. Hari Sekaran

Big Data analytics have been applied to evaluate the risk of readmission for diabetes patients. Predictive modeling has been employed by applying decision tree classification method. It has been observed that chance of readmission in diabetic patient is successfully predicted using the above analysis. Many analysis methods can be explored to improve the accuracy of the existing system.

The procedure involves integration of numerous factors such as clinical factors, socio-demographic factors, health conditions, disease parameters, hospitality quality parameters and various other parameters that can be specific to requirement of each individual health provider. The aim of this project is to determine the risk predictors that can cause

readmission among diabetic patients and detailed analysis has been performed to predict risk of readmission of diabetic patients.

4)Resampling Methods Improve the Predictive Power of Modeling in Class-Imbalanced Datasets
AUTHORS: P. Lee
In recent years, focus on applying a decision tree algorithm named as CART on the diabetes dataset after applying the resample filter over the data. The author emphasis on the class imbalance problem and the need to handle this problem before applying any algorithm to achieve better accuracy rates. The class imbalance is a mostly occur in a IJSER International Journal of Scientific & Engineering Research Volume 8, Issue 5, May-2017 1540 ISSN 2229-5518 IJSER © 2017 http://www.ijser.org dataset having dichotomous values, which means that the class variable have two possible outcomes and can be handled easily if observed earlier in data preprocessing stage and will help in boosting the accuracy of the predictive model. The study illustrates the effect of resampling means in field of medical the dataset used in this study was acquired from the National Health and Nutrition Examination Survey.
In this paper, Classification and Regression Tree (CART) being a recursive partitioning method aim at excruciating the data into different parts based on the maximum significant exposure variables carried out by this method. The tool used for experimentation is R software. The data was splinted into ratio of 70:30. Finally the maximum accuracy achieved by this study is 67%.

EXISTING SYSTEM

Diabetes Mellitus (DM) is commonly referred as Diabetes; it is a common chronic disease and poses a great threat to human health.
Diabetes is a long lasting disease and its affects people worldwide. It happens when the body is not capable of producing enough insulin.
Insulin which is secreted by pancreas, one of the most important hormones in the body, which is needed to maintain the level of glucose.
It may produce the symptoms of frequent urination, increased thirst and hunger. Diabetes can be controlled with the help of insulin injections, a healthy diet and regular exercise. Diabetes can also leads to other disease such as blindness, blood pressure, heart disease, and kidney disease etc. There are four types of diabetes.
Type 1 Diabetes: - Insulin isn't delivered sufficiently by pancrers, then type-1 diabetes may take place in body. It may arise any stage of life. e.g.: kids, youngsters [2].
Type 2 Diabetes: - Insulin doesn't in adequate amount and it is not sufficient for body need, then type-2 diabetes arises. Because of parent's inheritance, seniority, corpulence expands the danger of getting type diabetes. For the most part happens at 40 years old.
Gestational Diabetes: -It is the third principle shape, significantly happens with the pregnant ladies because of abundance glucose equal level in the body.
Pregestational Diabetes: It is another form of diabetes and it occurs when insulin-subordinate diabetes earlier getting to be pregnant [2].

Disadvantages

- Accuracy Difference
- Data Dependency.
- Model and Algorithm Interpretability.
- Generalization.

1. Limited Data Feature.

PROPOSED SYSTEM

The proposed system for predicting diabetes integrates a range of machine learning algorithms and deep learning methods including K-Nearest Neighbors (KNN) and Decision Trees, alongside deep learning methods such as Neural Networks.By leveraging these diverse techniques, the system aims to enhance accuracy and reliability in diabetes prediction. Additionally, the integration of Standard Scalar normalization ensures optimal preprocessing of the data, contributing to improved model performance.
This multifaceted approach allows for a comprehensive analysis of various features and patterns within the data, ultimately leading to more precise predictions of diabetes risk.
Through the fusion of traditional machine learning and cutting-edge deep learning methods, the proposed system endeavors to advance the field of diabetes

prediction, offering potentially groundbreaking insights for early detection and effective management strategies.

Advantages

- Accuracy: : It was seen that on this training set the Neural Networks algorithm provided the most accurate results.
- Personalized Medicine: ML can personalize risk assessments based on individual characteristics, lifestyle factors, and medical history, enabling tailored interventions.
- Scalability: Once trained, ML models can be deployed at scale to screen large populations for diabetes risk efficiently.
- Continuous Improvement: ML models can be continuously refined and updated with new data, improving their predictive capabilities over time.
  - By carefully considering these pros and a project can effectively leverage machine learning for predicting diabetes risks while addressing potential challenges and limitations.

CONCLUSION

In conclusion, the development and deployment of machine learning models for diabetes risk prediction require a comprehensive approach that encompasses data collection, preprocessing, feature selection, model training, evaluation, and deployment. By leveraging advanced algorithms such as KNN, decision trees, MLP, support vector machines, and deep learning architectures, researchers and healthcare practitioners can build robust predictive models capable of accurately assessing an individual's likelihood of developing diabetes.

It was seen that on this training set the Neural Networks algorithm provided the most accurate results.

FUTURE SCOPE

In future by using this project we can do early prediction of the diabetes risks by using more accurate information available and training the model using MLP Classifier

Additionally, integrating wearable devices and real-time monitoring could enhance early detection and intervention, leading to improved health outcomes. Moreover, collaborations with healthcare providers and researchers can further refine these models and contribute to preventive healthcare strategies.

REFERENCE

[1] Al-Majeed, Salah H., et al. "Deep learning techniques for diabetes disease prediction: A comprehensive review." IEEE Access 9 (2021): 55796-55813.

[2] Bansal, Aditi, and Ankush Mittal. "A review on application of machine learning techniques in diabetes prediction. "International Journal of Advanced Research in Computer Science 8.5 (2017): 146-149

[3] Chatterjee, Sayak, et al. "A survey on diabetes prediction using machine learning techniques." Procedia Computer Science 167 (2020): 1395-1405.

[4] Patel, Tejas B., et al. "Deep learning approach for predicting diabetes risk using support vector machine." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 8.8 (2019): 412-417.

[5] Rajalakshmi, P., and Dr R. Sivakumar. "A survey on machine learning approaches for diabetes prediction." International Journal of Computer Applications 138.1 (2016): 26-33.

[6] Singh, Ashutosh, et al. "Diabetes prediction using machine learning algorithms." International Journal of Computer Applications 139.18 (2016): 22-29.

[7] Tripathy, Bikash C., and Ram Bilas Pachori. "Machine learning techniques for diabetes detection and prediction." In 2020 IEEE International Conference on Power, Electrical, and Electronics & Communication Engineering (PEECE), pp. 1-5. IEEE, 2020.