IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

DIABETES PREDICTION USING MACHINE LEARNING

N. Viswanadha Reddy*,[1] P. Manoj Kumar, [2]V. Sri Rama Raju, [3]G. Sai Dharma, [4]D. Pavan Kalyan

*Assistant professor, Department of CSE (AIML)

[1]Student, Department of CSE(AIML)

Nadimpalli Satyanarayana Raju Institute of Technology, Visakhapatnam, AP, India

Abstract: Diabetes is an illness caused because of high glucose level in a human body. Diabetes should not be ignored if it is untreated then Diabetes may cause some major issues in a person like: heart related problems, kidney problem, blood pressure, eye damage and it can also affect other organs of human body. 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 K-Nearest Neighbor (KNN), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Gradient Boosting (GB) and Random Forest (RF). 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.

INTRODUCTION

Diabetes is a chronic condition that affects the body's ability to process blood sugar (glucose) properly. It is a major public health concern because of its high prevalence and the associated health complications. Machine learning is a type of artificial intelligence that involves training algorithms on large datasets to make predictions or take actions based on new data inputs. In the context of diabetes prediction, machine learning algorithms can be trained on existing patient data to identify patterns and risk factors that may be indicative of an individual's likelihood of developing the condition. This information can then be used to predict whether a person is at risk of developing diabetes, allowing for early intervention and potentially preventing the development of the condition. There are several different types of machine learning algorithms that can be used for diabetes prediction, including decision trees, random forests, and support vector machines. These algorithms can be trained on various types of data, including medical records, lifestyle and behavioral information, and genetic data. The accuracy of the predictions made by these algorithms can be improved by using larger and more diverse datasets, as well as by incorporating additional information such as family history and medical history.

NEED OF THE STUDY.

The main objective of diabetes prediction is to identify individuals who are at high risk of developing diabetes and to take appropriate action to prevent or delay the onset of the disease. This can include lifestyle changes, such as improving diet and increasing physical activity, or medical interventions, such as medication or blood sugar monitoring. By identifying individuals at high risk of developing diabetes, healthcare providers can take steps to prevent or delay the onset of the disease, which can improve the individual's health and reduce the burden of diabetes on society.
This experiment helps to detect the disease of the diabetes before it affects fully and as per the instructions, we can take care to avoid the disease. So, this helps to improve the identification and management of individuals at high risk of developing the disease.

Data and Sources of Data

In this study we used the PIMA Indian Diabetes (PID) dataset taken from the National Institute of Diabetes and kidney diseases center. Primary objective of using this dataset is to build an intelligent model that can predict whether a person has diabetes or not, using some measurements included in the dataset. are eight medical predictor variables and one target variable in the dataset. Diabetes classification and prediction are a binary classification problem. Details of the variables are shown dataset consists of 768 records of different healthy and diabetic female patients of age greater than twenty-one. Target variable outcome contains only two values, 0 and 1. Primary objective of using this dataset was to predict diabetes diagnostically. Whether a user has a chance of diabetes in the coming four years in women belongs to PIMA Indian. dataset has a total of eight variables: glucose tolerance, no. of pregnancies, body mass index, blood pressure, age, insulin, and Diabetes Pedigree Function. All eight attributes are used for the training classification model in this work.

RESEARCH METHODOLOGY

The research methodology for diabetes prediction using machine learning would depend on the specific goals and objectives of the study. In general, a machine learning study on diabetes prediction would involve the following steps:

- Identify the research question and define the goals and objectives of the study.
- Collect and preprocess the data. This typically involves cleaning the data to remove any errors or inconsistencies, and transforming the data into a form that can be used by the machine learning algorithm.
- Select and train a machine learning model. This involves choosing a suitable machine learning algorithm and using it to learn from the training data.
- Evaluate the performance of the trained model. This typically involves using a separate set of data (called the test data) to assess the accuracy and reliability of the model's predictions.
- Fine-tune the model, if necessary. If the initial performance of the model is not satisfactory, the model can be adjusted (or "tuned") to improve its performance.
- Apply the trained model to make predictions on new data. Once the model has been trained and evaluated, it can be used to make predictions on previously unseen data.

Statistical tools and econometric models

Statistical tools and econometric models are commonly used in the field of diabetes prediction. Statistical tools such as regression analysis can be used to identify relationships between different variables and to make predictions about the likelihood of an individual developing diabetes. Econometric models, on the other hand, are mathematical models that are used to analyze economic data and to make predictions about economic phenomena.

In the context of diabetes prediction, statistical tools and econometric models can be used to analyze data on factors that are known to influence the development of diabetes, such as age, gender, body mass index (BMI), and family history of diabetes. By analyzing this data, it is possible to identify patterns and trends that can be used to make predictions about the likelihood of an individual developing diabetes.

For example, a regression analysis could be used to identify the relationship between BMI and the likelihood of developing diabetes, while an econometric model could be used to analyze the relationship between healthcare spending and the prevalence of diabetes in a population. By using these tools and models, it is possible to gain a better understanding of the factors that influence the development of diabetes and to make more accurate predictions about the likelihood of an individual developing the disease.

Comparison of the Models

The research methodology for diabetes prediction using machine learning would depend on the specific goals and objectives of the study. In general, a machine learning study on diabetes prediction would involve the following steps:

Step1:

Identify the research question and define the goals and objectives of the study.

Step2:

Collect and pre-process the data. This typically involves cleaning the data to remove any errors or inconsistencies, and transforming the data into a form that can be used by the machine learning algorithm.

Step3:

Select and train a machine learning model. This involves choosing a suitable machine learning algorithm and using it to learn from the training data.

Step4:

Evaluate the performance of the trained model. This typically involves using a separate set of data (called the test data) to assess the accuracy and reliability of the model's predictions.

Apply the trained model to make predictions on new data. Once the model has been trained and evaluated, it can be used to make predictions on previously unseen data.
It is important to note that the steps described above are only a general outline of the research methodology for diabetes prediction using machine learning. The specific details of the study, such as the type of data used and the specific machine learning algorithm employed, will vary depending on the specific research question and goals of the study.
Acknowledgments in a research study on diabetes prediction using machine learning would typically include any individuals or organizations that provided support or assistance with the study. This could include funding agencies that provided financial support, research institutions that provided facilities or resources, and colleagues or mentors who provided advice or guidance.

In addition to thanking these individuals and organizations, the acknowledgments section of a research study on diabetes prediction using machine learning could also include any other individuals or groups that contributed to the success of the study, such as study participants or data providers. Acknowledging these contributions is an important way to recognize the efforts and contributions of others, and to build relationships and collaborations for future research endeavors.
It is difficult to provide a comparison of different models for diabetes prediction without more context. There are many different machine learning algorithms that can be used for diabetes prediction, and the best choice of algorithm will depend on the specific characteristics of the data and the goals of the study. Some common algorithms that might be used for diabetes prediction include logistic regression, decision trees, random forests, and support vector machines.

In general, when comparing different models for diabetes prediction, it is important to consider the following factors:
Accuracy: The accuracy of the model refers to how well the model is able to predict the outcome of interest. A model with high accuracy is more likely to make correct predictions, while a model with low accuracy is more likely to make incorrect predictions.
Precision: The precision of the model refers to how consistently the model is able to make correct predictions. A model with high precision is more likely to make consistent predictions, while a model with low precision may produce inconsistent results.
Recall: The recall of the model refers to the percentage of positive cases that are correctly predicted by the model. A model with high recall is more likely to identify all of the positive cases in the data, while a model with low recall may miss some of the positive cases.
F1 score: The F1 score is a measure of the overall performance of the model that takes into account both the precision and the recall. A model with a high F1 score is considered to be performing well, while a model with a low F1 score may not be performing as well.
Training time: The training time of the model refers to the amount of time it takes to train the model on the data. A model with a low training time is more efficient and can be trained more quickly, while a model with a high training time may require more computational resources and take longer to train.
By considering these factors, it is possible to compare different models for diabetes prediction and determine which model is the most appropriate for a given study. It is important to note, however, that the relative performance of different models can vary depending on the specific data and the goals of the study, so it is always important to evaluate the performance of a model on the specific data and tasks at hand.

IV. RESULTS AND DISCUSSION

One of the important real world medical problems is the detection of diabetes at its early stage. In this study, systematic efforts are made in designing a system which results in the prediction of diabetes. During this work, five machine learning classification algorithms are studied and evaluated on various measures. Experiments are performed on john Diabetes Database. Experimental results determine the adequacy of the designed system with an achieved accuracy of 99% using Decision Tree algorithm.

In future, the designed system with the used machine learning classification algorithms can be used to predict or diagnose other diseases. The work can be extended and improved for the automation of diabetes analysis including some other machine learning algorithms.

ACKNOWLEDGMENT

Acknowledgments in a research study on diabetes prediction using machine learning would typically include any individuals or organizations that provided support or assistance with the study. This could include funding agencies that provided financial support, research institutions that provided facilities or resources, and colleagues or mentors who provided advice or guidance.

In addition to thanking these individuals and organizations, the acknowledgments section of a research study on diabetes prediction using machine learning could also include any other individuals or groups that contributed to the success of the study, such as study participants or data providers. Acknowledging these contributions is an important way to recognize the efforts and contributions of others, and to build relationships and collaborations for future research endeavors.

REFERENCES

[1] R. Williams, S. Karuranga, B. Malanda et al., "Global and regional estimates and projections of diabetes-related health expenditure: results from the international diabetes federation diabetes atlas," Diabetes Research and Clinical Practice, Vol. 162, Article ID 108072, 2020.

[2] American Diabetes Association, "Diagnosis and classification of diabetes mellitus," Diabetes Care, vol. 37, no. Supplement 1,pp. S81–S90, 2014.

[3] G. Acciaroli, M. Vettoretti, A. Facchinetti, and G. Sparacino, "Calibration of minimally invasive continuous glucose monitoring sensors: state-of-the-art and current perspec-tives," Biosensors, vol. 8, no. 1, 2018.

[4] N. N. Tun, G. Arunagirinatha, S. K. Munshi, and. M. Pappachan, "Diabetes mellitus and stroke: a clinicalupdate," World Journal of Diabetes, vol. 8, no. 6, 2017.

[5] M. J. Davies, D. A. D'Alessio, J. Fradkin et al., "Management of hyperglycaemia in type 2 diabetes, 2018. a consensus report by the American diabetes association (ada) and the European association for the study of diabetes (easd)," Diabetology, vol. 61, no. 12, pp. 2461–2498, 2018.

[6] D. Bruen, C. Delaney, L. Florea, and D. Diamond, "Glucose sensing for diabetes monitoring: recent developments," Sensors, vol. 17, no. 8, 2017.

[7] S. Wadhwa and K. Babber, "Artificial intelligence in healthcare: predictive analysis on diabetes using machine learning algorithms," in Proceeding of the International Conference on Computational Science and Its Applications, pp. 354–366, Springer, Cagliari, Italy, July 2020.

[8] P. Tedeschi and S. Sciancalepore, "Edge and fog computing in critical infrastructures: analysis, security threats, and research challenges," in Proceeding of the 2019 IEEE European Sym-posium on Security and Privacy Workshops (Euros PW),pp. 1–10, IEEE, Stockholm, Sweden, June 2019.

[9] M. V. D. Schaar, A. M. Alaa, A. Floto et al., "How artificial intelligence and machine learning can help healthcare systems respond to COVID-19," Machine Learning, vol. 110, no. 1,pp. 1–14, 2021.