Mr.M.Thirunavukkarasu, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 13, Issue 5, May 2023, pp. 121-125

RESEARCH ARTICLE OPEN ACCESS

Diabetes Prediction Using Machine Learning

Mr.M.Thirunavukkarasu[1], J.Venkata Sharan[2], K.Sai Nishank Reddy[3]
[1]*Assistant Professor, Department of CSE, SCSVMV University Kanchipuram, Tamil Nadu, India*
[2,3]*UG Student, Department of CSE, SCSVMV University Kanchipuram, Tamil Nadu, India*

ABSTRACT
A number of disorders, such as heart disease, kidney disease, stroke, visual problems, nerve damage, etc., are made more likely by diabetes. The aim of this project is to develop a system that can more precisely predict a patient's risk of developing diabetes. The major objective of this research has been to develop a support vector machine-based system. Research articles were examined, along with a variety of machine learning algorithms, their applications, and studies. KNN was used on the medical data set, and it produced results with greater accuracy than earlier methods. KNN also offered more accuracy than other methods. Science's field of machine learning studies how machines pick up knowledge via experience. By using machine learning, it is possible to build adaptable computers that can gain knowledge from their mistakes. Additionally, early illness prediction enables treating patients before their condition deteriorates. To avoid illnesses, early disease prediction is crucial in the medical industry. Our dietary habits in modern lifestyles are often high in sugar and fat, which has raised the risk of diabetes. Understanding the disease's symptoms is crucial for making predictions about it. Machine-learning (ML) techniques are useful right now for identifying diseases. One of the most deadly illnesses, diabetes mellitus, affects a large number of individuals. Diabetes mellitus can be brought on by ageing, obesity, inactivity, inherited diabetes, a poor diet, high blood pressure, and other reasons.

Keywords :- K Nearest Neighbors Algorithm , Support vector machine, Machine learning ,Blood Pressure ,Body Mass Index.

Date of Submission: 05-05-2023 Date of acceptance: 16-05-2023

I. INTRODUCTION

Diabetes is one of the most common ongoing chronic metabolic diseases worldwide. The two subtypes of diabetes are type-1 and type-2. Internal immune system injury to pancreatic beta cells (cells), which results in very little or no insulin production, causes type 1 diabetes to develop. Type 2 diabetes is an autoimmune disease that develops when the body's cells do not react to insulin or when insufficient insulin is produced by the pancreatic cells to regulate blood glucose levels. Inadequate insulin causes type 1 diabetes by raising blood glucose levels and disrupting the metabolism of proteins, carbohydrates, and lipids. Polyuria, Polydipsia, Weakness, Polyphagia, Obesity, Rapid Diabetes symptoms include: loss of weight, genital thrush, blurred vision, itchiness, irritability, delayed healing, partial paresis, muscle swelling, alopecia, etc. Diabetes is a metabolic condition that results in millions of fatalities worldwide each year as a result of many health issues. Worldwide, a rise in the fatality rate from diabetes of 70% has been noted between 2000 and 2019.

A powerful machine learning (ML) based diagnostic system is required to detect these fatal diseases. With an expert decision system based on ML, patients with diabetes may be efficiently detected at an early stage. To attempt to predict diabetes, researchers looked at a variety of different datasets. For ML-based systems, an appropriate dataset with the necessary properties for training and validation is needed. By selecting significant and relevant variables from the dataset, the ML model's capacity to accurately predict outcomes is boosted.

These actions also assist in lowering the risk of health issues and blood pressure regulation.

Diagnosing the diabetic illness is made simpler by routine medical examinations. To find the illness, several laboratory tests are also carried out. Patients with type-2 diabetes require insulin for as long as they are alive. Hence, if this bad scenario is disregarded, resources will be depleted for individuals, families, and the entire nation. To live healthy lives, people with pre-diabetes must get symptomatic treatment and early detection. An intelligent medical diagnostic system based on symptoms, signs, laboratory tests, and observations will aid in the detection and prevention of diseases. Artificial intelligence (AI) has been employed by

medical diagnosis systems in a number of fascinating ways to identify disorders.

II. LITERATURE SURVEY

Arianna Dagliati et al. conducted a survey on the application of machine learning algorithms to predict diabetes [5]. This paper demonstrates how computational tools may be used effectively. Models that use quiet explicit data are used in clinical medicine to predict outcomes of interest. It was shown that glucose is the root hub, meaning that the glucose has the most data collected, using decision tree topologies and neural networks. The clinical diagnosis and common sense were backed by this. It examined the relationship between being receptive to various components and the risk of starting a particular entanglement, separated the patient population into a clinical focus for this risk, and created tools to support clinically informed treatment decisions.

In [9], Pisapia et al. used machine learning and image analysis to forecast hydrocephalus. The cerebral ventriculomegaly was employed, and 77 imaging characteristics were retrieved. Support vector machines, a machine learning method, was used to analyse the ventricular characteristics of 25 kids. Who need shunts and who did not was the key question. The outcomes were collected and contrasted. Findings indicate that every third kid needs shunts, with a sensitivity and specificity of 75% and 95%, respectively. In [10], a brand-new fuzzy rule-based classifier is presented. For the purpose of applying analytics and cluster creation, algorithms based on expectation maximisation and fuzzy-rule base classifier are created. Results were evaluated based on accuracy, reaction time, false positive rate, and calculation cost. Current practises and the suggested framework were contrasted.

In [11], Das et al. investigated the dengue and malaria cases in Delhi, India, and conducted a prediction analysis on the data. Simi et al. [12] explored the importance of early detection of female infertility. In their study, the authors used 26 criteria and 8 classes of female infertility; the findings revealed that the Random Forest methodology performed better than other methods and had an accuracy rate of 88%. An intelligent recommender system was proposed by Lafta et al. in [13] to assist people and medical professionals in calculating the short-term risk of developing heart failure. A model for forecasting cardiac sickness was proposed by scientists; depending on the results, the system also gives people advice on the importance of being checked and visiting a doctor.

Juyoung Lee and associates [15] discussed developing a type 2 diabetes predicting model. 1) In this study, clinical and genetic data were used. 2) assessed the misclassification rates of various models, including the Support Vector Machine, KNearest Neighbor, Quest (Quick, Unbiased,

Efficient, Statistical Tree), and Logistic Regression. The quantifiable Goal computation using body mass index and SNPvariables produced the fewest misclassifications when the models were tested, but overall, the strategic regression delivered the best results. As compared to statistical tree algorithms, the logistic algorithm was found to have reduced rates of misclassification. To incorporate genetic data classifiers, better research methodologies required to be created.

Using the Ensemble technique, Rahul Joshi and MinyechilAlehegn[16] have examined and forecast diabetic illnesses. The PIDD (Pima Indian Diabetes DataSet) data collection was utilised in this work to categorise and predict symptoms of diabetes using prediction algorithms including KNN, Random forest, J48, and Naive Bayes.

III. EXISTING SYSTEM

Over the past ten years, the number of diabetics has dramatically increased. The major cause of the rise in diabetes is the way people live nowadays. There are three common types of errors that might happen throughout the current medical diagnosis procedure. Drawbacks of Existing System: -

- A false-negative result occurs when a patient's test results show they do not have diabetes when in fact they do.

IV. PROPOSED SYSTEM

Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, and Age make up the dataset utilised for the study. A method for diabetes prediction using machine learning approaches is being developed

Advantages:

- It will deliver precise findings.
- For diabetes prediction, we have integrated a number of machine learning methods to prevent erroneous findings.

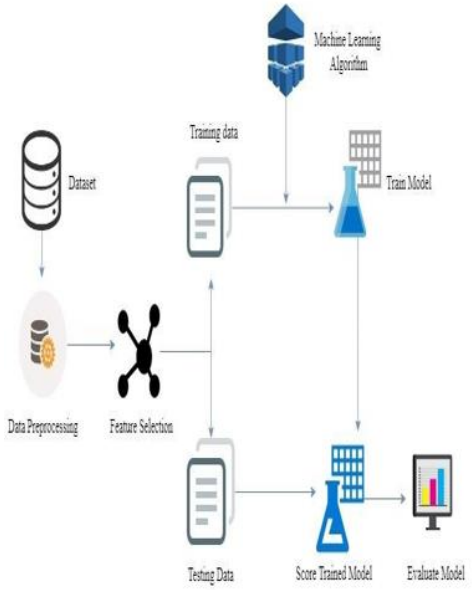
V. METHODOLOGY

A web application that has a Python script running in the background is used for the implementation. The front-end, or the web application, is created using Materialize CSS and HTML. The scripts for the front end and back end are connected using the Flask API. Google Collaboratory Notebooks, an online platform for developing ML models, was used to train the machine learning model. For more effective outcomes, we selected the SVM method.

Fig 1:-System Architecture.

- The second false-positive kind is. In this instance, test results falsely claim that the Step 5: Train-test separation.

Step 6: Discovering the dataset's accuracy score. Nevertheless, when the distribution of classes is patient has diabetes when they don't.



Step 6: Put x in class I if ki > kj I j

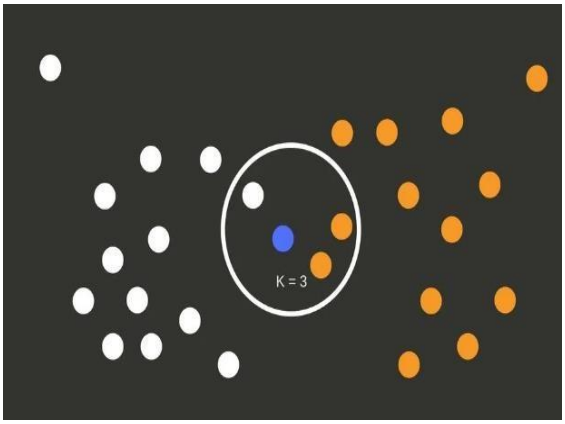


Fig 2 :- K Nearest neighbor Algorithm

VI. RESULTS

By using a number of Machine Learning algorithms on the dataset, we were able to achieve the accuracy displayed in the aforementioned figures and graphs. SVM offers the highest level of accuracy.

- The data you have prepared is now ready to be input to the machine learning model for training.
- Let's just use a very basic SVM on the model to demonstrate it.

unbalanced, overall accuracy might be deceptive, and it is crucial to accurately anticipate the minority class.
Step 7: For the entire model, our model had an accuracy of 0.7727272.
Step 8: This outcome appears to be quite positive.

Architecture

Step 1: The dataset's statistical measurements.
Step 2: Print the dataset's first five rows and count the number of rows and columns.
Step 3: Distinguishing data from labels.
Step 4: Data standardization

Algorithm:-

• Making a predictive system.
Step 1: Calculate "d(x, xi)" where I = 1, 2,..., n and d is the Euclidean distance between the points.
Step 2: Rather than descending, list the calculated n Euclidean distances.
Step 3: Let k be a positive integer and use the top k distances from this sorted list.
Step 4:. Find the k-points to which these k-distances correspond by using them.
Step 5: With a total of k points, let ki represent the number of points in the ith class, where k is 0.

Fig 3: Accuracy score of a test dataset

Fig 4:- The person diabetic or non diabetic

- Here, we created a mechanism to determine if a person has diabetes or not.

Fig-5 Values given for Diabetes prediction

VII. CONCLUSION

Diabetes is accompanied with a wide range of illnesses. One of its distinguishing features is chronic blood glucose elevation. We hope to apply predictive analysis to find and stop diabetes complications before they become serious by improving the classification algorithms. Moreover, our proposed study analyses the features in the dataset, and the best features are picked based on correlation values. Of all algorithms, the support vector machine provides the highest specificity and accuracy, making it the ideal choice for the analysis of diabetes data. Many models have been tried to predict diabetes, but researchers have always been more interested in how well the recommended models can predict illnesses.

Fig-6 Results for given values

REFERENCES

[1]. H. Zheng, H. W. Park, and K“Analysis and Prediction of Diabetes Using Machine Learning by S.Saru,S.Subashree ::SSRN.”
[2]. https://papers.ssrn.com/sol3/papers.cfm?abstract _id= 3368308 (accessed Oct. 22, 2021)
[3]. H. Ryu, “An efficient association rule mining method to predict diabetes mellitus: KNHANES
[4]. 2013–2015” in Advances in Intelligent Information Hiding and Multimedia Signal Processing, Cham, Switzerland: Springer, pp. 241-249, 2020.
[5]. J. M. Norris, R. K. Johnson and L. C. Stene, "Type 1 diabetes—Early life origins and changing epidemiology", Lancet Diabetes Endocrinol., vol. 8, no. 3, pp. 226-238, Mar. 2020
[6]. Classification and diagnosis of diabetes: Standards of medical care in diabetes— 2020", Diabetes Care, vol. 43, no. 1, pp. S14-S31, Jan. 2020
[7]. A. Jindal, A. Dua, N. Kumar, A. K. Das, A. V.
[8]. Vasilakos, and J. J. P. C. Rodrigues, “Providing Healthcare-as-a-Service Using Fuzzy Rule-Based
[9]. Big Data Analytics in Cloud Computing,” IEEE J. Biomed. Heal. Informatics, pp. 1–1, 2018. F. G. Woldemichael and S. Menaria, "Prediction of diabetes using data mining techniques", Proc. 2nd Int. Conf. Trends Electron. Informat. (ICOEI), pp. 414-418, May 2018.
[10]. J. B. Buse, D. J. Wexler, A. Tsapas, P. Rossing, G. Mingrone, C. Mathieu, et al., "2019 update to Management of hyperglycemia in type 2 diabetes 2018. A consensus report by the

American diabetes association (ADA) and the European association for the study of diabetes (EASD)", Diabetologia, vol. 63, no. 2, pp. 221-228, Feb. 2020.

[11]. Arianna Dagliati, Simone Marini, Lucia Sacchi, Giulia Cogni 2017, "Machine Learning Methods to Predict Diabetes Complications", Journal of diabetes science and technology 12(3):193229681770637.

[12]. S. Rouhani and M. MirSharif, "Data mining approach for the early risk assessment of gestational diabetes mellitus", Int. J. Knowl. Discovery Bioinf., vol. 8, no. 1, pp. 1-11, Jan. 2018.

[13]. E. W. Steyerberg, H. Uno, J. P. A. Ioannidis, B. van Calster, C. Ukaegbu, T. Dhingra, et al., "Poor performance of clinical prediction models: The harm of commonly applied methods", J. Clin. Epidemiol., vol. 98, pp. 133-143, Jun. 2018.

[14]. JuyoungLee, BhumsukKeam, EunJungJang, Mi SunPark 2011, "Development of a Predictive Model for Type 2 Diabetes Mellitus Using Genetic and Clinical Data", Osong Public Health and Research Perspectives Volume 2, Issue 2, ages 75-82.

[15]. S. Das and A. Thakral, "Predictive analysis of dengue and malaria," in 2016 International Conference on Computing, Communication and Automation (ICCCA), 2016, pp. 172–176.

[16]. M. S. Simi, K. S. Nayaki, M. Parameswaran, and S. Sivadasan, "Exploring female infertility using predictive analytic," in 2017 IEEE Global Humanitarian Technology Conference (GHTC), 2017, pp. 1–6.

[17]. R. Lafta, J. Zhang, X. Tao, Y. Li, and V. S.

[18]. Tseng, "An Intelligent Recommender System Based on Short-Term Risk Prediction for Heart Disease Patients," in 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2015, pp. 102–105.

[19]. S. T. Prasad, S. Sangavi, A. Deepa, F. Sairabanu, and R. Ragasudha, "Diabetic data analysis in big data with predictive method," in 2017 International Conference on Algorithms, Methodology, Models andApplications in Emerging Technologies (ICAMMAET), 2017, pp. 1–4.

[20]. JuyoungLee, BhumsukKeam, EunJungJang, Mi SunPark 2011, "Development of a Predictive Model for Type 2 Diabetes Mellitus Using Genetic and Clinical Data", Osong Public Health and Research Perspectives Volume 2, Issue 2, ages 75-82.