International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 6 ǁ 2022 ǁ PP. 1784-1787

Diabetes Disease Prediction Using Machine Learning Classifiers

Dr. D. Yugandhar [1]| Nakka Meghana[2]| Vavilapalli Gnana Prasanna [3]| Arasada Praveen Kumar [4]|Kota Gangadhar Rao

[1]*Professor and Associate Dean, Department of ECE, Aditya Institute of Technology and Management, Tekkali, Srikakulam, Andhra Pradesh, India*

[2,3,4 ,5]*student of Department of ECE, Aditya Institute of Technology and Management, Tekkali, Srikakulam, Andhra Pradesh, India*

ABSTRACT

Diabetes is an illness caused because of high blood glucose levels, insulin levels, BMI, blood pressure, uric acid, urine, plantar fasciitis, cholesterol, coronavirus diseasein a human body. Diabetes may cause some major issues in a person like blood pressure, eye damage, kidney problem and it is also affected on other organs of human body.If one can find this disease earlier, then can easily control it after being given the values for different attributes like glucose, BMI etc. By performing three machine learning classifiers, i.e., Random Forest, Decision tree and Logistic Regression on a dataset, an accuracy of 90% is obtained by the researchers.In this paper, by taking extra er parameters like uric acid, urine, plantar fasciitis, cholesterols and coronavirus disease severity,accuracy of finding diabetic disease prediction can be improved.The health care domain is one of the prominent research fields in the current scenario with the rapid improvement of technology. so, it is easy to handle the huge amount of data of the patient.

Keywords: *Random Forest, Decision Tree, Logistic Regression, Confusion matrix*

Date of Submission: 14-06-2022 Date of acceptance: 29-06-2022

I. INTRODUCTION

Diabetes is one of the highest percentages in growing countries. Diabetes is a disorder wherein blood sugar preserve to upward thrust due to a lack of insulin, which turned into impacted on metabolism. It isa noxious disorder. Diabetes takes place because of weight troubles also. It is a noxious disease. Some of the patients having diabetes due to over weight problem. It impacts the hormone insulin, ensuing in ordinary metabolism, and improves degree of sugar within side the blood. Diabetes attack lots of people worldwide and it is divided into two kinds diabetes. First diabetes is also called insulin-dependent diabetes is most frequently begins in childhood, second is also known as adult-onset diabetes or non-insulin dependent. According to World Health Organization (WHO) 442 billion people suffering from diabetes particularly from low-income countries [1]. And this could be increased to 490 billion up to the year of 2030. When it comes to glucose levels, Prediabetes occurred by having higher level of glucose than other different types of diabetes. The supervised learning algorithm is used for training the model, and evaluation is done using testing data. Diabetes is one of the important reasons for loss of life in developing countries. Diabetes is a disease in which blood sugar tiers maintain to rise due to a lack of insulin, which influences blood sugar metabolism. It is is a noxious disease. Diabetes was introduced approximately because of weight issues also. Diabetes influences the hormone insulin, resulting in metabolism, and improves of sugar inside facet the blood. First diabetes is likewise referred to as insulin-established diabetes is maximum regularly starts off evolved in childhood' 2nd is likewise called adult-onset diabetes or non-insulin established. According to World Health Organization 452 humans stricken by diabetes in particular from low-earnings countries. Prediabetes is characterised as having a better degree than different kinds of diabetes. The supervised learning is used in our project.

Machine learning:

Machine Learning is a machine of pc algorithms that may research from instance thru self-development without being explicitly coded with the aid of using a programmer[5].

The lifestyles of Machine Learning packages are easy and may be summarized withinside the following points:

1.Define a question
2.Collect data
3.Visualize data
4.Train set of rules
5.Test the Algorithm

6.Collect feedback
7.Refine the set of rules
8.Repeat step 4 to step 7
9. Use the version to make a prediction
Once the set of rules receives accurate at drawing the proper conclusions, it applies that expertise to new units of data.
1.Supervised learning:
A set of rules makes use of education records and remarks from human beings to examine the connection of given inputs to a given output. For instance, a practitioner can use advertising and marketing rate and climate forecast as enter records to expect the income of cans.You can use supervised studying whilst the output records are known. The rules will expect updated values[2].
There are classes of supervised studying:
Classification task
Regression task
2.Unsupervised mastering:
No labels are given to the mastering algorithm, leaving it on its personal to locate shape in its input. Unsupervised mastering may be an intention in itself (coming across hidden styles in data) or a way closer to an end (characteristic mastering).
3.Decision tree:
Highly interpretable class or regression version the splits data-function values into branches at selection node (e.g., if function isa shadeation, every viable shadeation becomenew branch) till a very last selection output is made [3].

4.Random woodland/Forest:
The set of guidelines is built upon a preference tree to beautify the accuracy drastically. Random woodland generates more easy choice bushes and makes use of the 'majority vote' method to determine which price we return. For the classification the very last prediction may be the only with the maximum vote; whilst for the regression task, the common prediction of all of the tree is the very last prediction [4].

II. RELATED WORK

For this future work it includes developing an Android application for the suggested hypothetical diabetes monitoring system, including the proposed categorization and prediction algorithms, and deploying it. Genetic algorithms, in conjunction with the suggested prediction mechanism, may be investigated for improved monitoring.Diabetes is considered an illness in which the glucose level with inside the blood gets extended because of its non-outreach to the cells present with inside the body. Blood glucose is shaped due to the meals you consume and is taken into consideration the primary supply of energy. However, elevated stage of blood glucose (additionally referred to as blood sugar) can result in many diabetic associated fitness problems. Similarly, the decreased level of blood sugar can also bring about hypoglycaemia or one-of-a-kind sugarproblems. To hold the most reliable glucose stage with inside the body, pancreatic endocrine hormones—Insulin and Glycogen play an essential role. Both of those hormones are secreted via way of means of islet cells gift with inside the pancreas in reaction to the blood sugar levels.

III. SYSTEM REQIREMENTS

HARDWARE REQUIREMENTS:
Monitor: 15'' LED
Input Devices: Keyboard, Mouse
RAM: 4 GB
SOFTWARE REQUIREMENTS:
system: Windows 10
Coding Language: Python.
Web Framework: Flask

IV. SYSTEM FLOWCHART

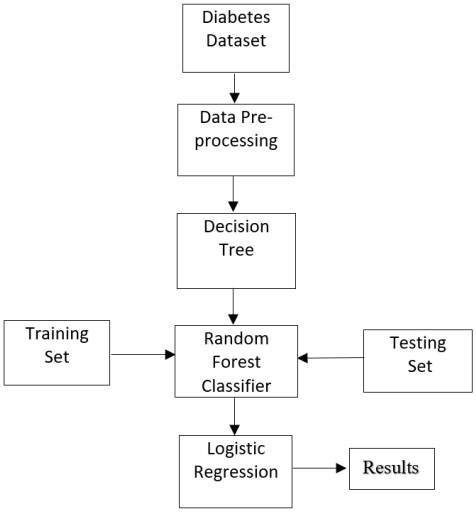

Fig.1: system architecture

Step1: Imported required libraries and diabetes dataset.
Step2: Pre-devised data to eliminate missing data.
Step3: In the whole dataset, 80% of the data taken for Training and 20% of the data set to Test.
Step4: Chosen Supervised Algorithm, Decision Tree, Random Forest and Logistic regression.
Step5: Tried to assemble a classifier that given the following set of symptoms.
Step6: Ensured use of Evaluation of the findings and shared instructions learned

V. INPUT DATA

In this data set we are taken columns and, in the dataset, which are described below.
Pregnancies: Number of times pregnant
Glucose: Plasma glucose concentration 2 hours in an oral glucose tolerance test
Blood Pressure: Diastolic blood pressure(mmHg)
Skin Thickness: Triceps skin fold thickness(mm)
Insulin:2-Hour serum insulin (mu U/ml)
BMI: Body mass index (weight / (height)^2)
Diabetes PedigreeFunction:Diabetes pedigree function
Age: Age (years)

VI. RESULT

PERFORMANCE ANALYSIS

Precision and recall

	Recall	Precision
Negative (0)	0.95	0.94
Positive (1)	0.88	0.89

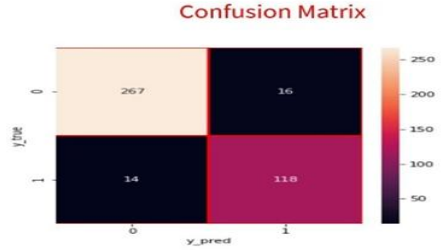

Fig. 2

PERFORMANCE ANALYSIS

Precision and recall

	Recall	Precision
Negative (0)	0.979	0.959
Positive (1)	0.909	0.952

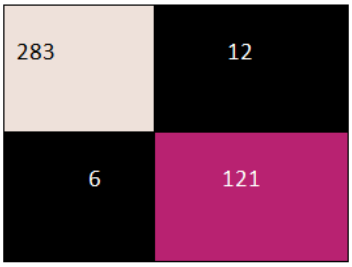

Fig. 3

VII. CONCLUSION

One of the huge impediments with the development of generation and medication is the early detection of a disease, that is on this case, diabetes. However, on this study, systematic efforts had been made into designing a version that is correct sufficient in figuring out the onset of the disease. With the experiments conducted on the Pima Indians Diabetes Dataset, we have readily predicted this disease. Moreover, the results achieved proved the adequacy of the system, with accuracies were shown in the Fig. 2 using the Random Forest Classifier. With this being said, it's far hopeful that we are able to put in force this version right into a gadget to are expecting different lethal sicknesses as well. By incorporating extraparameters,the results in the form of confusion matrix shown in Fig. 3 and the results were improved. Few more parameters to be incorporated so that it would result even more accuracy.

REFERENCES

[1]. Mir and S. N. Dhage, "Diabetes ailment prediction the usage of system studyingon huge recordsof healthcare," in 2018, Fourth InternationalConference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1–6.
[2]. D. Sisodia and D. S. Sisodia, "Prediction of diabetes the usage of type algorithms," Procedia Computer Science, vol. 132, pp. 1578 – 1585, 2018, global Conference on Computational Intelligence and Data Science.
[3]. P.S. Kohli and S. Arora," Application of machine learning in disease prediction," in 2018 4th International Conference on Computing communicationand automation,2018
[4]. smith, W. Knowler, and R. Johannes, "Using the Random Forest algorithm to forecast the onset of diabetes mellitus," Proceedings - Annual Symposium on Computer Applications in Medical Care, vol. 10
[5]. S. Yadav "Analysis of k-fold cross-validation over hold-out validation on enormous dataset fine classification," in 2016 IEEE sixth International Conference on Advanced Computing (IACC), 2016, pp.
[6]. F. Pedregosa, G. Varoquaux,A. Gram fort, V. Michel," Scikit-learn: Machine Learning in python,vol.12, no.85,p.28252830,2011.