IJITCE

International Journal of
Information Technology & Computer Engineering

www.ijitce.com

Diabetes Disease Prediction Using Machine Learning Algorithms

[1] P MOUNIKA, [2] V. ISAIAH

[1](Assistant Professor), **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

[2]MCA, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

ABSTRACT

Even though diabetes is one of the most prevalent illnesses globally, it is curable and preventable if caught in its early stages. Based on certain diagnostic metrics in the dataset, we create a model to predict whether a patient will acquire diabetes. We next investigate several methods to enhance the model's performance and accuracy. This article primarily makes use of logistic regression and conducts its analysis utilizing Python integrated development environments (IDEs). A dataset from Vanderbilt based on a study of rural African Americans in Virginia and the PIMA Indians Diabetes dataset from the National Institute of Diabetes and Digestive and Kidney Diseases are the primary data sources used in the experiment. There are two distinct approaches to function selection. To top it all off, we apply aggregation approaches, which boost speed by making more accurate predictions using only one model. The original datasets and datasets created later using feature selection and aggregation approaches are both documented in terms of accuracy and runtime. In addition, every example is accompanied with a comparison. For dataset 1, the highest accuracy achieved was about 78% when the aggregation approach Maximum Voting was used. For dataset 2, the highest accuracy was around 93% when the combined procedures of maximum polling and stacking were used. When it comes to developing predictive models, logistic regression is among the most successful algorithms.

1.INTRODUCTION

Diabetic, a condition characterized by elevated blood glucose levels due to insufficient or nonexistent insulin, is one of the most infamous illnesses that has recently swept the globe. Predicting and detecting this illness may be time-consuming and frustrating since there are so many factors that need to be considered for a person to be infected. The good news is that early detection is well within the realm of possibility. Federal - Israel Defense Forces. In nations where the median income was between 0 and 7 percent, 79 percent of the adults were residing. Diabetes will affect almost 700 million people by 2045, according to estimates (IDF).

Due to both hereditary and environmental causes, the prevalence of diabetes is steadily rising over the globe. There are a lot of reasons why these figures are quickly increasing, such as people eating unhealthy food and not getting enough exercise. In diabetes, a hormonal illness, a person's blood glucose levels rise because their body is unable to make insulin, which leads to improper sugar metabolism. Severe thirst, intense hunger, and the need to urinate often are among the noticeable symptoms. The condition is influenced by several risk factors, which include age, body mass index (BMI), glucose levels (GLUT), blood pressure (BP), and so on.

Every year, we see an increase in the number of cases, and the rate of new cases is also increasing. Concern about this matter is of the utmost importance, since diabetes has rapidly emerged as a leading cause of death worldwide.

The widespread availability of big data and the subsequent need to extract useful insights from it have contributed to machine learning's recent meteoric rise in popularity. While there are many other kinds of Machine Learning algorithms, the two most common are Non-Trained and Unlabeled Data: This is the Domain of Unsupervised Learning. To discover trends, if any, we simply put the data into action here.

In supervised learning, we use preexisting labels to train the model, and then we use those labels to assess how well the model performs on fresh data.

There have been major challenges with its identification in the past, but with the advent of Machine Learning and related techniques, these issues may be simplified while still producing thorough and precise results. It is

now known that Machine Learning along with the medical field has grown even more useful and successful. By analyzing a person's traits, machine learning may soon allow for the early diagnosis of a disease. Such preliminary efforts may lead to disease inhibition and prevent the illness from progressing to a critical stage. Early detection and treatment of diabetes is the goal of the study detailed in this article, which use machine learning algorithms to make such predictions.

2.LITERATURE SURVEY

Primary Stage of Diabetes Prediction using Machine Learning Approaches—IEEE International Conference on Artificial Intelligence and Smart Systems Issue Date: 12.April, 2021 [1]A Machine Learning Approach to Diabetes Prediction

[2] Thanishka 4, Dhanush Murthy 3, Viswanatha V1, and Ramachandra A.C. The author is an associate professor at India's Nitte Meenakshi Institute of Technology in the field of electrical and communication engineering in Bangalore. 2. Professor at the Nitte Meenakshi Institute of Technology in Bangalore, India, Department of Electronics and Communication Engineering 3,4Student

at India's Nitte Meenakshi Institute of Technology (EMIT) in Bangalore, studying electrical and communication engineering

1Referring to this work: Viswanatha V.

Here is the source: [1] https://www.researchgate.net/publication/350849420.
The second source is the following:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10107388/...
Review of Related Literature

Millions of individuals all around the globe deal with the long-term effects of diabetes. Treatment efficacy depends on prompt diagnosis. Diabetes prediction systems based on machine learning have been the subject of prior study. In order to diagnose diabetes, Singh et al. [4] used Random Forest, Multilayer Perceptron, and Naive Bayes. Researchers Kavakiotis et al. [5] looked at data mining techniques as they pertained to diabetes research. Diabetes was detected using an artificial neural network model by Jerjawi et al. [6]. Using ML algorithms, Maniruzzaman et al. [7] were able to forecast and categorize cases of diabetes.

Our research expands upon these previous efforts by zeroing in on diabetes risk assessment in its early stages. We preprocess data using a dataset obtained from Bangladesh's Sylhet Diabetes Hospital. Eleven classifiers, such as Support Vector Machine, Random Forest, and Logistic Regression, are trained and tested. Random Forest outperforms all other methods with an AUC of 98% and an accuracy of 98%. Important factors for diabetes prediction may be better identified according to this work.

To sum up, forecasting diabetes risk is greatly assisted by machine learning technologies. Feature selection methods and ensemble classifiers may be investigated in future studies to enhance performance.

Minhaz Uddin Emon, Maria Sultana Keyat, Md. Salman Kaiser, Md. Ariful Islam, Tabassum Tanha, and Md. Sabab Zulfiker's paper titled "Primary Stage of Diabetes Prediction using Machine Learning Approaches" describes how machine learning techniques can be used to predict the onset of diabetes in its early stages. Eleven distinct machine learning classifiers were used to construct prediction models in this work, which makes use of patient data

collected from Sylhet Diabetes Hospital in Bangladesh. After calculating the Area Under the Curve (AUC), the Random Forest classifier came out on top with a score of 98%.

An overview of diabetes as a common chronic illness is given in the paper's beginning. In order to effectively treat the condition and avoid consequences, early diagnosis is crucial. The study's overarching goal is to enhance patient outcomes by applying machine learning to make early diabetes predictions. Related efforts in the area of diabetes prediction using machine learning are also reviewed in the publication. Research on diabetes diagnosis using machine learning algorithms is cited. These algorithms include Random Forest, Multilayer Perceptron, Naive Bayes, and others. The review emphasizes the promising future of these methods for diabetes prediction and calls for further study into them.

Methodology details the steps used to gather data, prepare it for analysis, divide it into smaller pieces, tune hyperparameters, apply classifiers, and assess accuracy. Classifiers were trained using data collected from 520 patients, which included information on

symptoms and other variables. For all metrics measured, the Random Forest classifier performed best. This includes accuracy, area under the curve (AUC), precision, recall, and f1 score. This literature review summarizes the paper's main points, which center on the use of machine learning techniques for the prediction of diabetes in its early stages. It summarizes the study's relevance, discusses relevant literature, and describes the study's methodology and findings in depth.

3. EXISTING SYSTEM

They utilized the data analytics tool WEKA to forecast the occurrence of diabetes using healthcare Big Data. Various machine learning classifiers were applied to the publicly accessible dataset from UCI. They used the following classifiers: Naive Bayes, SVM, Random Forest, and Simple CART.

They began by gaining access to the dataset, which they then preprocessed using the Weka tool. They then used a 70:30 train/test split to apply several machine algorithms. They skipped the cross-validation stage, which is crucial for getting accurate and optimum findings.

In addition, the scientists conducted their experiment using the Pima Indians Diabetes Database, a publicly accessible dataset. Dataset selection and pre-processing are the first steps in their approach for making predictions. They used naive Bayes, support vector machine (SVM), and decision tree classification techniques after data preprocessing. They compared the various performance measures and examined the accuracy in a comparative manner because they used several assessment criteria. Their experiment yielded an accuracy of 76.30% at its peak. They have also not used cross-validation, as mentioned in [2].

Using the Indians Pima Diabetes Dataset, the scientists suggested a neural network for diabetes illness prediction. They anticipated the result by using patterns they discovered in the data, which they achieved by using many hidden layers. Adopting a proprietary neural network with numerous partitions and a set of association weights and units, they call their algorithms ADAP. With a sensitivity and specificity crossover point of 0.76, they are now attempting to refine their findings for future us

Disadvantages

1) There are no techniques and models for analyzing large scale datasets in the existing system.

2) Currently, we are unable to work with diabetic datasets in conjunction with hospitals or other medical institutions to improve our outcomes.

3.1 PROPOSED SYSTEM

We have used the Pima Indians Diabetes Database, a publically accessible dataset, to conduct our investigation. Several diabetes diagnostic metrics are part of this dataset. The National Institute of Diabetes and Digestive and Kidney Diseases collected the dataset in the first place. Every single case that has been documented involves individuals who are older than 21. The following figure shows the five stages that make up our suggested paradigm.

Advantages

➢ The system more effective due to fitting datasets for different ML Models by Applying Machine Learning Algorithms.
➢ Machine learning may examine an individual's features in the suggested system to enable early illness identification.

4. OUTPUT SCREENS

User login:

Service Provider Login Page

Remote User Dashboard

Service Provider Dashboard

All Remote Users

Uploaded Dataset
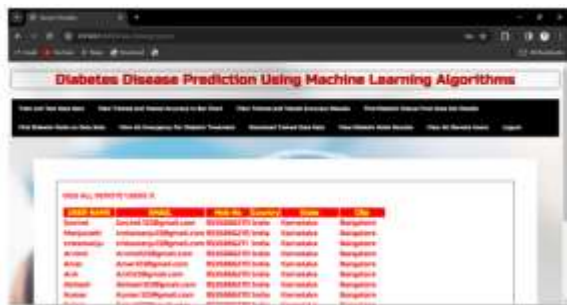
Train and Test Accuracy Bars

User Profile

Accuracy Results

Emergency Patients Report

Pie Chart

Working with a medical facility to compile a diabetic dataset is something we want to do in the future in the hopes of improving our current findings. To further improve outcomes, we will be increasing the use of ML and DL models.

5. CONCLUSION

The early diagnosis of an illness, in this instance diabetes, is one of the major hurdles to the advancement of medical technology. But in this work, researchers set out to build a model that can reliably predict when the illness may manifest. We have easily predicted this condition using the tests performed on the Pima Indians Diabetes Database. With a 76% accuracy rate utilizing the K-Nearest Neighbors classifiers, the findings further demonstrated the system's sufficiency. Be that as it may, we hold out hope that we may incorporate this model into a system that can anticipate the onset of other catastrophic illnesses. Automated analysis of diabetes, or any other condition, may one day have opportunity for improvement.

6.REFERENCES

[1]Gauri D. Kalyankar, Shivananda R. Poojara and Nagaraj V. Dharwadkar," Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop", International Conference On I-SMAC,978-1-5090-3243-3,2017.

[2] Ayush Anand and Divya Shakti," Prediction of Diabetes Based on Personal Lifestyle Indicators", 1st International Conference on Next Generation Computing Technologies, 978-1-4673-6809-4, September 2015.

[3] B. Nithya and Dr. V. Ilango," Predictive Analytics in Health Care Using Machine Learning Tools and Techniques", International Conference on Intelligent Computing and Control Systems, 978-1-5386-2745-7,2017.

[4] Dr Saravana kumar N M, Eswari T,

Sampath P and Lavanya S,” Predictive Methodology for Diabetic Data Analysis in Big Data”, 2nd International Symposium on Big Data and Cloud Computing,2015.

[5] Aiswarya Iyer, S. Jeyalatha and Ronak Sumbaly,” Diagnosis of Diabetes Using Classification Mining Techniques”, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.1, January 2015.

[6] P. Suresh Kumar and S. Pranavi “Performance Analysis of Machine Learning Algorithms on Diabetes Dataset using Big Data Analytics”, International Conference on Infocom Technologies and Unmanned Systems, 978-1-5386-0514-1, Dec. 18-20, 2017.

[7] Mani Butwall and Shraddha Kumar,” A Data Mining Approach for the Diagnosis of Diabetes Mellitus using Random Forest Classifier”, International Journal of Computer Applications, Volume 120 - Number 8,2015.

[8] K. Rajesh and V. Sangeetha, “Application of Data Mining Methods and Techniques for Diabetes Diagnosis”, International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 3, September 2012.

[9] Humar Kahramanli and Novruz Allahverdi,”Design of a Hybrid System for the Diabetes and Heart Disease”, Expert Systems with Applications: An International Journal, Volume 35 Issue 1-2, July, 2008.

[10] B.M. Patil, R.C. Joshi and Durga Toshniwal,”Association Rule for Classification of Type-2 Diabetic Patients”, ICMLC '10 Proceedings of the 2010 Second International Conference on Machine Learning and Computing, February 09 - 11, 2010.

[11] Dost Muhammad Khan1, Nawaz Mohamudally2, “An Integration of K-means and Decision Tree (ID3) towards a more Efficient Data Mining Algorithm”, Journal Of Computing, Volume 3, Issue 12, December 2011.