DIABETES PREDICTION USING MACHINE LEARNING

[1] Vikram Popat Deokate, [2] Suhana Momin, [3]Jasmeen Shaikh, [4]Misba Sayyad, [5]Muskan Nadaf

[1]Professor, [2]Student, [3] Student, [4] Student, [5]Student
[1] Department of Computer Engineering,
[1] AAEMF, Koregaon Bhima, Pune, India

Abstract: Diabetes Mellitus is among critical diseases and lots of people are suffering from this disease. Age, obesity, lack of exercise, hereditary diabetes, living style, bad diet, high blood pressure, etc. can cause Diabetes Mellitus. People having diabetes have high risk of diseases like heart disease, kidney disease, stroke, eye problem, nerve damage, etc. Current practice in hospital is to collect required information for diabetes diagnosis through various tests and appropriate treatment is provided based on diagnosis.

Big Data Analytics plays an significant role in healthcare industries. Healthcare industries have large volume databases. Using big data analytics one can study huge datasets and find hidden information, hidden patterns to discover knowledge from the data and predict outcomes accordingly.

In existing method, the classification and prediction accuracy is not so high. In this paper, we have proposed a diabetes prediction model for better classification of diabetes which includes few external factors responsible for diabetes along with regular factors like Glucose, BMI, Age, Insulin, etc.

Classification accuracy is boosted with new dataset compared to existing dataset. Further with imposed a pipeline model for diabetes prediction intended towards improving the accuracy of classification.

1. INTRODUCTION

Millions of people worldwide are adversely affected by diabetes mellitus, achronic metabolic disease marked by persistent hyper glycemia and has become a major public health concern. Early intervention and specified treatment plans are made possible by punctual and accurate diabetes prediction, which is needed in proactive healthcare management.

The artificial intelligence field of machine learning, which is evolving quickly, has shown great promising results in the area of healthcare, especially in terms of diagnosing and predicting complicated medical conditions.Tomakeamoreproactiveandpreventiveapproachtohealthcare, this study focuses on utilizing machine learning techniques to forecast when diabetes will manifest.

Age-long diabetes analysis techniques often depend on clinical risk factors and statistical models. These methods might not have the sensitivity and specificity required for accurately and precisely identifying people who are at risk, though .A favourable substitute is given by machine learning, which can identify patterns and relationships in large and varied datasets. Machine learning model scan analyze a wide variety of patient-related features, like clinical history, laboratory results, and demo-graphic data, RTML by utilizing highly accessible and developed algorithms.

2. LITERATURE SURVEY

1. **Gauri D. Kalyankar, Shivananda R. Poojara and Nagaraj V. Dharwadkar,” Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop”** It is necessary to collect, store and process this data to discover knowledge from it and utilize it to take significant decisions. Diabetic Mellitus (DM) is from the Non Communicable Diseases (NCD), and lots of people are suffering from it. Now days, for developing countries such as India, DM has become a big health issue. The DM is one of the critical diseases which has long term complications associated with it and also follows with various health problems.

2. **Ayush Anand and Divya Shakti,” Prediction of Diabetes Based on Personal Lifestyle Indicators** Diabetes Mellitus or Diabetes has been portrayed as worse than Cancer and HIV (Human Immunodeficiency Virus). It develops when there are high blood sugar levels over a prolonged period. Recently, it has been quoted as a risk factor for developing Alzheimer, and a leading cause for blindness & kidney failure. Prevention of the disease is a hot topic for research in the healthcare community. Many techniques have been discovered to find the causes of diabetes and cure it.

3. **B. Nithya and Dr. V. Ilango,” Predictive Analytics in Health Care Using Machine Learning Tools and Techniques.** When we have a huge data set on which we would like to perform predictive analysis or pattern recognition, machine learning is the way to go. Machine Learning (ML) is the fastest rising arena in computer science, and health informatics is of extreme challenge. The aim of Machine Learning is to develop algorithms which can learn and progress over time and can be used for predictions. Machine Learning practices are widely used in various fields and primarily health care industry has been benefitted a lot through machine learning prediction technique

4. **Dr Saravana kumar N M, Eswari T, Sampath P and Lavanya S,” Predictive Methodology for Diabetic Data Analysis in Big Data** Modernizing healthcare industry's move towards processing massive health records, and to access those for analysis and put into action will greatly increases the complexities. Due to the growing unstructured nature of Big Data form health industry, it is necessary to structure and emphasis its size into nominal value with possible solution. Healthcare industry faces many challenges that make us to know the importance to develop the data analytics.

5. **B.M. Patil, R.C. Joshi and Durga Toshniwal, "Association Rule for Classification of Type-2 Diabetic Patients"** The discovery of knowledge from medical databases is important in order to make effective medical diagnosis. The aim of data mining is extract the information from database and generate clear and understandable description of patterns. In this study we have introduced a new approach to generate association rules on numeric data. We propose a modified equal width binning interval approach to discretizing continuous valued attributes.

3.   BLOCK DIAGRAM:
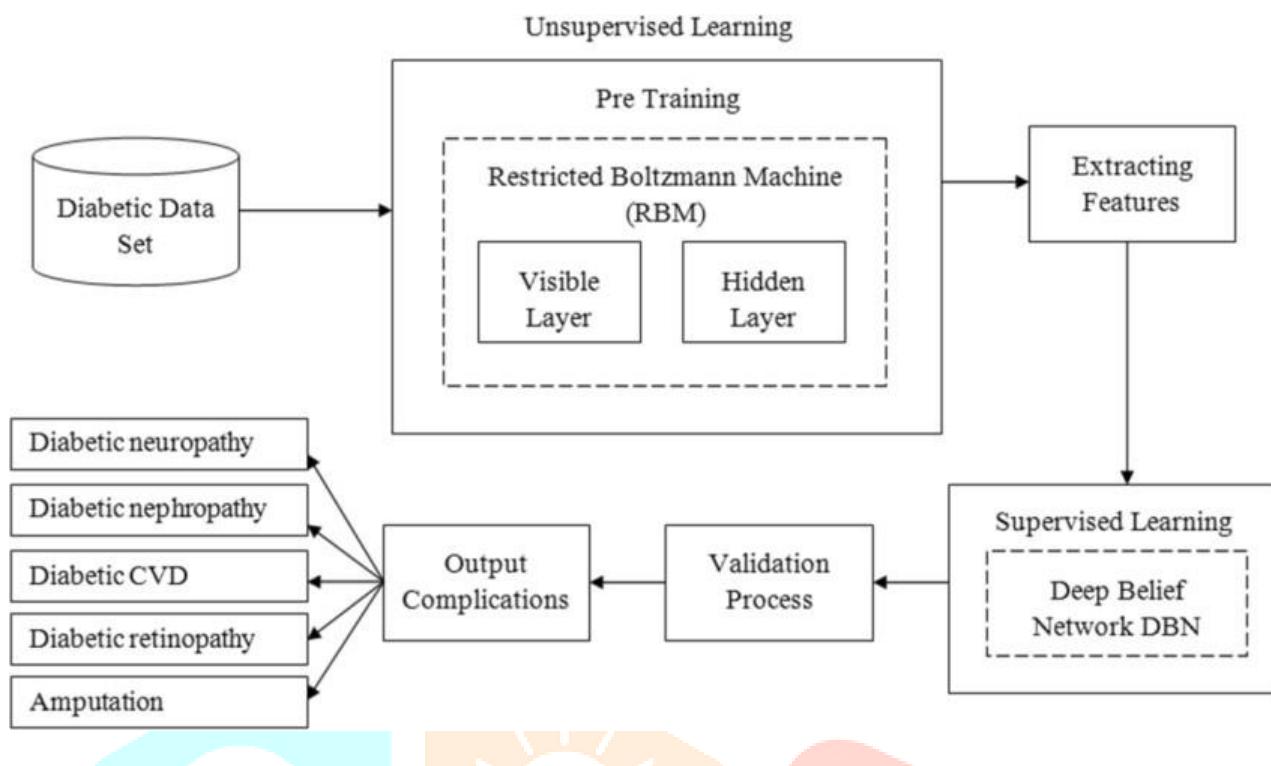
3.1 Working Principle:

Restricted Boltzmann machine (**RBM**):
•   A restricted Boltzmann machine (RBM) (also called a restricted Sherrington–Kirkpatrick model with external field or restricted stochastic Ising–Lenz–Little model)

Extracting features
•   The risk and the severity of diabetes can be reduced by predicting it at an early stage. In this research paper, we strive to propose machine learning algorithms namely KNN, SVM, Decision Tree, Random Forest, LGBM, and Adaboost that help predict diabetes.

Deep Belief Networks (DBN):

•   Deep Belief Networks (DBNs) are sophisticated artificial neural networks used in the field of deep learning, a subset of machine learning. They are designed to discover and learn patterns within large sets of data automatically. Imagine them as multi-layered networks, where each layer is capable of making sense of the information received from the previous one, gradually building up a complex understanding of the overall data.

•   **Validation process:**

The 5 stages of machine learning validation
•   ML Data validations.
•   Training validations.
•   Pre-deployment validations.
•   Post-deployment validations.
•   Governance & compliance validations.

Diabetic neuropathy:

•   Diabetic neuropathy is a type of nerve damage that can occur if you have diabetes. High blood sugar (glucose) can injure nerves throughout the body. Diabetic neuropathy most often damages nerves in the legs and feet.

Diabetic Nephropathy:

•   Nephropathy is the deterioration of kidney function. The final stage of nephropathy is called kidney failure, end-stage renal disease, or ESRD. According to the CDC, diabetes is the most common cause of ESRD.

Diabetic CVD:

•   When you have diabetes, you're more at risk of heart disease. This is also called cardiovascular disease (CVD) or coronary disease, and can lead to heart attacks and strokes. Cardiovascular disease affects your circulation too.

Diabetic retinopathy:

•   Diabetic retinopathy is caused by high blood sugar due to diabetes. Over time, having too much sugar in your blood can damage your retina — the part of your eye that detects light and sends signals to your brain through a nerve in the back of your eye (optic nerve).5 Aug 2024

Amputation:

•   Amputation is the loss or removal of a body part such as a finger, toe, hand, foot, arm or leg. It can be a life changing experience affecting your ability to move, work, interact with others and maintain your independence.

4.   Hardware Requirements

[1]   Processer: core i5

[2]   RAM: 8GB

[3]   HDD:256GB
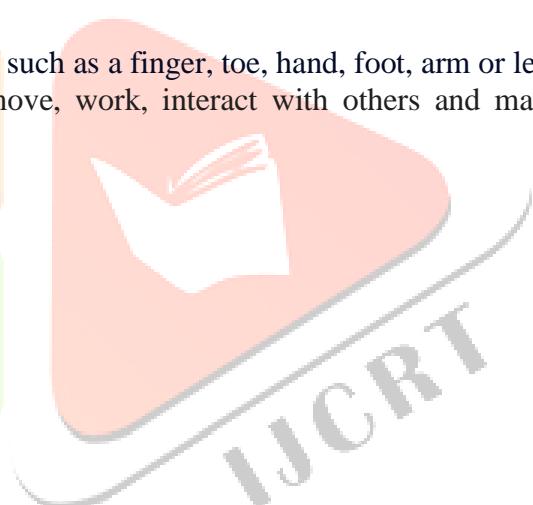
5.   SOFTWARE REQUIREMENTS:

•   Machine Learning
•   Python Libraries
•   KneighborsClassifier Algorithm
•   Sk learn
•   Jupyter Notebook
•   Kaggle

6.   ADVANTAGES AND DISADVANTAGES:

•   The advantage of this system is its ability to fit in linear and nonlinear functions by handling uncertainty in unknown functions also.

•   Probabilistic prediction can also be made through this model, but the main disadvantage of this model is to select a kernel for the correct representation of medical data.

•   The major disadvantage of the model was that it was not compared with the other diabetic prediction models and hence could not be validated.

• The use of machine learning methods and mathematical models to predict the risk of disease has been widely used in clinical medicine.

• Various machine learning classification algorithms are used for detecting diabetes.

• For diabetes prediction and monitoring, the recommended structural benefits of effective decision-making technique and helping in good outcome.

• Keeping in view the huge development in the ailment, the recommended prototype goal is to deal with efficiently through cloud computing solutions.

• If diabetes is not controlled, then they may cause serious health problems such as lower limb amputation, kidney failure, blindness that impact very badly on the quality of life. There are various research issues or challenges that are occurring while predicting the diabetes

• Researches using data mining procedures to study the patients' data which are beneficial to find important knowledge which is facilitating medical services and deeply study of disease

• In the diabetes prediction Machine learning methods are broadly used to get superior results. In the field of medical Decision support tool is the most prevalent mechanism procedures, which has grate sorting controls.

• Another famous machine learning method is a neural network and gives better results in different aspects. So Random decision forests ,Decision tree, and artificial neural system mostly uses methods to predict diabetes.

7.   CONCLUSION:

Diabetes can be a reason for reducing life expectancy and quality. Predicting this chronic disorder earlier can reduce the risk and complications of many diseases in the long run. In this paper, an automatic diabetes prediction system using various machine learning approaches has been proposed. The open-source Pima Indian and a private dataset of female Bangladeshi patients have been used in this work. SMOTE and ADASYN preprocessing techniques have been applied to handle the issue of imbalanced class problems.

This research paper reported different performance metrics, that is, precision, recall, accuracy, F1 score, and AUC for various machine learning and ensemble techniques. The XGBoost classifier achieved the best performance with 81% accuracy and an F1 score and AUC of 0.81 and 0.84, respectively, with the ADASYN approach. Next, the domain

adaptation technique has been applied to demonstrate the versatility of the proposed prediction system.

Finally, the best-performed XGBoost framework has been deployed into a website and smartphone application to predict diabetes instantly. There are some future scopes of this work, for example, we recommend getting additional private data with a larger cohort of patients to get better results.

Another extension of this work is combining machine learning models with fuzzy logic techniques and applying optimization approaches.

REFERENCES

1. Atlas, G. : *Diabetes. International Diabetes Federation*. 10th ed., IDF Diabetes Atlas.

2. Akhtar, S. , et al.: Prevalence of diabetes and pre-diabetes in Bangladesh: A systematic review and meta-analysis.(2020)

3. Prabhu, P. , Selvabharathi, S. : Deep belief neural network model for prediction of diabetes mellitus. In: International Conference on Imaging, Signal Processing and Communication, pp. 138–142 (2019)

4.   VijiyaKumar, K. , Lavanya, B. , Nirmala, I. , Caroline, S.S. : Random forest algorithm for the prediction of diabetes. In: International Conference on System, Computation, Automation and Networking, pp. 1–5 (2019)

5.   Hasan, M.K. , Alam, M.A. , Das, D. , Hossain, E. , Hasan, M. : Diabetes prediction using ensembling of different machine learning classifiers. *IEEE Access* 8, 76516–76531, (2020).

6.   Mohan, N. , Jain, V. : Performance analysis of support vector machine in diabetes prediction. In: *International Conference on Electronics*, Communication and Aerospace Technology, pp. 1–3 (2020)

7.   Smith, J.W. , Everhart, J.E. , Dickson, W.C. , Knowler, W.C. , Johannes, R.S. : Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In: *Annual Symposium on Computer Applications in Medical Care* pp. 261–265 (1998)

8.   Aurélien, G. : Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc., Sebastopol, CA

9.   Mitchell, T.M. : Machine Learning. McGraw-Hill, Inc., New York

10. Chatrati, S.P. , Hossain, G. , Goyal, A. , et al.: Smart home health monitoring system for predicting type 2 diabetes and hypertension. *J. King Saud Univ* (2020).