IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 6, June 2022

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106261|

A Review on Prediction of Diabetes in Females of Pima Indian Heritage: Using Supervised Machine Learning Approach

Teja H, Ashwini.C

MCA Student, UBDT College of Engineering, Davanagere, Karnataka, India

Dept. of Master of Computer Applications, UBDT College of Engineering, Davanagere, Karnataka, India

ABSTRACT: Diabetes is very dangerous disease and does not able to cure, if this disease affects once, it will maintain in your life time. At that same time, your blood having too much of glucose can cause health issues, like kidney heart and stroke problem, eye problem. Nowadays, diabetes is a common disease that affects millions of people over the world,an women are mostly affected by this disease. Recent healthcare studies have applied various innovative and advance technologies to diagnose people and predict their disease based on clinical data. One of such tech¬nologies in machine learning(ML) in which diagnosis and prediction can be made more accurately. In this paper, the designed model predict the diabetes of female of pima Indian heritage by taking the clinical dataset. Here, this problem is considered as a binary classification problem . Therefore, supervised learning algorithm have been used, such as classification tree(CT), support vector machine(SVM) k-nearest neighbor(k-NN), Naïve Bayes(NB), RandomFOrest(RF), Precision and recall result of all the supervised learning algorithms and compare them to determine the best algorithm that is suitable for prediction.

I. INTRODUCTION

Diabetes is a non-communicable disease, which affects healthcare severely by reducing the efficiency of a person [48, 34, 47, 9, 30, 49]. In this disease, the blood glucose level rises more than the normal glucose level in the body [35, 20, 5, 11, 41]. It is noteworthy to mention that glucose is the form of sugar that is needed by the body for better metabolism. All cells need glucose as a source of energy. However, if the blood glucose level increases due to lack of insulin hormone in the body, then it imbalances the blood glu¬cose in the body that results in severe damage to other parts of the body, such as eyes, heart, kidneys and many more [27, 12, 39, 44]. It is controlled by changing the lifestyle, such as food habits, medications, exercises to name a few. If the disease is diagnosed in time by prediction, then a person's health can be improved. Therefore, if the healthcare system uses intelligent 48
prediction mechanisms, then a person's life can be saved [40, 1, 23]. However, in this work, our main focus is to only predict whether a female of Pima Indian heritage has diabetes or not. Diabetes is of three types, namely type- 1, type-2 and gestational [24, 8, 42, 46, 17, 45]. In type-1, the immune system destroys the insulin cells. It generally happens to children and adolescents. In type-2, the pancreas makes very little insulin [36, 18, 37, 43]. It generally happens to adults. The former type is also called as insulin resistance, whereas the latter type is called as insulin deficiency. Recent healthcare studies have applied various technologies to diagnose people and predict their disease based on the collected clinical data. Nowadays, the healthcare system can predict diabetes more accurately using ML techniques. ML enables a computer to become intelligent by learning from the experiences or inputs (i.e., clinical data) and predict the output category (i.e., disease) [21, 6, 26, 2, 10]. The ML techniques are categorized into supervised, unsupervised and reinforcement learning. In supervised learning, the features, as well as the target class, are used as input for learning. In unsupervised learning, there is no such target class. Here, the input data is provided without the target class [13, 25, 28]. This data is further used for clustering using a similarity measure. Note that the input data may be unstructured. Reinforcement learning is an approach that uses the hit and trial method, where the computer or machine finds the best outcome. To get the best outcome,award and penalty values are used [16, 7, 15, 19]. In this paper, the diabetes of females of Pima Indians heritage problem has considered as a binary classification problem and mainly focused on the supervised learning algorithms. The supervised algorithms are best suits for classification problem. The main contributions, in this paper, are listed below.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106261|

II. OBJECTIVES

1. To analyze various parameter related to diabetics and get useful insights
2. To build a machine learning model which gives god accuracy
3. To develop an android app to accept new test data and give the result instantly

Scope
It finds useful for health sectors; one can use this product to check diabetic victim chances of pregnant women, this application could be used at hospitals or diagnostic laboratories to analyze the chances of being affected by diabetics

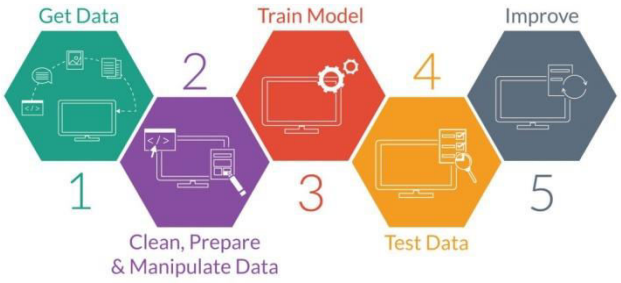
III. LITERATURE SURVEY

Shetty et al. used KNN and the Naïve Bayes technique has been used for the prediction of diabetes. Their technique was implemented as an expert software program, where users provide input in terms of patient records and the finding that either the patient is diabetic or not Singh et al. applied different algorithms on datasets of different types. They used the KNN, random forest and Naïve Bayesian algorithms. K-fold cross-validation technique was used for evaluation. Ahmed utilized patient information and plan of treatment dimensions for the classification of diabetes. Three algorithms were applied which were Naïve Bayes,
logistic, and j48 algorithms

Existing System
In the existing system analysis of various parameters are done at laboratories with help of various chemicals and conclusions is drawn, some facilities are manual and some are dine with the help of computers.

Proposed System
Proposed system is machine learning based approach, here data sets are basic building blocks, data sets are divided into

III. METHODOLOGY



Get Data
The first step in the Machine Learning process is getting data

Clean, Prepare the Data
Real-world data often has unorganized, missing, or noisy elements. Therefore, for Machine Learning success, after we chose our data, we need to clean, prepare, and manipulate the data.

Train model
This step is where the magic happens! The data set connects to an algorithm, and the algorithm leverages sophisticated mathematical modeling to learn and develop predictions. These algorithms commonly fall into one of three categories:
Binary – Classify into two categories Classification – Classify into many categories
Regression – Predict a numeric

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106261|

Test model
Now, it's time to validate your trained model. Using the test data from Step 3, we check the model's accuracy.
If the results are not satisfactory, you need to improve and retrain your ML model (Step 5).

Improve
Practice makes perfect! Here are a few things you can do to refine your model and improve accuracy, Review your model's results with your business stakeholders. Are there other data elements worth adding to your model to make it more accurate?
Reconsider your algorithm choice. Within each class of algorithm, there are dozens of algorithm choices. A different algorithm may perform better for you Adjust the
parameters of your chosen algorithm to improve performance. Sometimes small adjustments have a significant impact.
These are just high-level steps – ML is as complicated as you choose! That said, by understanding the basic process of developing ML models, you gain a solid foundation for further learning.

System requirements
In this and development of the architecture for the diabetes management system, the clinical requirements and design analysis of the system were on discussions with collaborators from the department of nutrition and food science of the university of Ghana and kwame Nkrumah
University of science and technology . from these discussion, the diet type
patients was detemined to bean essential approach suitable for diabetes management system
Following functionalities were mentioned. scheduling and reminding diabetic patients to take their medication and blood glucose readings, Recommending healthy meals for diabetics to keep their blood glucose levels in check encourage and tracking the activity of diabetic
patients

Functional Requirements
Management of medical staff users
Recording by doctors of the patients, they are attending, as well as, registration of the respective examinations that they will perform. Management of medical records of patients.

Non-Functional Requirements
Non functional requirements are attributes that either the system or the environment must have IEEE defines non functional requirements in software system Engineering.
The non-functional requirements are:

Usability
Usability gives the measure of much user friendly the system is. This also includes the internationalization requirements such as, languages, spellings, etc.It also refers to the mean time between failures, means what can be the maximum down time Ex, two hours in six months etc

Security
It refers to the ability to provide
confidentiality, data integrity, and data availability.

Scalability
It refers to maintain of the proposed software application to increase the
number of users or application associated with the product.

Use case diagram
Below diagram shows various processes being carried out by the end user.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106261|

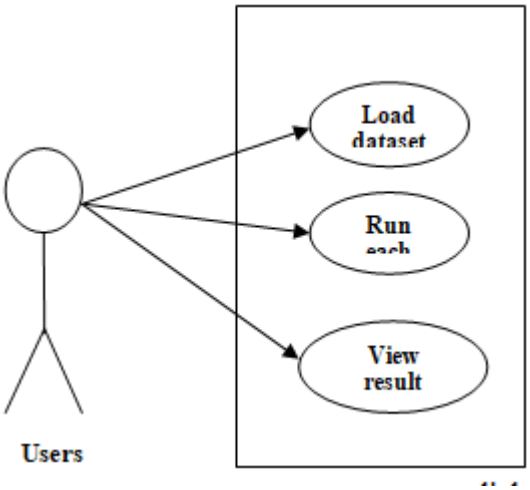


Fig: use case diagram

Datasets

The dataset is taken from the national institute of diabetes and digestive and kidney diseases. The data is stored in kaggle and UCI data repository. This dataset is mainly used to predict whether a pima Indians female has diabetes or not all the patients taken in the dataset are female in dataset as diabetic,
Following attributes are considered

Age: It shows age in years. The range is 21 to 81 and the average age is 33.

Pregnancies: It shows that the number of time a female gets pregnant. The range is 0 to 17 and the average 4
.
Glucose: It shows the plasma glucose concentration level (2hours) it is from 0 to 199 and
average is 121

Blood pressure: It shows the diastolic blood pressure in mm Hg. It is from 0 to 122 and
average is 69
.
Skin thickness: It shows the triceps skin thickness in mm. the range is 0 to 99 and the
average is 21.

Insulin: It ranges from 0 to 846. The average is 80.

BMI: It shows body mass index in Kg/m2. The range is 0 to 67.1 and the average is 32.

Outcome: It is either 0 or 1. Here, 0 means that a female has non-diabetic and 1 means that a female is diabetes.
In this dataset, the outcome used as the target class to predict that the pima Indians female is diabetic or not. It has 768 instances and columns.

IV. RELATED WORK

Many diabetes prediction algorithms have been proposed by the researchers to accurately predict the types 84 of diabetes as well as diabetes of Pima Indians
[23, 9, 49, 35, 20, 36, 15,19, 24, 85 8, 10]. These works are briefly discussed as follows.
Zolfagri et al. [48] have prposed a method to diagnose diabetes in females' populations of Pima Indians using an ensemble of neural network and SVM. Pham et al. [34] have predicted diabetes by a new data mining approach that balances using fitting and generalization. Wu et al. [47] have proposed a method to diagnose diabetes using a semi-supervised learning method that uses Laplacian SVM. Sanakal et al.[40] have diagnosed diabetes using the prognosis of fuzzy c-means clustering and SVM. Al et al. [1] have used decision tree technique to diagnose type-2 diabetes. Kumari

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 6, June 2022||

| DOI:10.15680/IJIRSET.2022.1106261|

et al. [23] have used SVM for classification of diabetes. Dey et al. [9] and Zou et al. [49] have implemented web-based approach to predict diabetes using ML approaches. However, none of the paper has addressed all the well-known supervised learning algorithms at a glance. Pradhan et al. have predicted diabetes using an artificial neural network (ANN) [35]. Karthikeyani et al. have proposed a comparison performance of data mining algorithm (CP-DMA) to predict the diabetes disease [97 20] .Karatsiolis et al. have proposed region-based SVM algorithm for medical diagnosis of Pima Indians [43] [18]. Guo et al. have used bayes network to predict type-2 diabetes [11]. Maniruzamman et al. [27] have performed a comparative analysis of diabetes mellitus data using ML techniques. Han et al. [12] have analyzed the data using rapid miner software. Saji et al. [39] have predicted diabetes using a multilayer perceptron. Jahangir et al. have proposed an expert system to predict diabetes using an autotuned multilayer perceptron [13]. Li et al. [25] have proposed a weight-adjusted approach to diagnose diabetes. In [28] have proposed an accurate diabetes risk stratification using ML techniques. Like Pradhan et al. [35], Sivastava et al. [43] have predicted diabetes using ANN approach. Kala et al. have proposed an intelligent hybrid system for a diabetes diagnosis [16]. Chen et al. have proposed a hybrid prediction model to diagnose type-2 diabetes using decision trees and k-means [7]. Kahramanli et al. have proposed ahybrid system for diabetes and heart disease [15]. In [19] authors have proposed a genetic algorithm approach using NN to diagnose Pima Indians diabetes. In [24] authors have proposed a fuzzy technique to diagnose diabetes accurately. Many such algorithms have been presented in [8, 42, 46, 17, 45, 21, 6, 26, 51, 10, 3, 50] However, they have not combinedly addressed most of the supervised learning algorithms.

V. CONCLUSION

The proposed project is an attempt to used Machine learning concept to analyze the dataset of samples being collected from kaggle repository, it has Indian women dataset with various attributes, we can conclude that BP, weight and age plays key role of becoming diabetic and main aim is to design and implements diabetes prediction using Machine Learning methods and performance analysis of that methods and it has been achieved successfully. The proposed approach uses various classification and ensemble learning method in which SVM , Knn ,Random Forest, Decision Tree, Logitic Regression and Gradient Boosting classifiers are used, And 77% classification accuracy has been achieved. The Experimental results can be assist health care to take early prediction and make early decision to cure diabetes and save human lifes.

REFERENCES

[1] K.Priyadarshini1, Dr.I.Lakshmi.2017. A Survey on Prediction of Diabetes Using Data Mining Technique from International Journalof Innovative Research in Science, Engineering and Technology, ISSN(Online): 2319-8753
[2] SonaliVyas , Rajeev Ranjan , Navdeep Singh , Arohan Mathur.2019. Review of Predictive Analysis Techniques forAnalysis Diabetes Risk.
[3] Kumari, Sonu, and Archana Singh .2013. A data mining approach for the diagnosis of
diabetes mellitus.
Intelligent Systems andControl (ISCO), 7th International Conference on. IEEE.
[4] T.Jayalakshmi and Dr.A.Santhakumaran.2010.A Novel Approach for Diagnosis of
Diabetes Mellitus Using Artificial Neural Networks", International Conference on Data
Storage and Data Engineering,159-163.
[5] J. Steffi, Dr.R.Balasubramanian.2018.Predicting Diabetes Mellitus using Data Mining Techniques from International Journal Of Engineering Development And Research. Volume 6, Issue 2 ,ISSN: 2321-9939
[6] Komi, M., Li, J., Zhai, Y., & Zhang, X. 2017, June. Application of data mining methods in diabetes prediction. In Image, Vision andComputing (ICIVC), 2017 2nd International
Conference on (pp. 1006-1010). IEEE.
[7] Rahul Joshi, Minyechil Alehegn.2017.Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach from International Research Journal of Engineering and Technology, p-ISSN: 2395-0072
[8] Perveen, S., Shahbaz, M., Guergachi, A., &Keshavjee, K. 2016. Performance analysis of data mining classification techniques topredict diabetes. Procedia Computer Science, 82, 115-121.
[9] Song, Y., Liang, J., Lu, J., & Zhao, X.2017. An efficient instance selection algorithm for k nearest neighbour regression.Neurocomputing, 251, 26-34.
[10] Pradeep, K. R., & Naveen, N. C. 2016. Predictive analysis of diabetes using J48 algorithm of classification techniques In Contemporary Computing and Informatics (IC3I), 2nd International Conference on (pp. 347-352).IEEE

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details