Diabetes Prediction Using Different Machine Learning Classifiers

[1]Sangeeta Waren [2]Nitesh Dubey
[1]Research Scholar, Department of Computer Science & Engineering
[2]Professor, Department of Computer Science & Engineering
Global Nature Care Sangathan's Group of Institutions, Jabalpur, India

Abstract— Diabetes Mellitus is critical and many people suffer from this condition. Diabetes mellitus may be the cause of age, obesity, lack of exercise, genetic diabetes, lifestyle, poor diet, high blood pressure, etc. Diabetes people have a high risk of heart failure, kidney disease, stroke, eye problem, nerve damage, etc. Current in-hospital procedure is to obtain the requisite diabetes diagnostic information through numerous tests and proper diagnostic care. In the healthcare sectors, big data analytics plays an important role. Healthcare industries have extensive databases. Big data analytics allow you to research large datasets to find hidden information, hidden trends to detect knowledge from the data and to forecast results accordingly. The classification and prediction accuracy in the current system is not so high. In this paper we have presented a model predicting diabetes for a better diabetes classification, including few external diabetes factors and common factors such as Glucose, BMI, age, insulin, etc. The accuracy of classification is increased with new datasets relative to existing datasets. In addition, a diabetes prediction pipeline model has been imposed to improve classification accuracy.

Keywords—*Diabetes Mellitus, Big Data Analytics, Healthcare Machine Learning.*

I. INTRODUCTION

There is a clear indicator of elevated blood sugar with diabetes, with some signs including increased appetite, increased malnutrition, weight loss, and repeated urination. At least 200 mg/dL of diabetes diagnosis is made over 2 hours post-load plasma glucose, and different diabetes diagnosis studies need timely detection of the call. Usually, diabetic patients need ongoing care, or it can lead to many life-threatening, dangerous consequences.

The early and quick diagnosis of diabetes plays an important role in curing diabetes. The proposed framework uses methods of machine learning for the identification of diabetes. The device proposed would be a medical field technology that will be useful in the detection of diabetes for patients and diabetes physicians. The proposed system is an automation system for diabetes detection using data from old diabetes patients.

II. LITERATURE REVIEW

Defusal Faruque and Asaduzzaman, Iqbal H.Sarker has discussed that diabetes is one of the most common disorder of the human body it is caused due the metabolic disorder .Hence that they used various and important ML algorithms that are Support Vector machine, NB,KNN and DT to predict the diabetes[1].

Sidong Wei,Xuejiao Zhao and Chunyan Miao presented that diabetes is commonly called as disorder in which glucose level in body is high. In this paper they use popular methods such as SVM and deep neural network for identify the disease and data processing. [2].

Lakshmi K.S and G.Santhosh Kumar according to them Hospital databases serve as wealthy information source for the fruitful medication diagnosis. IN this they used NLP tools along with combined with data mining algorithms for the extraction of rules [3].

Jian-xunChen , Shih-LiSu and Che-Ha Chang discussed about Ontology that generate a primary care planning to the medical professional's for the accustoming. The result of the research paper shows the model can be provided personalize diabetes mellitus care planning efficiently [4].

MM Alotaib, RSH.Istepanian, and A.Sungoor they are present a clever based mobile polygenic disease control system & tutoring model for the patients with diabetes. In this, system is able to store the clinical information about the diabetes system, such an often blood sugar level and BP measured and hypo glycaemia event [5].

Berina Alic and Lejila Gurbea,Almir Badnjevic they presented the overview of techniques in machine learning in the diabetes classification and cardiovascular diseases using BNs and ANN [6].

M.Durgadevi and Dr.R.Kalpana In this paper they estimate that risks, So gigantic cat goring and detection algorithms have been develop in the domain of DM. So, that this paper aim is to compare the fruition and 5 classification way are anti-miner, Ad boost, RBF network, CN2 and Bagging for the diabetes prediction [7].

ElliotB.Sloane, Nilmini Wickramasingle and Steve Goldberg they presented Wireless diabetes monitoring which is a cloud-based diabetes, it's a coaching platform for diabetes management and its a low cost, innovative, cloud-based diabetes support system [8].

Minyechil Alehegn and Rahul Joshi had present about the ML technology that help to identify a dataset at the elementary so that rescue the life.By implementing NB and K-nn algorithms.[9].

Umatejaswi and P.Suresh Kumar had discussed about algorithms such as SVM, NB, DT for identify the mellitus make use of technique like data mining [10].

III. PROPOSED WORK

The Proposed method use KNN algorithm for classification and prediction of diabetes using trained data. And, the proposed system also predicts the time of getting diabetes.
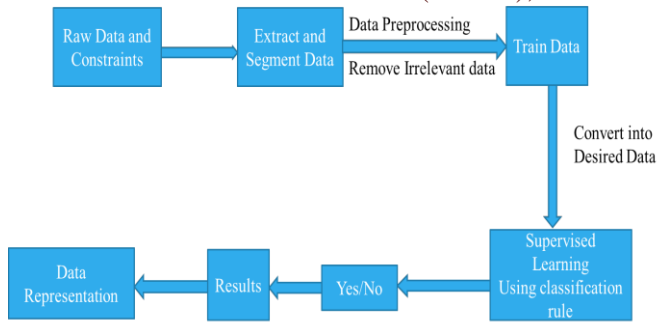
Data Constraints

Data is a collection global dataset. IN this system use Pima Indian data set is used for training a model. Data set contain

21 parameters and around 1000 dataset. The dataset feature/parameters are:

- Age
- Gender
- Relation
- DOB
- Sugar tested value
- Symptoms
- Family history etc.

This are data is trained to the model for the prediction of diabetes.

Train Dataset and Test Dataset

The training data is a initial set of data which is used to understand the program. This is the one in which we have to train the model first because to set the feature and this data is available on system. This data is used to teach the machine for do different actions. It is the data in which model can learn with algorithm to teach the model and doing work automatic.

Testing data is the input given to a software. It shows the data affects when the execution of the module that specifying and this is basically used for testing.

Pre-processing of data

Data preprocessing is a process in which that is actual use for converting the basic data into the clean data set. It is the step in which the data transform or an encode to the state that the machine can be easily parse. The major task of data preprocessing in learning process is to remove the unwanted data and filling the missed value. So that it help to machine can be trained easily.

Figure 2: Data Pre-processing.

Feature Extraction

Feature Extraction is the method in which it used for alter the key data for features of outcomes. This, trait square is used to compute the characteristics of designs given that facilitate in different amid the class of key pattern details. This method involving to decrease the counts of resource required to describe the huge set of data. Feature extraction is an attribute reduction process. This is also used to increasing the speed and effectiveness of supervised learning.

ML Algorithm: KNN

The k-nearest neighbor's is a ML algorithm is the non-parametric method proposed by Thomas Cover used for Regression and Classification. This algorithm is mainly used for the classification of problems in the industry. KNN algorithm is a type of instance-based learning method. This algorithm relies on the distance for objects classification, training data normalizing to the improve its accuracy dramatically. The neighbors are derived from the set of things for which classes or object property values are known. It can be thought of as a training set for the algorithm, although no explicit training steps are required.

System Design

Designing of system is the process in which it is used to define the interface, modules and data for a system to specified the demand to satisfy. System design is seen as the application of the system theory. The main thing of the design a system is to develop the system architecture by giving the data and information that is necessary for the implementation of a system. In this project three-tier architecture is used.
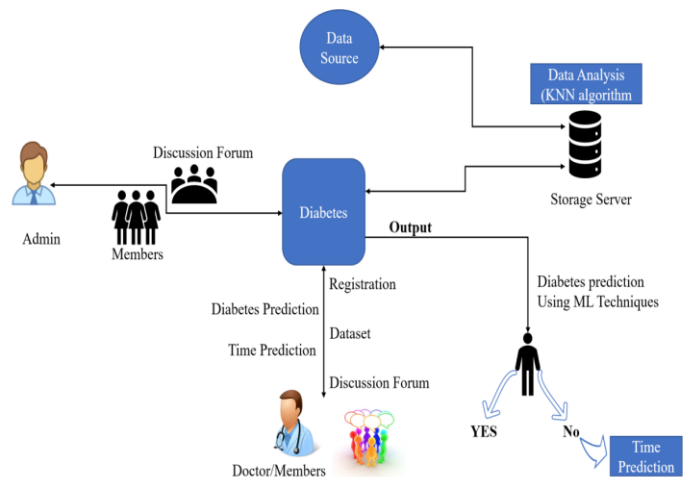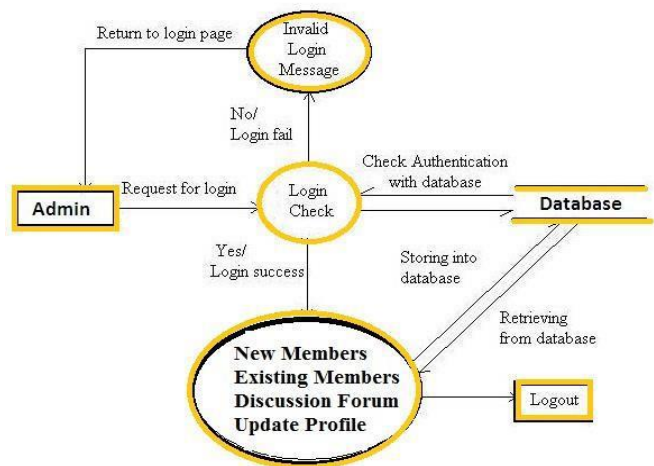

Figure : Architecture Design.
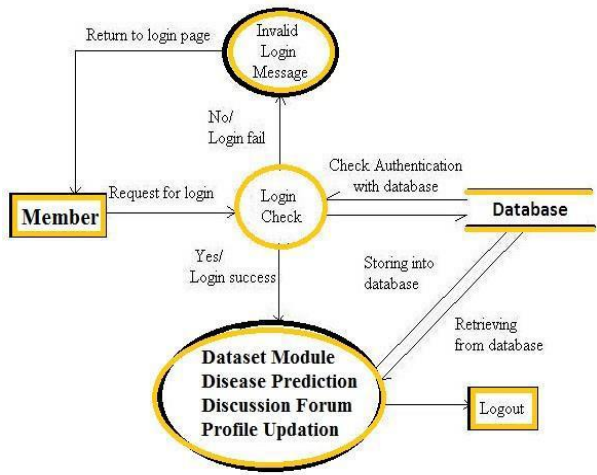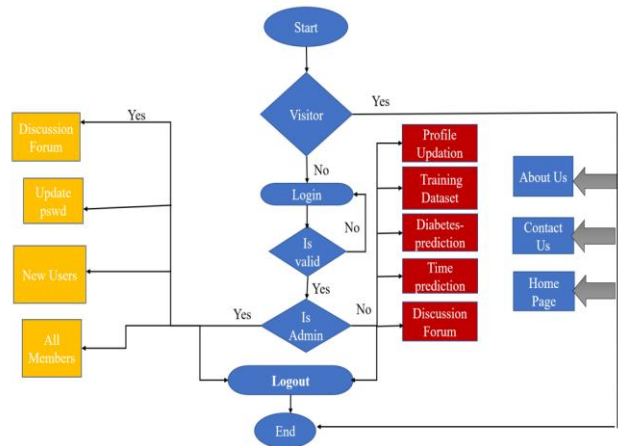

Figure: Data Flow for Admin

Figure : Data Flow design for Member

Sequence Diagram of a System

The sequence diagram of a system shows the entity interplay are ordered in the time order level. So, that it drafts the classes and object that are imply in the that plot and also the series of message exchange take place betwixt the body that need to be carried out by the purpose of that scenario.
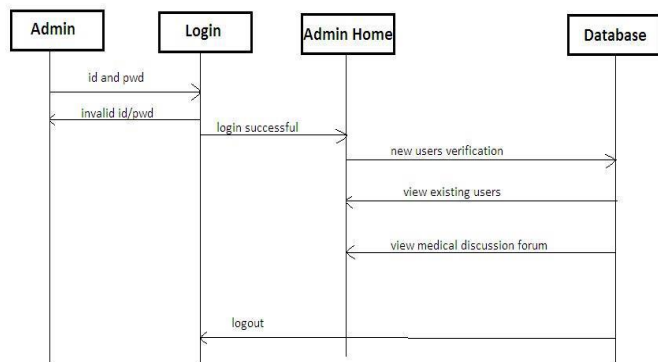
.

Figure: Sequence Diagram for Admin.

Figure : Sequence Diagram for Member/Doctor.

IV. RESULTS & ANALYSIS

Implementation can be described as the realization of an application, or execution of the plans, ideas, models, design and system development, specification of the model, standard, algorithms used in the system, or authority. In computer science, an implement is explained as the realization of technically specified or algorithms' as a programed, a software component, or any others computer systems through computer programming and deployment. Many of the implementations may existed for a given specification or standard.

Figure: Control Flow

In our project the result is classified into Yes or No. If the result is classified into No then we use time prediction module. Time Prediction - here we predict the "time" of getting the diabetes disease. We analyze the result of the diabetes prediction and check the accuracy of the diabetes prediction, time taken to compute the accuracy of the diabetes prediction, correctly classification and incorrectly classification of result of the diabetes prediction. We have used KNN Algorithm to predict the diabetes where result is classified into Yes or No and also for time prediction module same KNN Algorithm is used. We compared the testing data and actual data to get the accuracy of our project.

Constraint	KNN Algorithm
Accuracy	98%
Time	790 milli sec
Correctly Classified	98%
Incorrectly Classified	2%

CONCLUSION & FUTURE WORK

The prediction of diabetes is one the of great importance in today scenario, and concerning with its severe complications. Due to the biggest reason for the death in worldwide is diabetes. The System model is mainly focus to identification of diabetes using some of the parameters. System is useful to physicians to predict the diabetes in initial dais. So, that conventional treatments and solutions may be given to the patients. System used some of the techniques like ML for the prediction, so that to get the more precise results. There have been fortune of investigation on the diabetes imprint. Building diabetes disease prediction system is useful for hospitals and doctors. System predicts disease at early stages, so doctors can treat patients in a better way. Proposed model is the real time application in which is meant for multiple hospitals and predicts disease in less time. As we use machine learning algorithms for disease prediction, we will get more accurate and efficient results.

References

[1] "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus" Md Faisal Faruque, Asaduzzaman, Iqbal H. S arker, IEEE 2019.

[2] "A Comprehensive Exploration to the Machine Learning Techniques for Diabetes Identification" Sidong Wei1, Xuejiao Zhao, Chuny an Miao Shanghai Jiao Tong University, China.

[3]   "Association Rule Extraction from Medical Transcripts of Diabetic Patients" Lakshmi K S, G Santhosh Kumar, 2014.

[4]   "Diabetes Care Decision Support System" 2nd International Conference on Industrial and Information Systems IEEE 2010.

[5]   "An Intelligent Mobile Diabetes Management and Educational System for Saudi Arabia: System Architecture" M.M. Alotaibi, R.S.H. Istepanian, A.Sungoor and N. Philip, IEEE 2014.

[6]   "Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases" by BerinaAlic, Lejla Gurbeta, IEEE 2017.

[7]   "Performance Analysis of Classification Approaches for the Prediction of Type II Diabetes" by M. Durgadevi, M. Durgadevi, IEEE 2017.

[8]   "Cloud-Based Diabetes Coaching Platform for Diabetes Management" Elliot B. Sloane Senior Member IEEE, Nilmini Wickramasinghe, Steve Goldberg 2016.

[9]   Minyechil Alehegn and Rahul Joshi, "Analysis andprediction of diabetes diseases using machine learning algorithm":International Research Journal of Engineering and Technology Volume: 04 Issue: 10 | Oct -2017

[10]   P. Suresh Kumar and V. Umatejaswi, "Diagnosing Diabetes using Data Mining Techniques", International Journal of Scientific and Research Publications, Volume 7, Issue 6, June 2017 705 ISSN 2250-3153.

[11]   "Clustering Medical Data to Predict the Likelihood of Diseases" by Razan Paul, Abu Sayed Md. Latiful Hoque, IEEE 2010.

[12]   "Robust Parameter Estimation in a Model for Glucose Kinetics in Type 1 Diabetes Subjects" Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006.

[13]   Anjali C And Veena Vijayan V, Prediction and Diagnosis of Diabetes Mellitus, "A Machine Learning Approach" ,2015 IEEE in Intelligent Computational Systems (RAICS) | Trivandrum.

[14]   Ridam Pal ,Dr. Jayanta Poray, and Mainak Sen, ,"Application of Machine Learning Algorithms on Diabetic Retinopathy", 2017 2nd IEEE International Conference On Recent Trends In Electronics Information & Communication Technology, May 19-20, 2017, India.

[15]   Dr. M. Renuka Devi and J. Maria Shyla, "Analysis of Various Data Mining Techniques toPredict Diabetes Mellitus", International Journal ISSN 0973-4562 Volume 11, Number 1 (2016) pp 727-730.