



# Diabetes Prediction using Machine Learning

Avantika Nahar<sup>1</sup>, Dr. Ajay Lala<sup>2</sup>, Saurabh Sharma<sup>3</sup>

<sup>1</sup>Student, <sup>2</sup>Professor, <sup>3</sup>Assistant Professor, Computer Science and Engineering Department, Gyan Ganga College of Technology, India

**Abstract**— Obesity is a type of metabolic disorder that millions of people are affected. Diabetes is of a great significance and severe complications should be worried. Since there have been numerous research studies performed on diabetes, most of the research is based on the Pima Indian diabetes data set. This Pima Indian data set includes multiple studies of women in India's population conducted from 1900 to 2010.

The onset rate of diabetes is more frequent in 1965. Most research studies carried out before concentrating mainly on one or two specialized testing techniques while an inclusive research on a variety of techniques was missing. So we thoroughly utilized the machine learning and data pre-processing techniques to find the presence of diabetes in the sample. We will validate this by using data from the UCI Machine Learning repository.

**Keywords**--Machine learning, classification, KNN, diabetes keywords.

## I. INTRODUCTION

A direct manifestation of diabetes is high blood sugar, with symptoms including increased thirst, increased hunger, weight loss and frequent urination. The diagnosis of diabetes may be detected if an individual's blood sugar is at least 200mg/dL over a 2-hour post-load plasma glucose. Diabetic patients usually require constant treatments or else they will face dangerous consequences.

It's very important to reduce the number of diabetes in the community. Machine learning techniques can accurately discriminate specific faces against a background. The proposed system would be able to identify medical issues like diabetes better. The project is to detect diabetes in the old patients.

## II. LITERATURE SURVEY.

In diabetes, the likelihood that a person will get one is determined by factors including age, family history and physical examination of the individual.

This is commonly referred to as "insulin disorder", "diabetes" or "adult onset diabetes". This thesis uses the popular methods including Support Vector Machine and Deep Neural Network for diagnosis and data processing. [3].

Information garnered from hospital databases contain substantial information to prevent medication-related problems. IN this, they used NLP tools along with data mining to extract the rules [3].

Jian-xunChen, Shih-LiSu and Che-Ha Chang talked about helping in reviewing the patient's records of their prognosis and their condition. This paper shows the possibility to provide personalize diabetes mellitus care planning efficiently [4].

MM Alotaib, RSH.Istepanian, and A.Sungoor presented a sophisticated mobile environment based disease surveillance system.

A tutoring model has been successfully implemented to educate university students with diabetes. This system will allow for all clinical data to be stored about the diabetic condition and different anti-diabetic therapies.

Berina Alic and Lejila Gurbea presented examples of medical breakthroughs with BNs and ANN.

There was a risk that cat owners will not be able to deal with such danger. So this paper aims to compare five methods for diabetes prediction: AdBoost, RBFNet, CN2, Bagging and Classifier [7].

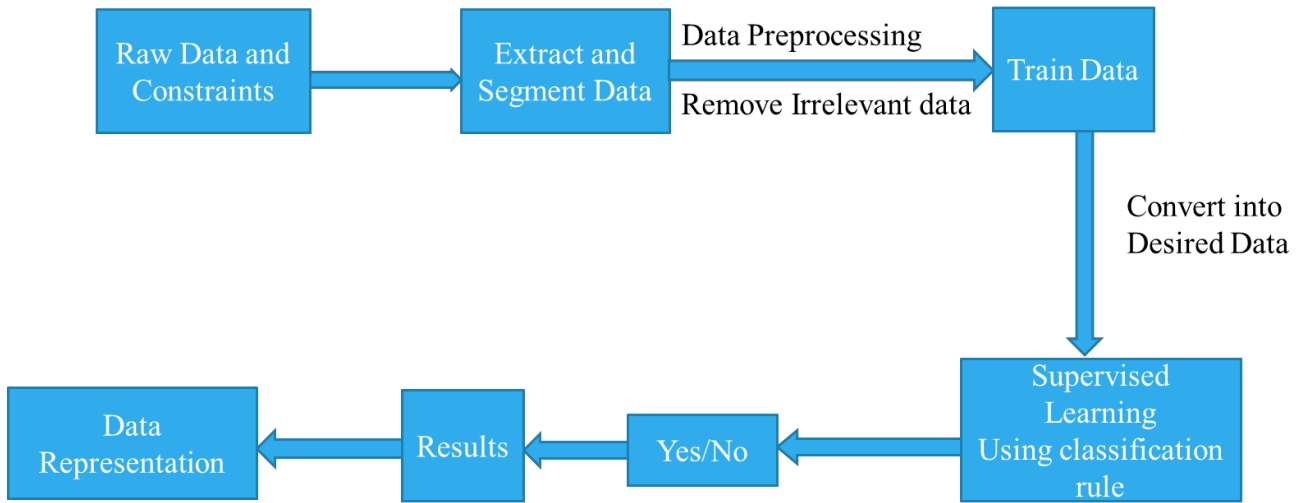
ElliotB.Sloane, Nilmini Wickramasingle and Steve Goldberg the researchers presented Wireless diabetes monitoring which is a cloud-based diabetes management consulting for patients with diabetes management.

Min Yakhudin and Rahul Joshi had presented about the ML technology that helps to identify a dataset at the elementary school, providing a situation to rescue the life. By implementing NB and K-nn algorithms.

Umatejaswi and P.Suresh Kumar had talked about algorithms such as SVM, NB, DT in order to find diseases through data mining technique.

## III. METHODOLOGY

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.



**Figure 1: Methodology**

*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.

*Result*

After taking that input data from the system will able to divine the statistics by appeal the ML algorithm & also provided the foremost output in the devise of different in between to detection the most accurate to treatment to diabetes millets.

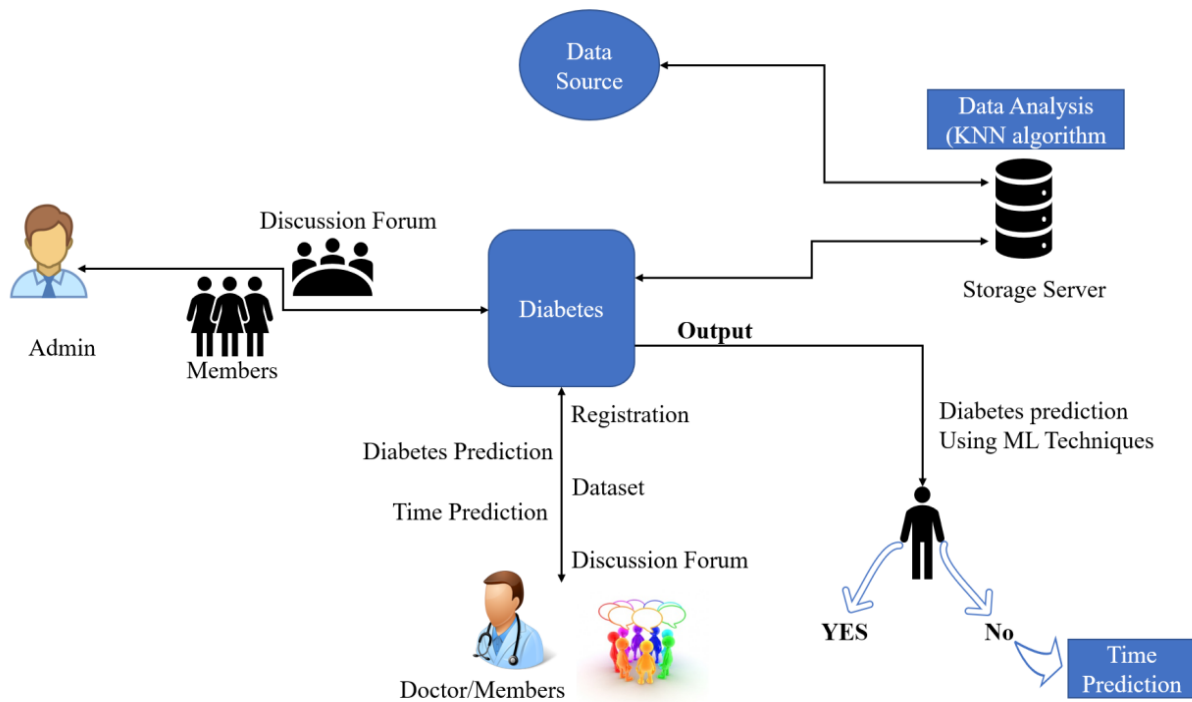
**IV. PROPOSED WORK**

The forecast or detection of diabetes is the major and concerning it is severe the complications. The diabetes complications showed in the below picture. Detection of mellitus in the starting phase and played a significant role in the heal the diabetes.

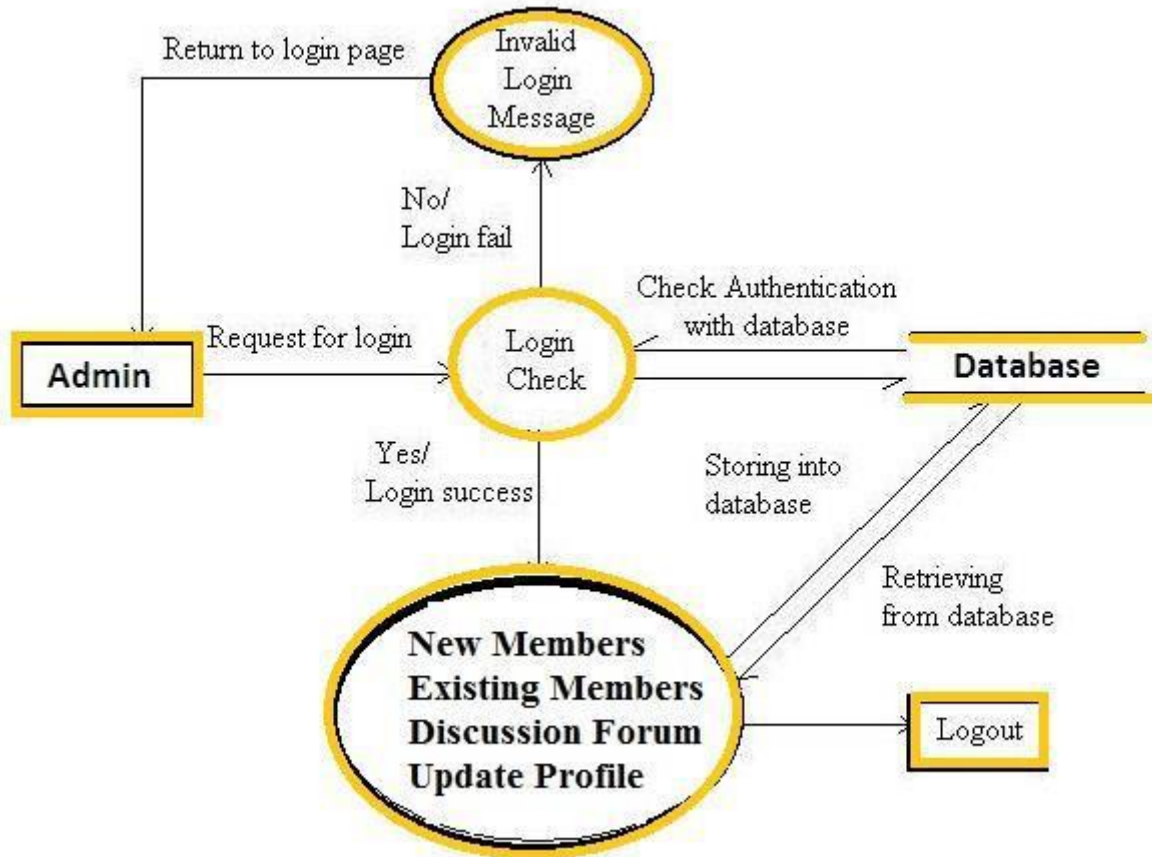
The detection diabetes is plays very important role for the human life because it leads to death. The off ered system is used to initial detection of diabetes and time prediction whereas time prediction means when the patients the diabetes it will be help to improve the habit of the patients. The proposed system is mainly centered on development of machine learning model and also it helpful in the medical sector to identify the diseases. This offer system is an automation to predicts the diabetes using old patient’s data.

*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 3: Architecture Design.**



**Figure 4: Data Flow for Admin.**

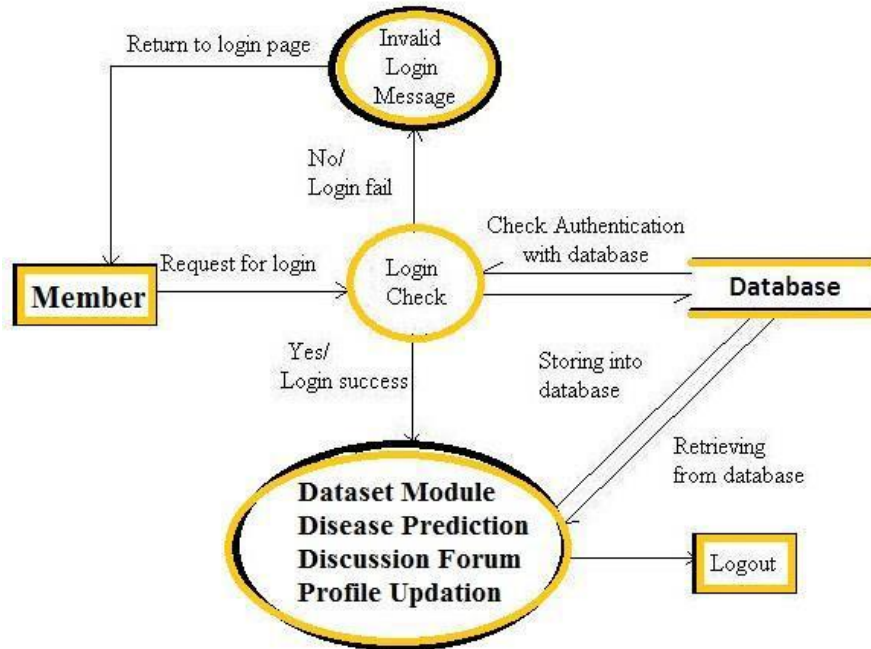


Figure 5: Data Flow design for Member.

*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.

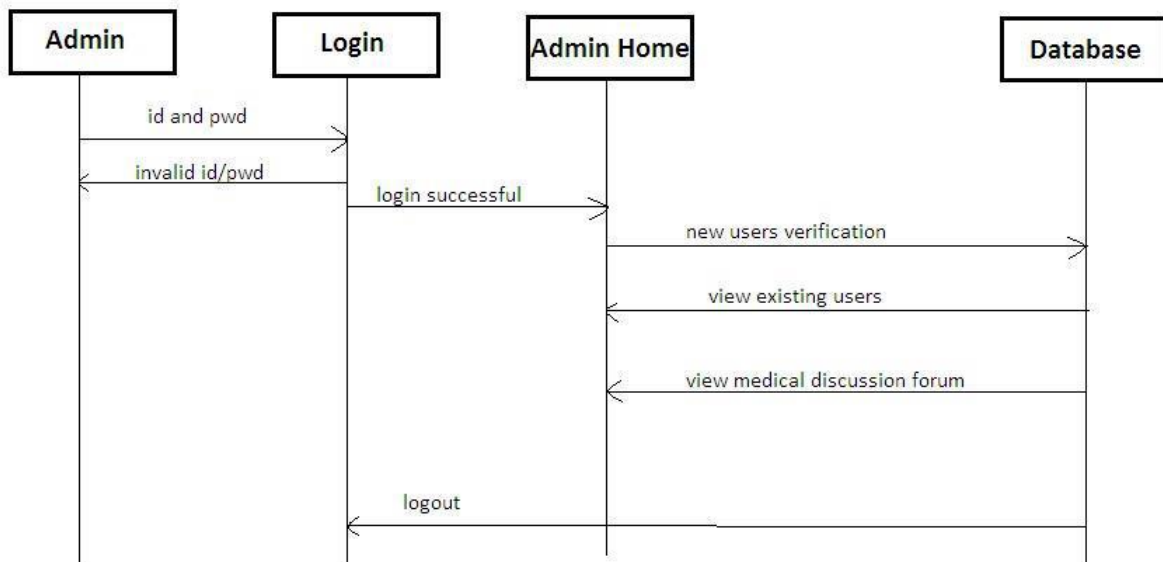
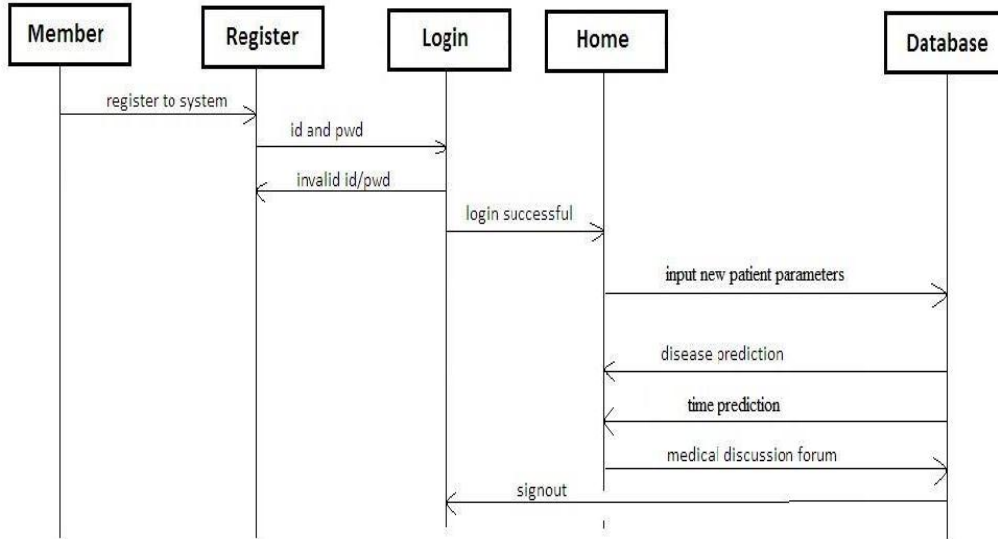


Figure 6: Sequence Diagram for Admin.



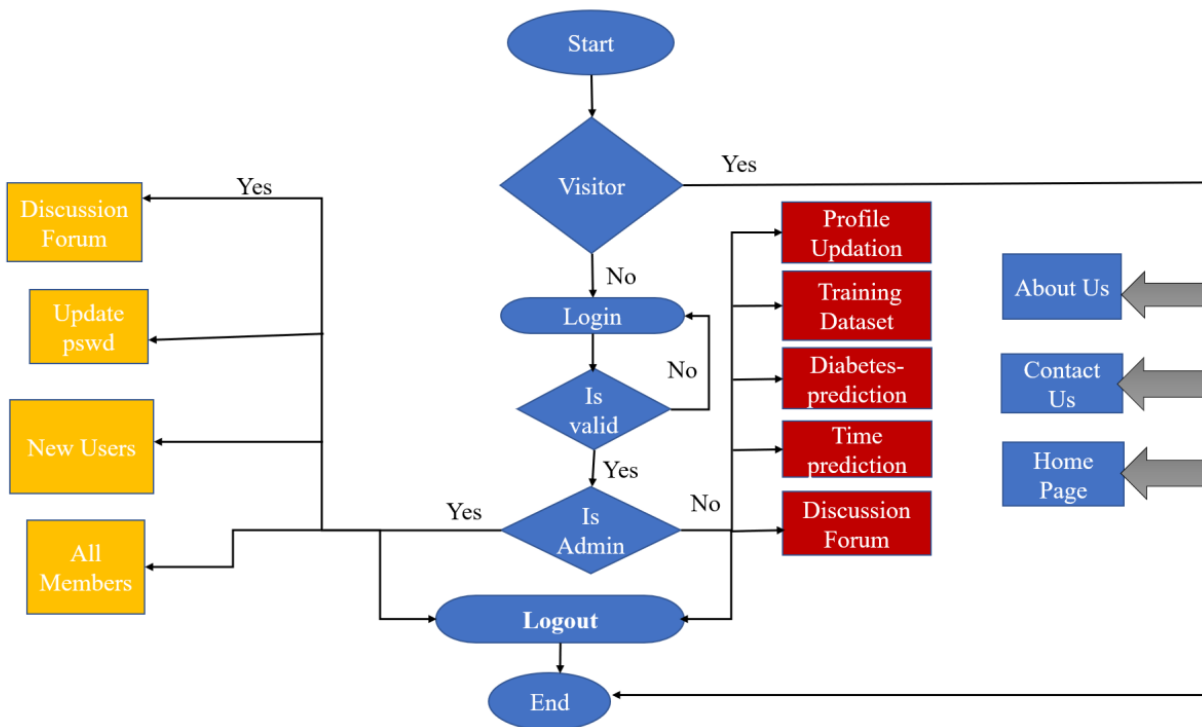
**Figure 7: Sequence Diagram for Member/Doctor.**

**V. IMPLEMENTATION AND RESULT**

*Implementation*

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 8: Control Flow**

### VI. RESULT DISCUSSION

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%

### VII. CONCLUSION AND FUTURE SCOPE

#### *Conclusion*

The prediction of diabetes is vital in today's scenario, and this is related to its causes and its threatening complications. The world's biggest cause of death is diabetes. The System model is designed to detect diabetes using a small number of indicators. System allows physicians to predict the likelihood of a person developing diabetes. So that medicines will be prescribed to the patients. System uses some of the machine learning strategies for the prediction, so as to obtain more accurate results. There has been a lot of investigation on diabetes profiles. Building a diabetes disease prediction system is really helpful for hospital administrations and physicians. When an illness is identified at an early stage, proper treatment becomes available. It is the multi hospital real time disease prediction system. Machine Learning algorithms will improve on disease prediction techniques.

#### *Future Scope*

"KNN algorithm" has been used to detect diabetes diseases, many different classification algorithms have been used and will continue to be in the future.

We can add module for visitors to ask questions for administrators and administrators can reply to those questions. We can include treatment module where doctors upload information about treatments for patients.

### REFERENCES

- [1] "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus" Md Faisal Faruque, Asaduzzaman, Iqbal H. Sarker, IEEE 2019.
- [2] "A Comprehensive Exploration to the Machine Learning Techniques for Diabetes Identification" Sidong Wei1, Xuejiao Zhao, Chunyan 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 and prediction 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 to Predict Diabetes Mellitus", International Journal ISSN 0973-4562 Volume 11, Number 1 (2016) pp 727-730.