

# A Review of Diabetic Prediction Using Machine Learning Techniques

M.Rajeswari<sup>1</sup>, Dr.P.Prabhu<sup>2</sup>

1(M.Phil research scholar, Department of Computer Applications, Alagappa University, Karaikudi  
Email: rajeswarisugan03@gmail.com)

2(Assistant professor in IT (DDE), Department of Computer Applications, Alagappa University, Karaikudi  
Email: pprabhu07@gmail.com)

## Abstract:

Diabetes is a very dangerous disease and does not able to cure. If this disease affects once, it will maintain in your life time. At the same time, your blood having too much of glucose can cause health issues. Like kidney disease, heart disease, stroke, eye problems, dental disease, foot problems, nerve damage. So you can take step to oversee your diabetes and avert these complications. The most general type of diabetes is type 1 and type 2. In this type of diabetes create problems like the body does not able to produce or use insulin. But there are also other kinds of diabetes, like gestational diabetes, which crop up during pregnancy. Gestational diabetes causes high blood sugar that can affect your pregnancy and you baby's health. Several machine learning and data mining methods are used to diagnoses diabetes and administering diabetes. This study focuses on new developments in machine learning which have made significant impacts in the detection and diagnosis of diabetes. In this work, the machine learning algorithms are used to classifying diabetes patients. The classification accuracy is achieved by the classifying diabetes patients.

*Keywords*— **Diabetes, classification, prediction, machine learning, accuracy**

## I. INTRODUCTION

Diabetes is a very dangerous disease and does not able to cure. If this disease affect once, it will maintain in your life time. At the same time, your blood having too much of glucose can cause health issues. Like kidney disease, heart disease, stroke ,eye problems, dental disease, foot problems ,nerve damage.so you can take step to oversee your diabetes and avert these complications.

The familiar types of diabetes:

- Type 1 diabetes
- Type 2 diabetes
- Gestational diabetes

### Type1 Diabetes

Body does not able to produce insulin. Its affect children and young adults. Also it can affect at any age. Peoples affected by this type of diabetes to take insulin every day.

### Type2 Diabetes

Body does not able to produce or use insulin. This type of diabetes mostly affected on middle-aged and up in years.

### Gestational Diabetes

Women's are mostly affected by this type of diabetes. This type of diabetes develops during pregnancy. Gestational diabetes causes high blood sugar that can affect your pregnancy and you baby's health.

### A. Challenges

Diabetes is also caused to vision problems. It reduces blood glucose level in retina in aged diabetes patients. In future it makes cataracts to the diabetes peoples, and it caused poor vision very easily. Vision problem makes a big trouble to the patients and it is affect their daily activity.

Diabetes caused hearing loss. Protracted high blood glucose levels may lead to hearing problem by affecting the supply of blood and oxygen to the tiny nerves and blood vessels of the inner ear. Over time, the nerves and blood vessels become damaged, affecting the person's ability to hear. It makes miscommunication between peoples. So every diabetes patients must check the hearing ability.

People who affected by diabetes they are literally very weak and health less. So the diabetic people need to increase and maintaining their daily

activity level. Food control is very important and another important part is physical activity. Diabetic patients must be kept the habit of daily exercise. Like brisk walking, bicycling, swimming, housework and gardening etc.

## II. CLASSIFICATION AND PREDICTION TECHNIQUES

Classification is a procedure to categories data into a desired and distinct number of classes where we can assign label to each class. And prediction models predict continuous valued functions.

Effectively diagnosing these problems are used to data mining and machine learning techniques. Classification is one of the significant procedures for disease prediction. Classification is the most well-known data mining tasks. Huge amount of business and medical data sets usually involves classification. Classification is a data mining function that can allocates the items in a collection to target categories. Expected accuracy is achieved by classifying diabetes patient datasets. Like J48, SVM, Naive Bayes, Decision Tree, logistic Regression, ANN etc. It is better to diagnose different diseases.

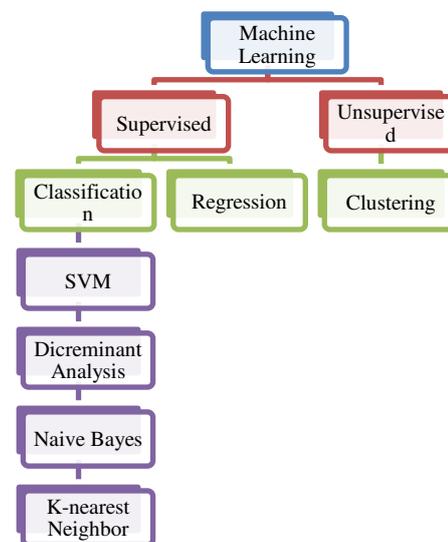


Fig.1 Flow of machine learning

Classification is one of the most significant decision making techniques in many real world problem. In proposed, using machine-learning algorithms. In this work, the key objective is to classify the data as diabetic or non-diabetic and improve the classification accuracy. For several

classification problem, the higher amount of samples chosen but it doesn't leads to higher classification accuracy. In many aspects, the performance of algorithm is high in the context of speed but the accuracy of data classification is low. The main objective of our model is to achieve high accuracy. Classification accuracy can be increase if we use much of the data set for training and few data sets for testing. This study has analysed various classification techniques for classification of diabetic and non-diabetic data.

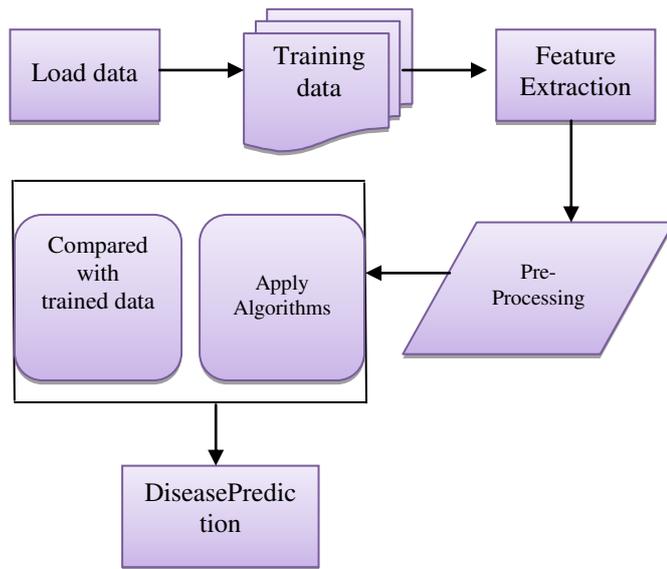


Fig.2 Fundamental architecture for prediction of disease

This study focused on prediction of diabetes at early stage. This is the step for avoid diabetes complications and stay healthy. In proposed, diabetes prediction using classification algorithms. Like PIDD (Pima Indian Diabetes Dataset) datasets are used for experimentations. Various measures are used to check the performance of algorithm or quality of algorithm. Hence it is achieved better accuracy.

**B. Classification Applications:** speech recognition, handwriting recognition, biometric identification, document classification, spams filtering, sentiment analysis, biological classification, pattern recognition, credit scoring etc.

### III. REVIEW OF LITERATURE

DeeptiSisodia<sup>a</sup>, Dlip Singh Sisodia<sup>b</sup>[1] this paper, work is diabetes prediction at early stage. Decision tree, SVM and Naive Bayes classification

algorithms are used to prediction. Accuracy is evaluated using WEKA tool. The Naïve Bayes gave the highest accuracy.

Xue-hui menget al.in [2] this study was comparison of three algorithms for predict the diabetes or prediabetes using common risk factors. The logistic algorithm, ANN and decision tree algorithms are compared. To test 735 patients, they are came from two communities in Guangzhou, china. The highest accuracy (77.87%) level is achieved by classification algorithm (C5.0).

V.AnujaKumariet al.in [3].the classifier is predict diabetes disease with optimal cost and better performance using SVM and pima indian diabetes dataset. Implemented by matlab R2010a. The result of the problem accuracy is 78.2%.

Monisha.A et al. in [4] in machine learning, different classifiers are used for predicting and diagnosing diabetes. Like Naive Bayes statistical modelling, logistic regression, Extreme Gradient Boosting. Pima Indians diabetes datasets are experimented. The accuracy for Extreme Gradient Boosting algorithm is 81%.which is greater than other two algorithms.

S.Selvakumar et al. in [5] Discussed about diabetes challenges. Data mining methods are used to predict the peoples whether diabetic or not. Binary Logistic Regression, Multilayer perception and K-Nearest neighbour algorithms are classified. The accuracy level in Binary Logistic Regression is 0.69, Multilayer Perception is 0.71 and K-Nearest neighbour is 0.80. K-Nearest neighbour is the highest accuracy than Binary Logistic Regression, Multilayer perception.

Aiswarya Iyar et al. in [6] worldwide 246 million peoples are affected by diabetes. These are increased over 380 million in 2025 reported by WHO .this paper in aim is finding solution to diagnosis the disease. Using decision tree and Naive Bayes algorithms. Weka tool is used for implementation. The naive Bayes algorithm is obtained 79.5652% of accuracy.

B.Tamilvanan et al. in [7] this paper in objective is predict diabetes with more accuracy. The three classification algorithms are compared for accuracy rate, namely Naive Bayes, Random Forest and NB-Tree. Implementation using weka tool. The result is Naive Bayes has the best predictive capacity with

highest accuracy rate (76.3%) and least error rate (23.7%).

Rahul Joshi et al. in [8] prediction of medical datasets at an early stage is safe for human life using machine learning techniques. To test the Pima Indians diabetes dataset. The applied algorithms are KNN, Naive Bayes, Random Forest and J48. We get the best result to ensemble approach, when combining individual techniques and methods. It is also called hybrid model. This provides the best performance and accuracy than the single one. Weka and java tools are used to predict diabetes.

Amina Azar et al. in [9] Diabetes affected among young peoples and ancient peoples. These are increased day by day and it does not curable. Data mining is used to early stage prediction. This paper in main aim is gives the differentiation and suggest best algorithm. The PID datasets are used. The Decision tree, Naive Bayes and K-Nearest neighbour algorithms are compared and used for predict the diabetes diagnosis at early stage with highest accuracy and efficiency. The WEKA is used for testing and validation using rapid miner. The result is the decision tree is the best prediction algorithm. It gives the accuracy level is 75.65%.

VeenaVijayan.V et al. in [10] Choosing appropriate algorithms for classification clearly increases the accuracy and efficiency of the system. The main objective of this study is to analysis the benefits of different pre-processing techniques for decision support systems for predicting diabetes which are based on Support Vector Machine (SVM), Naive Bayes classifier and Decision Tree. The pre-processing methods focused on this study are Principal Component Analysis and Discretization. The accuracy variation evaluated with and without pre-processing techniques. Weka tool is used in this study. The dataset was taken from the University of California, Irvine (UCI) repository of machine learning.

DeepikaVermaet al.in [11] in this paper take two disease datasets. That is breast cancer and diabetes dataset from UCI machine learning repository. In this paper used WEKA tool that is good classification tool. To classify breast cancer and diabetes dataset using Naive Bayes, SMO, REP tree, J48 and MLP algorithms on WEKA interface. After analysing the performance of all algorithms, J48

achieves 74.28% accuracy level than the other algorithms on breast cancer dataset and SMO achieves 76.80% accuracy level on diabetes datasets.

VeenaVijayan V etal.in [12] this paper focused on prediction of diabetes mellitus with low error rate. Datasets are obtained from UCI machine learning repository. Accuracy verification using Matlab and Weka tool. AdaBoost-decision stump classifiers are used to predict diabetes with low level error rate and its gives 80.729% of accuracy.

Ayan Mir et al.in [13] this paper focused on diabetes prediction. Pima Indians diabetes datasets are used. Naive Bayes, SVM, Random Forest and Simple CART algorithms are used for classification on Weka interface. The result is SVM provide the best accuracy than the others.

S M Hasan Mahmud et al.in [14] predict the condition of diabetes. In this paper considered five most important machine learning classification techniques for predicting diabetes.10-fold cross validation methods applied to find the performance measurements of the classification techniques. The result of the analysis shows that Naive Bayes achieved highest performance than the other classifiers, obtaining the F1 measure of 0.74.

AakanshaRathoreet al.in [15] Detect and predict diabetic disorder. Experimental using Pima Indians diabetes dataset and evaluated the performance measurements using R Studio. Two machine learning algorithms are used SVM and Decision Tree. The SVM provides 82% of accuracy.

TABLE 1: COMPARATIVE STUDY

S. NO	ORIGINATOR WITH TITLE	DATASET	ALGORITHM	TOOL	OUTCOME & ACCURACY
1	DeeptiSisodia, Dilip Singh Sisodia.” Prediction Of Diabetes Using Classification Algorithm Prediction Of Diabetes Using Classification Algorithm”	PIDD	Decision tree, SVM and Naive Bayesian.	Weka	Diabetes detection at early stage. Accuracy 76%.
2	Xue-Hui Meng, Yi-Xiang Huang, Dong-Ping Rao, Qiug Liu.” Comparison of three data mining models for predicting diabetes of prediabetes by risk factors”	To test 735 patients, they are came from two communities in Guangzhou, china	Logistic algorithm, ANN and Decision tree algorithms	Not Mentioned	Comparisons of three algorithms for predict the diabetes or prediabetes using common risk factors. The highest accuracy (77.87%) level is achieved by classification algorithm (C5.0).
3	Monisha.A, S.ShalinChistina, Nirmala Santiago. “Decision support system for a chronic disease-Diabetes”	PIDD	Naive Bayes statistical modelling, logistic regression, Extreme Gradient Boosting	R programming	The accuracy for Extreme Gradient Boosting algorithm is 81%
4	S.Selvakumar, K.Senthamarai Kannan and S.GothaiNachiyaar.” Prediction Of Diabetes Diagnosis Using Classification Based Data Mining Techniques”	Multi-dimensional healthcare dataset	Binary Logistic Regression, Multilayer perception and K-Nearest neighbor algorithms	Not mentioned	The accuracy level in Binary Logistic Regression is 0.69, Multilayer Perception is 0.71 and K-Nearest neighbour is 0.80.
5	Aiswarya Iyar, S. Jeyalatha and RonakSumbaly, “Diagnosis Of Diabetes Using Classification Mining Techniques”	Pima Indians Diabetes Database of National Institute of Diabetes and Digestive and Kidney Diseases	Decision tree and Naive Bayes algorithm	WEKA	The Naive Bayes algorithm is obtained 79.5652% of accuracy.
6	B.Tamilvanan, Dr.V.MuraliBhaskaran, “An Experimental Study Of Diabetes Disease Prediction System Using Classification Techniques”	Medical database for diabetes Disease dataset from UCI.	Naive Bayes, Random Forest and NB-Tree	WEKA	The result is Naive Bayes has the best predictive capacity with highest accuracy rate (76.3%) and least error rate (23.7%).
7	Rahul Joshi, MinyechilAlehegn, “Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach”	PIDD	KNN, Naive Bayes, Random Forest and J48	Weka 3.8.1 and java using NetBeans 8.2	To get the best result in ensemble approach, when combining individual techniques and methods. Also called hybrid model. This provide best performance and accuracy than the single one
8	Amina Azar, YasirAli, MuhammadAwais, KhurramZaheer, ” Data Mining Models Comparison for Diabetes Prediction”	PIDD	Decision tree, Naive Bayes and K-Nearest Neighbour algorithms	WEKA	The result of this paper is the decision tree is the best prediction algorithm. It gives the accuracy level is 75.65%.
9	VeenaVijayan.V, Anjali.C, ” Decision Support Systems for Predicting Diabetes Mellitus –A Review”	PIDD	Support Vector Machine (SVM), Naive Bayes classifier and Decision Tree.	WEKA	The accuracy variation with and without pre-processing techniques are also evaluated. Decision tree was constructed by including numerical discretization as a pre-processing technique in “Decision discovery for the diagnosis of Type IIDiabetes” the resulting accuracy was 78.1768
10	DeepikaVema ,Dr.Nidhi Mishra, “Analysis the prediction of breast cancer and diabetes disease dataset using data mining classification techniques”	Breast cancer and diabetes disease dataset from UCI.	Naive Bayes, SMO, REP Tree, J48 and MLP algorithms.	WEKA	J48 achieves 74.28% accuracy level than the other algorithms on breast cancer dataset and SMO achieves 76.80% accuracy level on diabetes datasets.
11	VeenaVijayan V, Anjali C, ”Prediction of diagnosis of diabetes mellitus –A machine learning approach”	Dataset obtained from UCI machine learning repository.	AdaBoost	Matlab – Weka	AdaBoost-decision stump classifiers are used to predict diabetes with low level error rate and its gives 80.729% of accuracy.
12	AakanshaRathore, Simran Chauhan, SakshiGujral, “Detecting and Predicting Diabetes Using Supervised Learning: An Approach towards Better Healthcare for Women”	Pima	SVM and Decision Tree.	R-Studio	SVM gives the 82% of accuracy.

#### IV. CONCLUSIONS

This study focused on machine learning classification algorithms for predicting diabetes disease with more accuracy. This literature study in SVM classification algorithm achieved highest accuracy. The different performance values of classification algorithms are calculated on various measures. Train and test the data like Pima Indians Diabetes Dataset. The classification algorithm achieved maximum testing accuracy. This study has collected various classification techniques and ensemble them for the purpose of improving accuracy, specificity and sensitivity.

#### REFERENCES

1. Deepti Sisodia, Dilip Singh Sisodia, "Prediction of Diabetes Using Classification Algorithm", [www.elsevier.com/locate/procedia](http://www.elsevier.com/locate/procedia), *Procedia computer science* 132(2018) 1578-1585.
2. Xue-Hui Meng, Yi-Xiang Huang, Dong-Ping Rao, Qiug Liu, 2013, "Comparison of Three Data Mining Models For Predicting Diabetes of Prediabetes By Rick Factos", *Kaohsiung journal of medical science*(2013) 29,93-99.
3. V.Anuja Kumari, R.Chithra. "Classification of Diabetes Disease Using Support Vector Machine".vol 3., Issue 2, March-April 2013, pp.1797-1801. [www.ijera.com](http://www.ijera.com).
4. Monisha.A, S.Shalin Chistina, Nirmala Santiago, "Decision support system for a chronic disease-Diabetes". *International Journal of Computer & Mathematical Science(IJCMS)*, ISSN 2347-8527 Volume 7, Issue 3, March 2018.
5. S.Selvakumar, K.Senthamarai Kannan and S.Gothai Nachiyar, "Prediction of Diabetes Diagnosis Using Classification Based Data Mining Techniques", *International Journal of statistics and Systems*, ISSN 0973-2675 Volume 12, Number 2(2017), PP.183-188. <http://www.ripublication.com>.
6. Aiswarya Iyar, S. Jeyalatha and Ronak Sumbaly, "Diagnosis of Diabetes Using Classification Mining Techniques", *International Journal of Data Mining & Knowledge Management Process (IJDMP)* Vol.5, No.1, January 2015.
7. B.Tamilvanan, Dr.V.Murali Bhaskaran, "An Experimental Study of Diabetes Disease Prediction System Using Classification Techniques", *IOSR Journal of Computer Engineering (IOSR-JCE)*, e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 19, Issue 1, Ver. IV (Jan.-Feb. 2017), PP 39-44, [www.iosrjournals.org](http://www.iosrjournals.org).
8. Rahul Joshi, Minyechil Alehegn, "Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach", *International Research Journal of Engineering and Technology (IRJET)*, Volume: 04 Issue: 10 | Oct -2017, e-ISSN: 2395-0056, p-ISSN: 2395-0072. [www.irjet.net](http://www.irjet.net).
9. Amina Azar, Yasir Ali, Muhammad Awais, Khurram Zaheer, "Data Mining Models Comparison for Diabetes Prediction", (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol. 9, No. 8, 2018.
10. Veena Vijayan.V, Anjali.C, "Decision Support Systems for Predicting Diabetes Mellitus –A Review", *Proceedings of 2015 Global Conference on Communication Technologies (GCCT 2015)*, 978-1-4799-8553-1/15/\$31.00 © s2015 IEEE.
11. Deepika Verma, Dr.Nidhi Mishra, "Analysis and prediction of breast cancer and diabetes disease dataset using data mining classification techniques", *Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017)*, IEEE Xplore Compliant - Part Number: CFP17M19-ART, ISBN:978-1-5386-1959-9.
12. Veena Vijayan V, Anjali c, "Prediction of diagnosis of diabetes mellitus –A machine learning approach", *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 10-12 December 2015 | Trivandrum*.
13. Ayman Mir, Sudhir N. Dhage, "Diabetes Disease Prediction using Machine Learning on Big Data of Healthcare", *2018 Fourth International Conference on Computing Communication Control and Automation (ICCCUBEA)*.
14. S M Hasan Mahmud, Md Altah Hossin, Md. Razu Ahmed, Sheak Rashed Haider Noori, Md Nazirul Islam Sarkar, "Machine Learning Based Unified Framework for Diabetes Prediction", *BDET 2018, August 25-27, 2018, Chengdu, China*. © 2018 Association for Computing Machinery. ACM ISBN 978-1-4503-6582-6/18/08...\$15.00. DOI: <https://doi.org/10.1145/3297730.3297737>.
15. Aakansha Rathore, Simran Chauhan, Sakshi Gujral, "Detecting and Predicting Diabetes Using Supervised Learning: An Approach towards Better Healthcare for Women", Volume 8, No. 5, May-June 2017, ISSN No. 0976-5697, Available Online at [www.ijarcs.info](http://www.ijarcs.info).

