Analysis of Diabetic Mellitus Using Predictive Algorithm – A Literature Review

Usha Nandhini A[1], Dr. K. Dharmarajan[2]
{ushanandhini.81090@gmail.com[1], dharmak07@gmail.com[2]}

Research Scholar, VISTAS[1], Associate Professor, Department of Information Technology, VISTAS[2]

Abstract. An enormous development of Technology growth of healthcare contains plenty number of data. To access a massive data, we are using Big Data Analytics. Diabetes is a leading disease and causes a death & economic development. In healthcare industries doctor prescribed medicine to the patient based on patient medical history and measurement of sugar level. Predictive algorithm helps to analysis, detect and predict disease at early period. It helps doctor to diagnosis disease and gives treatment to the patient respectively. The moto of the paper to research the diabetes prediction with help of Predictive analysis algorithm to predict the diabetes disease accurately.

Keywords: Diabetes, machine learning, datamining techniques.

1 Introduction

Diabetes is a familiar disease in universe. DM creates major health problem in the development countries. 5% of Indian peoples are affected by diabetic mellitus among 145 million people worldwide [1]. In traditional system research process can be executed on single window. In developed healthcare system contains large execution of irregular database to load these databases will create more complexities [2]. The known diabetes types are Type 1-DM, Type 2-DM, Prediabetes and Gestational diabetes. The youngster and grownup people are affected by Type1 and Type 2 respectively. The risk of diabetes is eye, kidney, liver, heart, and foot problem [3]. The healthcare system contains a large amount of information. The many details will be stored in the type of hardcopy [4]. To store and process the massive data we use Hadoop. Type1 test are C peptide test, a finger stick glucose test and urine test. Type2 test are blood sugar in empty stomach, 120 minutes post prandial test, selected blood sugar, HA1C test and oral glucose tolerance test. Women's are highly contributed by diabetes compared to men. During pregnancy diabetes will impact the mom and baby.

Female with sugar problem will face heart problem, miscarriage or children will reach the earth with abnormal issues. DM causes due to family heredity and environment factors. The family history shows a higher risk of diabetes in close family members. The classification is used to predict the diabetes results in accurate manner. The moto of analysis tasks were to understand the concepts of Predictive algorithm which is used to anticipate diabetics.

ICASISET 2020, May 16-17, Chennai, India

DOI 10.4108/eai.16-5-2020.2304019

2 Literature Review

(B.G. Mamatha Bai et al., 2019) described that Datamining techniques help to diagnose the disease of patient and intimate the doctor to provide treatment at early stage. It decreases the cost and gives better treatment. The techniques are Naïve Bayes, BIRCH, and OPTICS. Naïve Bayes techniques are imposed to anticipate the output based on diabetics patient. BIRCH is applied to the dataset where it will help to eliminate the error. OPTICS are grouped into three cluster are normal, prediabetes, diabetic. OPTICS algorithm gives better results compared to BIRCH algorithm.

(Sunil kumar et al., 2019) briefed that Datamining Techniques helps to examine the scientific record of diabetics people. The techniques are Decision Tree, Naïve Bayes Algorithm, Hadoop Framework, HDFS, MapReduce. To count the Diabetic patients those techniques are used. Based on date, time, code and value the result is predicted. Decision tree is higher to Naïve Bayes algorithm. Decision tree preforming well compared to Naïve Bayes algorithm.

(N.Sneha et al., 2019) briefed that Predictive Analysis Algorithm using optimal feature selection used to analysis a Diabetic Mellitus. The techniques are Decision Tree, Naïve Bayesian, Random Forest, K-Nearest Neighbor and Support Vector Machine. Predictive analysis algorithm used to identify and cure the diabetics at initial period with classification of techniques. To execute the research of the feature in the dataset and select the optimal feature origin on the correlation values.

(Muhammad Younus et al., 2020) described that Predictive analysis is used to predefine the type 2 Diabetic Mellitus. Predictive analyses are Data Pre-processing, Decision Tree, and Random Forest. Robust scaling and label encoding method are used in data pre-processing techniques. The tools are Scikit-Learn, Python, R. Dataset is collected from Non-government hospital in Dhaka city. According to the model, among the attributes the hba1c >7 and BMI level >20 which we are able to identify the complications of diseases.

(Uswar Ali Zia et al., 2017) stated that the machine learning techniques supports to get know the correct result. The techniques are Naïve Bayes, Decision Tree, K-Nearest Neighbor. To improve accuracy, Bootstrapping resampling techniques are used. The big data analysis is used to make effective healthcare data and reduce the treatment and unexpected cost. The Decision Tree algorithms provide a better accuracy compare to other algorithms.

(Wwifeng Xu et al., 2017) appreciated that the single classifier is not up to mark compare to ensemble classifier method. Random Forest nothing but ensemble classifier. Its collection of different decision trees. The random forest has high accuracy and good robustness. Data pre-processing and Random Forest are used. The main indicators of dataset are age, height, waist, hip. Based on Hemoglobin <7.0(normal) and >7.0(abnormal) the unwanted features are removed. K-means Discretization is used for better performance. 10-fold cross validation is designed for segregated output. Random Forest method assumes the impact of diabetics effectively with sufficient amount of data. It maintains the cost and re-admission of the patient.

(Madhuri Panwar et al., 2016) improved that the classification performance. The techniques are Data Pre-processing, PCA, K-NN. K-nearest neighbor diagnosis the result accurately. K-fold cross validation is designed for segregated output. One measure for testing and remaining measure for training set. It gives 100% result in accuracy, specificity, and sensitivity. It gives 100% result with equally convention and K-fold cross validation.

(T. Santhanam et al., 2015) implemented that the different predicting algorithms are K-means, Genetic algorithm and Support Vector Machine. K-mean algorithm is help to

eliminate the outside frame, inconsistent and noisy data. Genetic Algorithm is used to reduce data by applying features selection method. SVM Classifier gives better accuracy with help of 10-fold cross validation techniques. Random Forest algorithm is predicting the data more accurate and correctly.

(Saumendra Kumar et al., 2019) stated that to identify the diabetes with MLP neural network. Nonlinear problem can be solved using Multilayer perceptron neural network. MLP classifies input patterns and predict the result. It predicts less result but it more efficient by identifying the pattern.

(P. Prabhu et al., 2019) briefed that the Deep Belief Network is used to predict the data accurately. Deep belief network contains three phases they are pre-processing, pretraining and fine-tuning. The DBN method was effectively framed, skilled, authenticated and verified. The DBN is high effectively producing the value of recall, precision and F1 measure.

(S. Nivetha et al., 2019) resolved that clustering techniques used to predict a disease accurately. The clustering techniques are K-means, farthest first, Density based, Filtered and X-Means methods. SMOTE algorithm with oversampling. K-Means and X-Means algorithm predict better results. The accuracy high in F-measure 75.64%.

(Emrana Kabir Hashi., 2017) briefed that the approached loop helps doctor to predict disease correctly. The techniques are Decision Tree, K-NN. The decision tree contains very high accuracy. The system predicts 100% accuracy in decision tree training phase. The testing stage C4.5 and KNN produce 90.43% and 76.96%.

(Wenqian Chen et al, 2017) discovered that the prediction method is used to predict type2 diabetes. The techniques are Data Preprocessing, K-Mean, Decision Tree along with 10-Fold Cross-Validation Method. It gives 90.4% accuracy in hybrid prediction model.

3Methodology

3.1 Data Collection

In this research the authors have collected data from Medical Hospitals, UCI Machine Repository and Pima Indian Dataset. Maximum taken dataset has 768 features and 9 attributes from women diabetes. The features are number of times being pregnant, Glucose tolerance test, Blood pressure level, Body mass index, Triceps, Insulin, Age, Pedigree function and diabetes positive or non-diabetes. To predict diabetes 20 above years of people information were collected. Using BMI, Hba1c, and Age, hip, waist is used as a feature selection to predict a result accurately. Some other features also used other than attributes like hereditary, lifestyle, food habits, stress, etc.

3.2Data Preprocessing

In Data Preprocessing method the robust scaling is used to remove an outlier. The label encoding and one hot encoding is used to divide the label and encrypt the whole labels [4]. The information can be sensibly composed, unified as well as arranged used for examination. The data pre-processing is used to develop the value of mining results and the effectiveness of the method. Dimensionality Reduction is used to handle inappropriate attributes and noise. K-

means discretization is a finest method to split a data [6]. Exchange the unavailable value and not possible value with average [13]. Data cleaning method is used to find a misplaced value, eliminate irrelevant data and correct unreliable data in dataset. Data reduction is used to reduce data set [7]. It eliminates the inaccurately categorized data through K-means algorithm [13]. PCA expands Principal Component Analysis. PCA is a preprocessing method for dimensionality reduction. It reduces the redundant dimension from top to bottom dimensional dataset [7]. Wrapper method is used to select an attribute. It is used to select a best attribute in given dataset. Wrapper techniques are greedy forward direction, backward direction and bidirectional [11]. GA is a feature selection technique. It is used to select a feature [8].

3.3Predictive Algorithms

3.3.1 Naïve Bayes

Naïve Bayes algorithm is utilized for making the classifier. It is used to check the patient having diabetic or non-diabetic [1]. Naïve Bayes classifier is used through Bayes theorem. It is designed to find an expectation classification with independence of attributes [2]. The Naïve Bayes will take an input and produce the unknown class label. This classification technique used for bigger dataset [3].

3.3.2OPTICS

OPTICS expands as Ordering Points to Identify Cluster Structure. OPTICS is a clustering algorithm is used to group a data into cluster. OPTICS algorithm is designed for combining the data points to measure the density of object. It has two concepts they are Core Distance, Reachability Distance [1].

3.3.3BIRCH

BIRCH expands as balanced iterative reducing and clustering using hierarchies. This algorithm uses clustering techniques to build a CF tree. It supports hierarchy concepts. It will operate on higher dataset. BIRCH algorithm is used to remove a noise in the data. In this method for every scanning, it will create a CF tree to reduce an outlier. The clustering algorithm is used to group a data into cluster [1].

3.3.4Decision Tree

Decision Tree is a supervised learning algorithm. It is used to classify the categorical records. It uses a decision-making method. Decision tree is used for classification and regression. The tree contains two types they are classification tee and regression tree. A classification tree takes distinct values and regression tree take a continuous value. Decision tree is looks like a flowchart. In decision tree each internal node is a test on attribute, each branch produce the result of the test and leaf node denote a class label [2]. The decision tree will take an input and produce the output based on decision model [3]. The problem is arrived in decision model are splitting attribute, ending conditions, data trimming, train the example, direction of splits [3]. To improve the precision the Bootstrapping resampling technique is used [5]. The Information Gain is used to calculate the unidentified data [12]. The general decision tree algorithms containing ID3, CART, C4.5, C5.0, J48 etc. [13][16].

3.3.5Random Forest

The Random Forest is a supervised learning. RF is used in classification and regression model. It uses a random sample. It processes from the root node and split the attributes randomly. It uses bootstrapping or bagging techniques for selecting the attributes [3]. In Random Forest using Bootstrapping randomized resampling technique used to mine the model set from novel dataset then build a decision tree for each model set. Finally, combined the results to predict the classifications [6].

3.3.6Support Vector Machine

Support Vector Machine is a supervised learning. The SVM is used in classification and regression model. It creates hyperplane between various classes [3]. SVM care a many endless and definite values [8].

3.3.7K-Nearest Neighbor

K-Nearest Neighbor uses classification techniques. It classifies new cases based on similarity of measures [3]. K-nearest neighbor is an idle knowledge. In which all cases are stored and find new cases based on the limited neighbor. K-NN is used find nearest neighbor first and distance neighbor second [7]. KNN is a supervised learning technique. It defines the novel data depends on shortest length reference to novel data to the K nearest neighbor. Euclidean Distance is used to find the nearest data [12].

3.3.8Clustering Algorithm

Clustering algorithm is used to group a data into a cluster. Clustering algorithm contains different techniques; they are K-Means, Farthest First, Density Based, Filtered and X-Means. K-Means is a unsupervised clustering. It was found by MacQueen in 1976. In K-Means method choose randomly data first and if any data point is closer to the cluster use it or if any data point is not closer to the cluster leave and move to the nearest cluster and repeat the process for finding a meaningful cluster [8][17]. In this method the centroids is defined. Farthest First method is used for larger dataset. It is different from k-means techniques. It will locate the center of the cluster and move to furthest point of the nearby centre. Density Based clustering is unsupervised learning method. It is used to discover the non-linear structure depend on density. It contains two techniques are density reachability and density connectivity. Filtered clustering is used to reduce the cluster count. It will operate vital role in ramp-up the output. The order of split up between the starting cluster centre trigger the efficiency of filtered techniques. X-Means algorithm is stretched type of K-Means algorithm. It acts to manually find the cluster count through Bayesian Information Criterion. Bayesian Information Criteria supports in calculating the splitting choice [11].

3.3.9Multilayer Perceptron

MLP is a simple neural network. In MLP the Back-Propagation method is used for classification. Multilayer Perceptron is feed-forward neural network method. It has layers. The linear problem is solved using Single Layer Perceptron. The non-linear problems can be

solved using MLP. It has many unseen layers. It reorganizes the arrangement and classifies the input and expects the output [9].

3.3.10Deep Belief Network

Deep Belief Network contains three stages, they are preprocessing, pre-training, fine-tuning. In preprocessing stage, the normalization (min-max normalization) is applied in dataset. In pre-training stage train the layer from the bottom layer to top layer orderly. In fine-tuning stage all the hidden layers are tuned and provide the results [10].

Table 1. Assessment of varıous algorıthms.

S.no	Paper Title/Year/Author	Objectives	Techniques	Dataset	Accuracy/ Result
1.	Analysis and Detection of Diabetes using Data Mining Techniques-A Bigdata Application in Healthcare/2019/B.G. Mamatha Bai, B.M.Nalini and Jharna Majumdar	In modern world, data is ramping up unimaginably and Big Data Analytics is involving and impact the diagnosis platform. The feature of data mining in medical system is to frame the self-analysis tool to produce the report in accurate manner.	Naïve Bayes, BIRCH, OPTICS	UCI Machine learning repository	OPTICS: Precision-0.735849, Recall-0.795918, Fmeasure-0.7647058, Rand Index- 0.7108433. BIRCH: Precision-0.523809, Recall-0.523809, Fmeasure-0.5238095, Rand Index-0.53810762.
2.	Diabetes Data Analysis using MapReduce with Hadoop/2019/Sunil kumar and Maninder Singh	Bigdata is the reservoir of large amount of data. In medical system, Doctors direct the insulin to the sugar patients and the selecting the treatment is depends on the patient's past record and amount of sugar level at the normal intervals	Decision Tree, Naïve Bayes Algorithm	Dataset from UCI Machine learning. It contains 4 attribute date, time, code, value and 70 dataset files.	DECISOIN TREE: TPrate-1.00, FPrate-0.037, Precision- 1.0, Recall-0.79, Execution Time- 2. NAÏVE BAYES: TPrate-0.901, FPrate-0.072, Precision- 0.7, Recall-0.52, Execution Time-14.
3.	Analysis of diabetes mellitus for early prediction using optimal feature selection/2019/N.Sneha and Tarun Gangil.	To form the use of important techniques, frame a predictive algorithm using ML and find the optimal classifier to get us the nearby Value.	Decision Tree, Naïve Bayesian, Random Forest, Support Vector Machine, K-Nearest Neighbor	Dataset from UCI Machine Repository. It contains 15 attributes and 768 data items.	ACCURACY: Support vector machine-77.73, Random Forest-75.39, Naïve Bayes-73.48, Decision Tree-73.18, K-nearest Neighbor- 63.04
4.	Prediction Model for Prevalence of Type-2 Diabetic Mellitus Complication using	Maximum peoples are impact from the genetic problem due	Data Preprocessing, Decision Tree, Random Forest	Dataset for Non-government hospital in Dhaka city. 26000-	Accuracy Measurement MALE- CVD:

	Machine Learning Approach/2020/Muhammad Younus, Md Tahsir Alam, Shaikh Muhammad Allayear and Sheikh Joly Ferdous Ara	to improper lifestyle, eating habits and maintaining BMI. DM is the familiar genetic problem, being emerged by all age peoples.		megabyte, 2.1 populations consists 12824 male, 13176 females.	Training accuracy 87, test accuracy 83, FEMALE-CVD: Training accuracy 92, test accuracy 89 like this the accuracy was tested for all diseases.
5.	Predicting Diabetes Medical Dataset using Machine learning techniques/2017/Uswar Ali Zia, Dr. Naeem Khaan.	Medical system comprises more amount of data and make it more structuralized in defined process. The DM's features will be covered to make the segregation list aggressively.	Naïve Bayes, Decision Tree, K-Nearest Neighbor	Dataset from UCI Machine repository, 768 instances, 8 attributes, 1 class attribute	Decision tree have highest accuracy. Accuracy: Decision Tree(J48): Without Bootstrapping-78.43% After Bootstrapping-94.44%
6.	Risk prediction of typeII diabetes based on random forest model/2017/Wwifeng Xu, Jiaxin Zhag, Qiang Zhang, Xiaopeng Wei	The case study approaches a Type 2 Diabetes Prediction Model depends on random forest which targets at current indicators.	Data Preprocessing. Random Forest	373 instances and 10 features	Random Forest model can effectively predict the risk of diabetes. 10-Fold Cross Validation Random Forest: 85% ID3: 78.57% Naïve Bayes: 79.89% Adaboost: 84.19%
7.	K-Nearest Neighbor Based Methodology for accurate Diagnosis of Diabetes Mellitus/2016/Madhuri Panwar, Amit Acharya, Rishad A.Shafik2 and Dwaipayan Biswas	DM is most known and familiar disease in world to creates more deaths, disability and business loss among the famous diseases. Targets to improve the accuracy of analyze report, we have approached a mechanism origin on novel preprocessing method.	Data Preprocessing: Data Cleaning, Patter Matching Data Reduction, Classification: PCA, KNN	768 samples, 8 attributes from UCI Machine Learning Repository	K-NN: Accuracy:100%, Specificity:100%, Sensitivity:100%
8.	Application of K-means and Genetic Algorithms for dimension reduction by integrating SVM for Diabetes Diagnosis/2015/ T. Santhanam, M.S. Padmavathi.	Huge amount of data present in medical care system is creates complexity to manage. So, DM is essential to trace the required format and connectivity among the method presence.	K-Means Clustering, Genetic Algorithm, Support Vector Machine	768 samples, 8 attributes from UCI Machine Learning Repository	SVM ACCURACY: 98.67%
9.	Detection of Diabetes using Multilayer perceptron/2018/Saumendra	It is normally in unconscious mode but it will trigger	Data Preprocessing, Multilayer Perceptron	768 patient and 8 attributes	ACCURACY-77.5%

	Kumar, Mohapatra jaggit Kumar Swaim and Mihir Narayan Mohanty.	them whenever the people attack from other disease.			
10.	Deep Belief Neural Network Model for Prediction of Diabetes Mellitus/2019/ P. Prabhu, S. Selvabharathi	More existing methods available in world to analysis the diabetes. But still we required more accuracy tool and method to predict the diabetes at initial stage to reduce the treatment cost.	Deep Belief Network	768 patient and 8 attributes	Deep Belief Networks (DBN) Recall 1.0, Precision 0.6791, F1 Measure 0.808
11.	Detection of Type 2 Diabetes Using Clustering Methods – Balanced and Imbalanced Pima Indian Extended Dataset/2020/ S. Nivetha , B. Valarmathi , K. Santhi , and T. Chellatamilan	Insulin is being develops inside the body by pancreas. Diabetic disorder will affect the nerves system, eye vision, Kidneys and different parts of body.	CLUSTERING K-means, farthest first, Density based, Filtered and X-Means methods.	768 patient and 8 attributes	K-means, filtered clustering and X-means methods has maximum accuracy - 75.64%.
12.	An Expert Clinical Decision Support System to Predict Disease Using Classification Techniques/2017/Emrana Kabir Hashi, Md. Shahid Uz Zaman and Md. Rokibul Hasan	This case study remains put on eye to analysis the diabetic root cause since the diabetes diseases found high rate of percentage in the universe.	Decision Tree, K-NN	Pima Indian Diabetes Dataset	Decision tree algorithm accuracy- 90.43%
13.	A Hybrid Prediction Model for Type 2 Diabetes Using K-means and Decision Tree/2017/Wenqian Chen, Shuyu Chen, Hancui Zhang, Tianshu Wu	Type 2 Diabetes is the one of the high contributors in universe among the famous disease. So, predict the diagnosis result at initial stage is required for researching purpose of diabetes.	Data Preprocessing, K-Mean, Decision Tree With 10-Fold Cross-Validation Method	Pima Indian Diabetes Dataset	Model has better accuracy 90.04%

4 Existing Experimental Analysis

4.1Comparison of classification techniques based on Accuracy

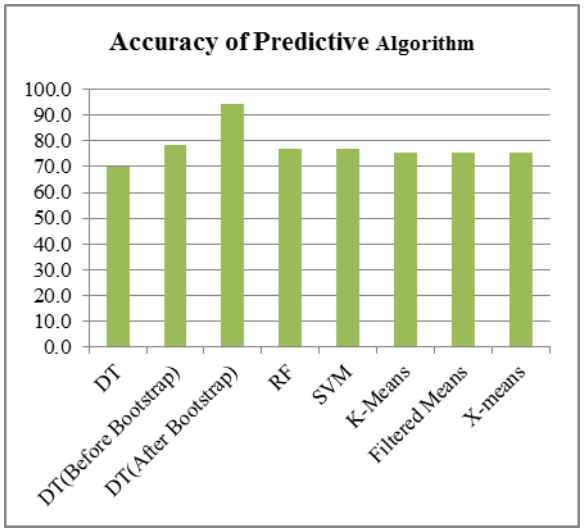

Fig. 1. Outcome obtained for Predictive Algorithm.

4.2Comparison of classification techniques based on 10- fold cross validation

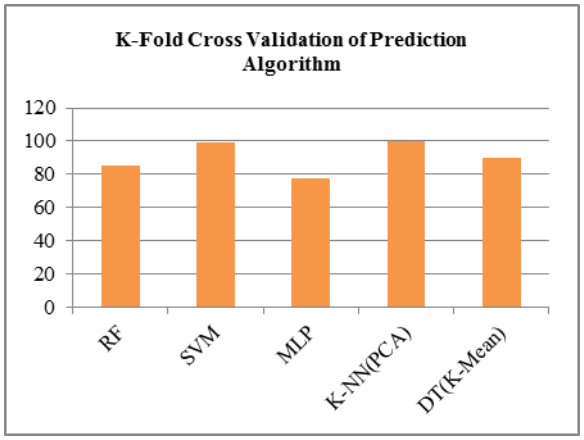

Fig. 2. Outcome obtained for Predictive Algorithm using 10-fold cross validation

5 Discussion

Predictive Algorithm is used to predict a disease accurately. The machine learning algorithms and data mining techniques are used predict an accurate result. The big data analytics is used to process larger data. From the reference of Decision Tree with Bootstrapping produces better results. The Support Vector Machine and Random Forest provides better results compare to other algorithms. The predictive algorithm along with PCA

provides excellent result e.g.-K-NN (PCA). Using these predicted results, the doctor can diagnose and treat a patient at early stage. It will reduce a treatment and unexpected cost. It keeps our county healthy and wealthy.

6 Conclusion

Predictive algorithm is used to cure and care the patient. Most of diseases are curable in this world. But through this study we would be found treatment against disease with Zero Side-effects. By providing treatment to the patient at early stage we would be reduced the impact of diabetes and as well as reduce the treatment cost. We would be utilized the existing equipment to get the accurate results rather than investment on new one. The study results would be reached the mid and low economy peoples to get away from the diabetes with low cost. In this paper the classification of predictive algorithms is studied for the purpose to sustain the healthy social environment. In future, using these techniques we would be found a new classifier to predict diabetes accurately.

References

[1] Bai BM, Nalini BM, Majumdar J. Analysis and detection of diabetes using data mining techniques—a big data application in health care. InEmerging Research in Computing, Information, Communication and Applications 2019 (pp. 443-455). Springer, Singapore.

[2] Kumar S, Singh M. Diabetes Data Analysis Using MapReduce with Hadoop. InEngineering Vibration, Communication and Information Processing 2019 (pp. 161-176). Springer, Singapore.

[3] Sneha N, Gangil T. Analysis of diabetes mellitus for early prediction using optimal features selection. Journal of Big data. 2019 Dec 1;6(1):13.

[4] Younus M, Munna MT, Alam MM, Allayear SM, Ara SJ. Prediction Model for Prevalence of Type-2 Diabetes Mellitus Complications Using Machine Learning Approach. InData Management and Analysis 2020 (pp. 103-116). Springer, Cham.

[5] Zia UA, Khan N. Predicting diabetes in medical datasets using machine learning techniques. International Journal of Scientific & Engineering Research Volume. 2017 May;8(5).

[6] Xu W, Zhang J, Zhang Q, Wei X. Risk prediction of type II diabetes based on random forest model. In2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB) 2017 Feb 27 (pp. 382-386). IEEE.

[7] Panwar M, Acharyya A, Shafik RA, Biswas D. K-nearest neighbor based methodology for accurate diagnosis of diabetes mellitus. In2016 Sixth International Symposium on Embedded Computing and System Design (ISED) 2016 Dec 15 (pp. 132-136). IEEE.

[8] Santhanam T, Padmavathi MS. Application of K-means and genetic algorithms for dimension reduction by integrating SVM for diabetes diagnosis. Procedia Computer Science. 2015 Jan 1;47:76-83.

[9] Mohapatra SK, Swain JK, Mohanty MN. Detection of diabetes using multilayer perceptron. InInternational conference on intelligent computing and applications 2019 (pp. 109-116). Springer, Singapore.

[10] Prabhu P, Selvabharathi S. Deep Belief Neural Network Model for Prediction of Diabetes Mellitus. In2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC) 2019 Jul 27 (pp. 138-142). IEEE.

[11] Nivetha S, Valarmathi B, Santhi K, Chellatamilan T. Detection of Type 2 Diabetes Using Clustering Methods–Balanced and Imbalanced Pima Indian Extended Dataset. InInternational conference on Computer Networks, Big data and IoT 2019 Dec 19 (pp. 610-619). Springer, Cham.

[12] Hashi EK, Zaman MS, Hasan MR. An expert clinical decision support system to predict disease using classification techniques. In2017 International Conference on Electrical, Computer and Communication Engineering (ECCE) 2017 Feb 16 (pp. 396-400). IEEE.

[13] Chen W, Chen S, Zhang H, Wu T. A hybrid prediction model for type 2 diabetes using K-means and decision tree. In2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS) 2017 Nov 24 (pp. 386-390). IEEE.
[14] Yuvaraj N, SriPreethaa KR. Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster. Cluster Computing. 2019 Jan 16;22(1):1-9.
[15] AlZubi AA. Big data analytic diabetics using map reduce and classification techniques. The Journal of Supercomputing. 2020 Jun;76(6):4328-37.15. AlZubi, Ahmad Ali. "Big data analytic diabetics using map reduce and classification techniques." The Journal of Supercomputing (2018): 1-10.
[16] M. Anathi, K. Vijayakumar , "An intelligent approach for dynamic network traffic restriction using MAC address verification", Computer Communications,Elesvier,5 February 2020.
[17] K. Pradeep Mohan Kumar, M. Saravanan, M. Thenmozhi ,K. Vijayakumar, " Intrusion detection system based on GA-fuzzy classifier for detecting malicious attacks", wiley, Feb 2019.