International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August 2019

Efficient Diagnostic System For Smart Diabetes

V.Prakash, Bhavani R, Anupriya A

ABSTRACT--- 5G networks, analytics of medical data and the internet of things are recent advances in big data technologies. Combining these advances with wearable computing and artificial intelligence, innovative diabetics monitoring system is implemented. In the existing system, it classifies the Diabetes 1.0 and Diabetes 2.0 methods, which show the intelligence and networking deficiencies. Thus, with personalized treatment, our goal is to design a sustainable, cost-effective, and smart diabetes diagnosis solution. Uses the machine learning algorithms in the proposed 5G smart diabetes- Naive Bayes, Logistic regression and artificial neural networks (ANN) are for the results.

Keywords - Preprocessing; K means clustering; Machine learning algorithm

I. INTRODUCTION

Diabetes mellitus could be a cluster of diseases that affect however your body uses blood sugar (glucose). It is the major issues in all over the world and classified into 3 types: Type 1 diabetes, Type 2 diabetes, Gestational diabetes which is due to insufficient insulin secretion [1].

To carry glucose into body cells, insulin is needed. Insulin and glucagon hormones produced by pancreas are supplied to the blood glucose [3]. To the cells of muscles and tissues glucose are very important. At the same time if the glucose get increased it will affect our various body parts like blood vessels , heart, kidneys, eyes, nerves and diabetes people may undergo high range of infections [4].

Diabetes reactions change dependent upon how much your glucose is raised. A few people, particularly those with Prediabetes or type2 diabetes, may not encounter indications at first. In sort type 1 diabetes, side effects will in general please rapidly and be increasingly serious [2].

Big Data Analysis improves human services framework through the decrease run time and the ideal expense. Its capacity to settle on the banking and medicinal services choice dependent on the correspondence made by framework.

Revised Manuscript Received on August 30, 2019.
* Correspondence Author
V.Prakash*, Assistant Professor/Mca, Sastra Deemed To Be University,Thanjavur, Tamilnadu, India.
. (Email: prakash125@gmail.com)
Bhavani R, Assistant Professor/CSE, Sastra Deemed To Be University, Kumbakonam, Tamilnadu, India.
. (Email: bhavani@src.sastra.edu)
Anupriya A, II MSC-CS, Sastra Deemed To Be University, Kumbakonam, Tamilnadu, India.
. (Email: amirthaanu14@gmail.com)

The manner in which that we utilize huge information is honestly exceptional [5]. For instance blood glucose meters, circulatory strain sleeves and scales that are associated with a bigger database that we can get to. We take the majority of that information and we take a look at the patterns. At that point we produce bits of knowledge to convey back to our individuals to drive conduct change. We utilize a support learning stage where we take a look at the information, produce numerous distinctive customized messages to send back to our individuals and after that realize what works best for which part. So our framework is consistently getting more intelligent.

II. RELATED WORK

Data mining is an important tool in medical science. It includes three representative data mining techniques for data mining prediction. It uses J48 decision tree and techniques bagging using J48 is used as base learner. For making the decision useful, these methods are combined to generate knowledge [2]. The results could also be influenced by insulin, devices and strips in diabetes education which does not seem to be the case as most participants in diabetes – educated and diabetes-uneducated groups had healthcare coverage. It shows that diabetes education in people with T1DM is associated significantly with effectivepeople [8]. A new model has been developed to classify treatment plans like Insulin, medication and diet for diabetic type 2. These treatment plans may help the diabetic patient to control the level of blood glucose. It was selected to develop the proposed model based on accuracy result after conducting comprehensive experiments between the J48 algorithms. This model was implemented using WEKA application [9]. For classifying, predicting and diagnosing diabetes, there are many algorithms have been proposed. It is a system that includes changing crude information into reasonable arrangement. But the structure of the layer is very complex [3]. Diabetes prediction and monitoring system is designed and implemented using ID3 and decision tree classification algorithm.

III. MODEL AND METHODOLOGY

WEKA is simple toolkit which consisting of machine learning algorithms [6]. Better accuracy was found after applying the Preprocessing, clustering, and classification techniques on different disease [7].

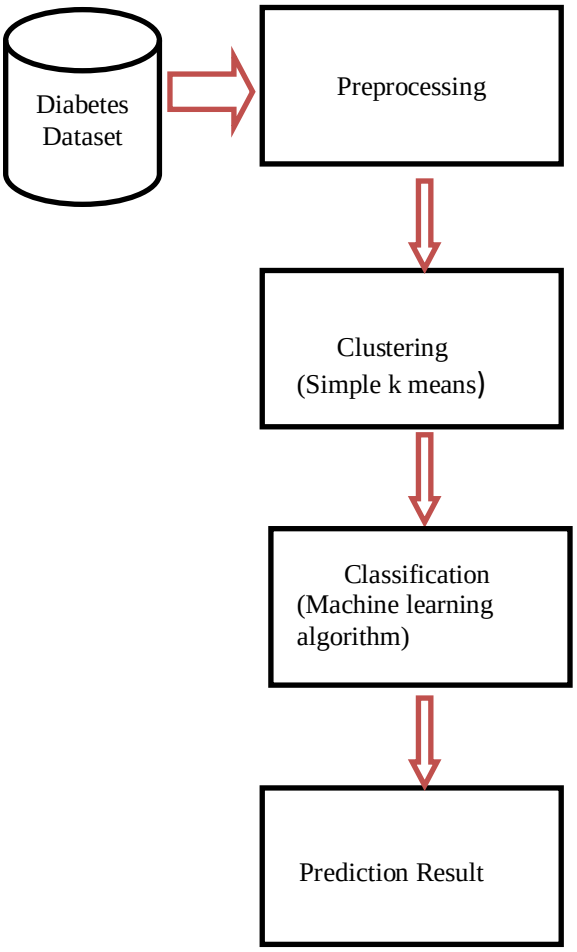


Fig.1 Architecture model

IV. DATASET DESCRIPTION

The patient information was stored in diabetes dataset. The diabetes and non-diabetes patient are differentiated here. It contains personal health data as well as results. The detailed dataset attributes are shown below:

- Age
- Gender
- Diet
- Height
- Weight
- Systolic Blood Pressure(SBP)
- Diastolic Blood Pressure(DBP)
- Random Blood Sugar(RBS)
- Class variable(class)

1	GEN	Age	Diet	Height	Weight	SBP	DBP	RBS	CL_VAR
2	0	48	1	159	46	140	70	160	1
3	0	67	1	155	79	180	110	120	0
4	1	60	1	163	59	140	70	130	1
5	1	59	1	155	58	110	70	130	1
6	1	57	1	163	49	110	70	120	0
7	0	60	0	159	68	130	70	140	1
8	1	64	0	163	71	140	70	120	0
9	1	56	1	165	71	140	90	130	1
10	0	31	0	160	47	133	90	140	1
11	0	60	0	158	46	100	70	140	1
12	1	75	1	165	71	90	70	140	1
13	1	53	1	159	55	100	70	130	1
14	0	50	1	158	49	100	70	120	0
15	0	50	1	159	59	130	70	172	1
16	1	40	1	171	62	130	90	150	1
17	1	65	1	153	60	130	100	110	0
18	0	70	1	154	63	150	90	140	1
19	0	89	1	154	51	150	70	130	1
20	1	52	0	165	73	120	70	130	1

Fig.2 Samples of dataset

V. PREPROCESSING

Preprocessing is the technique used to filter out or replace the mistaken data. Since we are collecting the raw data there may be high chances for noisy, missing value or inappropriate values. In order to sort out the mistakes we are doing theprocess of preprocessing is done to the raw data otherwise data become unfeasible.

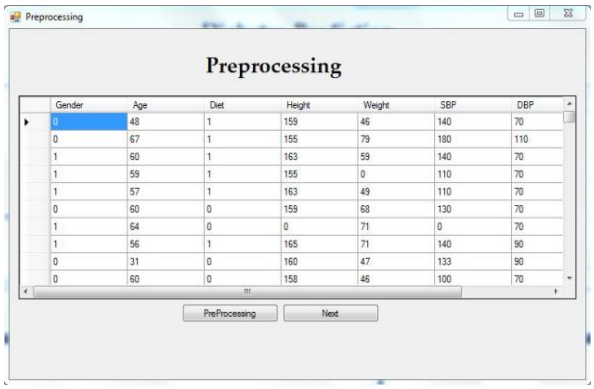


Fig 3. Before preprocessing

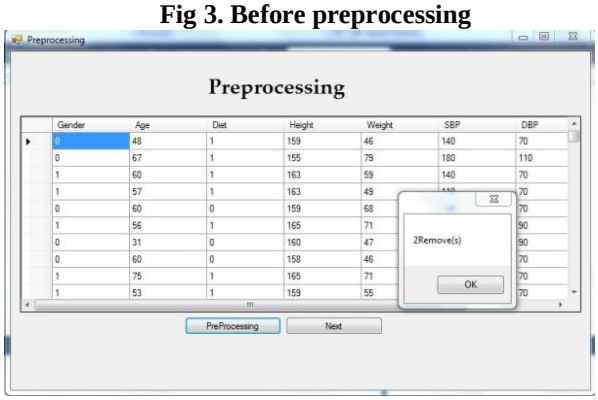


Fig 4. After preprocessing

VI. CLUSTERING

Clustering is the process of grouping or collecting the same group of objects or data together. In other words, the similar objects or points are formed as one cluster and different points are grouped in other cluster. The aim of the clustering is to group the same set of data as one cluster and others are as different cluster. It will bring the similar set of dataset together[8]. This will help to understand the structure or grouping the dataset together and various clustering algorithm will implemented for attaining this goal. K-means algorithm can do this type of clustering for massive data with high efficiency in short period of time. It is a simple iterative straightforward method used to partitioning the k-number of the datasetinto cluster [2]. This type of clustering technique is used to improve the accuracy of clustering process, before the applying this dataset into classification algorithm.

Pseudo code for K- means clustering
Input:

k: (the number of cluster)
D: CSV dataset

Output:

a set of k cluster as diabetes data and non-diabetes data

Method:

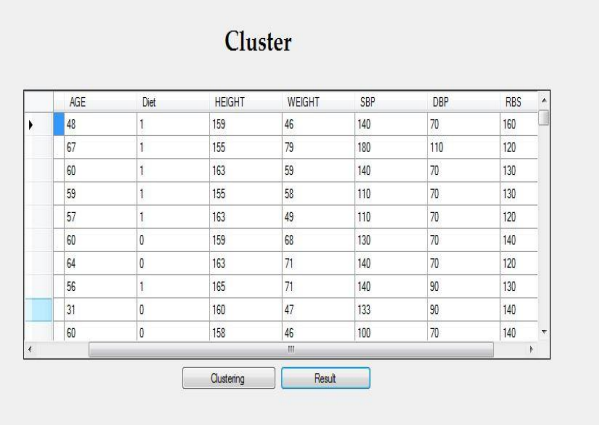
Select k objects randomly from d as initial cluster.

Repeat:

Based on the mean value (Re) assign the each object to select which object is the most similar.

The mean value for each cluster is calculated.

Until there is no change.



AGE	Diet	HEIGHT	WEIGHT	SBP	DBP	RBS
48	1	159	46	140	70	160
67	1	155	79	180	110	120
60	1	163	59	140	70	130
59	1	155	58	110	70	130
57	1	163	49	110	70	120
60	0	159	68	130	70	140
64	0	163	71	140	70	120
56	1	165	71	140	90	130
31	0	160	47	133	90	140
60	0	158	46	100	70	140

Fig 5 Clustering

VII. CLASSIFICATION

Classification is an important supervised machine learning algorithm. It has certain rules and uses various classification algorithms for training the specified dataset. By using classification we can separate the original data as a training data and a test data for evaluation process.

A. NAVIES BAYES

A Naive Bayes Classifier is one of the supervised machine-learning algorithms that use the Bayes' Theorem, where the features are statistically independent. It relies on the naïve assumption that variables of input are independent of each other. In simple terms, it is based on conditional probability.

Theorem in the language of probability:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Let's explain what each of these terms means.

- "P" indicates probability.
- P (A | B) = the chance of event A (hypothesis) occurring on condition that B (evidence) has occurred.
- P (B | A) = the chance of the event B (evidence) occurring on condition that A (hypothesis) has occurred.
- P (A) = the chance of event B (hypothesis) occurring.
- P (B) = the chance of event A (evidence) occurring.

B. LOGISTIC REGRESSION

Logistic regression is one of the predictive regressions. It will describe the data and to solve the two-category variable [6]. The relationship between binary, nominal, ordinals, ratio level independent variables is explained by this regression. It has only two possible outcomes like yes or no.

$$P = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m$$

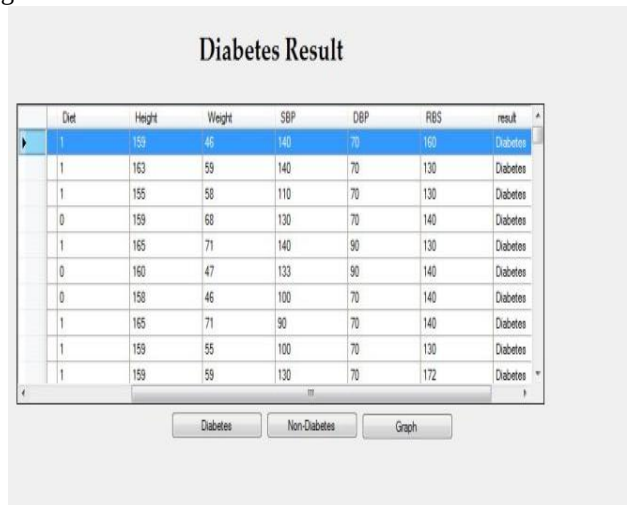
Most often, we would want to predict our outcomes as YES/NO (1/0)

C. ANN

Artificial Neural Networks is shortly called as ANN. It is designed to the process and analyzes the human brain. It will produce better result by using self-learning capabilities [5].The processing units of ANN are input and output units. Different forms and information are received by input unit and the information is processed based on the weighting system in order to produce the output in output unit.

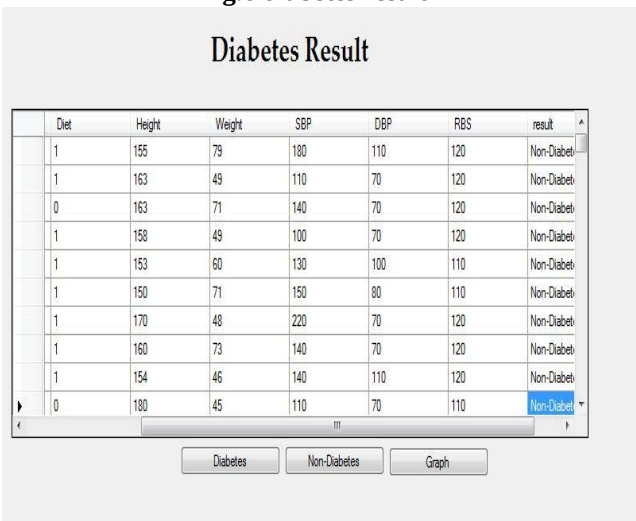
VIII. EXPERIMENTAL RESULT

In this section classification algorithm is evaluated for the training dataset. It evaluates predictive models by separating the original dataset as a training data so as to build the classifier, and a test data to evaluate the same. Pre-processing method is used to analyze the output visually and to find out the terms of values with efficiency and accuracy. K-means clustering techniques are used to improve the classifier accuracy, before applying the classification algorithm on the dataset.



Diet	Height	Weight	SBP	DBP	RBS	result
1	159	46	140	70	160	Diabetes
1	163	59	140	70	130	Diabetes
1	155	58	110	70	130	Diabetes
0	159	68	130	70	140	Diabetes
1	165	71	140	90	130	Diabetes
0	160	47	133	90	140	Diabetes
0	158	46	100	70	140	Diabetes
1	165	71	90	70	140	Diabetes
1	159	55	100	70	130	Diabetes
1	159	59	130	70	172	Diabetes

Fig.6 diabetes result



Diet	Height	Weight	SBP	DBP	RBS	result
1	155	79	180	110	120	Non-Diabet
1	163	49	110	70	120	Non-Diabet
0	163	71	140	70	120	Non-Diabet
1	158	49	100	70	120	Non-Diabet
1	153	60	130	100	110	Non-Diabet
1	150	71	150	80	110	Non-Diabet
1	170	48	220	70	120	Non-Diabet
1	160	73	140	70	120	Non-Diabet
1	154	46	140	110	120	Non-Diabet
0	180	45	110	70	110	Non-Diabet

Fig.7 non-diabetes result

Fig 8. Performance evaluation

IX. CONCLUSION

Diabetes are increasing day by day; early prediction of diabetes are used to identify the disease and for giving proper treatment in the initial stage. The diabetes prediction has using classification techniques useful for identifying in early stage. The work proposed to predict the diabetes and its complications using data mining techniques.

REFERENCES

1. "Performance Analysis of Classifier Models to Predict Diabetes Mellitus" J. Pradeep Kandhasamy, S. Balamurali (2015)
2. Performance Analysis of Data Mining Classification Techniques to Predict Diabetes" Sajida Perveen
3. "Prediction Of Diabetes Using Soft Computing Techniques- A Survey" M. Durairaj, G. KalaiselviINTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 4, ISSUE 03, (MARCH 2015)
4. "Comparison of Data mining algorithms in the Diagnosis of type II Diabetes" Sadri Sa'di1, Amanj Maleki1, Ramin Hashemi2, Zahra Panbechi1 and Kamal Chalabi1(2015)
5. "5G-Smart Diabetes:Toward Personalized Diabetes Diagnosis with Healthcare Big Data Clouds" Min Chen, Jun Yang, Jiehan Zhou, Yixue Hao, Jing Zhang, and Chan-Hyun Youn (2018)
6. "Type 2 diabetes mellitus prediction model based on data mining" - Han Wu, Shengqi Yang *, Zhangqin Huang, Jian He, Xiaoyi Wang (2018)
7. "A Tool for Diabetes Prediction and Monitoring Using Data Mining Technique" S. R. Priyanka Shetty ,Sujata Joshi (2016)
8. "Developing a Predicted model for diabetes type 2 treatment plans by using Data Mining" TARIG MOHAMED AHMED
9. R.Bhavani, V.Prakash, K.Chitra, "An Efficient Clustering Approach For Fair Semantic Web Content Retrieval Via Tri-Level Ontology Construction Model With Hybrid Dragonfly Algorithm", 2019, International Journal of Business Intelligence and Data Mining. Vol 14, pp62-88.
10. "Impact of Diabetes Education and Self-Management on the Quality of care for People with type1 diabetes mellitus in the middle East" Juan Jose Gagliardino