

Classification Of Gestational Diabetes Using Modified Fuzzy C Means Clustering And Machine Learning Technique

Geetha. V. R¹, Dr. Jayaveeran. N², Dr. A. Shaik Abdul Khadir N³

¹Assistant Professor, Department of Computer Science, A.V.C College Autonomous, Mannampandal

²Former Associate Professor and Head, Department of Computer Science

³Head and Associate Professor Department of Computer Science, Khadir Mohideen College, Adhiramapattinam, Affiliated to Bharathidasan University, Tiruchirappalli- 620 024, Tamil Nadu, India

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 28 April 2021

Abstract: In the recent past, Gestational Diabetes (GD) has become one of the major health issues in the domain of healthcare. GD is a slow and silent killer that slowly influences the physical body resistance mechanism and fundamental organs. The GD is a harmful threat of health that had occurred during pregnancy of women by affecting metabolism of glucose level which leads to health issues of pregnancy women and infant, so it is most important of effective and early prediction of GD. In this paper, intelligent diagnosis and prediction of GD is proposed based on processing of various attributes of patient. In the data mining analysis, the Linear Discriminative Analysis (LDA) technique is applied as data normalization process on patients attributes for dimension reduction. The Modified Fuzzy C Means Clustering (MFCM) algorithm is used for clustering data after preprocessing. The enhanced Naïve Bayes classifier is applied for classifying various stages of GD of the patient. The experimental results prove that the proposed methodology is efficient and accurate in terms of graphical and numerical representations.

Keywords: GD, Data Mining, Linear Discriminative Analysis (LDA), Modified Fuzzy C Means Clustering (MFCM), Enhanced Naïve Bayes classifier

1. Introduction

Diabetes disease is a health problem that slowly influences the spontaneous resistance mechanism of human being. The human body could fail to accept the food energy because of excessive level of sugar in blood. This kind of disease becomes the probable cause for kidney disease, cardiac attack, retinal diseases and bottom appendage amputation. World Health Organization (WHO) predicts that higher than 450 million people may be affected with diabetes disease throughout the world, the diabetes may be the reason for over higher than 2 million casualties diagonally the world, the state is higher widespread in average and low revenue countries [1]. Gestational mellitus diabetes (GDM) is a common human health issue. Beside with development in consumption of food behaviors, increased purchasing power and climate change, surrounded by various environmental parameters, the amount of female with pregnancy-complicated gestational diabetes is growing. GDM creates issues both for the mother and the fetal. In this circumstance, a hybrid paradigm can be suggested of a complicated structure build in the neural network based Bayesian networks, the conclusion sustain move toward to multi-criteria, and machine learning [2]. It is a continual health issue that leads to numerous difficulties that contain a group of public, substantial, and economic collision on public and civilization. Gestational diabetes mellitus (GDM) could be a kind of diabetic disease that happens in a little pregnant female though behind fetal delivery it typically proceeds to normalcy. Nevertheless, it is well known that dangers with GDM augment in improving DM at a later period in their life. The opportunity of utilizing analytical Machine Learning (ML) techniques to forecast incidence of DM after GDM is explored in very few works done in this area [3]. An international group of experts from the American Diabetes Association develops and updates biological requirements for the diagnosis. The Gestational diabetes is a glucose forbearance state that developed during pregnancy or was primarily accepted. The mother's excess glucose is transferred to the baby and that lead to big issue for fetal [4]. In data mining, a deep learning algorithm is proposed to investigate medical data in order to identify potential abnormalities and segregate information to various ways so that people may understand the character of human health issues. The development of the Convolutionary Neural Network (CNN) is used in element to separated data pattern resulting through an electrocardiograph (ECG). A dedicated CNN will be trained based on different information units accessed through different diabetic patients, called training information [5]. Medical industry plays a key role in human being health across the globe. When accepting a large quantity of healthcare information, the issues will occur on separating medical information. A fuzzy logic based convolutionary neural network (FCNN) technique is described in this research to guess the medical information class. This technique gathers data through the information set and constructs the conclusion table using the information sets collected features. Unrelated attributes are omitted by using the algorithm for key component analysis using Principle Component Analysis (PCA) [6]. The iteration adaptive dynamic programming (ADP) technique is applied to classify medical data, which is called an enhanced rate iteration ADP algorithm, to obtain the best approach for distinct stochastic progression. For the primary time we suggest a novel principle in the enhanced rate iteration ADP technique to ensure whether the policy estimated is constant for stochastic executions, or not. By examining the union possessions of the proposed methodology it is demonstrated that the roles of the iterative rate can unite to the best optimum. Furthermore, methodology allows the initial rate function to be an random semi-definite, optimistic gathering [7]. The ideal naïve Bayes classifier requires two steps,

discretization and structural development, which are frequent alternately awaiting the efficiency of the network classification may not be developed. Discretization is based on the principle of least length for definition. We use a general development approach to deal with dependent and inappropriate attribute that removes and/or links data attributes, based on joint and provisional information steps [8]. Naive Bayes methodology is one of the most accurate techniques in information classification, but it can only get a more accurate consequence in the huge training sample set. The necessity of a higher quantity of sample not only involves deep effort for earlier physical categorization, but in addition places onward a superior insist for storage and executing possessions throughout post-processing executions. In the huge information set Naive Bayes showed sufficient rapidity and efficiency, but in the type of small information set the consequence is obviously reduced. But, in reality it's frequently rigid to have a huge information training gathering. Furthermore, too huge a level of information will consequence in extremely elevated runtime complexity and space [9]. A new method for the classification of information streams, called the Rough Gaussian Naïve Bayes Classifier (RGNBC), is utilized. There are two recent approaches to tackle the issues of persistent idea float. The primary part to find the idea float is utilizing the coarse set hypothesis. Then, the Gaussian naïve classifier methodology is numerically modified to handle the active information without the old information. The classification is also carried out using the later probability and the objective function which obtains into description the numerous criterions. The proposed RGNBC methodology is implemented with two huge databases, and the outputs are examined beside the available MReC-DFS methodology using its parameters such as accuracy and efficiency [10].

2. Related Work

Bonaventura C. T. Mpondo et. al [11] has reviewed uncontrolled GDM provides morbidity and mortality on the maternal and fetal rates. Thus enhanced results are during early disease finding and powerful glycogenic rule. Although apparent protocol available in the common people for the showing and organization of diabetes mellitus, controlling GDM leftovers contentious with opposing rules and action protocols. Claire L. Meek et. al [12] has investigated Gestational diabetes mellitus (GDM) is incorporated with developed hazard to mother and fetal, but diagnostic criteria settled on internationally stay indescribable. It is significant to find women with GDM, as action stops poor effects such as prenatal death, accept dystocia and neonatal hypoglycemia. Recently, the United Kingdom National Institute for Health and Care Excellence (NICE) has proposed novel disease diagnostic standards for GDM that differ through the needs accepted by the World Health Organization's International Association of Diabetes and Pregnancy Study Groups (IADPSG). Diana Jaskolka et.al [13] has investigated that a boy baby may be incorporated with harmful function of the motherly cells of beta and an augmented hazard of gestational diabetes mellitus (GDM). Acknowledging that the overall contact of baby sex on the metabolism of motherly glucose is probable to be slight, we attempted to carry out a methodical evaluation and meta-analysis of observational cases in sort to find a reliable approximation of the incremental motherly hazard incorporated with the fetal sex. Cuilin Zhang et. al [14] has carried out important short-term and long-term poor physical condition results for together mothers and fetal, support the significance of considerate hazard issues, particularly adaptable factor, for GDM and stop the situation. Research from observational studies above the past decade has found a few diet and lifestyle reasons incorporated with the hazard of GDM, and has shown that time frames before and during pregnancy may be significant to GDM expansion. Findings from involvement research on the force of diet and lifestyle on GDM avoidance have been opposing and conflicting in large determine. Differences in revise people, kinds of involvement, timing and period of involvement and disease diagnostic criteria for GDM could all at least partially description for the huge heterogeneity in the detections through these involvement case studies. Sudeep Tanwar et. al [15] has proposed a methodology for new emerging area of research and development, Big Data (BD), has coined probable expansion of this information in terms of quantity, diversity and rapidity. Incessant storage, categorization, track (if necessary), real time investigation are presently a little of BD's present issues. Nevertheless, as information could be random, undependable and superfluous, those issues become more significant. Reducing the overall executing time-dimensionality (DR) decrease is therefore one of the successful methods. Hence, attractive into description the over, in this research, we used the methodologies of the principle component analysis (PCA) and the singular value decomposition (SVD) to execute DR over BD. Dunja Mladeni et. al [16] has proposed reduction of dimensionality is a widely used move in machine learning techniques, particularly while solving with a high-dimensional characteristic space. The unique space function is grouped onto a novel, dimensionally decreased space. The reduction of dimensionality is typically complete either by selecting a separation of the unique dimensions, or by building novel dimensions. This work contracts with the gathering of subsets of features for the reduction of machine learning dimensions. Yi Guo et. al [17] has proposed a novel technique called Sparse Dimensionality Reduction (SDR). It executes dimension assortment even as declining the dimensionality of the information. Unlike existing techniques of dropping dimensionality, this approach may not need an approximation of the dimensionality. The number of ultimate dimensions is the produce of that approach light fraction. The idea, in small, is to change input information into a satisfactory space where superfluous capacity is compressible. This approach arrangement is extremely supple that contains a numeral of alternatives along that

contour. In this work, Laplacian own maps execute the information conversion and the assortment of magnitude is completed by the 12/11 rule. Liang Pang et. al [18] has proposed one of the most extensively worn clustering techniques for modern clustering of medical data is the Fuzzy c-means algorithm (FCM). Nevertheless, the modern FCM technique has confident potential of the objective function to a local minimum, thereby leading to undesired effects of clustering. To begin this problem, this research proposes an improved FCM based on clustering centroids information using genetic algorithm. This methodology is proposed to maintain information on multidimensional features and is available during similar execution. Liangxiao Jiang [19] has proposed Naïve Bayes 'presumption of restricted independence attribute basically requires and is frequently harmed by quality dependencies. On the other hand, even as a Bayesian network can signify random dependencies of quality, learn from information an optimal Bayesian network classifier is heavy. Therefore, learning enhanced naive Bayes has involved a set of concentration from researchers and given many accurate and efficient methodologies to expand. In this research, some of those improved methodologies are used and recognize four key enhanced methods: 1) selection of features; 2) extension of structure; 3) local learning; 4) expansion of results. N. Sneha et. al [20] has proposed constant hyperglycemia of diabetes is related to long-haul harm, brokenness, and failure of various organs, particularly the eyes, kidneys, nerves, heart, and veins. The objective of this research is to make use of significant features, design a prediction algorithm using Machine learning and find the optimal classifier to give the closest result comparing to clinical outcomes. The proposed method aims to focus on selecting the attributes that ail in early detection of Diabetes Miletus using Predictive analysis. The result shows the decision tree algorithm and the Random forest has the highest specificity of 98.20% and 98.00%, respectively holds best for the analysis of diabetic data. Naïve Bayesian outcome states the best accuracy of 82.30%. N. Sneha et. al [20] has proposed stable diabetes disease hyperglycemia is incorporated with long-haul harm, brokenness, and collapse of assorted human body organs, especially of the retinal, kidneys, physical nerves, human heart, and vein layers. The objective of the research is to utilize significant features, extend a forecast methodology using deep Learning algorithm, and estimate the optimal classifier to carry the neighboring result evaluated to the medical results. The proposed system aims to use prognostic investigation to choose the qualities that are sick in early discovery of Diabetes Miletus. The experimental results gives the conclusion of tree methodology and the Random forest classifier has the maximum specificity of 97.10% and 97.00%, correspondingly keeps best for the investigation of diabetic disease information. Naïve Bayesian conclusion states that the best accuracy of 83.49% for proposed system.

of 83.49% for proposed system.

3. Proposed methodology

Gestational Diabetes (GD) smart diagnosis and prediction is proposed based on the execution of patient information with various attributes. In the field of data mining the Linear Discriminative Analysis (LDA) algorithm is applied on patient attributes for dimension reduction as a method of data normalization. Linear dimensionality reduction estimates, determined using Linear Discriminate Analysis (LDA) could be normally based on optimization of several output space reparability attributes. The resulting optimization issue is linear, but those reparability attributes are not straight corresponded to the accuracy of the output space classification. An experiment and error approach must therefore be executed, experimenting with various attributes of patient reparability that diverge in the weighting feature utilized, and selecting the one that estimated best on the training information set. Even the best weighting objective function between the trial options frequently results in sub-space classification of the information. After preprocessing the Modified Fuzzy C Means Clustering (MFCM) algorithm is used to cluster data. The compilation of same features attributes are grouped into similar collection and clusters of these modified objects known as Data Clustering (DC). It is an unsupervised method in data classification research. Fuzzy C Means (FCM) clustering algorithm is extensively used and is a famous algorithm for cluster investigation. The modification in FCM is done for grouping attributes of patients to improve its accuracy and efficiency by naming the algorithm as Modified FCM (MFCM). Number of information points in this technique is separated into 'k' clusters based on a few quantity criteria for correspondence. K-Means technique of MFCM has speedy and is thus extensively used to cluster algorithms. Vector quantization, cluster investigation objective function learning are just a few of MFCM use. However, the outputs produced with this algorithm depend typically on the collection of original cluster centroids. The major approach of this methodology is to give suitable amount of clusters. Condition of number of clusters earlier than applying the methodology is extremely practical and requires deep information of cluster field. The improved classifier Naïve Bayes is used for the description of the patient's various stages of GD. Naive Bayes is a commonly used system of classification, based on the theory of Bayes. The posterior class probability of a test information position can be estimated on the source of class provisional mass inference and class preceding probability, and the test information will be dispensed to the class with the maximum position class likelihood. The key difficulty solved with naive Bayes method is the estimation of class provisional density. The class provisional density is typically calculated based on the information points. Bayes theorem evaluates in probability theory, the provisional and increasing likelihood of two random

occurrences. Also, it is used to measure future likelihood given observations. The figure 1 shows the Naïve Bayesian structure that collects various nodes information through probabilistic approach.

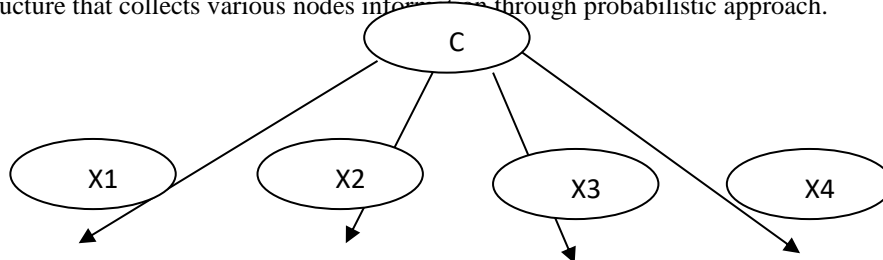


Figure 1: Naïve Bayesian Structure

Naïve Bayes network keeps a theoretical method and a likelihood focus to state. The structural representation is a graph, in which nodes signify attributes and curve signify dependencies of attributes. Dependencies of the attributes are enumerated by node condition likelihood. Naïve Bayes networks are often used for classification issues, in which a learner attempts to construct a categorization. An instance E is a array $\langle a_1; a_2; ; a_n \rangle$, where the a_i is the A_i rate. Let C correspond to the group changeable (in a Bayesian network, corresponding to the class node). We use c to signify the name C takes, and accurate and effectively to represent the E class. Enhancement its arrangement is a straight method to defeat the disadvantage of existing naive Bayes, since curve can clearly signify attribute dependencies. Enhanced Naive Bayes (ENB) is a comprehensive tree-like naive Bayes in which the group node peaks straight to all attribute nodes and an attribute node can contain only single attribute node parent from an additional. The presumption is impractical in applications for real-world data mining. An expected way to assuage this idealistic assumption is thus to enlarge the arrangement of naive Bayes by using directed arcs to directly represent attribute dependencies. The method that shows is basically a Bayesian network. Therefore it is essential to appreciate the arrangement of Bayesian networks. Learning the optimal arrangement however is an NP-hard issue. In practice it is important to impose restrictions on the bayesian network arrangement

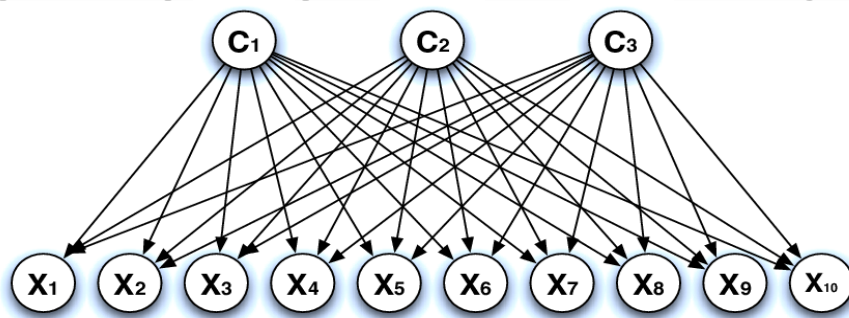


Figure 2: Supervised learning probabilistic approach of Naïve Bayes classifier

The figure 2 shows supervised learning probabilistic approach of Naïve Bayes classification technique. Learning Enhanced Naive Bayes for instance leads to a suitable computational difficulty and a significant development over naive Bayes. Unlike the collection of features which totally removes the slightest significant attributes from the space of the attribute, attribute weighting to all attribute differently depending on its classification input. How to know the weights is perceptibly significant and has tired some researcher attention. The essential design of the limited learning technique is to build a naive Bayes on a separation of the training information set (called local training information), as a substitute of on the complete information set.

Naïve Bayes formulation:

$$B(Ca, br_1, br_2, \dots, br_M) = B(Ca) \prod_{i=1}^M B(ar_i | bC) \quad (1)$$

Then, the joint probability for the rating patterns of user y , i.e. $\{aR_y(xx_1), bR_y(xx_2), \dots, xR_y(xx_M)\}$, can be expanded as:

$$P(R_y(x_1), R_y(x_2), \dots, R_y(x_M)) = \sum_C P(C) \prod_{i \in X(y)} P(R_y(x_i) | C) \quad (2)$$

This method would first pick an application category 'C' from the $P(C)$ spectrum as shown by Formula (2) and then score all objects using the specific defined class 'C'. In other words, this model suggests a single user class

added to the grades of all the objects and thus avoids the situation where a user is of different user groups and separate user groups are added to the grades of different items.

4. Attribute Development (AD) for NaïveBayes Network

One sensible issue in learning a classifier for a Naïve Baye Network is the elevated difference due to lack of training information. Specifically, estimates of the likelihood are subject to unpredictability. Instinctively, the underlying distribution would not be clearly reproduced when the training information are not sufficient. If there are more training information even with the same succession, the primary sharing in the training information would be reinforced, and thus could be learned more simply. That would progress the learning methodology by adding more examples to the training information set, called information development for Naïve Bayes Network. To illustrate more training information examples, however, the underlying sharing should be recognized, which not the case in actuality is. One way to rise above this issue is to replica the obtainable instance. The idea is addition a number of replicas of training information instances to the training information set, based on correspondence. The figure 3 shows the machine learning approach techniques classification as supervised and un supervised learning methods. The Naïve Bayesian classification technique comes under probabilistic approach which is derived from supervised learning algorithm.

There are various strategies, for instance cloning, such as cloning all the training instances (global cloning), cloning a subset of training instances (local cloning), and cloning instances to boost a predefined achieve (such as AUC) on the ensuing enhanced results of Naïve Bayes classification. Remember that we decide the methodologies through the attribute development approach, because attribute development is in fact a pre-processing technique that may be given to any algorithm. Our experiments goal to evaluate the essential (core) output in the continuation of Naive Bayes. We include failed to comprise to our facts the most recent attribute-weighted Naive Bayes algorithms, as they are presently not very aggressive in terms of accuracy.

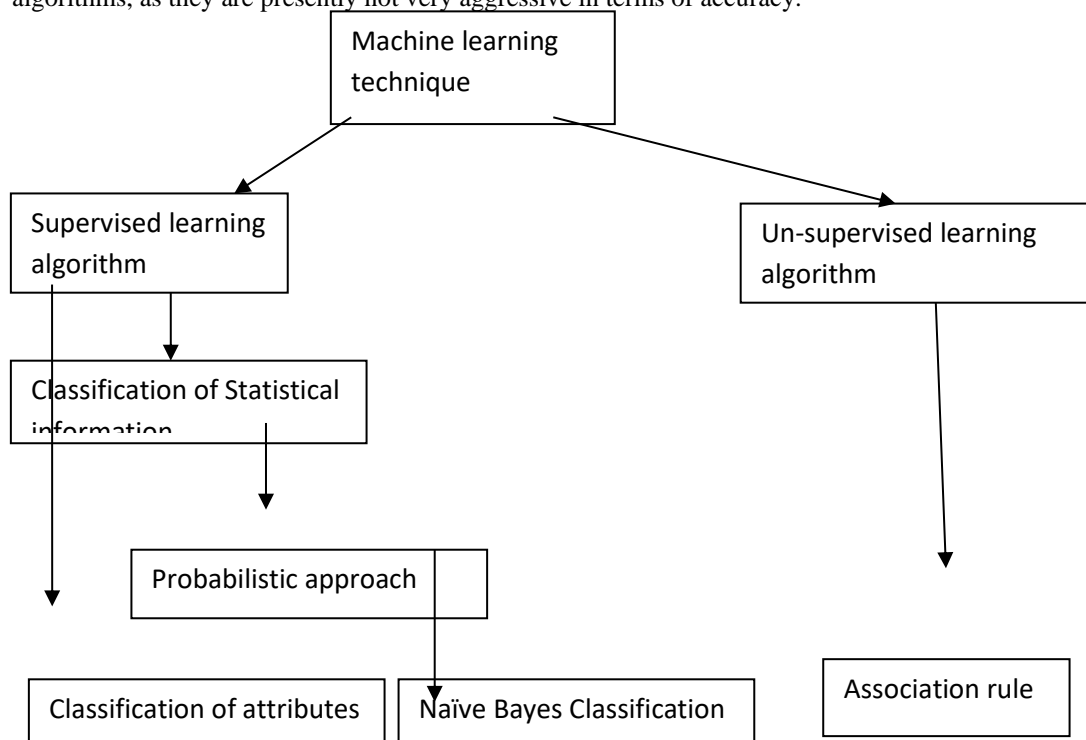


Figure 3: Naïve Bayes Classification architecture

5. Data Analysis and Interpretation

The input data is the collective information of patient attributes. There are various attributes such as Glucose, INS, BMI (Body Mass Index), Age and Number of pregnancy is considered as input arguments collected from various women patients during pregnancy. The proposed methodology considers on early finding of GD for women who are pregnant (patients). The growing demand of machine learning applications for diabetes disease forecast

is showing better presentation in healthcare decision-making. For instance the attributes are given as input for preprocessing in the order of [148 0 33 .92 50 3] attributes respectively. LDA is a multivariate arithmetical method employed to reduce the dimensionality of information. Not only can it hold a huge quantity of information fast but it can also avoid computationally powerful estimations. The basic idea of LDA is to re-build the unique variables with a positive involvement into a novel set of uncorrelated, incorporated variables. First we regulate the unique information in order to concern LDA algorithm.

The LDA can be represented as group A's own structures are used irrespective of the indication of the main variable. Group B's own statuses are only used if, and ultimately,

The LDA formula can be formed as,

$$Lx_{ki}^{et} = a\bar{x}_i + \sum_{f=1}^{p_1} B V_{h_{1i}} z_{kh_2} + \sum_{f=1}^{p_2} \begin{cases} B V_{h_{2i}} bz_{kh_2} & \text{if } bz_{kh_2} \geq 0 \\ 0 & \text{if } bz_{kh_2} \leq 0 \end{cases} + \sum_{f=1}^{p_3} \begin{cases} B V_{h_{3i}} bz_{kh_3} & \text{if } bz_{kh_3} \leq 0 \\ 0 & \text{if } bz_{kh_3} \geq 0 \end{cases} \quad (3)$$

Here $h_1 = l_1(f)$, $h_2 = l_2(f)$, $h_3 = l_3(f)$, where $l_1(\dots)$, $l_2(\dots)$, $l_3(\dots)$ represent the list of eigen-states of A, B and C groups, p_1 , p_2 , p_3 are principal components numbers of A, B and C groups.

Therefore, the matrix for association is calculated. After that, we estimate the individual principles and the entity vectors. In order to shorten the outcomes, we set down the individual values and estimate the accumulative role ratio. In our experiment Table 1 shows the association between accumulative part and number of PCs. It can be simply seen that the superior the numeral of PCs, the better the role they create and the additional they can signify the unique quality information. After preprocessing, the dimensional reduction information of different patients is given to clustering processing using Modified Fuzzy C means clustering technique. The system developed by Fuzzy Expert helps to solve several challenging tasks related to diabetes application and support for the medical practitioner's foundation. MFCM definition is to convert the information into the knowledge required. Each case has attributes and each attribute can be constructed as a fuzzy variable with a certain fuzzy number. The key significance of the MFCM algorithm is to present an enough amount of clusters. Before implementing the proposed methodology, the condition of amount of clusters is extremely unfeasible and need deep clustering data. We can propose algorithm for improvement in the initialization of MFCM algorithm centroids.

6. Pseudo code for MFCM clustering

Let S be the system,
 S= I, Fn, C, S, F where, I= set of input M,
 Fe M= matrix of term- n*mn= number of datapoints.
 M= Attribute Number.
 Fe= initial centroid set c1, c2, ck
 k= centroid number.
 Fn= f(D1), f(D2), f(D3)
 D1= checks whether the datapoint present in Cluster
 C1 D2= checks whether the datapoint
 Present in Cluster C2 D3= checks
 the data point present in Cluster C2 D3 is high, low and medium.

7. MFCM Proposed Algorithm

8.

1) Choose two centroids, i.e. the lowest centroid point and the highest centroid point from the dataset.

2) After selecting the centroids we create two clusters with members that are different from each other.

Input: T: The set of n Data points with T1,T2,Tn where attributes, n = no attributes. Each attribute is numeric.
 Output: Adequate No of Clusters with correctly distributed n Data points.

A) Calculate the sum of the values of each Data Point attribute (to find the points in the datasets that are the farthest apart).

b) Take Data points as the initial centroids with the lowest and highest total values.

C) Create initial divisions (clusters) using Cosine, Sorensen-Dice and Manhattan Distance measuring formulas between each Data points and the original centroids.

D) Find the distance of each Data points from the centroid in both the initial partitions.

- E) Find new partition centroids generated at step c.
- f) Calculate the distance between each Data points and the new cluster centers and figure out the outliers according to the following objective function.
- (G) If the Data points distance from the cluster implies; l then not an outlier.
- H) Calculate Cluster New Centroids.
- I) Calculate the difference between each outlier and the new cluster centroids, and figure out the outliers that are not sufficient in phase f.
- J) Let the set of outliers obtained in step h be $O = \{OL1, OL2, \dots, OLp\}$ (the value of K depends on the number of outliers).
- K) Repeat until $(A == NULL)$ Create a new cluster for the B group, taking the mean value as the centroid of its members. Find outliers of these clusters, in phase 6, depending on the objective function. If there are no outliers $= n$ then create a new cluster with one outliers as its members and check each outlier.

9. MFCM Formula

We consider, $X = \{x_1, x_2, \dots, x_n\}$ be a set of specified information. The modified fuzzy pseudo-partition or fuzzy c-partition of X is a relations of fuzzy subsets of X, indicated by $P = \{A_1, A_2, \dots, A_c\}$, which gives

for all $k \in N_n$ and

for all $i \in N_c$, where c is a integer in positive values

A group of given information $X = \{x_1, x_2, \dots, x_n\}$, where x_k , in common, is a matrix array,

$$x_k = \{x_{k1}, x_{k2}, \dots, x_{kn}\} \in R^p$$

for all $k \in N_n$, The issue of Fuzzy clustering is to consider a Fuzzy quasi- and related group clusters for the best possible interpretation of the information framework. This needs some criteria which express the general idea that cluster associations are powerful and weak among groups.

Provided a pseudo-partition $P = \{A_1, A_2, \dots, A_c\}$, the c groups, v_1, v_2, \dots, v_n combined with the separation are estimated by the formula for MFCM is given as,

$$v_i = \frac{\sum_{k=1}^n [A_i(x_k)]^m x_k}{\sum_{k=1}^n [A_i(x_k)]^m}$$

for all $i \in N_c$, where $m > 1$ is a real number v_i is the influence of membership grades. Notice that the above determine v_i as the cluster core of the fuzzy category A_i , is in reality the fixed percentage of data i

A datum x_k 's weight is the m th strength of x_k 's participation rating in Fuzzy set A_i .

The efficiency index for a fuzzy pseudo-partition P, $J_m(P)$, is then described by the equation in terms of the group clusters

where $\|\cdot\|$ is some inner product $J_m(P) = \sum_{k=1}^n \sum_{i=1}^c [A_i(x_k)]^m \|x_k - v_i\|^2$ v_i represents the distance between x_k and v_i .

10. Performance measures

Methodology	Accuracy in %	Sensitivity in %	Specificity in %
MFCM Naïve	98.3833	97.3833	96.3888
Bayes			
FCM	92.3938	91.3933	92.3873
K means	80.3933	73.9338	76.3983

11. Results and Discussion

Classification set also occurs to organize, summarize, and generalize the information about the variables of input and output. Naïve Bayes classification is gathering is used in glucose {medium, high}, INS {medium, high}, BMI {medium, high}, DPF {medium, high}, {medium, old} and DM {medium, high}. Fuzzy numbers are special type of fuzzy sets which boundary the probable membership objective function types. The fuzzy operator makes a substantial difference to the classification expert system's overall performance. Connection and union are clustering operators. Classification are identified by normalization and conform represent the unions. The figure 1 shows the Enhanced Naïve Bayes classification approach with target vector and output obtained with respect to data points. The figure 4 (a) shows attributes Glucose classification of training information set of patients. It is classified into low medium and higher range. The final classification of a testing patient is given as any one of low

or medium or high classification. The various attributes provides different classification levels. The testing information set of patient will give whether the patient is affected by GD as low, medium or high.

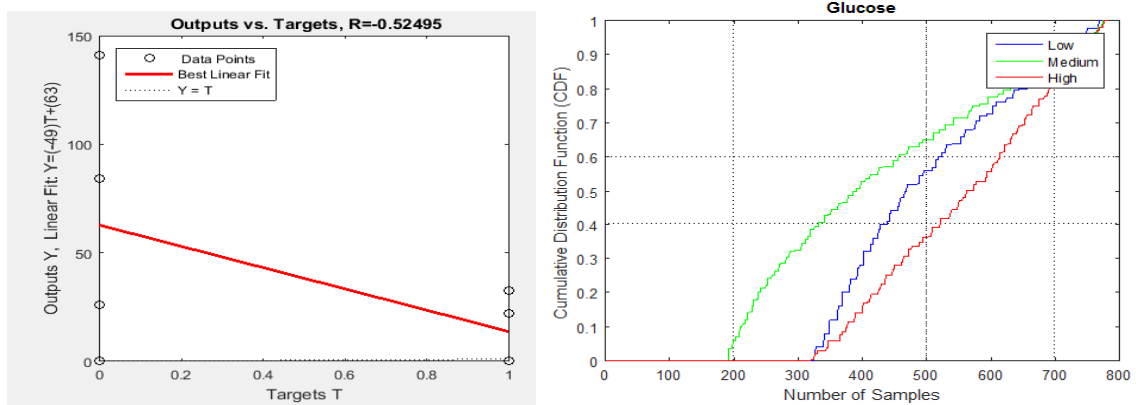


Figure 4 (a) Enhanced Naïve Bayes classifier Target vector, (b) Attribute Glucose classification of patients

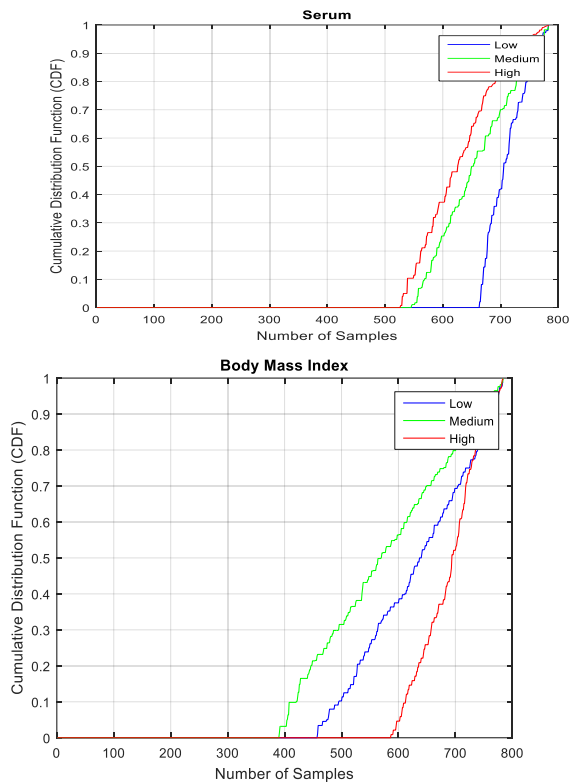


Figure 5 (a) Serum attribute classification of patients, (b) BMI attribute classification of patients

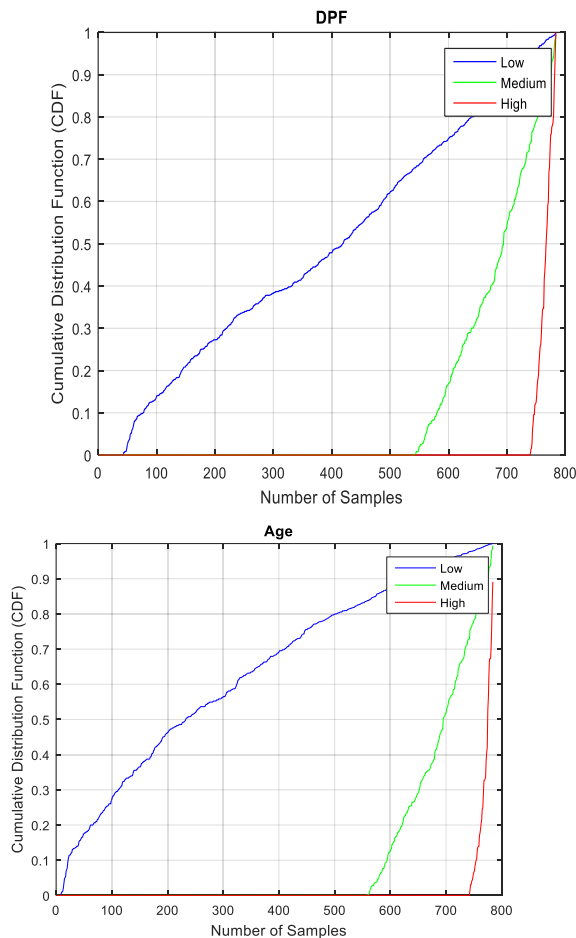


Figure 6 (a) DPF attribute classification of patients, (b) Age attribute classification of patients

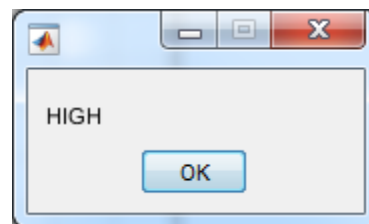
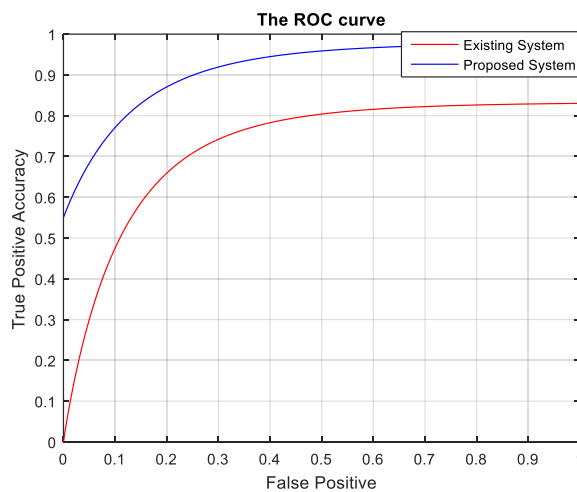


Figure 7 (a) ROC Curve, (b) Classification of Enhanced Naïve Bayes classifier

The figure 7 (a) shows the ROC curve. ROC Curves describe the trade-off for an analytical reproduction using dissimilar likelihood thresholds among the true positive rate and the false positive rate. Precision-Recall curves illustrate the trade-off for a analytical model using different likelihood thresholds between the true positive rate and the positive predictive value. ROC curves are appropriate when the clarification is reasonable between each class, while precision-recall curves are appropriate for imbalanced datasets.

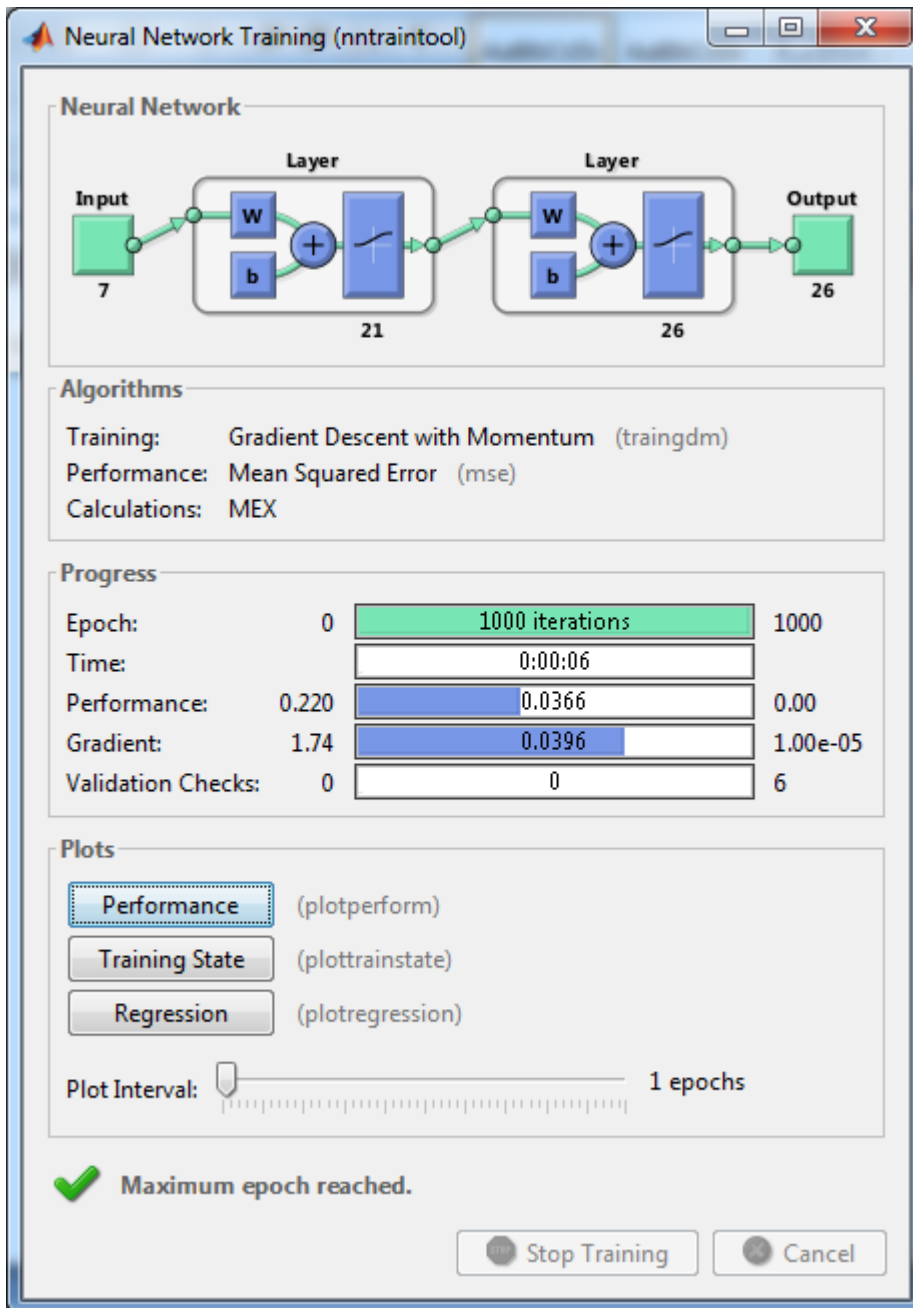


Figure 8. Enhanced Naïve Bayesian Neural Network Simulation

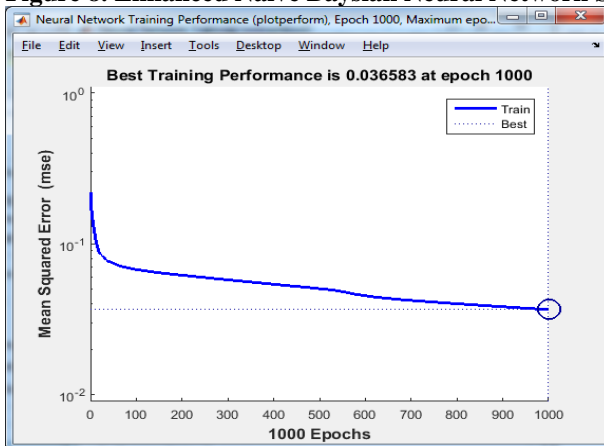


Figure 9 Reduction of Mean Square Error with respect to number of Epochs of Enhanced Naïve Bayesian Classifier

Table 1: The accumulative involvement of LDA analysis

Discrimination	Sigma in Percentage
20	100
19	98
18	96
17	94
16	92
15	90
14	88
13	86
12	84
11	82
10	80
9	78
8	76
7	74
6	72
5	70
4	68
3	66
2	64
1	62

12. Conclusion

GD's intelligence diagnosis and prediction is proposed based on analyzing patient data with different attributes. In the field of data mining, the Linear Discriminative Field (LDA) technique is applied on patient attributes for dimension reduction as a method of data normalization. After preprocessing the Modified Fuzzy C Means Clustering (MFCM) algorithm is used to cluster data. The Enhanced classifier Naïve Bayes is used for the description of the patient's various stages of GD. The experimental results indicate that the approach suggested is efficient and reliable in terms of both graphical and numerical representations.

References

1. Yashi Srivastava¹, Pooja Khanna², Sachin Kumar³, "Estimation of Gestational Diabetes Mellitus using Azure AI Services", 978-1-5386-9346-9/19/\$31.00 ©2019 IEEE.
2. Egidio Gomes Filho, Plácido Rogério Pinheiro, (Member, Ieee), Mirian Caliope Dantas Pinheiro, Luciano Comin Nunes, And Luiza Barcelos Gualberto Gomes, "Heterogeneous Methodology to Support the early Diagnosis of Gestational Diabetes", 2169-3536 2019 IEEE..
3. Devi R Krishnan, Gayathri P Menakath, Anagha Radhakrishnan, Yarrangangu Himavarshini, Aparna A, Kaveri Mukundan, Rahul Krishnan Pathinarupothi, Bithin Alangot, Sirisha Mahankali, Chakravarthy Maddipati, "Evaluation of predisposing factors of Diabetes Mellitus post Gestational Diabetes Mellitus using Machine Learning Techniques", 2019 IEEE.
4. Nassim Doualia*, Julien Dollonb and Marie-Christine Jaulenta, "Personalized prediction of gestational Diabetes using a clinical decision support system", *Microsoft, One Microsoft Way, Redmond, WA, 98074, USA*.
5. Muhammad Irfan, Ibrahim A. Hameed, "Deep Learning based Classification for Healthcare Data Analysis System", 978-1-5386-2366-4/17/\$31.00 ©2017 IEEE.
6. Balamurugan Ramasamy¹ Abdul Zubar Hameed², "Classification of healthcare data using hybridised fuzzy and convolutional neural network", *Healthcare Technology Letters*, 2019, Vol. 6, Iss. 3, pp. 59–63.
7. Mingming Liang ^a, Ding Wang^{b,c}, Derong Liu ^d, "Improved value iteration for neural-network-based stochastic optimal control design", *Neural Networks* 124 (2020) 280–295.
8. Miriam Mart´inez-Arroyo, L. Enrique Sucar, "Learning an Optimal Naive Bayes Classifier", *The 18th International Conference on Pattern Recognition (ICPR'06)*0-7695-2521-0/06 \$20.00, 2016 IEEE

9. Yuguang Huang, Lei Li, "NAIVE BAYES CLASSIFICATION ALGORITHM BASED ON SMALL SAMPLE SET", 978-1-61284-204-2/11/\$26.00 ©2011 IEEE.
10. D. Kishore Babul · Y. Ramadevi2 · K. V. Ramana3, "RGNBC: Rough Gaussian Naïve Bayes Classifier for Data Stream Classification with Recurring Concept Drift", research article - computer engineering and computer science, 2016.
11. Bonaventura C. T. Mpondo1,2, Alex Ernest1,3* and Hannah E. Dee4, "Gestational diabetes mellitus: challenges in diagnosis and management", Mpondo et al. Journal of Diabetes & Metabolic Disorders (2015).
12. Claire L. Meek1,2,3 & Hannah B. Lewis1 & Charlotte Patient4 & Helen R. Murphy1,2 & David Simmons2,5, "Diagnosis of gestational diabetes mellitus: falling through the net", Springer 2015.
13. Diana Jaskolka1 & Ravi Retnakaran1,2,3,4 & Bernard Zinman1,2,3,4 & Caroline K. Kramer1,2,3, "Sex of the baby and risk of gestational diabetes mellitus in the mother: a systematic review and meta-analysis", Springer 2015.
14. Cuilin Zhang1 & Shristi Rawal1 & Yap Seng Chong2,3, "Risk factors for gestational diabetes: is prevention possible?", Springer 2016.
15. Sudeep Tanwar1(B), Tilak Ramani1, and Sudhanshu Tyagi2, "Dimensionality Reduction Using PCA and SVD in Big Data: A Comparative Case Study",
16. ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2018
17. Dunja Mladenić, "Feature Selection for Dimensionality Reduction", Springer-Verlag Berlin Heidelberg 2006.
18. Yi Guo1, Junbin Gao2, and Feng Li3, "Dimensionality Reduction with Dimension
19. Selection", Springer-Verlag Berlin Heidelberg 2013.
20. Liang Pang, Kai Xiao, Alei Liang, and Haibing Guan, "An Improved Clustering Analysis Method Based on Fuzzy C-Means Algorithm by Adding PSO Algorithm", Springer-Verlag Berlin Heidelberg 2012
21. Liangxiao Jiang1, Dianhong Wang2, Zhihua Cai1, and Xuesong Yan1, "Survey of Improving Naive Bayes for Classification", Springer-Verlag Berlin Heidelberg 2007.
22. N. Sneha1* and Tarun Gangil2, "Analysis of diabetes mellitus for early prediction using optimal features selection", Springer 2019.