I J C T A, 10(9), 2017, pp. 679-686 ISSN: 0974-5572
© International Science Press

Correlate gestational diabetes with juvenile diabetes using Memetic based Anytime TBCA

Payal Sutaria, Rahul R. Joshi* and Preeti Mulay*

ABSTRACT

Diabetes is a lingering disease and it is increasing day by day in the worldwide and to be existing in all age group. This study aims to the novel concept of machine learning to diagnosis the Gestational Diabetes and correlate it with the juvenile Diabetes by using memetic based Anytime Threshold based clustering algorithm(TBCA). Diabetes Diagnosis through Memetic based TBCA algorithm. Memetic is an algorithm for feature extraction by identifying the reports of diabetic patients. In TBCA, the threshold is set by a distance between cluster center and data point. If data point has less the distance than the htreshold value, then it is assigned to cluster otherwise not. We are going to use this technique to diagnosis the diabetes and finding out the correlation between Gestational and Juvenile Diabetes.

Keywords: Memetic algorithm, clustering, data mining, threshold based clustering algorithm TBCA.

1. INTRODUCTION

Nowadays diabetes is a perennial illness to everyone's life and needs continuous medical assistance to prevent it. In this paper, trying to describe the correlation between the Gestational Diabetes (GD) and Juvenile Diabetes (JD).

A simply diabetes is a disease in which a person has high blood sugar, eitherbecause the pancreas does not produce enough insulin, orbecause cells do not respond to the insulin that is produced.This high blood sugar produces the humanistic symptoms ofpolyuria (aka. frequent urination), polydipsia (increased dryness) andpolyphagia (increasing hunger) [1].There are three main typesof diabetes mellitus (DM).In Type 1 DM, the body'sfell down to produce insulin, and it requires the patient toinject insulin or wear an insulin pump. This type of diabetes was adverted to as "insulin-dependent diabetes mellitus"(IDDM) or "juvenile diabetes". Type 2 DM in which body resists the insulin. In this condition cells fail to use insulinproperly, this form was adverted to as non-Insulin-dependent diabetes mellitus (NIDDM) or "adult-onset diabetes". The third form is Gestational diabetes which occurs when pregnant women without a previous diagnosis of diabetes mellitus develop a high blood glucose level and suffer from gestational diabetes. It may precede thedevelopment of type 2 DM. Type 2diabetes may be controlled with medications. diabetes are chronic conditions to be cured. In type 1 DM, Pancreas transplants have been tried with limited success and in type 2 DM, gastric bypass surgery has been efficacious in many with morbid obesity. Gestational diabetes usually repealafter delivery. Diabetes without proper treatments can cause many complications. In this paper, we are going to use memetic based TBCA algorithm and try to mapping between the T1D (Juvenile diabetes) and T2D (Gestational diabetes).

1.1. Memetic algorithm

The memetic algorithm was introduced by both Darwinian natural evolution principles, and in the "meme" was introduced by Dawkins (1976). Dawkins defines a "meme" as an information unit whichreproduces

* Department of Computer Science, Symbiosis Institute of Technology (SIT), affiliated to Symbiosis International University (SIU), Pune, India, *Emails: {payal.sutaria, preeti.mulay, rahulj}@sitpune.edu.in*

with exchange of people's ideas. This term is the cultural evolution equivalent of the gene concept, and this memetic algorithm is used to improve the individual's knowledge.

1.2. Threshold Based Clustering Algorithm(TBCA)

TBCA is somehow like a closeness factor based clustering problem. It is based on the simple assumption that data arrives in chunks. Thus closer the data points, higher the probability that they related to same data chunks.

1.3. Anytime algorithm

Anytime algorithm computes decision policies and gives an incremental output. It is an algorithm which returns a valid solution to the problem even it is interrupted at any time before it confines. This algorithm is expected to find better and better output by more time executing.

In data mining scalability is a major problem because of massive dataset. To tackle this problem, anytime is used with TBCA clustering algorithm. By using anytime algorithm user interaction will be handling during runtime.

Now day's, technology is being used for making a decision about diagnosis the disease and the proposed exploratory research work aims to Diagnosis Gestational diabetes and correlate it with juvenile diabetes.

Memetic and TBCA algorithm are used to analyze Diabetes and it is hoped that this research will provide proper analysis to find the correlation between the gestational and juvenile Diabetes. So in previous research studies, there were different techniques and algorithms are being used for diagnosisdiabetes and the results from those techniques were accurate. We can improve it by using New Combination of memetic and TBCA. Till these days memetic algorithm is mostly used in medical research field to diagnosis the different disease. The proposed study represents a Memetic based TBCA and carried out by considering diabetic patient's medical reports as a primary input dataset.

Table 1
Classification of diabetes during pregnancy [2].

Gestational Diabetes Class	*Onset*	*Fasting Plasma Glucose (mg/dL)*	*Two-Hour Postprandial Glucose (mg/dL)*	*Treatment*
A1	Gestational	<105	<120	Diet
A2	Gestational	>105	>120	Medications

Table 1: shows the classes of Gestational diabetes during pregnancy. In which it shows the Fasting Plasma Glucose (FPG) is main criteria for diagnosisdiabetes [2].

2. RESEARCH METHODOLOGY

2.1. Memetic Algorithm

It is based on local search optimality. It is based on replication and imitation of data. The below figure shows the working of the memetic algorithm. The memetic algorithm is mostly used in pattern recognition, artificial neural networks, robotic motion planning and circuit design, graph coloring and feature extraction etc.

Figure 1: it shows working of amemetic algorithm and it is the local search algorithm.

Figure 2 : it shows the memetic based TBCA cycle and how every step is connected to each other.

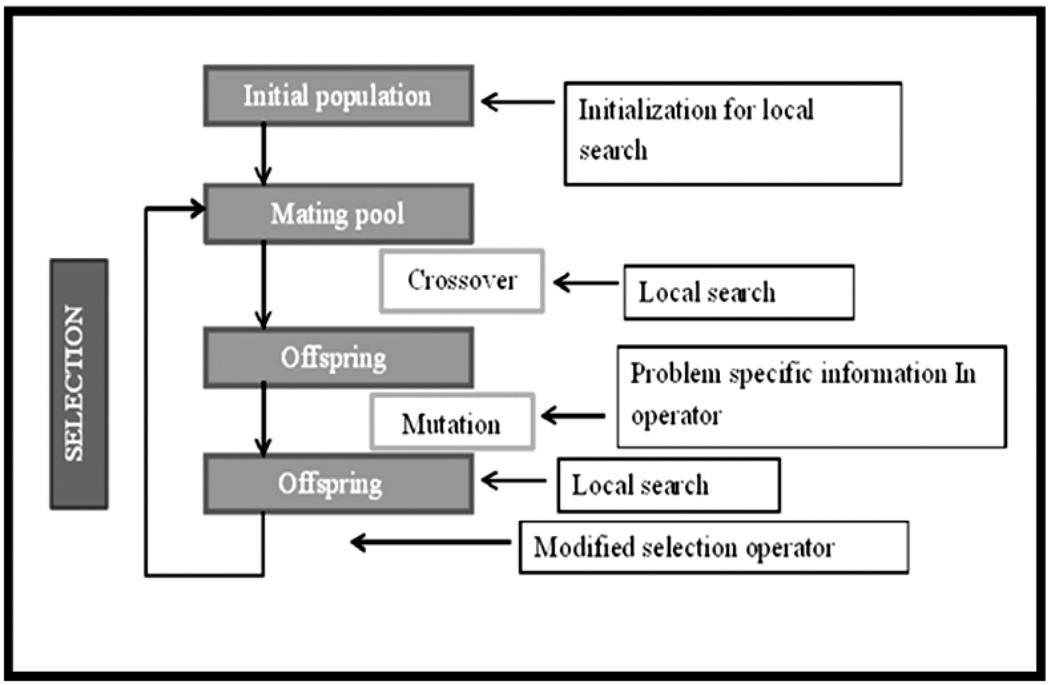

Figure 1: Working of a Memetic algorithm

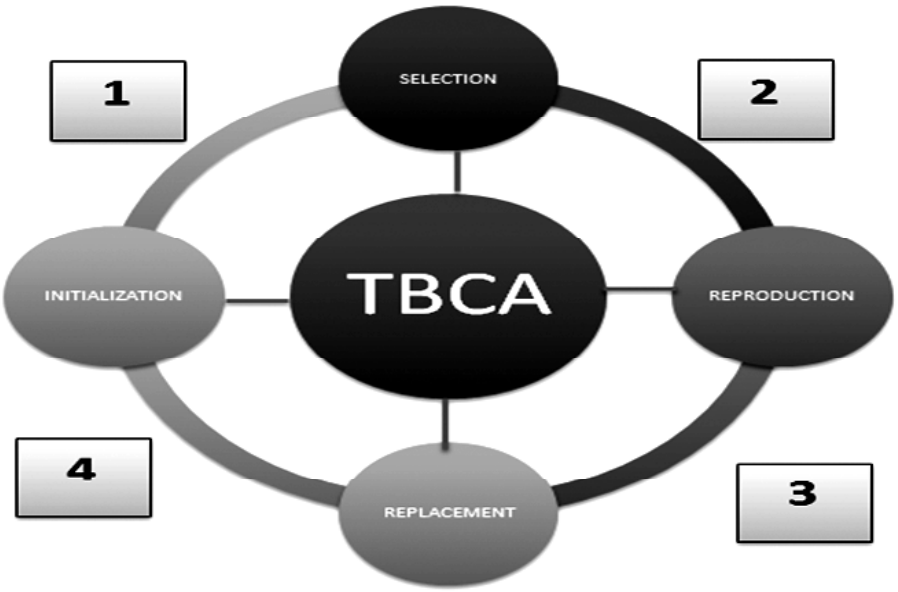

Figure 2: The Memetic based TBCA cycle

It consists of four steps.

[1] Initialization: Randomly generated population from raw data or chromosomes.

[2] Selection: Individuals are selected to create mate pool for reproduction according to selection methods.

[3] Reproduction: On selected individuals cross-over and mutation are applied to it.

[4] Replacement: Individuals are replaced by new ones comes from cross-over and mutation.

2.2. Memetic based Threshold based clustering algorithm TBCA

Table 2: It shows the comparison of memetic and TBCA.(Initialization-Selection-Reproduction(Cross-over and Mutation), Replacement).

Table 2
Comparison of memetic and TBCA

	MEMETIC		TBCA
1	Initialization • Input dataset (population)	1	Insert data series by some probability based • Sum • Weight • Error
2	Selection • Select random generation	2	PCA • Principal component analysis
3	Cross-over • Create off-spring for next generation (clustering) combining the chromosomes.(randomly chooses a point & exchanges the sub sequences before and after that point between two chromosomes to create off-spring.) (global search)	3	Fitness • Deep Learning Algorithm-Gray wolf Optimization(GWO). • Generate off-springs for next generations. And Order crossover • Choose subsequence of nodes from one parent & preserving the order from other. And Cycle crossover • Simply for searching.
4	Mutation • It changes the new off-spring by flipping bits 1 to 0 or 0 to 1.(Local search)	4	Compare PCA & fitness and generate one mutation based dataset. (closeness factor)
5	Replacement • Make new population using global and local search.	5	Cluster dataset & original dataset (comparison).Original dataset+ new dataset = threshold dataset.

3. DATA COLLECTION

The random data collected from online UCI repository dataset of gestational diabetes and juvenile diabetes and it will be analyzed using proposed techniques.

4. LITERATURE REVIEW

The literature review has been from the related papers of memetic algorithm and some techniques which are used in past to diagnosis diabetes.

The researcher umamaheshwariet al. (2016) [3], used the memetic algorithm in machine learning way for diagnosis the cancer type. They have used this algorithm for feature extracting of cancer and diagnosis the cancer and then this Algorithm hybridize by non genetic Local Search to refine solution quality with CFSSubsetEvaluator along with Rank Search are known as Memetic algorithms (MAs). There are many ways to design the Memetic Algorithm. Iterative dynamic programming (IDP), based on by integrating IDP, which is an efficient solution identification method, and anCFSSubset with genetic search algorithm (CGA), which is a global optimizer, a new MA called MA-IDP which has been applied successfully to solve the multidrug scheduling optimization problem.

NavidMoshtaghi**,**ArezooYazdaniSeqerlou et al.(2014) [4], used the XCS classifier systems to diagnosis the diabetes. In this research work they used xcs classifier system and DM is characterized by recurrent or persistent hyperglycemia, and they diagnosis the diabetes by demonstrating FPG (Fasting Plasma Glucose)level. They have also used decision support tree, XCS, AD TREE, SVM, C4.5, K STAR, DEMPSTER-SAFER Theory, Baysian method and borda count method and analyzed the results coming out from that and diagnosis the Gestational diabetes.

According to TarikRashid, SamanAbdullah and RezhnaMirza Abdullah et al. (2016) [5], they used AI techniques for prediction and Classifying the diabetes types and diagnosis the diabetes. They used fuzzy

logic classifier model for classifying the T1D and T2D and compared to ANN. fasting blood sugar level was mapped by them to classifying the T1D and T2D.

Asma A. AlJarullah et al. (2011)[6]revealed the decision tree discovery for diagnosis the type2 diabetes. This study consists of two stages, data pre-processing and decision tree construction. The data pre-processing phase aims at preparing the dataset for the second phase. The second phase includes using one of the decision tree algorithms to construct a decision tree model for the prediction of patients with developing diabetes. In Data Pre-processing they have used attribute identification and selection, handling missing values, and numerical discretization. Identification and selection they consider the number of times pregnant and age, Plasma- Glucose, Diastolic BP (blood pressure),Triceps SFT (skin fold Thickness), Serum–Insulin, BMI (Body mass index), DPF (Diabetes Pedegree Function) Age, Class (type of diabetes within last 5 years).

According to preetimulay,Rahul joshi, adityakumaranguriya, alishagonsalvis, dakshyaadeepankar, dipankarghosh et al.(2016) [7] elaborates the knowledge Management System and same for clustering techniques based on closeness factor and Threshold known as Threshold Based Clustering Algorithm (TBCA) is presented in this paper. This study resulted by considering by diabetic patients medical reports as standard input dataset. The result obtained from this TBCA proves that by considering Threshold, Eight attributes extracted by analyze their effect on patients diabetes mellitus(DM) and out of this eight , Four are impactful and another four are non-impactful attributes and validated by threshold based clustering algorithm(TBCA).

According to SuhasGaikwad, Dr. PreetiMulay, Rahul Joshi (2015-A)[8], describe the use of analytical hierarchy process AHP as a mathematical tool to recommend sugar based ice-creams to diabetic patients. Results of AHP are verified by considering different weights ratios. By considering sugar level as an impactful criteria, for both, In an ice creams and patient's blood sugar level, This system recommend the ice-creams to the diabetic patients.

According to [14] if mother is having Gestational diabetes then her baby is larger than normal and he or she is at higher risk for some complications. They have some complications like obesity and breathing problem and higher risk of having T1D diabetes.

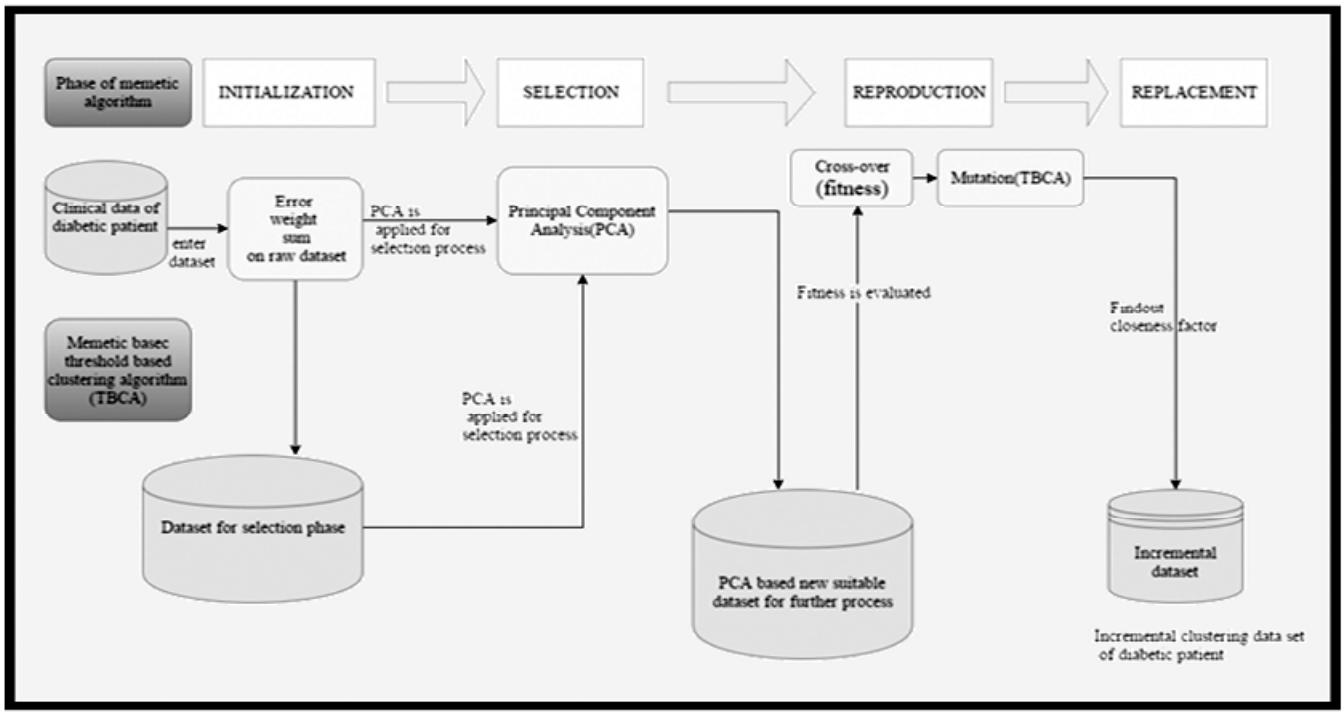

Figure 3: Research Design Architecture

5. RESEARCH DESIGN ARCHITECTURE

Figure 4 shows architecture for research design, Clinical data shows the dataset for the research. In INITIALIZATION phase, Sum, Error and Weight are going to apply and generating one dataset which will be used in the next phase. In SELECTION phase, Principal Component Analysis (PCA) is used for feature extraction and it will be applied on the dataset which is generated in first initialization phase. Then Cross-over and Mutation is applied in REPRODUCTION phase and in Cross-over the off-spring is generated and it is compared with PCA based dataset and then the generated dataset will be replaced with new generated dataset in REPLACEMENT phase.

Figure 4: It describes the flow of how input data is initialized and selected and how the further processes are going to apply on those dataset. And how the new incremental recommended dataset for diabetic patient is generated.

Figure 5: Shows an algorithm,In the first step, this is an initialization step. In this step the sum, error and weight is applied on primary dataset. And generating one DS initializedataset. In next step, it is a selection step. In that step Principal component Analysis is performed and generate a PCA based dataset $DS_{PCA.}$ Then in 3rdstep, that is Reproduction step and in it fitness is evaluated by Mean Square Error (MSE) and DSfitnessdataset is generated. And after that in same step cross-over is performed and generate the off-springs and by using that dataset DS_{PCA} will be updated. And after it In Replacement step closeness will be find out and generate the closeness factor based incremental dataset.

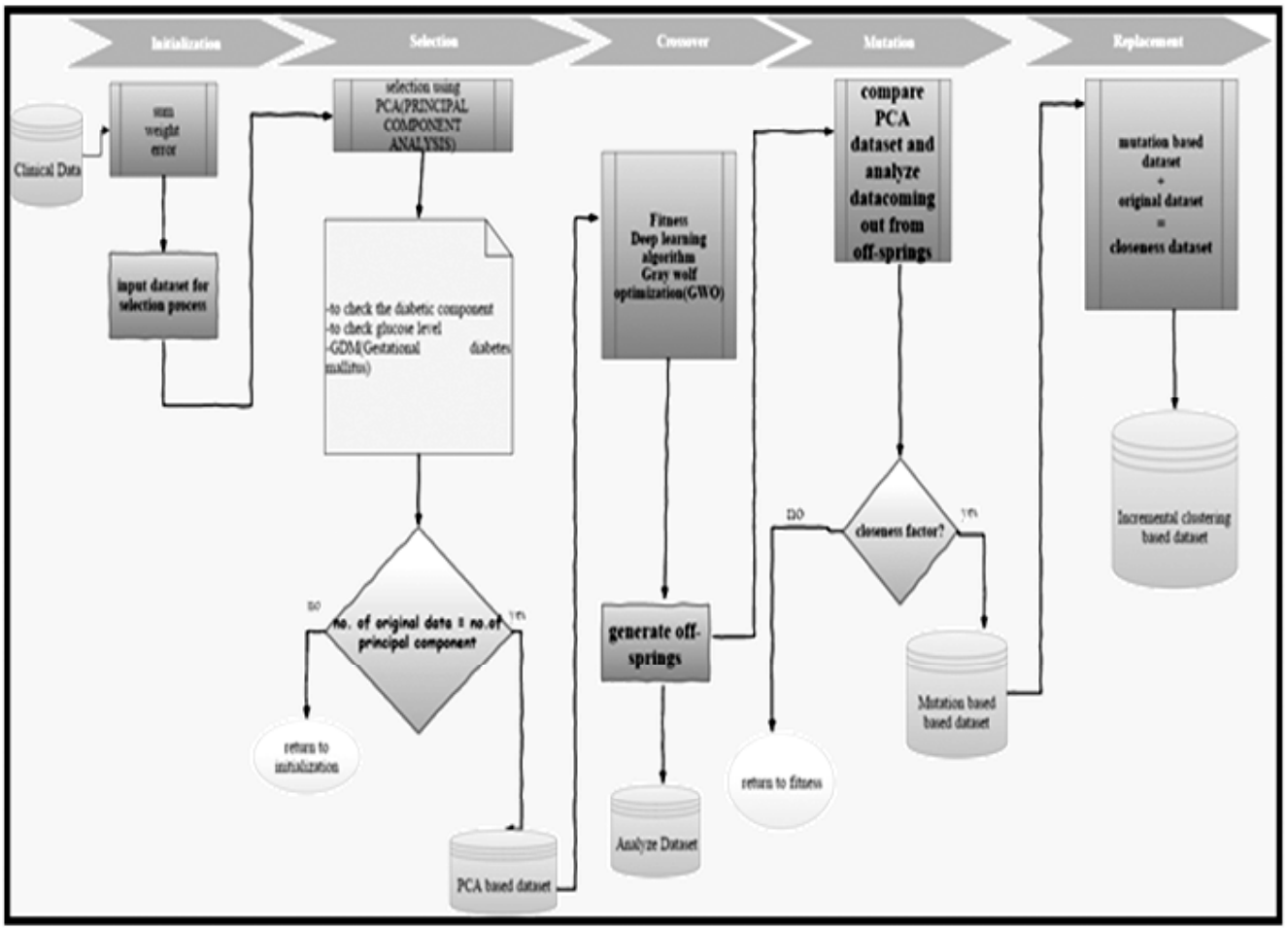

Figure 4: Flowchart for research

ALGORITHM:

Input: Data Sample, Data series.

Output: Impactful and Non Impactful attributes.

Step 1: INITIALIZATION

Enter the dataset for initialization phase.

For i = 0 to n ; i = n no. of population.

$D_s = d_{s1} + d_{s2}$; d_s = data series.

$w = w_1 + w_2 + w_3 + \ldots w_n$; w = weight.

$E = (x - x_i^2)$; E = Error.

$= \sqrt{(x - x_i^2)}$

Generate dataset $DS_{initialize}$. And it is become an input for next selection phase.

Step 2: SELECTION

PCA

Apply Principal component Analysis (PCA) in SELECTION step.

Let x be a matrix for n observation by P variable & covariance matrix is S.

$Z_1 = \sum_{i=1}^{p} a_i \cdot x_i \qquad a_i = 1,2,3,\ldots p.$

Generate dataset DS_{PCA} and it will be an input for next reproduction step.

Step 3: REPRODUCTION

FITNESS

Fitness is finding out by mean square error

$MSE = f(x) = \frac{1}{n} \sum_{i=1}^{p} f_i(x_i - t)^2 - \sum_{i=1}^{p} P_i(x_i - t)^2$

Where t = measure of center of distribution.

Generate dataset $DS_{fitness}$ and it will be compared with DS_{PCA} in next step.

Cross-over

If $DS_{PCA} = DS_{fitness}$

Then $DS_{PCA} = DS_{PCA}$

Else

$DS_{PCA(i)} = DS_{PCA(i+1)}$

Update DS_{PCA}.

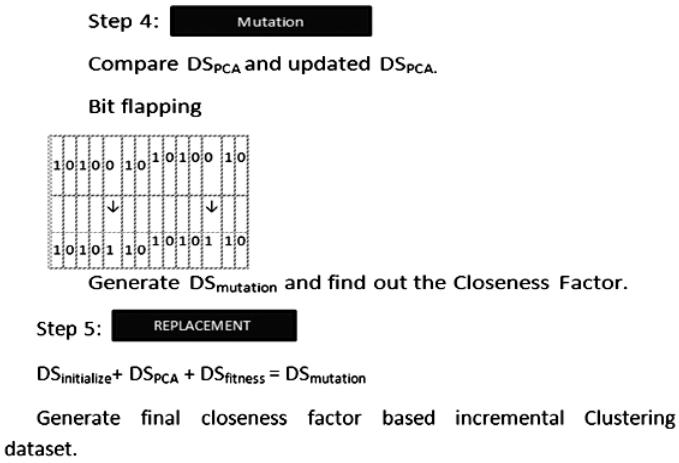

Figure 5: Algorithm

6. CONCLUDING REMARKS

The proposed exploratory research aims to diagnosis the diabetes and find out whether the Juvenile Diabetes and Gestational Diabetes are related to each other or not. Till this day it is proved that if mother have Gestational diabetes then there is higher risk for Juvenile diabetes so the aim is to find out the relation between Gestational Diabetes and Juvenile Diabetes.

7. FUTURE ENHANCEMENT

The proposed work is related to diagnosis the diabetes and shows Gestational Diabetes and Juvenile diabetes are related to each other or not. After it we can prove it how they are related and which are the impactful criteria for Gestational diabetes and juvenile Diabetes and how it can prevent.

REFERENCES

[1] http://www.npr.org/blogs/health/2012/06/21/155505445/how-to-spot-a-neglected-tropical-disease.

[2] By ruthtoiba, phd, rd, cde, gestational diabetes — diagnosis, treatment, and outcomes

[3] B. umamaheswari, v. steffi joy, machine learning approach for identifying and diagnosing the cancer types 2016.

[4] Navidmoshtaghi, arezooyazdaniseqerlou, diabetes diagnosis via xcs classifier system 2014.

[5] Tarik a. rashid, saman m. abdullah and rezhnamirzaabdullah,an intelligent approach for diabetes classiûcation, prediction and description.2016.

[6] decision tree discovery for the diagnosis of type ii diabetes.2011

[7] Dr. preetimulay ,RahulR joshi, adityakumaranguriya, alishagonsalvis, dakshyaadeepankar, dipankarghosh , analysis via incremental clustering; data mining of diabetes mellitus using threshold based clustering algorithm(tbca) 2016

[8] Suhasgaikwad, dr. preetimulay, rahulraghvendrajoshi, analytical hierarchy process to recommend an ice cream to a diabetic patient based on sugar content in it 2015.

[9] Anjali khandegar, a review on the diagnosis of diabetes mellitus 2015.

[10] Akashladdha,rahulraghvendrajoshi, dr.peetimulay, enriching process of ice-cream recommendation using combinatorial ranking of ahp and montecarloahp 2016.

[11] B. ganeshprabu, m. geethanjali, dynamic monitoring and optimization of faultdiagnosis of photo voltaic solar power system using ann and memetic algorithm2016.

[12] Messaoudanekkaa,dalilaboughaci, a memetic algorithm with support vector machine for feature selection and classiûcation.

[13] A. zibakhsh, m. sanieeabadeh, gene selection for cancer tumor detection using a novel memetic algorithm with a multi-view ûtness function 2013.

[14] http://www.webmd.com/baby/gestational-diabetes-you#1