International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075, Volume-8 Issue-8 June, 2019

A Novel Partitioning Driven Differential Evolution based Epileptic Seizure Cluster Analysis

Maninder Kaur, Meghna Dhalaria

Abstract*: *Epilepsy is amongst the most widely recognized neurological ailments that reveals in redundant epileptic seizures because of an anomalous, synchronous movement of an enormous gathering of neurons.This disease comes at third level among the prevalent brain ailment that can cause detrimental effects on the day-to-day life of the victim. The early recognition of epileptic seizures is of main concern for the identification of victims with epilepsy. The present work proposes a novel partitioned based differential evolution clustering approach for detecting the epileptic seizures. The cluster process aims to arrange a collection of data samples into different clusters, in such a way that the objects belonging to a cluster are too close to one another than the objects belonging to different clusters. The partition method is primarily based on greedy heuristics approach that is used iteratively to obtain an optimal set of clusters. The proposed Partition Based Clustering using Differential Evolution (PCDE)approach is compared with Differential evolution based clustering based on cluster validity measure DB index. The empirical analysis is carried out onEpileptic Seizure medical datasets. The results revealed that PCDE approach required less computation time in comparison to DE based clustering and outperformed DE based clustering with DB Index value of 0.8626.

Index Terms*: *Epileptic Seizure,Differential Evolution, Clustering,Partitioning.

I. INTRODUCTION

Epilepsy is a typical interminable nerve issue, which is set apart by seizures. It is generally instigated by extreme ejection of cortical cells from the brain. In epilepsy syndrome, the victimgo through seizures triggered by the functionality disorder of the brain of the patients [1]. Out of 0.05 billion folks in the United States detected with epilepsy syndrome, around thirty lac patients are influenced by epilepsy syndrome. [2].

This disease comes at third level among the prevalent brain ailment. It has four different stages. The first one is preictal state that shows up before the seizure activates. Next is the ictal state, which initiates by the commencement of the attack and finishes with a seizure.The next state is the postictal state. The last state is interictal state that commences next to postictal state of first attack and finishesprior to the beginning of preictal state of consecutive seizure. The seizures may be easily forecasted by identifying the commencement of the preictal state.

Observing the activity of the brain by the EEG (electroencephalogram) has turned out to be an imperative instrument in the field of epilepsy detection. The electroencephalogram recordsof the patients affected by epilepsy demonstrate two classes of a typical activity. One activity is Ictal that is the activity recorded at the time of seizure. Another one is inter-ictal, which is theanomalous signals noted during seizures.

Among numerous reasons of this disease, one is molecular transformation that causesasymmetrical neuronal movement of neurons. However, the underlying reason of epilepsy is still mysterious;the initial identification may be helpful for the treatmentof epilepsy. The sufferers of epilepsy may undergo treatment through medicine or clinical processes [3]. Nevertheless, these procedures are not entirely efficacious. Unluckily, the seizures, which cannot be entirely curedthrough medical treatment, shows detrimental effects on the day-to-daylife of the victim. In such cases, it is difficult for the patients work independently work and perform their daily- activity. This further result in social loneliness of folks and financial hitches. The early detection of the seizures gives sufficient time to circumvent the outbreak by proper medication.

Until now, both supervised and unsupervised machine learning approaches have been exploited in epileptic seizure disease analysis. The authors Rasekhi et al. [4] used univariate linear features and anticipated an approach for prediction of epileptic seizure. The work assumed the beginning of preictal time 10 to 40 minutes before the ictal state with a difference of 10 minutes. The authors utilized six EEG channels to extract twenty-two univariate linear features thereby creating a feature space with 132 features. The results obtained the 73.90% prediction sensitivity after the application of proposed approach. Bandarabadi et al. [5] applied Support Vector Machines for classification on the extracted spectral power features.The results showcased sensitivity of 75.8%.Zandi et al. [6] utilized the EEG signals of scalp based on zero crossings. The authors calculated the histogram of all interims in a sliding window and chosen values from specific bins for interpretations.Paschalis et al. [7] proposed an automated unsupervised method of detecting the epileptic attacks based on Ensemble Empirical Decomposition and k-means clustering . The proposed approach gave promising results with accuracy 98% in comparison with related methods. Mormann et al. [8] proposed a new mutation operation for rapid feature selection through Genetic Algorithm (GA) relying on elitism of the allele. The fittest chromosomes were preserved by normal elitism in GA, which were then evaluated by utilizing the fitness function. The highest fit allele was conserved and the evaluation of fitness of the allele performed based on the occurrences of frequency. The chromosome that underwent this mutation approach was found to have the highest fitness as it was created based on the fittest alleles. The proposed approach increased the search for the fittest chromosomes, and minimized the time taken for optimal convergence. Guo et al. [9] implemented Genetic Programming to perform automatic feature extraction from the original database in the epileptic EEG classification process. The discriminatory performance of k-Nearest Neighbors (k-NN) classifier was enhanced and the input feature dimensionality was reduced. On the other hand, the keydisadvantage of the technique is that the Genetic Programming based feature extraction system is costly. Orhan [10] used wavelet to extract the features of the

Revised Manuscript Received on June 11, 2019

ManinderKaur, Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala, India

Meghna Dhalaria,Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala, India

epileptic EEG signals during seizures and seizure free intervals. The log variance of wavelet coefficients of epileptic EEG data was used in classifying the epileptic EEG data. The computational results revealed significant difference between the log variance of wavelet coefficients of EEG data of seizures and that of non-seizure periods. Based on the results the authors have concluded that the region of the brain where epileptic EEG data were recorded has an influence on the analysis and diagnosis of epilepsy. Acharya et al. [11] examined the performance of five different classification model namely Decision tree,Fuzzy Neural Network, Support Vector Machine and k-Nearest Neighborfor prediction of epileptics seizures. They concluded that support vector machine performs better as comparison to other model.

The major work in literature has been done on the EEG signal, using classification models namely support vector machine, Fuzzy neural network etc and clustering with k-means. No attention has been paid on the application of evolutionary approaches in this field.This is the first step towards the application of evolutionary approach for creating a a novel partitioning based clustering using differential evolution approach i.e **P**artitioning based **C**lustering using **D**ifferential **E**volution (PCDE)approach that outperforms the simple differential evolution based clustering in terms of DB index and computational time.

II. PROPOSED METHODOLOGY

Partitioning method is tends to be more efficient and easily adapted for huge datasets. These methods primarily based on greedy heuristics approach that are used iteratively to obtain a local optimum solution. In partitioned technique, the data set is divided into the clusters using some technique. For the given database of n objects, a partitioned algorithm partitions the data into ‘k’ (number of clusters, which are pre-decided), so that optimization of an objective function can be done. Each cluster has at least one object and each object belongs to only one cluster. The maintarget of partition based clustering is to partition the set of data points into K parts. Each partcorrespond to one cluster. The technique of partitioning strategy relies on specific objective function. Partitioned clustering algorithms decide a grouping solution by way of maximize the similarities among objects inside the same groups while minimize the dissimilarities between distinct groups.

The process utilizes Differential Evolution approach for partitioning the data into different cluster. This technique was developed by R. Storn and K. Price in 1996[12]. It is robust and fast optimization technique that uses a stochastic, populace based search methodology. DE is used for the real value problems. DE includes three main operator namely mutation, crossover, and selection. Firstly create a mutant vector (donor vector) using mutation by randomly selected three vectors; secondly, crossover is done between donor vector and parent vector to create trail vector. While the trial vector is generated, the selection step is applied to maintain either of the vectors. The simplest way is to preserve the best fitness vector.

A. The proposed PCDE Approach

The proposed methodology divides the existing dataset into distinct parts within the feature space. Suppose X = {x_1, x_2, x_3,...., x_n} be the set of sample space. Let $\bar{x}_m$ denotes the mean of mean of X as in eq (1).

$$\bar{x}_m = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

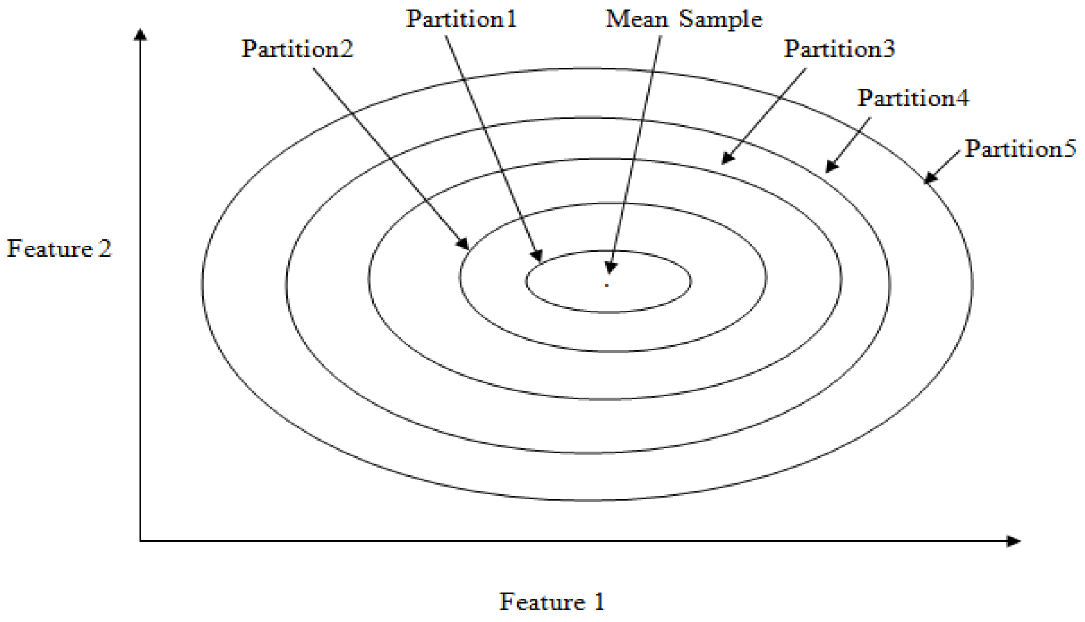

Fig 1: Partitioning Methodology
(Comparing the distance of objects from its mean sample)

Algorithm 1: Partitioning based Clustering using Differential Evolution approach (PCDE)

```
Input : InputFile, K
Begin
    X.Data=Read(InputFile)
X.Mean=mean(X.Data)
    X.D= Euclidean_distance(X.Data, X.Mean)
X.D= Sort(X.D, 'ascending')
    min= X.D(1)
    max = X.D(Data_Size)
    [BestSol, BestCost]= DE(X.Data,min, max,K)
End
```

DE(X.Data,min, max, K)

Input:- Pop_Size- Population Size, Max_Iteration-Maximum Iterations,pCR-Probability of crossover,β_min-minimum limit of scaling factor, β_max-maximum limit of scaling factor.
Output:-BestSolution, BestCost along with the partition range of cluster.
Begin

```
Initialize(nVar, VarSize, Max_Iteration)
  Set VarMin= min_tuple(X)
  Set VarMax= max_tuple(X)
for(i=1 to Pop_Size)
             Pop.solution=Initialize_Population(min,max,K-1)
             Pop.fit= clusteringcost(pop.solution,min,max,X)
if(Pop(i).fit<BestSolution.fit)
                      BestSolution=Pop(i)
                 Endif
          Endfor
for(it=1 to to Max_Iteration)
for(i=1 to Pop_Size)
                    x=Pop(i).Pos;
                    Y=randperm(Pop_Size);
                    Y(Y==i)=[]
                    p=Y (1)
                    q=Y (2)
                     r=Y (3)
β= random_generation(β_max ,β_min ,VarSize)
                      a=Pop(p).Pos+β.*(Pop(q).Pos-Pop(r).Pos)
                      a=maximum (a,VarMin)
                      a=minimum(a,VarMax)
                      b[1:x]=[0]
                       i0=random_integer([1 , numelements(x)])
for(l=1to. numelements(x))
if(l==i0 || random_no<=pCR)
                                  b(l)=a(l)
                              Else
                                  b(l)=x(l)
                              Endif
                         Endfor
                   NewSolution.Pos=b
                  [NewSolution.Fit, NewSolution.Out]=clusteringcost (NewSolution.Pos)
if(NewSolution.Fit<Pop(i).fit)
                                   Pop(i)=NewSolution;
if(Pop(i).fit<BestSolution.fit)
                                        BestSolution=Pop(i)
                                   Endif
                              Endif
               Endfor
       BestCost(it)=BestSolution.Fit;
End
```

Fig 2: Illustration of the Proposed Methodology

Let D = {d_1, d_2,..., d_n} represents the distance calculated by calculating the deviation of each xi ∈ X from the its mean $\bar{x}_m$as in eq(2)

$$d_i = euclidean\_distance(x_i,\bar{x}_m) \quad (2)$$

Let d_{min} and d_{max} represent the minimum and maximum value in D respectively that will be used to partition the samples in X. The process is initiated by sorting the distances D in ascending order and then this sorted list is divided into k sets using differential evolution algorithm where k denoted the count of clusters. The figure 1 represents the partitioning strategy where the partitions are decided by finding the distance of objects from its mean sample.

For Partitioned based DE- clustering, mean of the data *X.Mean* is calculated and measures the distances using Euclidean Distance *(X.D= X.Data - X.Mean),* sort the *X.D* in ascending order .Here d_{min} and d_{max} represnt the maximum and minimum D values.Then the split the data values D € [d_{min} ,d_{max}] in k cluster values using Differential evolution approach keeping DB index as fitmess function.For partitioning method, if there is n clusters (number of clusters) then n-1 near optimal partition values within the range of [d_{min} ,d_{max}] are generated by Differential Evolution algorithm.

Figure 3 shows the solution representation for ‘n’ clusters where Pi denotes the ith partition value.

P_1	P_2	..	P_{n-1}

Fig 3: Solution representation for PCDE

For eg. If the value of n=4 ,then there are n-1 partitions found by DE algorithm .Thus the resultant partitioning points will be [x_{min}, P_1, P_2, P_3, x_{max}] and the objects lying within the distance ranging from [x_{min}, P_1) belongs to first cluster,[P_1, P_2) to second cluster ,[P_2, P_3) to third cluster and [P_3, x_{max}] to fourth cluster.

Fitness Function

The clustering process partitions the original data into an optimal set of clusters bearing the subsequent properties: more similarities within the clusters and more dissimilarities between clusters. DB and Dunn index are used for measuring the validity of cluster. As the lower DB index, better the quality of cluster by making more compact and separated clusters where as higher the Dunn value better is the cluster.In this work, DB index is used as fitness function.

Mutation

For each vector α_i at time t, first randomly choose three different vectorsx_a,x_b and x_c at *t* .The mutation is done to generate donor vector v_j. Donor vector are generated by including a weighted difference of two populace vector to a third vector as in eq(3).

$$v_i^{t+1} = x_a^t + F(x_b^t - x_c^t) \quad (3)$$

Here F is a parameter, represents the differential weight where F ϵ [0, 2], v_j is donor vector generated by mutation.

Crossover

Donor vector does crossover with the target vector (current generation) to generate the trial vector. pCR (Crossover Probability) parameter is used for controlling the rate pCR ϵ [0, 1]. In this binomial crossover is used to perform crossover on each d variables or component by generating a consistently appropriated random number r ϵ [0, 1], if random number r, is greater than crossover probability then value of target vector become trial vector otherwise value of donor vector become trial vector as in eq(4)

$$u_{j,i}^{t+1} = \left\{ \begin{matrix} v_{j,i} & if\ r_i \leq pCR, \\ x_{j,i}^t & otherwise, \end{matrix} \ j = 1,2,3 \ldots .d. \right\} (4)$$

Selection

Selection is same as that used in genetic algorithms. Trail vector is compare with the parent vector and choose the fittest value as in eq(5)

$$x_i^{t+1} = \left\{ \begin{matrix} u_i^{t+1} & if\ f(u_i^{t+1}) \leq f(x_i^t) \\ x_i^t & otherwise. \end{matrix} \right\} (5)$$

Termination Criteria

The procedure is repeated for a fixed range of iterations and optimization results gives the best recorded solution and cost at the end of each run.

Figure 2 illustrates the detailed steps of the proposed PCDE approach that utilizes Differential evolution for generating optimal values of partitions.

B. Differential Evolution Based Clustering (DEC)

Solution Representation

Given a dataset (*m x n* size) containing ‘*m*’ samples with ‘*n*’ attributes, the solution for clustering the data in *k* clusters is encoded as *n x k* size linear array representing the *k* cluster centroids in sequence.

Let *n*=2 where *n* be the number of attributes in a dataset and *k*=4 where *k* be the number of cluster.Then the solution is encoded as shown in figure 4.

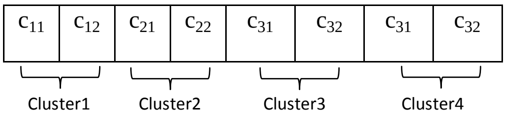

Figure 4: Solution Representation of DEC approach

C. Validity measure used

Validity Measures are used to check the quality of the clusters formed by the algorithm. The validity measures quantify the quality of clusters on the basis of the characteristics that are intrinsic in the data sets.

Davies-Bouldin (DB) Index

Donald W. Bouldin and David L. Davies introduced DBI in 1979. It is a used for evaluating clustering algorithms [13]. DB Index evaluates the inter cluster differences and intra cluster similarity. It is defined as in eq(6).

$$\mathrm{DB} = \frac{1}{c}\sum_{i=1}^{c} Max_{i \neq j} \left\{ \frac{d(x_i)+d(x_j)}{d(c_i c_j)} \right\} (6)$$

Where C denotes the number of clusters, $d(x_i)$and $d(x_j)$ are clusters sample i and j to their appropriate clusters centroid, i and j are cluster label and $d(c_i c_j)$ distance between the centroids.

Dunn Index

It was introduced by J.C. Dunn in 1974 [13]. It is a metric that aims to recognize the set of clusters that are closely packed, having small variance value between the data points of a cluster, are well-separated, far apart from the dissimilar data points, i.e. means of clusters values have sufficient distance among them [13]. For the given data set, higher the value of DI, better is the clustering approach. It is defined as in eq(7).

$$Dunn = min_{1 \leq i \leq c} \left\{ \min \left\{ \frac{d(c_i c_j)}{max_{1 \leq k \leq c}(d(X_k))} \right\} \right\} (7)$$

Where d $(c_i c_j)$ represents the inter-cluster distance between cluster, c is the cluster count and $d(X_k)$ denotes the intra-cluster space.

III. SIMULATION RESULTS

The proposed Novel Partitioning based clustering using Differential evolutionary approach and Differential evolution based clustering is implemented using MATLAB (R2016a) on Intel core i7 processor,5th Gen with 8 GB RAM under 64-bit Operating System. The proposed approach is executed for 10 runs, where each run consists of 100 iterations of the proposed Novel Partitioning based clustering using Differential evolutionary approach algorithm. The Dataset is collected from UCI Repository of Epileptics Seizures Recognition Dataset[14]. It contains 11500 numbers of instances and 179 numbers of attributes with no missing values.

Table 1: The control parameters of Partition Based clustering using DE approach

Partition Based clustering using DE approach (PCDE)	
Parameters	**Values**
Max_It	50
PopSize	100
β_min	0.2
β_max	0.8
pCR	0.2

Various control parameters used in partitioning clustering using evolutionary algorithm are: PopSize-Population size, Max_It-Maximum iterations, β_min-Lower bound, β_max-Upper bound, pCR-Crossover probability (parameters values tabulated in table 1).

Table 2: Comparative results of proposed PCDE approach and DE Clustering
(In terms of DB index(DBI), Dunn index(DI) and Computational Time(CT)in seconds)

	DE-Clustering			**PCDE**		
No. of Runs	**DB Index**	**Dunn**	**CT**	**DB Index**	**Dunn**	**CT**
Run 1	0.8987	0.8670	57.545	0.8674	0.9441	45.237
Run 2	0.8796	0.8756	46.822	0.8770	0.9234	53.896
Run 3	0.9952	0.9342	58.527	0.8515	0.9548	48.532
Run 4	0.8584	0.7944	48.354	0.8242	0.9704	44.121
Run 5	0.8417	0.9575	56.285	0.8686	0.8125	38.959
Run 6	0.9266	0.8935	48.974	0.8516	0.8737	55.237
Run 7	0.8802	0.8541	58.367	0.8905	0.9537	39.667
Run 8	0.86 17	0.8834	46.762	0.8472	0.9615	49.750
Run 9	0.9584	0.9032	59.868	0.9103	0.8737	57.363
Run 10	0.8643	0.8467	57.684	0.8381	0.9704	57.316
Average	0.9003	0.8809	53.918	**0.8626**	0.9238	49.007

The table 2 depicts the DB index, Dunn index and computational time of proposed PCDE and DE based clustering. The lesser the value of DB index, better the cluster is. It is clear from the table that PCDE approachgives better results in terms of average DB index, in comparison to DE based clustering. The results shows lower DB index value of 0.8626 for PCDE approach. Similarly, the computation time for PCDE is about 49 seconds,comparatively lesser than DE-Clustering approach.

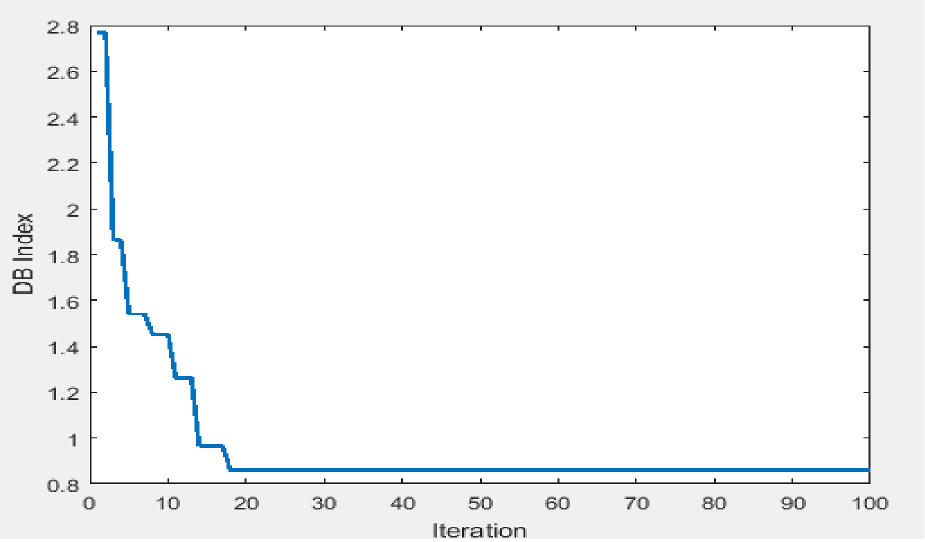

Fig 5: depicts the DB-index vs number of iterations forPCDE approach

.

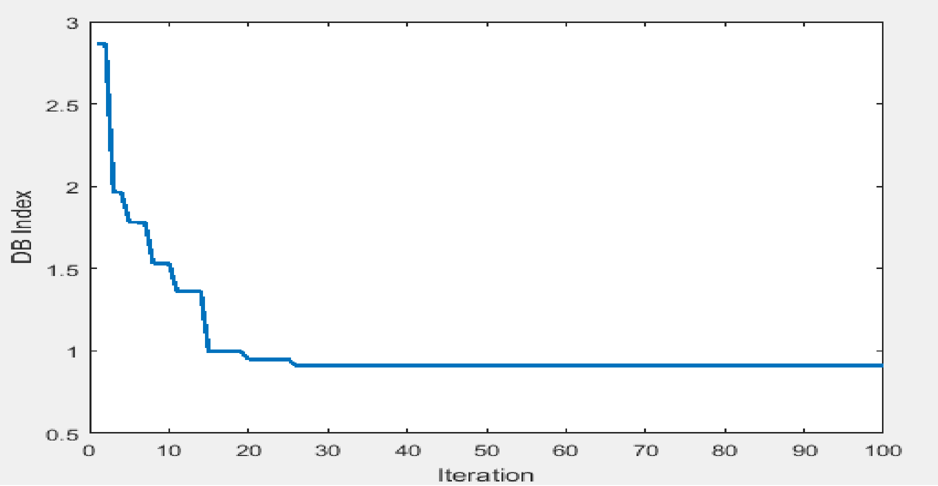

Fig 6: depicts the DB-index vs number of iterations forDEC approach

Figure 5 shows that graph plotted between number of iteration and DB Index values for PCDEapproach. It is clear from the graph that from 17th iteration onwards, the PCDE approach gives stagnant results with DB index value of approx. 0.8. Figure 6 shows that graph plotted between number of iteration and DB Index values for DE based clustering. In this case, the results become stagnant with DB index of 0.9 only after 25thiteration, which is comparatively farther in comparison to the results of the proposed PCDE approach. From the results, it can be concluded that the proposed PCDE approach outperforms the DE based clustering approach both in terms of DB index and in terms of computational time.

IV. CONCLUSION AND FUTURE DIRECTIONS

In this paper, a novel PCDE (Partitioned based clustering using differential evolution approach) is proposed for cluster analysis of epileptic Seizure. The work utilizes Differential Evolution algorithm to generate optimal set of clusters. The proposed approach is implemented on Epileptic Seizures dataset. Theexperimental result concludes that PCDE approach performs better than DE based clustering algorithm in terms of DB index and computational time. The work can be extended by applying the proposed approach on large-scale datasets using Hadoop environment.

REFERENCES

1. U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, “Automated EEG analysis of epilepsy: a review,” Knowledge-Based Systems, vol. 45, pp. 147–165, 2013.
2. R. S. Fisher, W. Van Emde Boas, W. Blume et al., “Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE),” Epilepsia, vol. 46, no. 4, pp. 470–472, 2005.
3. M. Guenot, “Surgical treatment of epilepsy: Outcome of various surgical procedures in adults and children,” Revue Neurologique, vol. 160, no. 5, pp. S241–S250, 2004.
4. J. Rasekhi, M. R. K. Mollaei, M. Bandarabadi, C. A. Teixeira, and A. Dourado, “Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods,” Journal of Neuroscience Methods, vol. 217, no. 1-2, pp. 9–16, 2013.
5. C. A. Teixeira, B. Direito, M. Bandarabadi et al., “Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients,” Computer Methods and Programs in Biomedicine, vol. 114, no. 3, pp. 324–336, 2014.
6. Zandi, Ali Shahidi, et al. "Predicting epileptic seizures in scalp EEG based on a variational bayesian gaussian mixture model of zero-crossing intervals." IEEE Transactions on Biomedical Engineering 60.5 (2013): 1401-1413.
7. Bizopoulos, Paschalis A., et al. "EEG epileptic seizure detection using k-means clustering and marginal spectrum based on ensemble empirical mode decomposition." 13th IEEE International Conference on BioInformatics and BioEngineering. IEEE, 2013.
8. Mormann, Florian, et al. "On the predictability of epileptic seizures." Clinical neurophysiology 116.3 (2005): 569-587.
9. Guo, Ling, et al. "Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks." Journal of neuroscience methods 191.1 (2010): 101-109.
10. Orhan, Umut, Mahmut Hekim, and Mahmut Ozer. "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model." Expert Systems with Applications 38.10 (2011): 13475-13481.
11. Acharya, J. Lina, and J. Gotman, “Comparison Machine Learning Algorithms for Recognition of Epileptic Seizures in EEG.” Computational and Mathematical Methods in Medicine,2012.
12. S. Das, S. S. Mullick, P. N. Suganthan, "Recent Advances in Differential Evolution - An Updated Survey," Swarm and Evolutionary Computation, doi:10.1016/j.swevo.2016.01.004, 2016.
13. Davies, David L., and Donald W. Bouldin.A cluster separation measure.;Pattern Analysis and Machine Intelligence, IEEE Transactions on 2 (1979): 224-227.
14. Michalski,R.S. Learning“UCI repository of machine learning databases,”1987.[Online].Available:https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition.

AUTHORS PROFILE

Maninder Kaur is holding an academic position as Assistant Professor in Department of Computer Science and Engineering, Thapar Institute of Engineering & Technology, Patiala. She received her Bachelor Degree from Sant Longowal Institute of Engineering and Technology and Master’s degree from Punjabi University. She completed her Ph.D. in the field of VLSI physical design automation using Evolutionary approach. Her major research experiences and interests include IOT, Big Data Analytics, Data Mining and Swarm intelligence. She has 25 publications including prominent journals, conferences. She has also acted as mentor in various capstone projects in the field of IOT and machine learning. She is currently supervising two Ph.D. and two M.E. students. She has also supervised more than 11 M.E./M.Tech. thesis. She is an associate member of The Institution of Engineers, India. Her current research includes the application of machine learning and swarm intelligence techniques for big data analytics, cyberbullying and sarcasm detection.

Meghna Dhalaria is pursuing Ph.D in computer Science Department, Jaypee University of Information and Technology, Waknaghat. She received her Bachelor’s degree from Baddi University of Emerging Sciences and Technologies. She completed her Master’s degree from Thapar Institute of Engineering and Technology, Patiala. Her current research includes the applications of Machine learning and Deep learning.