Volume: 14 Issue: 01 Pages: 5277-5286(2022) ISSN: 0975-0290

Bio Inspired Algorithms for Dimensionality Reduction and Outlier Detection in Medical Datasets

Dr. S. Vijayarani
Assistant Professor, Department of Computer Science, School of Computer Science and Engineering, Bharathiar University, Coimbatore
E-mail: vijimohan_2000@yahoo.com
Dr.C.Sivamathi
[1]Assistant Professor, Department of Computer Science, PSG College of Arts & Science, Coimbatore
E-mail: c.sivamatghi@gmail.com
Mrs.S. Maria Sylviaa
Assistant professor in Department of computer science, Nirmala college for women, Coimbatore, India
E-mail: mariasylviaa1991@gmail.com

ABSTRACT

Dimensionality Reduction is one of the useful techniques used in number of applications in order to reduce the number of features to improve the productivity and efficiency of the task. Clustering is one of the influential tasks in data mining. Dimensionality reductions are used in data mining, Image processing, Networking, Mobile computing, etc. The elementary intention of this work is to apply dimensionality reduction algorithms and then cluster the datasets to detect outliers. A bio-inspired ACO (Ant Colony optimization) algorithm has been proposed to reduce dimensionality. Also another bio-inspired algorithm FA (Firefly Algorithm) has been proposed to detect outliers. The three distinct medical datasets: thyroid dataset, Oesophagal dataset and Heart disease dataset are used for experimental results.

Keywords**: Dimensionality reduction, Clustering, Outlier detection, ACO (Ant Colony optimization) algorithm, FA (Firefly Algorithm).**

Date of Submission: May 03, 2022 Date of Acceptance: May 25, 2022

I. INTRODUCTION

To perform assorted data mining tasks like classification, association rule mining, clustering, neural networks, datasets play a vital role. Every dataset is having number of features and instance in it and the number of features required for performing data mining tasks differ from application to application [1]. Numbers of features are required for the performance of data mining tasks. Not entire features are required for the performance. If at all the attributes are used it will affect the efficiency and consumes more memory space. Thus, it is mandatory to select the significant features or to reduce the number of features for performing the further tasks.

Dimensionality Reduction is one of the useful techniques used in number of applications in order to reduce the number of features to improve the productivity and efficiency of the task. There are two different sectors in dimensionality reduction they are Feature Extraction (FE) or Feature Reduction and Feature Selection. Feature Extraction is the mechanism of leveling the high dimensional data into minor dimensional space, whereas, Feature Selection is the mechanism of recruiting only the profitable data and eliminating the other. All the attributes are used for performing data mining tasks and it will affect the efficiency and consumes more memory space. So, the dimensionality reduction technique is used by the data set to reduce the number of attributes. Some of the applications of feature extraction are semantic analysis [2], data compression, data decomposition, projection and pattern recognition [2].

After performing the reduction task the clustering task is performed in order to detect outliers. The output features of the reduction process are given as input to perform clustering. An outlier is an object that is significantly dissimilar or inconsistent to other data object [3]. Clustering is the practice of grouping the identical objects together. Outlier detection is the dominant step in data mining. Numerous methods have been developed for the clustering tasks to detect outliers. The Clustering algorithms are widely divided into two groups they are hierarchical and partitioning [29]. Data clustering procedures are highly supportive to detect outliers and detecting outliers is one of the data mining tasks and it is also called as outlier mining.

There are many algorithms for outlier detection in static and stored data sets which are based on a variety of approaches like nearest neighbor based, density based outlier detection, distance based outlier detection and clustering based outlier detection [4].The identical data are grouped together to form a cluster and the outliers are detected based on cluster based outlier detected.

The elementary intention of this work is to correlate the performance of dimensionality reduction algorithms and then to perform clustering task to detect outliers. An

Volume: 14 Issue: 01 Pages: 5277-5286(2022) ISSN: 0975-0290

ACO (Ant Colony optimization) have been proposed for the reduction task. Also FA (Firefly Algorithm) have been proposed to detect outliers by clustering the data.

The remaining sectors are formulates as follows. Section 2 illustrates the literature review. Section 3 explains proposed ACO algorithm for dimensionality reduction and also illustrates FA algorithm for outlier detection. Section 4 gives an outline about the experimental results and conclusion is described in section 5.

II. LITERATURE REVIEW

Dr. S. Vijayarani et.al. [4] discussed two different clustering algorithms namely BIRCH with K-means and CURE with K-means to detect outliers. Two performance factors such as clustering accuracy and outlier accuracy are used for observation. Through examining the experimental results, it has been found that the CURE with K-means clustering algorithm performance is more accurate than the BIRCH with K-means algorithm.

S. D. Pachgade et.al. [5] proposed an algorithm to group the data items into number of clusters. The computational time is reduced by reducing the size of the dataset. Threshold value is used by the users in order to calculate the outliers from the data in each cluster. Here, it has been proved that the hybrid clustering approach takes less time for clustering task.

Elahi M. Kun Li et.al. [12] discussed about a data clustering based approach, the data are divided into fixed number of cluster using k-means algorithm. In, this work the outliers are calculated by taking the mean value for every cluster and taking the output of previous cluster mean value as the input to the next cluster value. Then, k-means algorithm have been performed for different datasets to find the better outlier detection techniques which makes less cost consumption.

Behera Abhishek, et.al. [13] discussed about several methods for detecting outliers like density based method and distance based method. A new outlier detection technique has been proposed and compared with existing PAM clustering algorithm. bupa dataset has been used for this experiment. This data set has 2 classes and 6 dimensions. In, PAM clustering algorithm it is found that 19 data points were outlier in class1 and 17 outliers in class2. From proposed algorithm it is found that 19 data points were outlier in class1 and 20 outliers in class2.

Rekha Aswathi et.al [15] have used normalization filtering which is used to standardize all the features in the dataset the dimensionality reduction and clustering is performed. Here, the diabetic's dataset which contains 768 instances and 8 attributes and PCA algorithm is used to reduce the features of the dataset. Out of 8 features, 4 features are selected. WEKA3.7 tool is used for the investigation. After dimensionality reduction density based clustering algorithm is used to find the maximal set of density. Dimensionality reduction is used to increase the accuracy of the clustered data.

Rajendra Pamula et.al. [8] presented the clustering based outlier detection method, the k-means clustering algorithm to splits the data set into number of clusters and the points which are nearby the centroid of the cluster are not feasible data for outlier and it can be pruned out from each cluster. In this work distance based outlier score is calculated for remaining datapoints which reduces the consumption due to the pruning of some points. Based on the score of the top n points are declared the highest score is considered to be the outliers and the local distance based outlier factor(LOF) is used to measure the degree an object which deviates from its neighborhood region points. The experimental results demonstrate the number of computations for detecting outliers. Here, the proposed method performs better than the existing method.

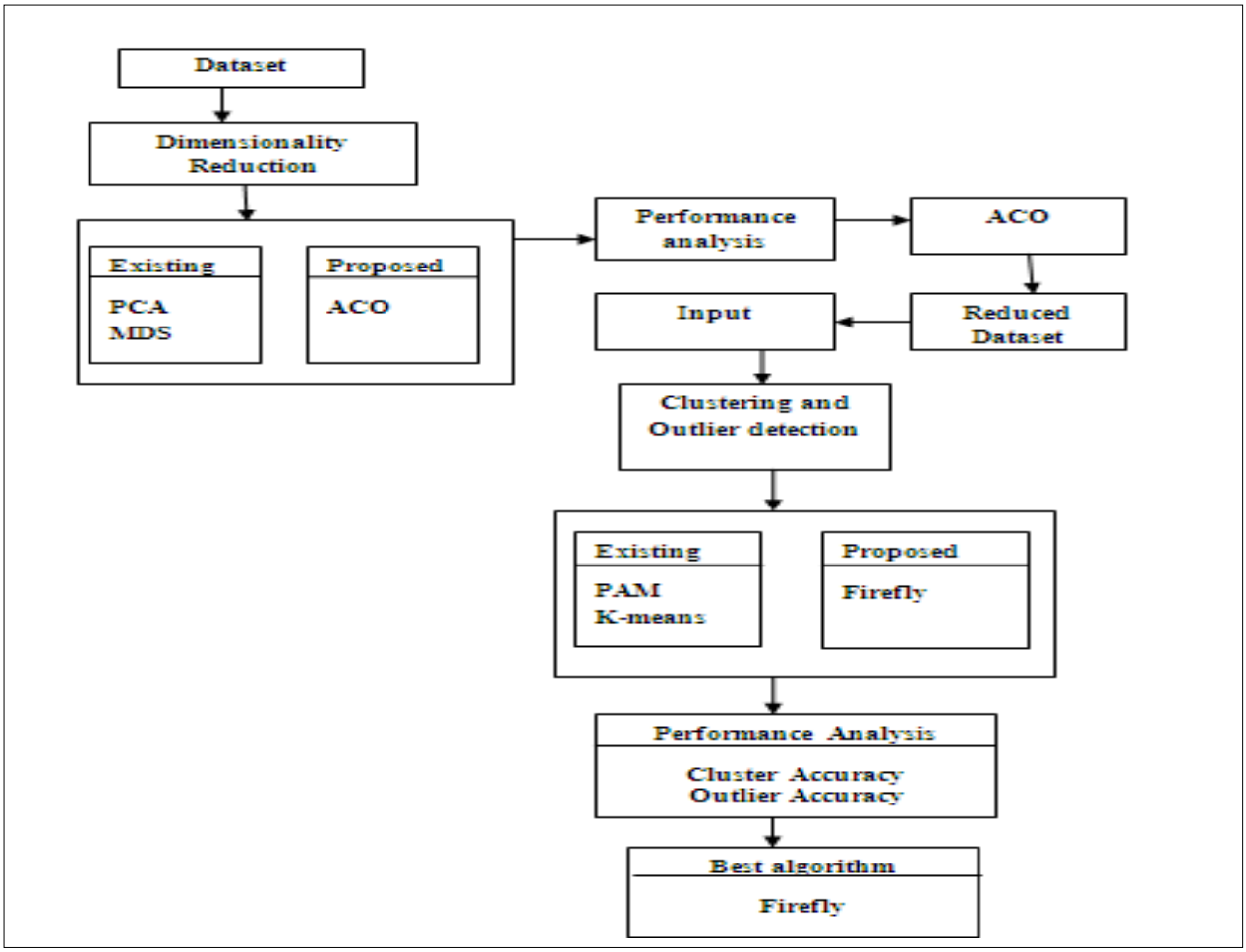
Garima singh, et.al. [14] proposed a novel clustering algorithm termed as I-Clarans algorithm which is compared with other clustering algorithms like PAM, CLARA and CLARANS. The work has been explained in two stages. In the first stage, clustering technique is executed and in the second stage outliers are detected. The datasets used are Bupa class1, Bupa class2 and Iris datasets. And it has been proved that I-Clarans works more efficiently than other clustering algorithms.

III PROPOSED ALGORITHMS

1. ACO (Ant Colony Optimization)

Ant colony optimization (ACO) algorithm is a probabilistic method for solving computational problems and it can be used to find the path [24]. This algorithm is an associated with ant colony family. Ant colony optimization algorithms have been applied to many combinatorial optimization problems [25] , extending from quadratic task to protein compaction or steering vehicles and a lot of ensuing methods that have been amended active problems in real variables, multi-targets and parallel operations. Feature subset extraction is a process of extracting a subset of features from the original features that characterizes the full dataset. There may be thousands of features present in a real world datasets and each feature may carry only a little bit of information, it would be very difficult to treat all the features. Therefore it is very crucial to mine or select significant features from the dataset.

Volume: 14 Issue: 01 Pages: 5277-5286(2022) ISSN: 0975-0290



The dataset D= [d_1, d2….d_n] is taken as input. Then, the ACO parameters like population of ants (p) and Intensity of pheromone trial (b_i) are initialized. The ants are generated using the intensity value by traversing the path of every data. Ants are randomly choosed for traversing. Pheromone value (τ_{ij}) is calculated for traversing the path.

Where p is the number of attribute and b is the number of data. Then the traversing paths are found and shortest path is calculated. From this the new subset is created then by using ACS (Ant Colony System) ranking the new subset is ranked to create a new solution. ACO algorithm is given in below.

$$\tau_{ij} = 1 / \sum_{i=1}^{p} b$$

Steps in ACO Algorithm

STEP 1: Input the dataset D= [d_1, d_2, d_3……d_n]

STEP 2: Initialize the ACO parameters
a.) Population of ants (p) =number of attributes
b.) Intensity of pheromone trial (b_i) =number of data

STEP 3: Generate the ants by evaluating the features. Every ant is assigned to visit every feature to build the solution.

STEP4: Calculate the Pheromone value (τ_{ij}) to visit all the data by the random ant (p_i).

STEP 5: Find the traversing path of the ant and set the starting node to be 1.calculate the distance of the path from the node to the destination.

STEP 6: Create the subset of the features by reaching the data and completing the visit.

STEP 7: ACS Rank algorithm is used to rank (ρ_{ij}) the subset of new feature solution and finally the features are extracted.

Volume: 14 Issue: 01 Pages: 5277-5286(2022) ISSN: 0975-0290

Ant Colony System Ranking is done by the formulae.

$$\rho_{ij} = \tau_{ij}^{\alpha} \cdot \eta_{ij}^{\beta} / \sum \tau_{ij}^{\alpha} . \eta_{ij}^{\beta}$$

Where τ_{ij} is the amount of pheromone on the edge i, j. η_{ij} is the desirability of edge I, j typically $1/d_{ij}$. α is the parameter to control the influence of τ_{ij}. β is the parameter to control the influence of η_{ij}.

2. FIREFLY ALGORITHM (FA)

Short and rhythmic flashes for communication and attracting the potential hunt are used by most of the fireflies [23]. All fireflies are unisex; the high light emitting fireflies are attracted by the less light emitting fireflies. The attractiveness is proportional to the illumination and they both declines as their distance rises. If no other brighter firefly is available, it will move randomly. The size of the fireflies is initialized by n. Set the range of the fireflies by initializing it as 1. Four values for the boundary [x_{min}, x_{max}, y_{min}, y_{max}] are set. The initial location of the fireflies is found and then they are grouped together based on the light intensity (l_i). Make a move to all less light emitting towards high light emitting fireflies and end the iteration. Randomness of the flies is reduced by the iteration to make sure whether they are within the range. Attractiveness parameter (β) is set to find the ranges.

$$\beta = \exp(-\gamma * r)$$

Where γ is the coefficient of light absorption and r is the range.

A Firefly Algorithm assumes three basic rules which are described as follows:

- Every firefly will be attracted to other fireflies irrespective to their gender because they are unisexual
- They attract each other proportionally to their illumination intensity and reversely proportional to their search spaces, the brighter flashing firefly will attract the other less bright ones, the more the distance the less attractiveness, if no brighter firefly nearby they will move randomly
- The brightest firefly cannot be attracted and it will travel randomly.

The cluster centers are the decision variables when FA is used to solve the grouping problems, In N dimensional space there will be a connection between the objective function and the value of all Euclidean distance. At the beginning and based on the objective function all object (fireflies) will be randomly propagated in whole search distance (space).

FA procedure consists of two phases:

- The first is the difference in the light intensity thus; the light intensity is linked with the objective values. Considering the maximization or minimization case problem, the firefly with either lower or higher light intensity will attract another individual with either lower or higher light intensity.
- The second phase is traveling to the direction of the attractive fireflies; the firefly attractiveness will be balanced with light intensity gained by the neighbor fireflies.

Steps in Firefly Algorithm

STEP 1: Input the dataset D

STEP 2: Determine the size of variables (n) = number of fireflies.

STEP 3: Compute the average of each column and set the range for the data

Max generation=number of times

For i=1 to n where n-number of fireflies; Range = [x_{min}, x_{max}, y_{min}, y_{max}]

STEP 4: Generate the initial location of n fireflies and display the path for every fireflies

STEP 5: Group the fireflies by their light intensity (l_i) and trace the path of all roaming

Fireflies (i.e) the fireflies with high light emitted are grouped together and less light emitting in another group.

STEP 6: Make a move to all less light emitting towards high light emitting fireflies and end the iteration.

STEP 7: Reduce the randomness during the iteration and make sure the fireflies are within range.

If x_n (i) < = range (1); otherwise End;

STEP 8: Attractiveness parameter (β) is set to the group. The data which is out of the range are detected as outliers.

Volume: 14 Issue: 01 Pages: 5277-5286(2022) ISSN: 0975-0290

IV. EXPERIMENTAL RESULTS

The experimental results are implemented in Matlab 2013a. The work is performed in PC Intel Pentium Processor, 2GB RAM, OS Windows 7 Ultimate 32-bit. Higher dimensional data are transformed into lower dimensional data in order to increase the efficiency of dataset. Three different medical dataset have been used in this work they are thyroid dataset which contains 7200 instances and 21 attributes which is collected from keel data repository, Oesophagal data set contains 979 instances and 13 attributes which is collected from SMCR repository, Heart disease dataset contains 366 instances and 12 attributes and it is collected from UCI repository.

Table 1 Performance Measures of Dimensionality Reduction

Data set	Algorithm	Original Feature	Reduced Feature	Execution time (milliseconds)
Thyroid	PCA	21	7	1006
	MDS		11	1019
	ACO		14	854
Oesophagal	PCA	13	6	1012
	MDS		5	1011
	ACO		8	992
Heart	PCA	12	5	869
	MDS		6	1008
	ACO		7	756

It could be seen that ACO algorithm performs better and reduces more number of feature when compared with other algorithms. Table 1 Performance Measures of number of features reduced for three different datasets (Thyroid, Oesophagal and Heart).

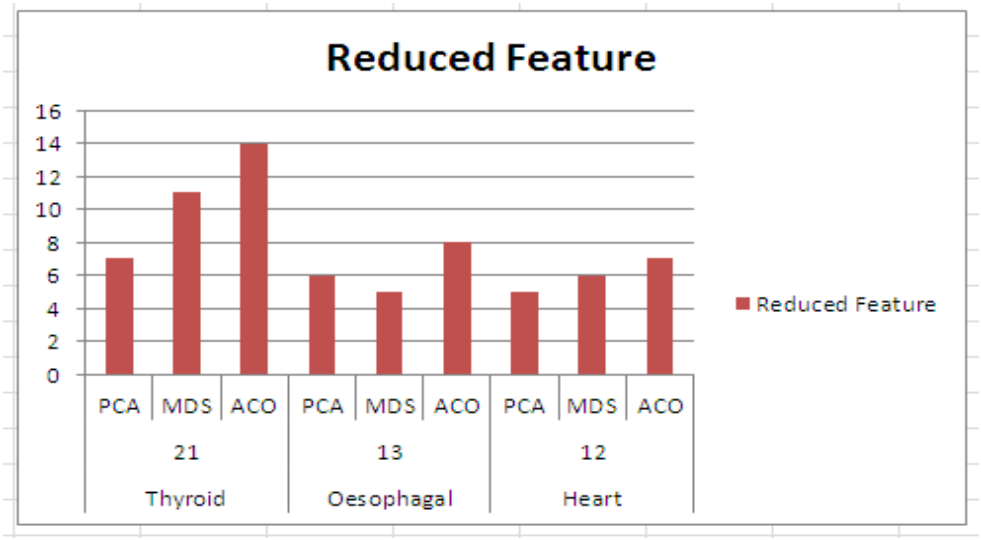


Figure 2 Number of Attributes Reduced

Fig 2 shows the reduced features by using the dimensionality reduction algorithms like PCA, MDS and ACO. Here, ACO is the proposed optimization algorithm which performs efficiently in reduction.

1. Time Complexity of Dimensionality Reduction

Performance of time complexity is measured for reducing number of features from the original features of the dataset. The time complexity is the amount of time required to reduce the features. In Matlab, TIC TOC time is calculated in milliseconds. Table.1 shows the proposed optimization algorithm (ACO) consumes less time for reducing features when compared with other existing algorithms.

Volume: 14 Issue: 01 Pages: 5277-5286(2022) ISSN: 0975-0290

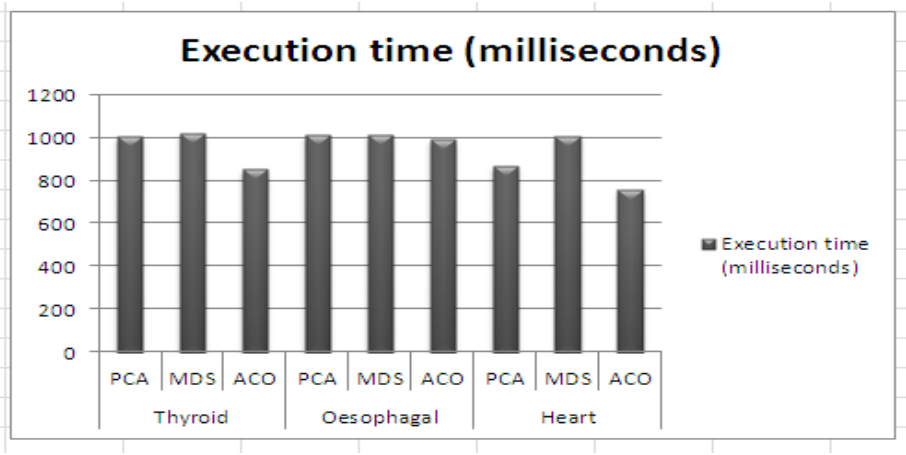


Figure 3 Time complexity

2. Accuracy of the Cluster

The dataset which is reduced and the new features are extracted and the new features are taken as input to the second part of the work (i.e.) clustering. Clustering is done to the reduced data set in order to detect the outliers. Three different datasets namely thyroid, Oesophagal, heart disease have been used. Accuracy is calculated by three measure clustering accuracy, precision, recall. The clustering algorithms K-means, Clarans, PAM are used to find the accuracy of clusters is shown in Table 2. The result of clustering can be calculated by accuracy, precision, recall. Accuracy can be calculated using the formula [21].

ACCURACY = (TP+TN) / (P+N)

Where TP is defined to be True Positive, TN is True Negative, P is positive, N is Negative. The accuracy determines about the measurement of the value.

- TP-Correct input with correct output
- TN-Correct input with false output
- FP-False input with correct output
- FN-False input with false output

Table 2 Performance measures of clusters

Data set	Algorithm	Clustering Accuracy (%)	Precision	Recall	No. of outliers detected	False Alarm Rate	Execution Time (milliseconds)
Thyroid No. of Features Reduced - 14	K-means	92	62	67	71	40	1687
	PAM	91	65	83	50	35	312
	Firefly	95	76	85	80	20	134
Oesophagal No. of Features Reduced - 8	K-means	85	66	56	124	45	6076
	PAM	85	78	85	70	30	794
	Firefly	93	80	87	130	25	365
Heart No. of Features Reduced - 7	K-means	80	77	54	141	50	1744
	PAM	82	72	75	20	60	1751
	Firefly	90	82	78	150	35	554

3. Precision

Precision determines about the uncertainty measurement. The precision is calculated using the formula [21].

PRECISION (P) = TP / (TP+FP)

Precision in Matlab is said to be relevant retrieved items/retrieved items. The relevant items are the data which are relevant for the specific cluster and the retrieved items are the data which are collected from specified cluster.

Volume: 14 Issue: 01 Pages: 5277-5286(2022) ISSN: 0975-0290

4. Recall

The recall is a measure used to identify the condition which makes the measurement positive. The recall is calculated using the formula [21].

$$\text{RECALL(R)} = TP / (TP+FN)$$

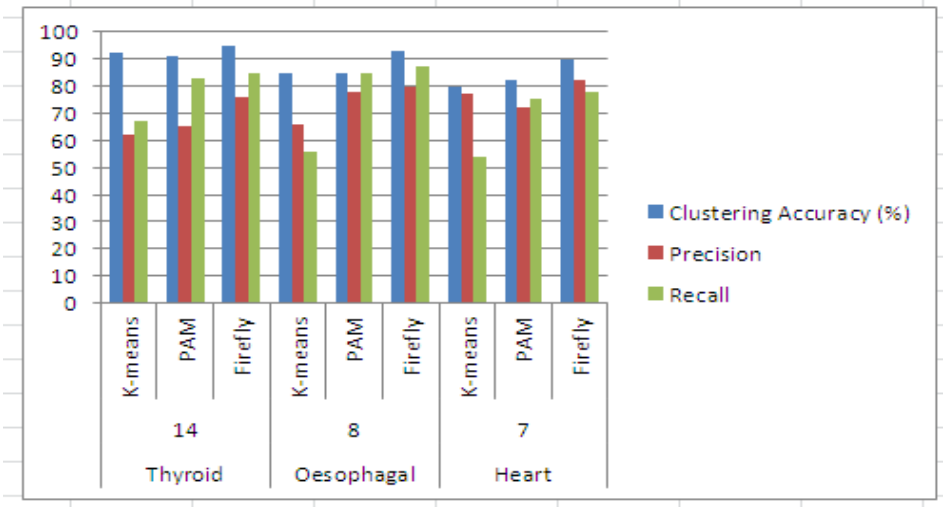


Figure 4 Accuracy of the clusters

5. Outlier Accuracy

Outlier detection accuracy is calculated using three factors (i.e.) number of outliers, Outlier accuracy and time taken K-means, CLARANS, PAM are the algorithms used to find the accuracy.

It could be seen that the proposed firefly algorithm detects more number of outliers when compared to k-means, PAM, Clarans. These algorithms are used for identifying the number of outliers from the original data in the dataset. The detection of outliers is shown in table 2. From this chart it has been proved that the proposed algorithm detects more number of outliers.

6. Number of outliers detected

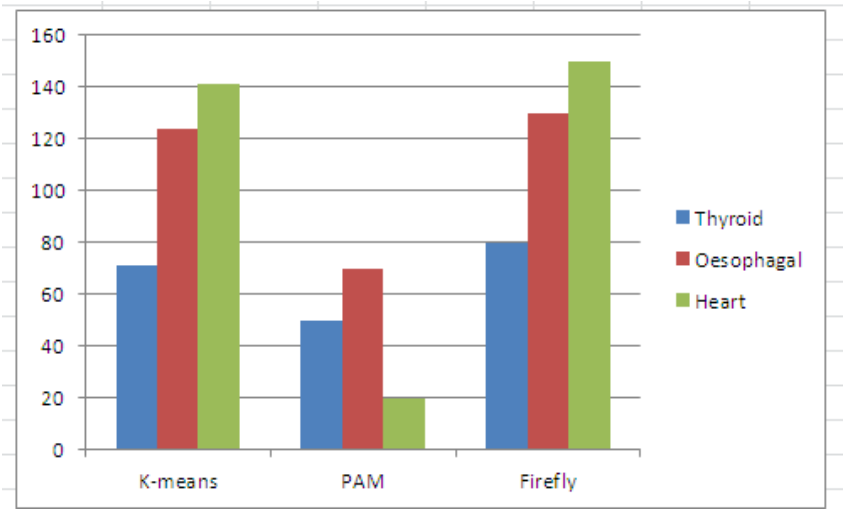


Figure 5 Number of outliers detected

7. False Alarm Rate (FAR)

False Alarm Rate (FAR) is illustrated in Table 2, FAR can be calculated using the formula.

$$\text{FAR} = FP/ (TP + FP)$$

Volume: 14 Issue: 01 Pages: 5277-5286(2022) ISSN: 0975-0290

The performance factor is measured in terms of detecting outliers. It has been proved that the proposed firefly algorithm performs well because it contains low false alarm rate when compared with existing algorithms. Table 2 shows the FAR.

- FP is False positive
- TP is True Positive
- P is the precision

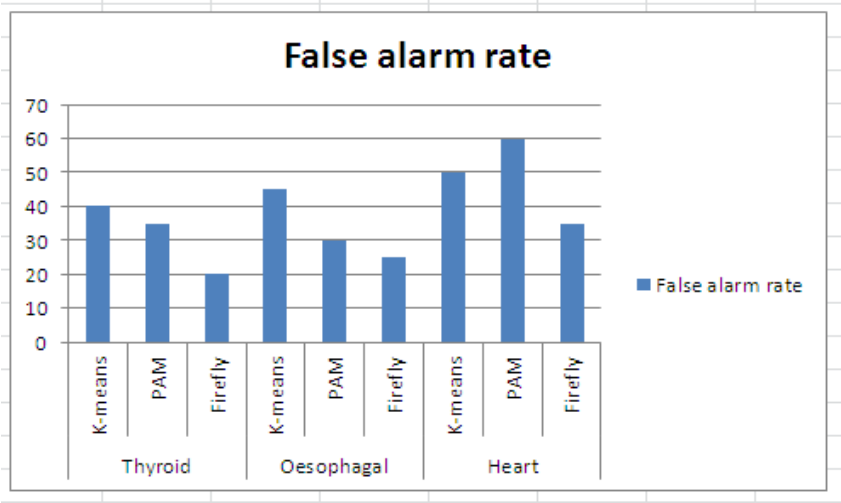


Figure 6 False Alarm Rate

8. Time Complexity

Time complexity is measured in terms of time required for detecting outliers. The performance factor is compared with the existing clustering algorithms like K-means, Clarans, PAM and proposed firefly algorithms. It has been proved that firefly algorithm consumes less time. Time is measured in milliseconds. The result of time complexity is shown in table 2.

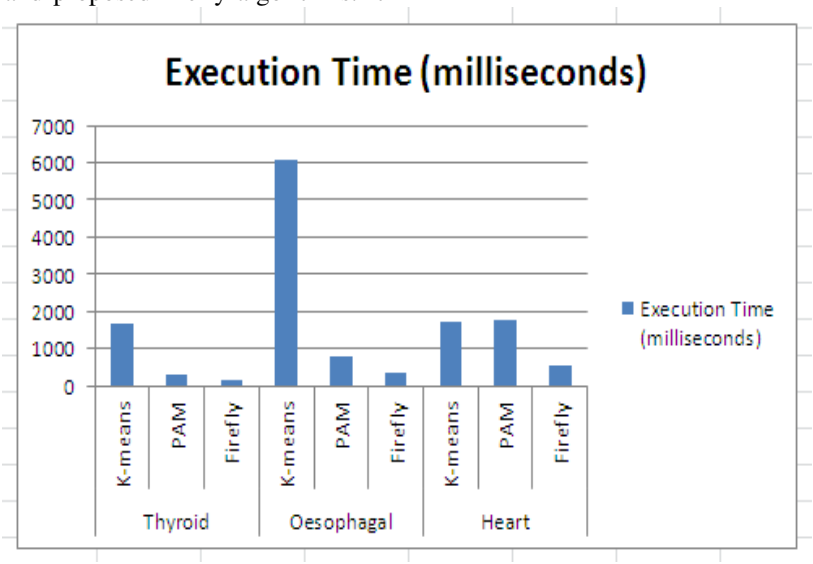


Figure 7 Execution Time for Detecting Outliers

From the above fig 7, it is observed that the proposed firefly algorithm takes less time to detect outliers when compared with existing algorithms. Therefore, the firefly algorithm performs better in clustering and detects outliers because it contains high detection rate and low false alarm rate and less time. Hence, the clustering accuracy and outlier accuracy is very efficient and the proposed firefly is found to be the best algorithm.

Volume: 14 Issue: 01 Pages: 5277-5286(2022) ISSN: 0975-0290

V. CONCLUSION

In the first phase dimensionality reduction is performed and in the second phase, the reduced data are clustered to find outliers. The accuracy of the cluster has been found after the reduction of attributes. ACO and FF bio-inspired algorithms are used in this work. Experimental results show that the proposed work performs better. In future, this research work can be enhanced by applying other dimensionality reduction techniques and optimization techniques to reduce number of features. Number of clustering algorithms and optimization algorithms can be used to detect outliers with reduced execution time and the accuracy can be increased. Here, the medical dataset have been used and in future it can be applied to other high dimensional data like transactional data, weather forecast data and data streams. Here, the numerical attributes are used for reduction and outlier detection. In future, new dimensionality reduction and outlier detection techniques are to be proposed for handling categorical attributes.

REFERENCES

[1]. Dr. S. Vijayarani, S. Maria Sylviaa-Comparative Analysis of Dimensionality Reduction Techniques. International Journal of Innovative Research in Computer and Communication Engineering Vol. 4, Issue 1, January 2016.

[2]. http://www.comp.dit.ie/btierney/Oracle11gDoc/datamine.111/b28129/feature_extr.html

[3]. Larose D.T, "Discovery knowledge in data-Introduction to Data mining, ISBN 0-471-66657-2, ohn Wiley & Sons, Inc., 2005.

[4]. Dr. S. Vijayarani, Ms. P. Jothi- Hierarchical and Partitioning Clustering Algorithms for Detecting Outliers in Data Streams- International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 4, April 2014

[5]. S. D. Pachgade, S. S. Dhande, "Outlier Detection over Data Set Using Cluster-Based and Distance-Base Approach", International Journal of Advanced Research in Computer Science and Software Engineering ISSN: 2277, Volume 2, Issue 6, June 2012.

[6]. Cormen, Thomas H, Charles E, "Introduction to Algorithms, 2nd edition", McGraw- Hill, New York.

[7]. Edwin M. Knox and Raymond T. Ng, "Algorithms for Mining Distance", Based Outliers in Large Datasets. http://www.vldb.org/conf/1998/p392.pdf

[8]. Rajendra Pamula, Jatindra Kumar Deka, SukumarNandi "An Outlier Detection Method based on clustering", Second International Conference on Emerging Applications of Information Technology, 2011.

[9]. Grubbs, F. E. (February 1969). "Procedures for detecting outlying observations in samples". Technometrics 11 (1): 1–21. doi:10.1080/00401706.1969.10490657.

[10]. Hodge.V and J. Austin, "A Survey of Outlier Detection Methodologies", Artificial Intelligence Review, Vol. 22, pp. 85-126, 2003.

[11]. Aastha Joshi, Rajneet Kaur, "A Review: Comparative Study of Various Clustering Techniques in Data Mining" International Journal of Advanced Research in computer Science and Software Engineering, Volume 3, Issue 3, March 2013 ISSN: 2277 128X

[12]. Elahi, M. KunLi, Nisar, W. XinjieLv, HonganWang, "Fuzzy Systems and Knowledge Discovery", Fifth International Conference on Vol.5, andVol .3, pp. 23-27, 2002

[13]. Behera Abhishek, T., Johnson, T., and Chadderdon, G.: 1998, 'Classification and Novelty Detection using Linear Models and a Class Dependent - Elliptical Bassi Function Neural Network '. In: Proceedings of the International conference on neural networks. Anchorage, Alaska.

[14]. Garima Singh, Vijay Kumar, "An Efficient Clustering and Distance Based Approach for Outlier Detection", International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 7–July 2013

[15]. Rekha Awasthi, Anil Kumar Tiwari and Seema Pathak, "An Analysis Of Density Based Clustering Technique with Dimensionality Reduction For Diabetic Patient" International Journal of Computer Engineering and Applications, Volume IX, Issue IV, April 15 www.ijcea.com ISSN 2321-3469.

[16]. Zhang, T., Ramakrishnan, R., and Livny, M. 1997. BIRCH: A new data clustering algorithm and its applications. Journal of Data Mining and Knowledge Discovery, 1, 2, 141-182

[17]. Sadia Patka1, M. S. Khatib2, Kamlesh Kelwade, "Recent Trends and Rapid Development of Applications in Data Mining", International Conference on Advances in Engineering & Technology, IOSR Journal of Computer Science 2014, page no 73-78.

[18]. Charu C. Aggarwal, Phillip S. Y, An effective and efficient algorithm for higher dimensional outlier detection.

[19]. Ishida, E.E.O & de Souza, R.S, Hubble parameter reconstruction from a principal component analysis: minimizing the bias. Astronomy & Astrophysics, Volume 527, id.A49 (2011)

[20]. http://www.slideshare.net/kompellark/t19-factor-analysis

[21].Dr. T. Christopher, T. Divya, "A Study of Clustering Based Algorithm for Outlier Detection in Data streams". Proceedings of the UGC Sponsored National Conference on Advanced Networking and Applications, 27th March 2015.

[22]. Caruana, R., Niculescu-Mizil, A. "An empirical comparison of supervised learning algorithms". Proceedings: 23rd International Conference on Machine Learning, 2006. http:/ / citeseerx. ist. psu. edu/ viewdoc/ download? doi=10. 1. 1. 122. 5901& rep=rep1& type=pdf

[23].http://www.academia.edu/14545523/IRJET-A HYBRID FIREFLY BASED APPROACH FOR DATA CLUSTERING Volume No.3 Issue No. 4, August 2015

[24]. Ajeet Pandey, Akhilesh Kumar Singh- Ant Colony Optimization Based Routing Algorithm in Various Wireless Sensor Network- A Survey- Journal of Advanced Computing and Communication Technologies

Volume: 14 Issue: 01 Pages: 5277-5286(2022) ISSN: 0975-0290

[25].https://books.google.co.in/books?id=7KuqCAAAQBAJ&pg=PA15&lpg=PA15&dq=%22been+applied+to+many+combinatorial+optimization+problems,%22&source=bl&ots=LxB2gjBdyW&sig=FNup4pJps2J8sqQpvsYRa2bQYA&hl=en&sa=X&redir_esc=y#v=onepage&q=%22been%20applied%20to%20many%20combinatorial%20optimization%20problems%2C%22&f=false
[26]. D. Asir Antony Gnana Singh, P.Surenther, E. Jebamalar Leavline- Ant Colony Optimization Based Attribute Reduction for Disease Diagnostic System- International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.55 (2015)© Research India Publications; httpwww.ripublication.comijaer.htm
[27]. Fahim. A. M., A. M. Salem, F. A. Torkey and M. A. Ramadan, "An Efficient enhanced k-means clustering algorithm", journal of Zhejiang University, Vol.10 (7), 2006 page no 1626-1633.
[28]. Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.
[29]. http://www.thearling.com/text/dmwhite/dmwhite.html
[30]. Jonathon Shlens, "A Tutorial on Principal Component Analysis" https://en.wikipedia.org/wiki/Principal_component_analysis

BIOGRAPHIES:

1. Dr. S.Vijayarani M.C.A., M.Phil., Ph.D., DCSE., working as Assistant Professor in the Department of Computer Science, Bharathiar University, Coimbatore. Her research interests include Privacy-Preserving Data Mining, Utility Mining, Text Mining, Web Mining, Image Mining, and Health Care Analytics. She has published more than 200 research articles in International / National journals and conferences. She has also written 20 book chapters. She has produced 33 M.Phil./Ph.D. research scholars. She is a member of various professional bodies like CSI, IAENG, UACEE, INSA, etc.

2. Dr.C.Sivamathi M.Sc., MPhil., Ph.D : She is working as an Assistant Professor in Department of Computer Science with Data Analytics, PSG College of Arts & Science, Coimbatore. She has 6 years of teaching experience. Her research are include data mining, big data analytics. She has published many journal articles and chapters in edited book.

3. S.Maria Sylvia: Mrs S Maria Sylviaa, completed her MCA.,M.Phil in computer science.Currently working as a Assistant professor in Department of computer science, Nirmala college for women, Coimbatore.Her field of research interest are Data mining, network Security. She has published and presented papers in journals and conferences