DOI: 10.21917/ijsc.2019.0267

FUZZY CLUSTERING ALGORITHMS - COMPARATIVE STUDIES FOR NOISY SPEECH SIGNALS

H.Y. Vani[1], M.A. Anusuya[2] and M.L. Chayadevi[3]

[1]*Department of Information Science and Engineering, JSS Science and Technology University, India*
[2]*Department of Computer Science and Engineering, JSS Science and Technology University, India*
[3]*Department of Computer Science and Engineering, JSS Academy of Technical Education, India*

Abstract

In the area of speech signal processing and recognition, application of soft computing techniques is one of the prominent techniques for clustering the overlapping data. Kernel FCM technique is one of the efficient method to cluster the data by computing the cluster centroids. This paper presents and compares the most important clustering techniques like k-means, Fuzzy C means and Kernel Fuzzy C Means algorithms for clustering noisy speech signals. The clustering performances of these techniques are tabulated for homogeneous and heterogeneous speech data sets. This paper highlights the importance of KFCM algorithm for clustering the overlapping data. It also demonstrates the computation time and recognition accuracies of each technique. Our study identifies the KFCM technique performs better than k-means and FCM techniques.

Keywords:

Additive Noise, Clustering, Convolved Noise, Fuzzy C Means (FCM), Heterogeneous Data, Homogeneous Data, K-means, Kernel Fuzzy C Means (KFCM), Principal Component Analysis (PCA), Validity Measures

1. INTRODUCTION

Clustering is the process of grouping similar objects into their respective groups. Clustering using fuzzy concepts play a major role in solving the complex problems in the area of pattern recognition and speech signal processing. It partitions the data sets into subsets of different clusters with common features. Cluster eliminates the differences among the data members by computing their centroids. Clustering finds its applications in many fields like data mining, pattern recognition, speech signal processing, bio-informatics, machine learning, image processing etc. Speech datasets are non-linear and non-stationary in nature. Hence fuzzy methods are well suited for better modeling the data with uncertainties. Various types of clustering methods exist in the literature using fuzzy technique. We have adopted FCM and KFCM techniques for the concise representation of related data. This paper presents a comparative study on these techniques considering the parameters like fuzzier, stopping criteria, time taken for execution and recognition accuracy. These parameters are demonstrated for convoluted and additive noisy speech signals applied on homogeneous and heterogeneous data sets. This paper demonstrates the importance of KFCM technique in clustering noisy speech data. As per the state of the art much work has been done in the area of image processing using FCM and *k*-FCM techniques and minimal work has been carried out for speech signal processing using above techniques it is not observed in the literature the application of these techniques for speech data by varying the various parameters of the above techniques. Since no experimental work is observed in this paper, we have attempted to execute this for the different speech data sets and reported the results and the performances of the above techniques. We feel it will be useful for the beginners in to select and initialize the various parameter values of FCM and *k*-FCM techniques application of speech signal processing. Presently, the application of fuzzy clustering techniques is applied in many fields like image processing [18]-[20], agriculture [21], medical imaging [25], speech processing [27] [28], data mining [29] [30] and also in other fields [22]-[24], [26]. But all these papers discuss the concepts of fuzzy clustering techniques briefly. But experimentation of these techniques is not carried out in depth related to speech recognition application using speech dataset. Hence this paper is unique in this direction, presenting the efficiency of all the fuzzy clustering algorithms for various speech data sets, realizing the influence of the various parameters and their initial values

2. CLUSTERING TECHNIQUES

Clustering has been a popular approach to unsupervised pattern recognition [1] problems. Clustering defines the metric data compactness for the knowledge relevance of the data. Clustering is used to identify the compact relationships among the data. The Fuzzy is used to identify and handle the imprecision in the data. In this section *k*-means, FCM and KFCM techniques are discussed. Clustering can be classified as: Soft Clustering (Overlapping Clustering) and hard Clustering (or Exclusive Clustering): In case of soft clustering techniques, concept of fuzzy membership functions is used to cluster data, so that each point may belong to, two or more clusters with different degrees of membership value. In this case, data will be associated to an appropriate membership value. In many situations, fuzzy clustering is more natural than hard clustering. FCM [2] allows speech samples to belong to two or more clusters than belonging to one cluster. And these samples are assigned with fuzzy membership value between 0 and 1 indicating partial belongingness to the clusters.

2.1 K-MEANS CLUSTERING TECHNIQUE

The *k*-means algorithm [3] is one of the simplest unsupervised and hard clustering algorithms. This method is used to classify a given data set in to various clusters.

Algorithm

Step 1: Choose random centroid’s.

Step 2: Calculate the distance between centroid’s and data points

Step 3: *k*-means assigns the data point to the cluster using minimal Euclidian distance measure.

$$J_{KM}\left(X;V\right)=\sum_{i=1}^{c}\sum_{j=1}^{n}D_{ij}^{2} \tag{1}$$

Step 4: Calculate new centroid's

$$V_i=\sum_{i=1}^{n_i}\frac{D_{ij}}{n_i} \quad 1\le i\le c \tag{2}$$

Step 5: Check if new centroid's are equal to old centroid's

Step 6: If new centroid's are equal to old centroid's then program ends else go to step2

Input: *V*-centroid number, *x* and *y* - centroid values distance between centroid and data points, x_1, y_1 - values of the data point, x_{11}, y_{11} are the new centroid's values with the levels, D_{ij} - Euclidian distance between each data point and centroids, and *n* - number of iterations

Output: Number of clusters.

Advantages: *k*-Means is easier to understand and simple to implement

Disadvantages:

- It is not effective for overlapping clusters
- Fails to cluster effectively the heterogeneous data.
- It provides the local optima of the squared error function.
- Randomly choosing the cluster center may not yield good results.

Observations: *k*-means algorithm is simple to implement and works better for the homogeneous data. It fails to produce good results for the heterogeneous data set due to the inefficiency in handling speech variability's present in the speech data [4]. To handle the speech variability's in a better way FCM and *k*-FCM can be used as suggested from the literature.

2.2 FUZZY C MEANS CLUSTERING TECHNIQUE (FCM)

The most well-known fuzzy clustering algorithm is Fuzzy C-Means clustering technique. It was introduced by Bezdek [5] [6] by modifying the crisp clustering function. He introduced the idea of a fuzzification parameter (*m*) whose value ranges between [1,*n*] where *n*=2, that determines the degree of fuzziness in the clusters. This technique sounds good when data has much speech variability's [7] [8].

Algorithm

Step 1: Randomly initialize the clusters centers

$$J_{KM}\left(X;V\right)=\sum_{i=1}^{c}\sum_{j=1}^{n}D_{ij}^{2} \tag{3}$$

Step 2: Create the distance matrix from a data point to each of the cluster centre using Euclidean distance using Eq.(3).

$$V_i=\frac{\sum_{j=1}^{n}u_{ij}^{m}x_j}{\sum_{j=1}^{n}u_{ij}^{m}} 1\le i\le c \tag{4}$$

Step 3: The membership matrix is computed using fuzzification parameter with Eq.(5).

$$u_{ij}\frac{1}{\sum_{k=1}^{c}\left(\frac{D_{ijA}}{D_{kjA}}\right)^{\frac{2}{(m-1)}}};1\le i\le c,1\le j\le n \tag{5}$$

$$J_{KM}\left(U,\lambda;X\right)=\sum_{i=1}^{C}\sum_{t=1}^{T}u_{it}^{m}d_{it}^{2}\sum_{j=1}^{n}D_{ii}^{2} \tag{6}$$

Step 4: Values of the U_{ij} matrix should be less than or equal to one ($U_{ij}\le 1$)

Step 5: Compute new centroid's

Step 6: Optimize cluster centers by generating new centroids

Step 7: Cluster assignment for the data points.

Input: x_1,- data vector, V_i - centroids of fuzzy clusters, *c* - number of fuzzy clusters, *m* - fuzzification parameter, *U* - assigns Fuzzy membership value to each sample indicating the membership value from one data sample to the n^{th} cluster, ε - stopping criteria, D_{ij} -distance measure and *n* - number of data points.

Output: Data points are (overlapping) assigned to clusters

Advantages:

- For overlapped dataset FCM gives better results than the *k*-Means.
- Each data point is assigned with a membership value to each cluster center, as a result data point may belong to more than one cluster center.

Disadvantages:

- FCM requires a priori specification of the number of clusters (but this can also be chosen by mountain or modified mountain algorithms).
- Takes more iterations even with the lesser value of fuzzifier '*m*'

2.3 KERNEL FUZZY C MEANS CLUSTERING TECHNIQUE (KFCM)

Chen and Zhang [9] enhanced this technique by replacing the Euclidean distance with a new kernel induced distance function. It is a kernel based objective function that maps the data into kernel space by using Gaussian Radial Basis function [10]. It first maps the data into high dimensional space to gain high discriminate capability. It calculates the measure of the samples in their original data space with kernel function. It overcomes the limitation of data intrinsic shape dependency and is sensitive to initial values of algorithm. This improves the algorithm robustness.

Algorithm:

Step 1: Initially, select the cluster centroids at random and initialize the value of sigma using trial and error procedure.

Step 2: Update the membership matrix using Eq.(7)

$$u_{ik} \frac{\left(\dfrac{1}{1-K\left(x_k, v_i\right)}\right)^{\frac{1}{(m-1)}}}{\sum_{k=1}^{c}\left(\dfrac{1}{1-K\left(x_k, v_i\right)}\right)^{\frac{1}{(m-1)}}} \tag{7}$$

Step 3: Select the centroids using Eq.(8)

$$v_i = \frac{\sum_{k=1}^{n} u_{ik} K\left(x_k, v_i\right) x_k}{\sum_{k=1}^{n} u_{ik}^{m} K\left(x_k, v_i\right)} \tag{8}$$

Step 4: Reiterate step 2 to 3 till the termination criteria is satisfied.

$$\|V_{new} - V_{old}\| \leq \varepsilon \tag{9}$$

where V is the vector of cluster centers and ε is the termination criteria

Input: x_1- data vector, v_i - centroids of a fuzzy clusters, c - number of fuzzy clusters, m - fuzzification parameter, U - Fuzzy membership function indicating the membership from one sample to the n^{th} cluster, ε - stopping criterion, σ - kernel parameter determining the geometrical structure of the mapped samples in the kernel space, $K(x_k,v_i)$ - distance between mapped kernel value and centroids.

Output: x_1 - data points (overlapping) assigned to clusters.

Advantages

- It does not require prior knowledge to determine the system topological structure.
- It is efficient in dealing with noise and outliers.

Disadvantages

- It is very difficult to select the optimal parameter values of kernel function based on the type of problem.

2.4 KERNEL VALIDITY MEASURES [17]

To identify the suitability of the clusters, kernel-based validity measures are used. Generally, these measures are classified into two classes: (a) The classes using membership value (b) classes using geometrical characteristics. When kernel-based clustering is used, kernel-based validity measures are preferred. In the proposed work Gaussian kernel-based validity measure is used. Performance of the validity measures depends on the parameters of the clusters. In our study we have computed the recognition accuracies for various clustering techniques considering the time taken by each function for clustering the data. The parameters used to evaluate the FCM and KFCM clustering techniques are i) the number of clusters, ii) fuzzifier iii) stopping criterion. The strength and weakness of all the clustering algorithms as per the literature is tabulated in Table.1 in brief.

Table.1. Strength and weakness of all the clustering algorithms

***k*-Means [16] [3]**	**FCM [7] [8] [16]**	**KFCM [9] [10] [16]**
Suitable for trained data	Suitable for trained and untrained data	Trained and Untrained data
Clusters better for Clean data	Considers both Clean and Noisy data sets	Considers both clean and Noisy data
Only suitable for Homogeneous dataset	More Suitable for Homogeneous dataset compared to heterogeneous dataset	Suitable for both Homogeneous and Heterogeneous datasets (better than FCM)
Not suitable for Overlapped data	Suitable for Overlapped data	Suitable for Overlapped data
Algorithm performance is dependent on Cluster initialization parameters that plays a major role in better clustering	Cluster initialization parameter and fuzzy membership function plays a major role in clustering the data	Cluster initialization parameter, fuzzy membership function and kernel function plays a vital role in clustering the data

3. PRINCIPAL COMPONENT ANALYSIS

Features are extracted using PCA technique accounts for a maximal amount of total variance in the observed variables. The first principal component identified accounts for most of the variance in the data. The second component identified accounts for the second largest amount of variance in the data and is uncorrelated with the first principal component and so on. PCA is preferred because it possesses higher correlation features with Minimal solution space.

Algorithm

PCA algorithm (X,k): Top k Eigen values, where X: $N \times m$ data matrix each data point x_i = column vector =1,m

$$x = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{10}$$

X - subtract mean x from each column vector x_i in X

$\sum \rightarrow XX^T$ co-variance matrix of X

$\{\lambda_i, u_i\}_i$ = 1,2,...,N are Eigen vectors/Eigen values of

$$\sum \ldots \lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots \lambda_N \tag{11}$$

It returns $\{\lambda_i, u_i\}_i$ = 1,2,..k top k principal components.

4. DATASET

For training and testing the isolated speech signals are recorded using Praat software. The signals are recorded in mono channel with 8KHz frequency as shown in Table.2. These signals are used for simulation purpose. Speech data is collected from 10 male and 10 female speakers. Each speaker is asked to utter each word 10 times. Totally 2000 speech data samples are recorded from both male and female speakers. Two types of data sets are created namely Homogeneous and Heterogeneous data sets.

Homogeneous data set consists of all the words uttered from the single speaker and heterogeneous data set consists of the words uttered from different speakers. These speech samples are used for evaluation of the clustering technique.

The signals are treated as noisy signals by adding Gaussian and Babble noise with the SNR of 10dB and 15dB to create additive and convoluted noisy speech signals respectively.

Table.2. Isolated Words

Number	Kannada	Symbol used
1	ONDU	One
2	ERADU	Two
3	MOORU	Three
4	NALKU	Four
5	AIDU	Five
6	AARU	Six
7	ELU	Seven
8	ENTU	Eight
9	OMBHAT	Nine
10	HATHU	Ten

5. SYSTEM MODEL FOR THE PROPOSED APPROACH

This section explains the complete procedure adopted to cluster the speech signal using various techniques. The additive and convoluted noisy speech signals are processed to extract the features using MFCC technique [13] [14]. These features are further processed to select the prominent features using PCA algorithm by reducing the dimensions of the number of features. Obtained reduced features are used for clustering the data as discussed in section 2 and section 3.

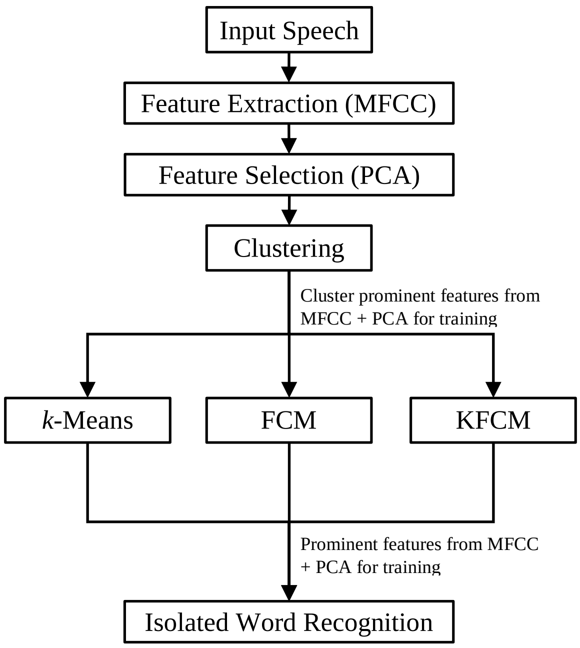

Fig.1. Flowchart of the proposed work

The algorithms are validated by initializing the fuzzification parameter m as to "2" and stopping criteria to "0.0001". The results of the respective algorithms are tabulated individually along with the performance of homogeneous and heterogeneous speech datasets. The normal procedure is adopted for selecting and extracting features. These features are clustered using various clustering technique as shown in Fig.1.

- MFCC-PCA-KFCM
- MFCC-PCA-FCM
- MFCC-PCA-*k*-Means

6. PERFORMANCE ANALYSIS OF VARIOUS CLUSTERING ALGORITHMS WITH EXPERIMENTAL RESULTS

Comparison between *k*-means, FCM and KFCM algorithms is tabulated as shown in the Table.3-Table.6. The simulations are performed using MATLAB simulation tool. The Table.3 and Table.3(a) present the results based on the time taken by the clustering algorithms to cluster the homogeneous and heterogeneous convoluted noisy speech data. The parameter considered for the efficiency of the algorithm is based on the number of clusters and time taken by clustering algorithm.

Table 3. Execution time for homogeneous data

No. of Clusters	*k*-means (s)	FCM (s)	KFCM (s)
5	0.127	0.035	0.014
10	0.168	0.115	0.014
15	0.15	0.126	0.014
20	0.153	0.152	0.019

Table.3(a). Execution time for heterogeneous data

No of Clusters	*k*-means (s)	FCM (s)	KFCM (s)
5	0.151	0.058	0.014
10	0.159	0.103	0.014
15	0.179	0.167	0.02
20	0.243	0.24	0.02

From Table.3 and Table.3(a), it is observed that as the number of clusters increases the execution time taken by KFCM technique is less than FCM and k-means techniques. The Table.4 and Table.5 shows the recognition rates for male and female speakers.

Table.4. Recognition rate for homogeneous convoluted noisy speech data

Algorithms	Gender	Word Recognition Accuracies		
		Convoluted Noisy Signal		Clean
		15dB	10dB	
KFCM	Male	58	50	90
	Female	50	50	85
FCM	Male	55	50	90
	Female	50	40	85
K-Means	Male	50	45	80
	Female	40	40	78

Table.5. Recognition rate for homogeneous additive noisy speech data

Algorithms	Gender	Word Recognition Accuracies		
		Convoluted Noisy Signal		Clean
		15dB	10dB	
KFCM	Male	60	55	90
	Female	50	50	85
FCM	Male	60	55	90
	Female	50	40	85
K-Means	Male	46	45	80
	Female	40	40	78

Considering the recognition accuracy and the type of clustering technique the sigma value for clean and noisy speech is initialized as 0.26 and 10 respectively. The recognition results for male and female speakers are tabulated in Table.4 and Table.5. If these parameter values are varied to higher values, the samples will over fit and fails to cluster properly. It is also observed that KFCM and FCM performs relatively same for homogeneous, heterogeneous, clean, additive noise and convoluted noise for male and female speakers. Considering heterogeneous data as shown in Table.6 with 10dB of SNR KFCM performs better than FCM and *k*-Means. As depicted from the Table.6, *k*-Means is very poor in handling the heterogeneous data under the clean environment. It shows that speech variability's existing in the speech data are not clustered properly. Hence KFCM tries to cluster effectively compare to others. Whereas KFCM outperforms in clustering the convoluted noisy speech data in the heterogeneous environment. This proves that KFCM technique better handles the variability's present in the speech signal efficiently by increasing the recognition accuracy up to 2 to 5% than FCM.

Table.6. Heterogeneous Data set with their recognition rate

	K-Means		FCM		KFCM	
Data	Clean	Conv-10dB	Clean	Conv-10dB	Clean	Conv-10dB
Aaru	20	60	90	90	100	90
Aidu	0	70	80	40	70	30
Alli	70	30	70	50	70	50
Arda	59	10	40	50	0	70
Aramane	60	60	60	70	60	50
Bharatha	70	0	80	10	0	60
Balagade	80	50	60	50	20	50
Belligge	60	50	40	60	70	50
Bengalore	30	30	40	40	30	50
Chikkadu	80	40	70	80	100	100
Average Accuracy	52.9	40	63	54	65	60

Table.7. Average performance of clustering algorithms with execution time

Techniques	Accuracy	Execution time
K-means	56	0.1495
FCM	62	0.099
KFCM	64	0.015

The Table.7 shows the overall recognition rate for different clustering techniques. The recognition rate for heterogeneous clean and convoluted noisy speech is tabulated. Here the speech samples of both male and female speakers are combined to explore the suitability applicability and the performance of all the three techniques.

- **Observations:** For the various data sets i.e. clean and noisy (additive and convoluted noises) speech signals the following observations are drawn on the various clustering techniques:
- **Clean speech signal:** For homogeneous data set all the three algorithms works better with 95% of accuracies. But for heterogeneous clean speech data set the recognition performance is approximately to 50-60% from *k*-FCM, FCM and *k*-Means.
- **Noisy speech signals:** Adding additive and convoluted noises to homogeneous data 60% and 50-55% of recognition accuracies is achieved for all the clustering algorithm.

For heterogeneous convoluted data 50% of recognition accuracies is obtained considering FCM and *k*-FCM. Among this KFCM performs well.

Generally, for homogeneous data FCM and KFCM performs better than *k*-means having the sigma value 0.26. Using additive noise all the clustering techniques performs equivalently good equivalently good yielding 50% accuracy, whereas for heterogeneous convoluted noisy speech data the performance is reduced 5% yielding 50% of recognition accuracy compared to additive noise, whereas for heterogeneous data when less recognition accuracies to 2-5% is observed by KFCM technique. Hence KFCM better for all types of data set.

7. CONCLUSIONS

In this work, an experimental comparative study has been carried out between *k*-means, FCM and KFCM clustering algorithms to evaluate the recognition accuracies for the homogeneous and heterogeneous speech data under clean and noisy signals. The results are tabulated as discussed in the section 6. It is clear that in all the cases KFCM clustering technique performs better than other techniques. It has been analyzed that the performance of the algorithm depends on the values of fuzzifier value, stopping criteria, number of clusters and sigma parameter. From this study it is realized that for any type of speech data KFCM technique performs s better than two techniques.

7.1 LIMITATIONS

- Only isolated word data set is used.
- Only two parameters of fuzzy clustering technique are used.
- Gaussian kernel is used.

7.2 FUTURE ENHANCEMENTS

The KFCM performance can be enhanced by considering the following

- By replacing Gaussian by polynomial kernel function
- Adopting noisy feature extraction algorithms
- By adopting various fuzzy indices
- By using trial and error method to identify fuzzifier and sigma parameter values
- Our future work is to enhance the performance of KFCM technique for heterogeneous data using the above methods.

REFERENCES

[1] A.K Jain, M.N Murty, "Data Clustering", *ACM Computing Surveys*, Vol. 31, pp. 372-378 ,1999.

[2] Dibya Jyoti Bora and Anil Kumar Gupta, "A Comparative Study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm", *International Journal of Computer Trends and Technology*, Vol. 10, No. 4, pp. 108-113, 2014.

[3] Xin-Guang Li, Min-feng Yao and Wen-Tao Huang "Speech Recognition based on K-Means Clustering and Neural Network Ensembles", *Proceedings of 7th International Conference on Natural Computation*, pp. 26-28, 2011.

[4] Shi Na Liu Xumin and Guang Yong, "Research on K-Means Clustering Algorithm: An Improved k-means Clustering Algorithm", *Proceedings of 3rd International Symposium on Intelligent Information Technology and Security Informatics*, pp. 1-5, 2010.

[5] James C. Bezdek, Robert Ehrlich and William Full, "FCM: The Fuzzy C-Means Clustering Algorithm", *Computers and Geosciences*, Vol. 10, No. 2-3, pp. 191-203, 1984.

[6] J.C. Bezdek, "A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, No. 1, pp. 1-8, 1980.

[7] S. Miyamoto and K. Umayahara, "Methods in Hard and Fuzzy Clustering", *Soft Computing and Human-centered Machines*, pp. 85-129, 2000.

[8] D. Tran, M. Wagner and T. Zheng, "A Fuzzy Approach to Statistical Models in Speech and Speaker Recognition", *Proceedings of International Conference on Fuzzy Systems*, pp. 22-25, 2000.

[9] Zhong-Don Wu, Wei-Xin Xie and Jian-Ping Yu, "Fuzzy C-Means Clustering Algorithm based on Kernel Method" , *Proceedings of 5th International Conference on Computational Intelligence and Multimedia Applications*, pp. 1-5, 2003.

[10] Dao-Qiang Zhang and Song-Can Chen, "Clustering Incomplete Data Using Kernel-Based Fuzzy C-means Algorithm", *Neural Processing Letters*, Vol. 18, No. 3, pp. 155-162, 2003.

[11] Lindsay I. Smith, "A Tutorial on Principal Components Analysis", Available at: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

[12] Shaham Shabani and Yaser Norouzi,"Speech Recognition using Principal Components Analysis and Neural Networks", *Proceedings of 8th International Conference on Intelligent Systems*, pp. 112-117, 2016.

[13] B.J. Mohan and N.R. Babu, "Speech Recognition using MFCC and DTW", *Proceedings of International Conference on Advances in Electrical Engineering*, pp. 6-12, 2014.

[14] M.A. Anusuya and S.K. Katti, "Front End Analysis of Speech Signal Processing: A Review", *International Journal of Speech Technology*, Vol. 11, No. 2, pp. 99-145, 2011.

[15] H.Y. Vani and M.A. Anusuya, "Isolated Speech Recognition using K-means and FCM Technique", *Proceedings of International Conference on Emerging Research in Electronics, Computer Science and Technology*, pp. 78-86, 2015.

[16] H.Y. Vani and M.A. Anusuya, "Noisy Speech Recognition using KFCM", *Proceedings of International Conference on Cognitive Computing Information Processing*, pp. 231-236, 2017.

[17] Jeen-Shing Wang and Jen-Chieh Chiang, "A Cluster Validity Measure with Outlier Detection for Support Vector Clustering", *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, Vol. 38, No. 1, pp. 78-89, 2008.

[18] A. Yang, Y. Zhou, X. Li and M. Tang, "A Region-Based Image Segmentation Method with Kernel FCM", *Proceedings of International Conference on Fuzzy Information and Engineering*, pp. 902-910, 2007.

[19] T. Xinting, Z. Xiaofeng, G. Hongjjang and K. Lun, "FCM-Based Image Segmentation with Kernel Functions", *Proceedings of International Conference on Computational Science and Engineering*, pp. 21-24, 2017.

[20] Guang Hu and Zhenbin Du, "Adaptive Kernel-Based Fuzzy C-Means Clustering with Spatial Constraints for Image Segmentation", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 33, No. 1, pp. 1954-1964, 2019.

[21] N. Vinushree, B. Hemalatha and Vishnu Kumar Kaliappan, "Efficient Kernel-Based Fuzzy C-Means Clustering for Pest Detection and Classification", *Proceedings of International World Congress on Computing and Communication Technologies*, pp. 1-8, 2014.

[22] S. Ramathilagam, R. Devi and S.R. Kannan, "Extended Fuzzy C-Means: An Analyzing Data Clustering Problems", *Cluster Computing*, Vol. 16, No. 3, pp. 389-406, 2012.

[23] Dao-Qiang Zhang and Song-Can Chen, "Kernel-Based Fuzzy and Possibilistic C-Means Clustering", Available at: parnec.nuaa.edu.cn/zhangdq /icann03.pdf

[24] Yi Ding and Xian Fu, "Kernel-Based Fuzzy C-Means Clustering Algorithm Based on Genetic Algorithm", *Neurocomputing*, Vol. 188, pp. 233-238, 2015.

[25] K. Venkatesh Sharma, "Invasive Weed Optimization and Kernel Fuzzy C-Means Based MRI brain Tissue Segmentation", *International Journal of Computer Sciences and Engineering*, Vol. 6, No. 11, pp. 43-50, 2018.

[26] Masoumeh Khanlari, "An Improved KFCM Clustering Method used for Multiple Fault Diagnosis of Analog Circuits", *Circuits, Systems, and Signal Processing*, Vol. 36, No. 9, pp. 3491-3513, 2017.

[27] Jian Wang and Yuanyuan Zhang, “Speaker Recognition Based on KPCA and KFCM”, *Proceedings of International Conference on Mechatronics, Electronic, Industrial and Control Engineering*, pp. 12-18, 2015.
[28] X. Zang, F.P. Vista and K.T. Chong, “Fast Global Kernel Fuzzy C-Means Clustering Algorithm for Consonant/Vowel Segmentation of Speech Signal”, *Journal of Zhejiang University Science C*, Vol. 15, No. 7, pp. 551-563, 2014.
[29] Ireneusz Czarnowski and Piotr Jedrzejowicz, “An Approach to Data Reduction for Learning from Big Datasets: Integrating Stacking, Rotation, and Agent Population Learning Techniques”, *Complexity*, Vol. 2018, pp. 1-13, 2018.
[30] Ireneusz Czarnowski and Piotr Jedrzejowicz, “Learning From Examples with Data Reduction and Stacked Generalization”, *Journal of Intelligent and Fuzzy Systems*, Vol. 32, No. 2, pp. 1401-1411, 2017.