DOI: 10.21917/ijsc.2018.0249

INTERPRETATION OF ECG USING MODIFIED INTUITIONISTIC FUZZY C-MEANS CLUSTERING FOR ARRHYTHMIA DATA

C.K. Roopa[1] and B.S. Harish[2]

[1]*Department of Computer Science, JSS Technical Institution Campus, Mysuru, India*

[2]*Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, India*

Abstract

An electrocardiogram (ECG) is defined as a measure of variation in the electrical activity of the heart and is broadly used in detection and classification of heart-related diseases. The abnormalities present in the heart can be easily analyzed through the variation in electrical signal captured from the heart through impulse waveforms which are generated by certain specialized cardiac tissues. Different authors have developed various clustering models and classification techniques for detecting heart-related diseases. However there still exists a limitation in terms of accuracy. In this article, we proposed a new modified unsupervised clustering algorithm for effective detection of heart diseases. To select the best discriminate feature for effective learning, this article make use of feature selection methods such as principal component analysis, linear discriminative analysis, and regularized locality preserving indexing. The reduced features set are clustered using modified intuitionistic Fuzzy C-means clustering (mifcm) method. The experiment results proved that the proposed method effectively identifies the discriminative features. Further the obtained accuracy is also better when compared to other existing method.

Keywords:

Electrocardiogram, Heart Diseases, Feature Selection, Intuitionistic Fuzzy c-means.

1. INTRODUCTION

Medical image or signal processing is an efficient, powerful tool adopted in the field of human healthcare services to gain insight over the normal and pathological processes that affect the human health. In modern health care systems, medical image processing acts as an innovative core field for specific applications such as neuroinformatics, bioinformatics and medical informatics [1] [2]. The recent survey conducted by the World Health Organization (WHO) shows that Cardiovascular Diseases (CVD) are the primary cause of the increase in death rate. Electrocardiogram (ECG) is found to be an essential non-invasive clinical tool which comprises details regarding the rhythmic behavior and functionality of the human heart. It is frequently used by cardiologists to detect and diagnose heart-related diseases. There are various types of CVD, among them Arrhythmia is one type which requires more attention. Several stages are involved in classification and detection of arrhythmia. The important phases are: feature selection stage and selection of clustering model [3]. Feature selection is a vital stage during the ECG signal processing. Therefore, it is necessary to extract precise features and dimensionality of the feature should be reduced to obtain optimal classification results.

Arrhythmia (Captured using ECG Signal) can be stated as a set of variable conditions captured in the form of electrical signals [4]. The variation in the irregular behaviour of the heartbeat is proportional to the difference in output electrical signals. The limitations observed in the detection of arrhythmia is that symptoms do not show up all the time and the presence of noise due to mechanical components have a more significant impact on the accuracy of the anomaly detection. It is seen that the ECG signal characteristic is different for every person and also varies with respect to time and their conditions [5]. Thus, there is a huge requirement for the effective model to diagnose arrhythmia diseases with high accuracy rate. Numerous research works are reported in terms of arrhythmia detection. Each method has its own advantages and disadvantages.

In this article, an effective arrhythmia classification method is proposed. The proposed model comprises different feature selection algorithms and modified Intuitionistic Fuzzy C-Means clustering method.

The remainder of the manuscript is organized as follows: section 2 presents literature work. Section 3 presents the proposed methodology, section 4 presents the experimental results. The conclusions are drawn in section 5.

2. LITERATURE REVIEW

In the Literature, several techniques and methodologies are reported to analyze ECG arrhythmia. The reviews of different existing techniques are as follows,

An improved modular learning vector quantization (LVQ) based neural network and integrated response from fuzzy systems for classification of arrhythmia is developed in [11]. Input dataset comprises 48 ECG signals which are further distinguished into 15 classes. The initial 23 records are captured from random population and the other 25 records are randomly calculated through clinically significant arrhythmias. Further, the proposed system architecture has been divided into three modules in which individual module deals with 5 different types of arrhythmias and the matrix is developed. Overall 16 experimentation is carried out and the study on results show that an accuracy rate of 95.33% is achieved with18 cluster. The research work reported that the developed method can be further used to calculate optimal parameters in LVQ networks. Furthermore, an efficient ECG arrhythmia classification using Fuzzy C-means clustering technique is developed by [12]. Fuzzy C-means clustering technique has been developed along with Heart Rate Variability (HRV). 20 files are captured from ECG record and are used for training phase and 18 files are considered for testing phase. The results are calculated in terms of error classification. A comparative analysis is carried out to prove the efficiency of the system. The experiment results using FCM clustering techniques has resulted with recognition rate of 98.50% to 99.60% with an average accuracy rate of 99.05% and error rate of less than 0.6%. Similarly a Mixed Fuzzy Clustering (MFC) algorithm has been developed by the author [19] to algorithm is used to design Takagi-Sugeno (TS) fuzzy system models. Mixed fuzzy models have the capability to withstand multivariate time variant and time invariant features and the weight of individual component can be controlled as per the user requirement through the clustering process. At the initial stage, MFC algorithms are developed to obtain the TS model antecedent fuzzy sets followed by construction of fuzzy c-means algorithm. The developed model can be effectively used to analyze the several healthcare issues, where it is proportional to time series with unequal length. The result obtained from the study on analysis shows that the proposed model outperformed fuzzy based models and the existing *k*-nearest neighbour classifiers.

Author in [9] presents a robust and efficient unsupervised method for ECG arrhythmia classification. 6 classes of arrhythmia are clustered using a Robust Spatial Kernel FCM (RSKFCM) method. RSKFCM is variant of FCM method which uses Gaussian kernel as distance metric and incorporates neighborhood information.

A hybrid prediction model comprising Improved Fuzzy C Means (IFCM) clustering, support vector machine and principal component analysis is used to detect type 2 diabetes and cardiovascular risk prediction model is developed by the author [8]. Algorithms deployed during the development of proposed system are. IFCM is used to validate the class label of selected input data. The SVM based classifier is used for classification purpose by applying a k-fold cross-validation set, and principal component analysis is used for dimensionality reduction. Information gain parameters and entropy are used for the measurement of attributes and performance evaluation is done in terms of specificity and sensitivity. It is observed from the study that higher predictive accuracy is obtained from the developed model compared to other existing techniques. A novel efficient algorithm comprising a machine learning algorithm for the classification of ECG arrhythmia is developed in [6]. Morphological filter based pre-processing technique is considered for the removal of the high-frequency noise component and the presence of baseline drift. This technique is selected because of the requirement to preserve the ECG morphology and preservation of QRS complex sharpness. Principal component analysis technique is used as a dimensionality reduction technique to minimize the dimensionality of the multidimensional data in free space. Extreme machine learning based back propagation neural network, radial basis function and support vector machine are used for optimal weight calculation. The author has conducted experimentation on the developed technique, and the study demonstrates that better results are achieved in terms of sensitivity, accuracy and average specificity. Further, the results are compared with the existing techniques, and it is observed that developed technique outperformed other techniques, and it is efficient in classifying six types of variation in heartbeat namely normal beat, branch block, left bundle, premature ventricular contraction, bundle branch block and paced heartbeat. The author summarized the research stating that the combination of SVM algorithm along with PCA recognizes and classifies the ECG signal accurately compared to other existing techniques. A novel system comprising unique data mining technique and a new procedure to classify the heart-related diseases are developed [7]. Initially, the input data is captured comprising the number of diseases related to arrhythmia. ECG signal is further pre-processed for the removal of noise in the signal and extracted specific features such as QRS complex, polarity, R-peak, gradient and duration from the ECG signal. Further, the size of the input is reduced through principal component analysis and minimized features are clustered through fuzzy c means clustering algorithm. Support Vector Machine (SVM) is employed for the classification purpose. From the study, it has been observed that 88.4% accuracy is achieved for the training classification and 90% accuracy is obtained for testing classification purpose.

An efficient technique comprising LDA and SVM method to classify the big data comprising ECG signals through cloud computing is proposed in [10]. Initially, noise is removed using FIR and IIR filters. Later SVM is employed for pattern recognition. The experiment considered mobile health dataset as input and twenty attributes are extracted through the process. Performance of the system is evaluated in terms of specificity, mean square error and study on results show that the developed technique outperforms other existing techniques.

From the above literature, it is ascertained that various machine learning algorithms are deployed to determine and classify ECG arrhythmia. However, still there exists a research gap in terms of accuracy rate and precision. In this research, different feature selection methods are explored along with new intuitionistic fuzzy c-means clustering algorithm. The proposed algorithm is based on intuitionistic fuzzy set theory for calculating the hesitation degree that arises while defining the membership function.

3. METHODOLOGY

The methodology of ECG arrhythmia classification is presented in this section. The proposed method comprises of 3 phases viz: Feature Extraction, Feature selection and Clustering. To cluster the data we explore the use of Intuitionistic fuzzy set theory . The main advantage of the proposed method is able to calculate the hesitation degree that arises while defining the membership function. The block diagram of the proposed arrhythmia classification is presented in Fig.1.

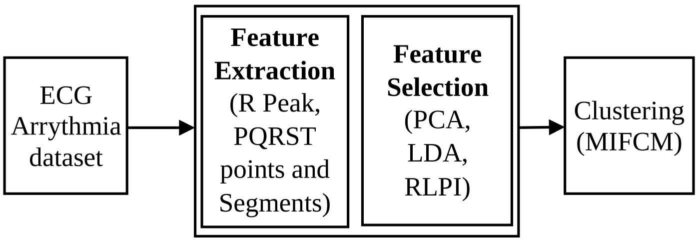

Fig.1. Block diagram of the proposed ECG arrhythmia classification method

3.1 FEATURE EXTRACTION

In this section, a part of the signal is selected and extracted from the R peak as a time domain based feature. The R peak obtained across the ECG signal is considered as a key index for cardiac disease detection. Further, other parameters such as P, Q, R, S and T are considered to ensure the increase in performance accuracy of ECG arrhythmia classification. The feature extraction is further followed by feature selection. During feature selection, more importance is given to identify the unique feature such as amplitude of the signal, time length between the individual interval and period of component detection.

3.2 FEATURE SELECTION

The performance of the proposed method is mainly affected by the huge dimensionality of feature matrix. In this paper to alleviate high dimensionality problem three widely used feature selection methods are employed viz. Principle Component analysis (PCA), Linear Discriminant Analysis (LDA) and Regularized Locality Preserving Indexing (RLPI). A brief description about these feature selection method is presented in next subsections.

3.2.1 Principal Component Analysis (PCA):

Principal Component Analysis (PCA) [13] is a very well-known linear feature selection technique. In PCA, the correlation among the features are computed by projecting the input data into feature space. Compared to other feature selection methods, PCA is simple and non-parametric. PCA have the ability to preserve larger percentage of total variance by using only a few components.

3.2.2 Linear Discriminant Analysis (LDA):

In machine learning application, Linear Discriminant Analysis (LDA) [14] is one of the commonly used feature selection method. Similar to PCA, LDA also projects input data to feature space. In LDA, the axes direction which maximizes the separation between the classes are computed using the class label information.

3.2.3 Regularized Locality Preserving Indexing (RLPI):

Regularized Locality Preserving Indexing (RPLI) [15] is a variant of LPI. To address the high computation problem of LPI, RLPI breaks LPI problem into two parts: graph embedding and regularized square problem. RLPI incorporates different regularizes naturally and this makes it more flexible and computation inexpensive.

3.3 CLUSTERING

In this stage reduced ECG data is clustered by employing Modified Intutionistic Fuzzy C-Means (MIFCM). MIFCM is based on Intutionistic fuzzy set and variant of original IFCM [18]. Unlike FCM, where data points are grouped based on membership value, in MIFCM, it considers non membership along with membership value to decide the belongingness of the data to a cluster. Let $X = \{x_1, x_2, x_3,\ldots,x_n\}$ be the n ECG arrhythmia samples which are to be clustered among c clusters.

The main aim of the MIFCM is to minimize the objective function shown in Eq.(1),

$$J = \sum_{i=1}^{c}\sum_{j=1}^{n} w_{ij}^{m} d\left(x_j, v_i\right). \tag{1}$$

where, c is the number of clusters, n is the number of samples, m is a fuzzifier value, which controls the fuzziness of the resulting partition, w_{ij} is the MIFCM membership degree of x_j in i^{th} cluster. v_i is the i^{th} cluster center. $d(x_j,v_i)$ is the modified hausdroff distance between x_j sample and v_i cluster center. Similar to FCM, MIFCM optimizes the objective function iteratively by updating the membership and cluster centers. In MIFCM, the belongingness of samples depends on the membership, non-membership and hesitation degree values. Thus, the MIFCM membership value is the combination of membership and hesitation degree. The MIFCM membership is computed as:

$$w_{ij} = u_{ij} + \pi_{ij} \tag{2}$$

where, u_{ij} and π_{ij} is the FCM membership and hesitation degree values respectively. The membership is computed as:

$$u_{ij} = \frac{\left(1-\left(x_{j,}v_i\right)\right)^{\frac{-1}{(m-1)}}}{\sum_{k=1}^{c}\left(1-\left(x_{j,}v_k\right)\right)^{\frac{-1}{(m-1)}}} \tag{3}$$

The hesitation degree is computed as:

$$\pi_{ij} = 1 - u_{ij} + \eta_{ij} \tag{4}$$

where, η_{ij} is the non membership value. In MIFCM, the non membership value is computed using two complement generator functions viz., Sugeno's [16] and Yager's [17]. Sugeno's complement generator function is computed as:

$$\eta_{ij} = \frac{1-u_{ij}}{1+\alpha u_{ij}} \tag{5}$$

The Yager's complement generator function is computed as:

$$\eta_{ij} = \left(1-\left(u_{ij}\right)^{\alpha}\right)^{\frac{1}{\alpha}} \tag{6}$$

where, α is a constant.

Similar to FCM, MIFCM also works in an iterative process to update the membership and cluster center values. The cluster centers are updated using Eq.(7)

$$v_i = \frac{\sum_{k=1}^{c} w_{ij}^{m} d\left(x_{j,}v_i\right)x_j}{\sum_{k=1}^{c} w_{ij}^{m} d\left(x_{j,}v_i\right)} \tag{7}$$

MIFCM is an iterative process and it stops when convergence criteria is satisfied (i.e., the difference between objective function value of successive iterations is less than the user specified stopping criteria value). When MIFCM converges, each ECG sample is associated with the membership values. Based on the membership values, ECG samples are clustered where it exhibits maximum value. Algorithm 1 presents the individual steps involved in MIFCM method.

Algorithm 1: Modified Intuitionistic Fuzzy C-Means for ECG Arrhythmia Classification

Data: ECG Arrhythmia data $X = \{x_1, x_2, x_3,\ldots,x_n\}$, Number of clusters ($c$), Stopping criteria ($\varepsilon$), Window size N_k, p, q, m

Result: Membership matrix, Cluster centers

Initialize cluster centers $V = \{v_1, v_2, v_3,\ldots,v_n\}$, t = 0 (iteration), $J(t = 0) = 0$ (initial objective function value);

Repeat

// *Calculate membership value*

for $i \leftarrow 1$ to c do

    for $j \leftarrow 1$ to n do

        calculate membership value u_{ij} using Eq.(3)

    end

end

// *Calculate non membership value*
for $i \leftarrow 1$ to c do
    for $j \leftarrow 1$ to n do
        calculate η_{ij} using Eq.(5) or Eq.(6)
    end
end
// *Calculate hesitation degree value for* $i \leftarrow 1$ *to* c *do*
for $j \leftarrow 1$ to n do
    calculate π_{ij} using Eq.(4)
    end
end
// *Calculate MIFCM membership value*
for $i \leftarrow 1$ to c do
    for $j \leftarrow 1$ to n do
        calculate w_{ij} using Eq.(2)
    end
end
// *Update cluster center V*
for $i \leftarrow 1$ to c do
    calculate v_i using Eq.(7)
end
$t=t+1$;
// *Calculate objective function*
calculate objective function $J(t)$ using Eq.(1) until $\{J(t)-J(t-1)\} < \varepsilon$;

4. EXPERIMENTAL EVALUATION, RESULTS AND DISCUSSION

This section describes the details about experimental setup, dataset used, results followed by discussion.

4.1 DATASET

To evaluate the proposed method the experimentations are carried out on UCI machine learning arrhythmia dataset. In this dataset, there are total 16 classes where class 01 refers to normal and class 02-16 refers to different arrhythmia types. This dataset has 279 attributes and 452 samples. This dataset contains a classes which as a very minimal samples, thus in this paper only six classes are used to conduct the experiments. The Table.1 presents the description of the class and number of samples used for experimentation.

4.2 RESULTS

The performance of the proposed method is evaluated in terms of accuracy. To corroborate the efficacy of the proposed method, we compared the results with traditional FCM (FCM), Kernel FCM (KFCM), Robust Spatial Kernel FCM (RSKFCM) and original IFCM. We have used both sugeno's and yager's complement generator functions. The Table.2 presents the results comparison of the proposed method. In Table.2, MIFCM_s and MIFCM_y refers to proposed method with sugeno's and yager's complement generator function respectively. All the methods in comparison and proposed methods are implemented using Matlab 2017a.

Table.1. Tabulated for input dataset

Class	Class name	No. of instances
1	Normal	237
2	Ischemic changes (Coronary Artery Disease)	36
3	Old Inferior Myocardial Infarction	14
4	Sinus Bradycardy	24
5	Right bundle branch block	48
6	Others	18

Table.2. Comparing performance in terms of accuracy

Method	Alpha value	Accuracy (%)			
		With Original Features (279)	PCA (70)	LDA (30)	RLPI (10)
FCM	-	67.38	67.93	70.66	74.96
KFCM	-	68.81	70.57	74.46	77.58
RSKFCM	-	76.13	86.94	88.95	90.35
IFCM_S	0.6	83.30	87.02	89.16	91.05
IFCM_Y	0.4	84.01	87.68	90.18	92.13
MIFCM_S	25	85.52	88.14	91.23	94.75
MIFCM_Y	20	86.12	89.05	92.27	95.36

Originally the dataset contains 279 features. We employed three feature selection methods and found the optimal feature numbers through empirical evaluation for all three methods. From empirical evaluation it is found that 70, 30 and 10 are optimal feature numbers for PCA, LDA and RLPI respectively.

The performance of the MIFCM mainly depends on the value of α. We varied α value from 0 to 50 with an increment of 0.1. The optimal value for α is found through empirical evaluation. The Table.2 present the results comparison of the proposed method with other clustering methods. The Fig.2 presents the percentage data reduced using all the three feature selection methods with different clustering methods. The Fig.3 depicts the comparative analysis of the proposed method with other existing methods.

From Table.2, we can observe that compared to PCA and LDA, RLPI feature selection method performed better with all clustering methods. RLPI produced better accuracy improvement. In addition 74.91%, 89.24% and 96.41% data reduction is achieved through PCA, LDA and RLPI respectively. The same is depicted in Fig.2.

In addition to feature selection method, the clustering method also plays an important role in achieving better performance. Hence after feature selection the selected features are further fed to different clustering algorithms. Initial analysis is carried out through FCM, KFCM, RSKFCM and IFCM with Sugenos and Yagers complement generator functions. IFCM and the proposed algorithm are varied with respect to α value to obtain better optimized weight component. As previously mentioned α value is varied between 0 and 50. However, the best values of α achieving more results are considered and presented in Table.2. The MIFCM handles the uncertainty to better extent by taking the non-membership and hesitation degree in to account. The results obtained from the proposed algorithm are compared with other algorithms and the same is depicted in Table.2. From the analysis, it is observed that the better result with high accuracy rate of 95.36% is obtained from RLPI feature selection method and proposed MIFCM clustering algorithm.

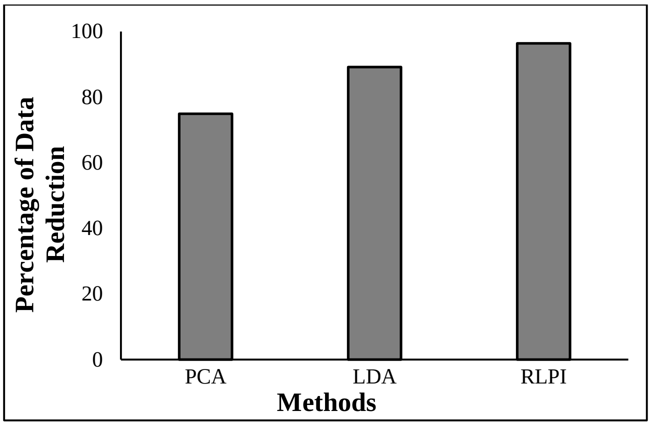

Fig.2. Comparison of Data Reduction in percentage

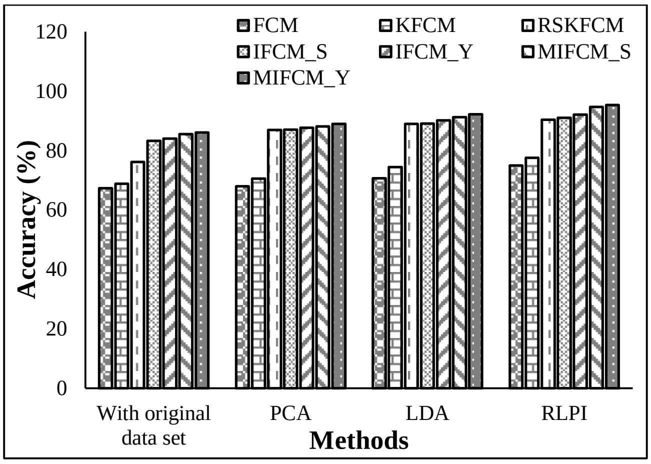

Fig.3. Comparative analysis of Accuracy between the proposed method and other existing methods

5. CONCLUSIONS

In this paper, we have proposed a modified intuitionistic fuzzy c-means algorithm (MIFCM) which is based on intuitionistic fuzzy sets (IFSs) theory. The main objective of the new proposed algorithm is to handle uncertainty and vagueness associated with real data. To incorporate hesitation degree, two parametric intuitionistic fuzzy complements generators namely Sugeno's negation function and Yager's negation function are investigated. MIFCM was applied on UCI arrhythmia dataset, for clustering of 6 classes of arrhythmia. In addition, feature selection methods like PCA, LDA and RLPI are explored to improve the performance of the proposed method. The performance of the MIFCM method is compared with other variants of fuzzy C Means clustering methods. Experimental results demonstrate the superior performance of the MIFCM method with RLPI over others.

REFERENCES

[1] C. Kulikowski et al., "Medical Imaging Informatics and Medical Informatics: Opportunities and Constraints", *Methods of Information in Medicine*, Vol. 41, No. 2, pp. 183-189, 2002.

[2] J. Wiemer et al., "Informatics United", *Methods of Information in Medicine*, Vol. 42, No. 2, pp. 126-133, 2003.

[3] U.R. Acharya et al., "Automated Detection of Arrhythmias using Different Intervals of Tachycardia ECG Segments with Convolutional Neural Network", *Information Sciences*, Vol. 405, pp. 81-90, 2017.

[4] S.U. Kumar and H.H. Inbarani, "Neighborhood Rough Set based ECG Signal Classification for Diagnosis of Cardiac Diseases", *Soft Computing*, Vol. 21, No. 16, pp. 4721-4733, 2017.

[5] B. Boashash, "*Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*", Academic Press, 2015.

[6] J. Kim, H.S. Shin, K. Shin and M. Lee, Robust Algorithm for Arrhythmia Classification in ECG using Extreme Learning Machine, *Biomedical Engineering*, Vol. 8, No. 1, pp. 1-31, 2009.

[7] P. Raman and S. Ghosh, "Classification of Heart Diseases based on ECG Analysis using FCM and SVM Methods", *International Journal of Engineering Science*, Vol. 67, No. 1, pp. 31-39, 2016.

[8] P. Radha and B. Srinvasan, "Hybrid Prediction Model for the Risk of Cardiovascular Disease in Type-2 Diabetic Patients", *Expert Systems with Applications*, Vol. 2, No. 10, pp. 23-29, 2014.

[9] C.K. Roopa, B.S. Harish and S.V. Arun Kumar, "A Novel Method of Clustering ECG Arrhythmia using Robust Spatial Kernel Fuzzy C-Means", *Proceedings of 8^{th} International Conference on Advances in Computing and Communications*, pp. 221-234, 2018.

[10] R. Varatharajan, G. Manogaran and M.K. Priyan, "A Big Data Classification Approach using LDA with an Enhanced SVM Method for ECG Signals in Cloud Computing", *Multimedia Tools and Applications*, Vol. 77, No. 8, pp. 1-24, 2018.

[11] A. Dallali, A. Kachouri and M. Samet, "Classification of Cardiac Arrhythmia Using WT, HRV, and Fuzzy C-Means Clustering", *Signal Processing: An International Journal*, Vol. 5, No. 3, pp. 101-109, 2011.

[12] A.N. Benaichouche., H. Oulhadj and P. Siarry, "Improved Spatial Fuzzy C means Clustering for Image Segmentation using PSO Initialization, Mahalanobis Distance and Post-Segmentation Correction", *Digital Signal Processing*, Vol. 23, No. 5, pp. 1390-1400, 2013.

[13] S. Wold, K. Esbensen and P. Geladi, "Principal Component Analysis", *Chemometrics and Intelligent Laboratory Systems*, Vol. 2, No. 1-3, pp. 37-52, 1987.

[14] R.A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems", *Annals of Human Genetics*, Vol. 7, pp. 179-188, 1936.

[15] D. Cai, X. He, W.V. Zhang and J. Han, "Regularized Locality Preserving Indexing Via Spectral Regression", *Proceedings of 6th ACM Conference on Information and Knowledge Management*, pp. 741-750, 2007.
[16] M. Sugeno and T. Terano, "A Model of Learning Based on Fuzzy Information", *Kybernetes*, Vol. 6, pp. 157-166, 1977.
[17] R.R. Yager, "On the Measure of Fuzziness and Negation Part I: Membership in the Unit Interval", *International Journal of General Systems*, Vol. 5, No. 2, pp. 221-229, 1979.
[18] S.V. Arun Kumar and B.S. Harish, "A Modified Intuitionistic Fuzzy Clustering Algorithm for Medical Image Segmentation", *Journal of Intelligent Systems*, Vol. 27, No. 4, pp. 593-607, 2017.
[19] M.S. Catia et al., "Takagi-Sugeno Fuzzy Modeling using Mixed Fuzzy Clustering", *IEEE Transaction of Fuzzy Systems*, Vol. 25, No. 6, pp. 1417-1429, 2017.