International Journal of Artificial Intelligence and Neural Networks– IJAINN
Volume 5: Issue 2 [ISSN : 2250-3749]
Publication Date : 30 October, 2015

An Echocardiogram Diagnosis System based on LDA and Fuzzy C-Mean Unsupervised Classifier

Murat KARABATAK

***Abstract*—Today, many people are at risk of heart attack. Echocardiogram imaging techniques, looking at the eye doctors to view patients with heart attack risk is difficult to determine. In this case, the possibility of making an incorrect diagnosis by doctors rises. Recently, many computer-based decision support systems have been developed to decrease wrong diagnosis possibility in the field of echocardiogram. The main purpose of these decision support systems designed to help expert doctors in the field of medical diagnostics. In this study, an echocardiogram diagnostic (ED) based on Linear Discriminant Analysis (LDA) and Fuzzy C-Mean (FCM) clustering algorithm (LDA-FCM-ED) system is presented. This echocardiogram diagnostic based on Linear Discriminant Analysis and Fuzzy C-Mean clustering algorithm system is composed of two stages. First step is the feature extraction and feature reduction stage by using LDA and second step is clustering stage by using FCM unsupervised classifier. In stage of feature reduction, LDA is used to reduce the attributes from 12 to 4. In clustering stage, features obtained in first stage are fed into the FCM unsupervised classifier. The correct diagnosis performance of the LDA-FCM-ED clustering system for diagnosis the echocardiogram is calculated by using classification accuracy, sensitivity and specificity values respectively. The classification accuracy of this LDA-FCM clustering system for diagnosis of the echocardiogram (LDA-FCM-ED) was obtained 93.44 %. This result is better than FCM and other existing unsupervised classifier methods.**

***Keywords*— Linear Discriminant Analysis (LDA), Fuzzy C-Mean (FCM) unsupervised classifier, Echocardiogram data, confusion matrix.**

I. Introduction

Echocardiography, the ultrasonic sound waves, allowing examination of the heart various structures is a diagnostic and research methods. In vascular system, an ultrasonic diagnostic method is used. In the today world, vibrations, which are the sound wave frequencies of millions per second in the vicinity. These vibrations are used for medical ultrasonic sound. Ultrasonic sound waves of the body tissue in a certain direction are rapidly advancing the 5140 feet per second on average. While these advances have come across according to the tissue characteristics they undergo reflection and breaking. Reflected in the different structures of the heart that sound waves with a special receiver (piezoelectric transducer) is received and converted into electrical signs. These marks will be converted into pictures, display or paper record is reflected by reflection of ultrasonic waves. Artificial heart valve that is inserted into the structure and functioning of the state also can be evaluated by echocardiography. Ventricle by echocardiography (the heart ventricles), and the movement of the wall cavity, the heart muscle and heart valves can be analyzed for their growth. This process is not a question of no harm to patients. Be done is quite simple. However, the evaluation must be experienced to be able to do. There are several types of echocardiograms. These can be ordered as below:

Transesophageal echocardiogram: This transesophageal echocardiogram requires that the transducer be inserted down the throat into the esophagus [1]. The esophagus is placed close to the heart, clear images of the heart structures can be obtained without the interference of the lungs and chest.

Transthoracic echocardiogram: Transthoracic echocardiogram is the standard echocardiogram. Transthoracic echocardiogram is a painless test similar to X-ray, but without the radiation. It uses the procedure by the same technology used to evaluate a baby's health before birth [2]. A hand-held device named a transducer is placed on the chest and transmits high frequency sound waves (ultrasound). These high frequency sound waves (ultrasound) bounce off the heart structures, producing images and sounds that can be used by the doctor to detect heart damage and disease.

Intravascular ultrasound: Intravascular ultrasound is a form of echocardiography done during cardiac catheterization. The transducer is threaded into the heart blood vessels by a catheter in the groin during this procedure. The transducer is often used to provide detailed information about the atherosclerosis (blockage) inside the blood vessels.

Stress echocardiogram: This stress echocardiogram is performed while the person exercises on a treadmill or stationary bicycle. This stress echocardiogram can be used to visualize the motion of the heart's walls and pumping action when the heart is stressed. This stress echocardiogram may expose a lack of blood flow that isn't always apparent on other heart tests. This stress echocardiogram is realized just prior and just after the exercise.

Dobutamine or adenosine/sestamibi stress echocardiogram: This Dobutamine stress echocardiogram is another form of stress echocardiogram. Moreover, instead of exercising to stress the heart, the stress is obtained by giving a drug that stimulates the heart. This Dobutamine stress echocardiogram is used to evaluate your heart and valve function when you are unable to exercise on a treadmill or stationary bike. At the

Murat KARABATAK

Faculty of Technology / Firat University
Turkey

same time, this Dobutamine stress is also used to determine how well your heart tolerates activity.

In literature, there are many studies about automatic diagnosis of echocardiogram. In one of these studies, authors presented a new classification system that is based on fuzzy level sets sub grouping classification system [3]. This new classification system permits a fast classification method with quite accurate results. In this study, classification results are compared to some existing results in the literature.

In the other study about automatic diagnosis of echocardiogram, it is introduced an automatic method for classification and segmentation of different intracardiac masses in tumor echocardiograms [4]. Classification and segmentation of selected benign tumor echocardiograms were performed. In that study, oscillator network was used with network weights defined for both whole texture region and texture boundary detection for the tumor segmentation. This method has similar accuracy performance and its segmentation duration is sufficiently. The obtained results by using oscillator network classification method were discussed and compared with the artificial neural network classifier.

In this study, an echocardiogram diagnosis system based on LDA feature reduction and FCM clustering algorithm was proposed. LDA was used for feature reduction stage and FCM unsupervised classifier was used for clustering stage. Results of experiments shows that proposed LDA-FCAM-ED system is more efficient than FCM and some other existing methods.

II. The Feature Reduction Process

In this study, The Linear Discriminant Analysis (LDA) method was used for feature reduction. LDA is a class specific discriminative [5,6]. This technique benefits supervised learning to find a set of base vectors. These base vectors are demonstrated as t_k. The t_k vectors are ratio of the between- and withinclass scatters of the training sample set, which is maximized. For find t_k base vectors, the following generalized eigenvalue problem is solved.

$$Topt = \underset{t}{\arg\max} \frac{\left|T^T G_C T\right|}{\left|T^T G_V T\right|} = [t_1, t_2, ..., t_L] \qquad (1)$$

here, $\{t_k | 1 \le k \le L\}$ are the LDA subspace base vectors. L is the dimension of the subspace. G_C and G_V are the between and within class scatter matrices. These matrices can be given as below:

$$G_C = \sum_{k=1}^{a} B_k (\mu_k - \mu)(\mu_k - \mu)^T \qquad (2)$$

$$G_v = \sum_{k=1}^{a} \sum_{x_u \in X_k} (x_u - \mu_k)(x_u - \mu_k)^T \qquad (3)$$

here, a is the number of classes and $x \in R^N$ is the data sample. X_k is the set of samples with class label k. μ_k is the mean for the all the samples with the class label k. B_k is the number of samples in the class k. If G_v is non-singular, the base vectors t_k sought in Eq.(1) are the first L largest eigen values $\{\psi_k | 1 \le k \le L\}$. It can be obtained its representation in LDA subspace by a simple linear projection $T^T x$ for a given test sample x due to the LDA base vectors are orthogonal to each other [5, 6].

III. Echocardiogram Database

In this paper, the echocardiogram data set from the UCI machine learning repository was used. All patients in this dataset suffered from heart attacks in the past [7-9]. Some of these heart attack patients are still alive and some of these heart attack patients are not. One patient had previously suffered a heart attack again against the risk of heart attack should be followed for at least a year. Echocardiogram database information used in this study was organized for this problem to be unpredictability.

Attributes of the echocardiogram database used in this study are given below [7]:

Attribute 1: Label of this attribute is “The survival”. The number of month’s patient survived.

Attribute 2: Label of this attribute is “The still-alive”. It has binary value. In here, 0 is dead at end of survival period and 1 is still alive.

Attribute 3: Label of this attribute is “The age of patient at heart-attack”. In here, age in years when heart attack occurred.

Attribute 4: Label of this attribute is “The pericardial effusion”. It has binary value. In here, pericardial effusion is fluid around the heart. 0 is no fluid and 1 is fluid.

Attribute 5: Label of this attribute is “The fractional shortening”. It is a measure of contracility around the heart lower numbers are increasingly abnormal.

Attribute 6: Label of this attribute is “The epss”. Larger numbers of they are increasingly abnormal.

Attribute 7: Label of this attribute is “The lvdd”. It means left ventricular end-diastolic dimension. This is a measure of the size of the heart at end-diastole. Large hearts tend to be sick hearts.

Attribute 8: Label of this attribute is “The wall motion score” In here, a measure of how the segments of the left ventricle are moving.

Attribute 9: Label of this attribute is “The wall motion index”. It equals wall-motion-score divided by number of segments seen. This variable is used instead of the wall motion score.

Attribute 10: Label of this attribute is “The mult”. It means “a derivate var which can be ignored”.

Attribute 11: Label of this attribute is “The name”. It is name of the patient. This Attribute was not used in the LDA-FCM-ED system.

Attribute 12: Label of this attribute is “The group”. In here, values of this attribute are meaningless or ignore it.

Attribute 13: Label of this attribute is "Alive". It has Boolean value. 0 means patient was either dead after 1 year or had been followed for less than 1 year. 1 means patient was alive at 1 year.

IV. Echocardiogram Diagnostic based on LDA and FCM Clustering System

In this section, an echocardiogram diagnostic based on Linear Discriminant Analysis (LDA) and Fuzzy C-Mean (FCM) clustering algorithm (LDA-FCM-ED) system is introduced. This echocardiogram diagnostic based on LDA and FCM clustering algorithm (LDA-FCM-ED) system is composed of from two stage: The feature reduction stage by using LDA and clustering stage by using FCM unsupervised classifier. In stage of feature extraction and feature reduction, LDA is used to reduce the attributes from 12 to 4. In clustering stage, features obtained in first stage are given to inputs of FCM unsupervised classifier. The block diagram of the LDA-FCM clustering system for diagnosis of echocardiogram used in this study is given in Fig. 1.

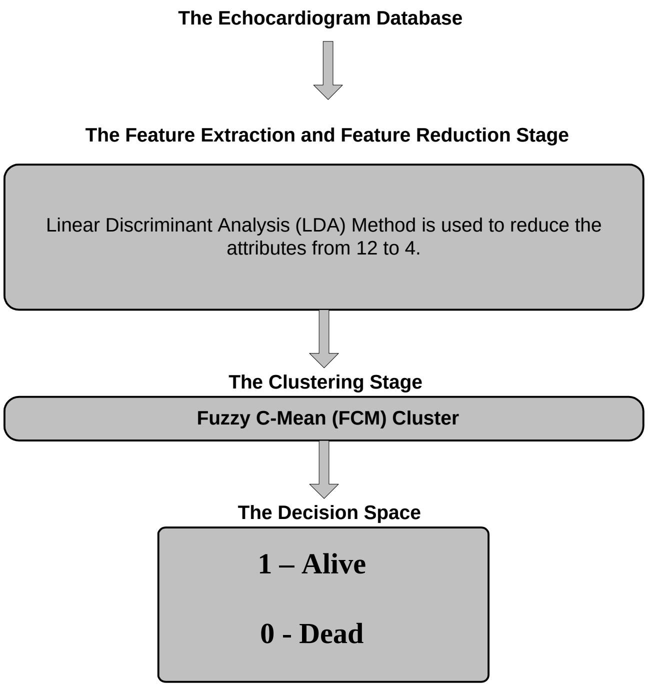

Fig.1. The block diagram of the LDA-FCM clustering system for diagnosis of echocardiogram.

A. *The Feature Reduction Stage of LDA-FCM clustering system*

In this stage, the Linear Discriminant Analysis (LDA) was conducted to this database. Echocardiogram database with 13 attributes and 132 records were used in the experimental works. It is worth to mention that 71 records were discarded because of the missing values and 1 attribute (The name) was discarded. Thus, 61 samples with 12 attributes were used in LDA-FCM clustering system. All attributes are explained in Section 3. These 12 attributes are the survival, the still-alive, the age of patient at heart-attack, the pericardial effusion, the fractional shortening, the epss, the lvdd, the wall motion score, the wall motion index, the mult, the name, and the group respectively.

In this database, there are two class labels: 0 (0 means patient was either dead after 1 year or had been followed for less than 1 year) and 1 (1 means patient was alive at 1 year). The class value number of samples are 44 for 0 (patient was either dead after 1 year or had been followed for less than 1 year) and 17 for 1 (1 means patient was alive at 1 year).

B. *The Clustering Stage by Using Fuzzy C-Mean of LDA-FCM Clustering System*

The FCM is the one of the unsupervised classification (clustering) technique [10, 11]. It is a clustering technique that is different from hard k-means. The FCM uses fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1. Also it has an iterative algorithm. The goal of it is to find cluster centers, which are called as centroids that minimize a dissimilarity function. Therefore, the membership matrix (U) is randomly initialized as appropriate to Eq.(4) for providing of fuzzy partitioning, [12].

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1,\ldots,n \qquad (4)$$

Moreover, the dissimilarity function of FCM can be given as below:

$$J(U, c_1, c_2, \ldots, c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}{}^{m} d_{ij}{}^{2} \qquad (5)$$

The symbols and meaning of these used in FCM can be given as below:

Symbol	Meaning
c_i	the centroid of cluster i
u_{ij}	membership and between 0 and 1
$m \in [1,\infty]$	a weighting exponent
d_{ij}	the Euclidian distance between i_{th} centroid(c_i) and j_{th} data point

In FCM algorithm, there are two processes to reach a minimum of dissimilarity function. These processes are Eqs.6 and 7.

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}{}^{m} x_j}{\sum_{j=1}^{n} u_{ij}{}^{m}} \qquad (6)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad (7)$$

The readers may find the more knowledge about the algorithm of FCM proposed by [12, 13].

V. Experimental Works and Results

In the experimental works, the echocardiogram dataset was considered that was mentioned in Section 3. All experiments were conducted on Matlab [17]. Due to nature of the unsupervised clustering methods, all dataset was used for testing the performance of the proposed LDA-FCM clustering system.

The performance of the proposed LDA-FCM clustering system was evaluated by using several metrics such as classification accuracy, sensitivity and specificity analysis, and confusion matrix performance evaluation methods respectively [14]. For this reason, firstly, the correct diagnosis rate of echocardiogram was calculated for obtaining the results of clustering accuracy analysis for LDA-FCM clustering system for diagnosis of echocardiogram. For estimating of this clustering accuracy, the equation of classification accuracy in [14-16] was used.

Secondly, results of sensitivity, specificity and accuracy analysis of LDA-FCM clustering system for diagnosis of echocardiogram were calculated. These values of this LDA-FCM clustering system were performed by using equation in [14]. True positive (TP), true negative (TN), false negative (FN), and false positive (FP) terms are commonly used along with the description of sensitivity, specificity and accuracy. Sensitivity, specificity and accuracy are described in the following equations [18-20];

$$Sensitivity = \frac{TP}{TP + FN} \quad (8)$$

$$Specificity = \frac{TN}{TN + FP} \quad (9)$$

$$Acuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

Thirdly, confusion matrix of LDA and LDA-FCM clustering system for diagnosis of echocardiogram were given in Table 1 and Table 2, respectively.

TABLE 1. THE OBTAINED CONFUSION MATRIX BY USING FCM CLUSTERING SYSTEM.

Output / Desired	Number of alive	Number of dead
Number of alive	17 (TP)	6 (FN)
Number of dead	0 (FP)	38 (TN)

TABLE 2. THE OBTAINED CONFUSION MATRIX BY USING LDA-FCM-ED CLUSTERING SYSTEM

Output / Desired	Number of alive	Number of dead
Number of alive	17 (TP)	4 (FN)
Number of dead	0 (FP)	40 (TN)

As one can see in Table 2, 17 alive samples were classified as alive. Thus, the true positive of the confusion matrix was 17. In addition, 0 alive samples were detected as dead which indicated the false positives. There were 4 false negative samples and 38 dead samples were classified as dead which symbolized the true negatives. The calculated sensitivity, specificity and the accuracy values are 80.95%, 100% and 93.44% respectively. K-Mean, FCM and LDA-FCM evaluation results were given in Table 3.

TABLE 3. PERFORMANCE COMPARISON OF FCM, LDA-FCM AND K-MEANS CLUSTERING SYSTEMS

Folds	Number of Data	K-Means	FCM %	LDA-FCM
Sensitivity	61	77.27	73.91	80.95
Specificity	61	100.00	100.00	100.00
Accuracy	61	91.8	90.16	93.44

VI. Discussion and Conclusion

In this study, an echocardiogram diagnostic based on Linear Discriminant Analysis (LDA) and Fuzzy C-Mean (FCM) clustering algorithm (LDA-FCM-ED) system is introduced. This echocardiogram diagnostic based on LDA and FCM clustering algorithm (LDA-FCM-ED) system is occurred from two stage: The feature extraction and feature reduction stage by using LDA and clustering stage by using FCM unsupervised classifier. In stage of feature extraction and feature reduction, LDA was used to reduce the attributes from 12 to 4. In clustering stage, features obtained in first stage are given to inputs of FCM unsupervised classifier. The correct diagnosis performance of the LDA-FCM clustering system for diagnosis of the echocardiogram is estimated by using classification accuracy, sensitivity and specificity analysis and confusion matrix respectively.

These results of this LDA-FCM-ED clustering system used in this study were given in Section 4. According to these results LDA-FCM clustering system is more efficient than previous methods used for diagnosis of the echocardiogram in literature, which are mentioned in Section 2.

Moreover, as shown from obtained results, the used LDA-FCM-ED clustering system can be very helpful to the physicians for their final decision about their patients. The physicians can determine more accurate decisions than previous by using LDA-FCM clustering system for diagnosis of the echocardiogram.

References

[1]. http://www.medicinenet.com/echocardiogram/article.htm, (last accessed: September 11, 2010).

[2]. http://www.medicinenet.com/echocardiogram/article.htm, (last accessed: September 11, 2010).

[3]. P. Kukkurainen and P. Luukka, Classification method using fuzzy level set subgrouping, Expert Systems with Applications, Volume 34, Issue 2, February 2008, Pages 859-865.

[4]. M. Strzelecki, A. Materka, J. Drozdz, M. Krzeminska-Pakula and J. D. Kasprzak, Classification and segmentation of intracardiac masses in cardiac tumor echocardiograms, Computerized Medical Imaging and Graphics, Volume 30, Issue 2, March 2006, Pages 95-107.

[5]. F. Tang and H. Tao, (2006). Fast Linear Discriminant Analysis Using Binary Bases The 18th International Conference on Pattern Recognition (ICPR'06).

[6]. P.N., Belhumeur, J. Hespanda, and D. Kiregeman, (1997). Eigenfaces vs.Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE PAMI.

[7]. UJI Reporsitory of Machine Learning Databases Available from http://www.ics.uci.edu./~mlearn/MLReporsitory.html, (last accessed: September, 19, 2010).

[8]. Salzberg, S. (1988). Exemplar-based learning: Theory and implementation (Technical Report TR-10-88). Harvard University, Center for Research in Computing Technology, Aiken Computation Laboratory (33 Oxford Street; Cambridge, MA 02138).

[9]. Kan, G., Visser, C., Kooler, J., & Dunning, A. (1986). Short and long term predictive value of wall motion score in acute myocardial infarction. British Heart Journal, 56, 422-427.

[10]. S. Albayrak, F. Amasyali, (2003). Fuzzy C-Means Clustering on Medical Diagnostic Systems, International XII. Turkish Symposium on Artificial Intelligence and Neural Networks – TAINN 2003.

[11]. P. Dulyakarn and Y. Rangsanseri, (2001). Fuzzy C-Means Clustering Using Spatial Information with Application to Remote Sensing, 22nd Asian Conference on Remote Sensing, 5-9 November 2001, Singapure.

[12]. J. C. Bezdek. Fuzzy Mathemathics in Pattern Classification. PhD Thesis, Applied Math. Center, Cornell University, Ithaca, 1973.

[13]. J.-S. R. Jang, C.-T. Sun, E.Mizutani, Neuro-Fuzzy and Soft Computing, p.p. (426-427) Prentice Hall, 1997.

[14]. K., Polat and S. Gunes, (2007). An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease, Digital Signal Processing, Volume 17, Issue 4, Pages 702-710.

[15]. K. Polat and S. Gunes, (2008), Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer, Expert Systems with Applications, Volume 34, Issue 1, Pages 214-221.

[16]. Sengur, A. (2008). An expert system based on linear discriminant analysis and adaptive neuro-fuzzy inference system to diagnosis heart valve diseases. Expert Systems with Applications, 35(1), 214-222.

[17]. MATLAB 6.5 version Wavelet Toolbox, MathWorks Company, (2008).

[18]. Avci, E. and Akpolat, Z. H. (2006). Speech Recognition Using A Wavelet Packet Adaptive Network Based Fuzzy Inference System, Elsevier Expert Systems with Applications, vol. 31(3), p.p. 495-503.

[19]. Avci, E., (2008). Comparison of Wavelet Families for Texture Classification by Using Wavelet Packet Entropy Adaptive Network Based Fuzzy Inference System", Applied Soft Computing, 8 (1), 225-231.

[20]. Karabatak, M., "Mammographic Mass Classification by Using a New Naïve Bayesian Classifier", Second Int. Conf. on Advances in Computing, Electronics and Electrical Technology –CEET 2014, Kuala Lumpur, Malaysia, 20-21 Dec. 2014

About Author:

Mr. Dr. Murat Karabatak received BS degree in 1999, Master degree in 2002 and PhD title in Electric and Electronics Engineering Department at Firat University in 2008. He worked for University of Sam Houston State (Texas-USA) as visitor researcher in 2014. Assist. Prof. Dr. Karabatak is vice president of Software Engineering Department now. His interest areas are Data Mining, Big Data, intelligent systems, image processing and e-learning.