International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 03, 2021

Detection Of Abnormal Liver In Ultrasonic Images From Fcm Features

Harikumar Rajaguru[1], R.Karthikamani[2]

[1] Professor, Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam, India
[2] Assistant Professor, Electronics and Instrumentation Engineering, Bannari Amman Institute of Technology, Sathyamangalam, India

Email: harikumarr@bitsathy.ac.in[1], karthikamani@bitsathy.ac.in[2]

Abstract.. The objective of this paper is to detect the liver abnormalities from ultrasonic image database. The liver cancer is one of the major health issuesnow a day. Medical Imaging techniques are proposed to diagnose the abnormalities in the earlier stage. In this paper extracted Fuzzy C means (FCM) Clustering features of liver images and five classifiers like Expectation maximization, Gaussian Mixture Model, Linear Discriminant Analysis,Bayesian Linear Discriminant Analysis Classifier, Logistic regression classifier are used to detect the normal or abnormal condition of the liver. The classifier performance are analyzed by the bench mark parameters Sensitivity, Specificity, Accuracy, Precision, Error Rate, Mathew Correlation Coefficient (MCC), and Classifier Success Index (CSI) and compared. The Logistic regression achieved a higher accuracy of 80.95% and outperformed other four classifiers.

Keywords:FCM, GMM, LDA, BLDC.

1. INTRODUCTION

The liver is most essential organ in the human body that performs various function such as eradicating waste materials from the blood, maintains blood sugar and producing important nutrient for the human body. Unlike other diseases liver abnormalities will not have noticeable symptoms in the initial level [1]. According to the analysis of WHO, liver cancer is also one of the major reasons of the death around the world. More over liver cirrhosis is the second prominent type of cancer in men and the seventh cancer type in women. Therefore, earlier identification of disease plays a vital role for proper treatment. Several imaging techniques are available for the diagnosis of liver diseases but most frequently used modality is ultrasound imaging, because of low cost and it does not produce any radiation. So the image processing and classification algorithms are mostly used in the detection of liver abnormalities [2]. The following is the structure of the paper: The paper is introduced in subdivision 1. The materials and processes are explained in subdivision 2. The classifiers are discussed in Subdivision 3. Subdivision 4 gives insight into the results. This article concluded in Subdivision 5.

2. MATERIALS AND METHODS

This paper comprises of four stages, they are image acquisition, and preprocessing, fuzzy c means clustering and classification. The noises are removed with filter in preprocessing and fuzzy c means clustering is employed then the liver abnormalities are classified as normal and cirrhosis using five classifiers. The performances of the classifiers are compared using standard bench mark matrices.

The Methodology Of Work Flow

Figure 1 illustrates the work flow of fuzzy c means clustering based liver image classification. The first stage for image classification is image acquisition. The images are collected from signal processing database**.**The signal processing lab databaseconsists of 84 liver ultrasound images, in that 42 images are normal and another 42 are abnormal images.

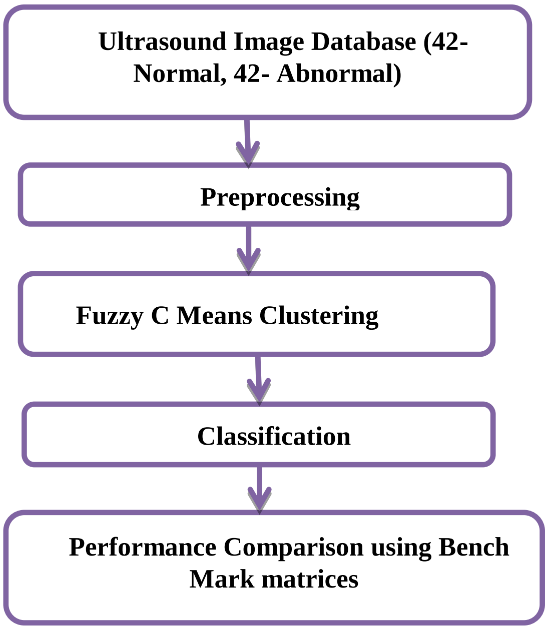

Fig. 1. Methodology of Work Flow

Preprocessing

The ultrasonic images are affected by noises both in acquisition and processing time. These noises are removed by applying the appropriate filters in the preprocessing stage itself.The speckle noise [3] is the major contributor as the noise sources in the ultrasonic images, which induces the blurring effect in it. In this work a 12 th order median filter with window size of 5x5 was used to remove the effect of speckle noise.

Fuzzy C Means Clustering

Data clustering has received a lot of attention in recent years because it is reliable data analysis tool [4]. Clustering, also known as data grouping, is an unsupervised classification technique that groups pattern data (most commonly vectors in multidimensional space) into clusters (or groups) [4].The application areas of clustering includes image compression, data analytics, data mining and pattern recognition [5]. FCM is used to solve a wide range of problems, from data processing to image segmentation. In FCM, a data sample can be assigned to several clusters at the same time. The membership value indicates the degree of resemblance. A membership value is assigned to a data sample in FCM is based on its similarity to the cluster centre. Membership values range from 0 to 1, with the stronger the similarity, the higher the membership value [6][12][13].To determines clustering, defuzzification is used at the completion of the clustering process. FCM is a recurrent algorithm that achieves a result by adjusting the cluster centre and membership value repeatedly. Solving the cost function gives these updated equations. Consider M data samples with data X, where X = {$x_1, x_2, x_3, \ldots x_M$}.By decreasing the succeeding cost function, it will be partitioned into c-clusters.

$$Z = \sum_{p=1}^{M} \sum_{q=1}^{c} u_{pq}^{N} \left\| x_p - v_q \right\|^2 \tag{1}$$

Where u_{pq} indicates the membership function of x_p with p cluster center and N is constant. The Norm vector is represented as $\|.\|$, the fuzziness of the subsequent partition is determined by the parameter N. The below equation are obtained by using Lagrange method

$$v_q = \frac{\sum_{q=1}^{M} u_{pq}^{N} x_q}{\sum_{q=1}^{M} u_{pq}^{N}} \tag{2}$$

$$u_{pq} = \sum_{k=1}^{c} \left(\frac{\| x_q - v_p \|}{\| x_q - v_k \|} \right)^2 \tag{3}$$

The statistical parameters such as mean, variance, Skewness, kurtosis, Pearson Correlation coefficient (PCC) is depicted in Table 1. The canonical correlation coefficient for FCM cluster based feature is 0.765521; it shows that the features are highly correlated for both the classes.

Table 1. Statistical Parameters for FCM Cluster Features in Abnormal and Normal Liver Images

S.No	Statistical Parameters	Abnormal	Normal
1	Mean	0.150911	0.156967
2	Variance	0.001051	0.001482
3	Skewness	-0.65167	-0.76419
4	Kurtosis	0.70954	0.935349
5	Pearson Correlation Coefficient	0.583968	0.545351
6	CCA	**0.765521**	

Figure 2 indicates the histogram plot for FCM cluster feature for abnormal case. From figure 2 it is observed that the features are nonlinear. Norm plot of FCM Clusters Features for Abnormal Liver Images is depicted in figure 3 in that the features are overlapped. Sothat thenonlinearclassifiers are used for further classification of ultrasonic images.

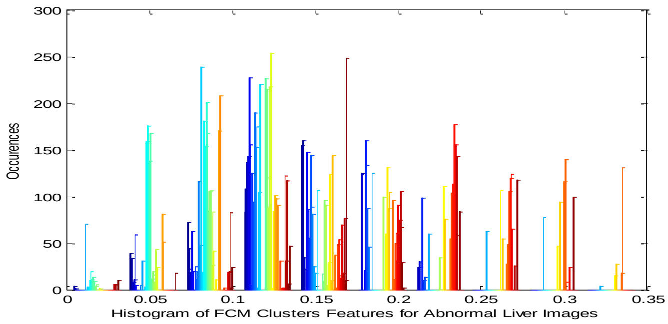

Fig 2.Histogram plot of FCM Clusters Features for Abnormal Liver Images

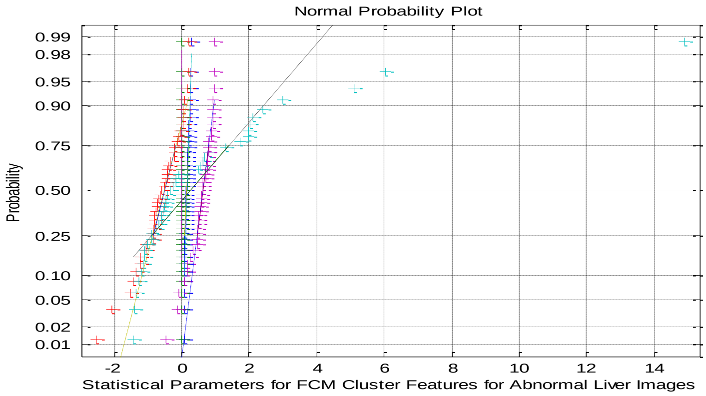

Fig 3.Norm plot of FCM Clusters Features for Abnormal Liver Images

1 *Performance of Classifiers in detection of normal and abnormal liver*

Expectation Maximization (EM)

The EM algorithm is a simple method to determine maximum likelihood approximations in the incidence of hidden variables and incomplete data problem.The EM algorithm has two steps, which are as follows.

1. Expectation Step (E step):The expected value is computed first, supposes the model have an estimate of the parameter and observed data for data y. The expected value of y_1 is calculated as follows for a given measurement z_1 and based on the present approximation of the parameter [7].

$$y_1^{[k+1]} = E[y_1 | z_1, p^k] \tag{4}$$

2. Maximization Step (M step): The likelihood function is optimized by assuming that the threshold values are considered as known parameters. The errors are minimized by combining E step and M step.

Gaussian Mixture model (GMM)

GMM is a simplest form of algorithm used for classification. The acquired feature vector is effectively modeled using a mixture of M Gaussians in GMM. Maximum likelihood model parameters can be calculated from a set of training feature vectors using an iterative expectation-maximization (EM) approach. Maximum likelihood modeling values are computed using an iterative expectation-maximization (EM) process for a given set of training vectors [8].The probability density function of GMM as given as

$$p(b)=\sum_{c=1}^{N}V_c\ p_c(b)=\sum_{c=1}^{N}V_c\ Q\left(b|\mu_c,\sum_c\right) \quad (5)$$

Linear Discriminant Analysis (LDA) as a classifier

Linear discriminant analysis is used for dimensionality reduction method as well as a reliable classification approach. To differentiate two or more classes LDA is used. Principal Coefficient Analysis (PCA) coefficients are placed on a unique LDA projection axis to generate and interpret the projected features more effectively. The scatter matrixes are computed with in the class or between the classes. The matrix represented between the classes is represented by variance.[9]

$$C_B=\frac{1}{f}\sum_{c=1}^{f}(p_k-p)(p_k-p)^T \quad (6)$$

The variance matrix within the class is

$$C_W=\sum_{k=1}^{f}\sum_{Z_l\in W_i}(Z_l-p_i)(Z_l-p_i)^T \quad (7)$$

Bayesian Linear Discriminant Analysis Classifier(BLDC)

The Bayesian rule is used for classification in BLDC to reduce the errors. Consider g is a feature vector with two classes m, n and the decision threshold is Y. The discriminant function is represented as

$$h_i(g)=\ln P(i|g) \quad (8)$$

According to Bays rule the covariance matrixes are similar for all the classes [10].

Logistic Regression (LR)

LR is one of the most widely used statistical modeling techniques. The regression model represents the probability of the ratio of the natural log of an event occurring to the possibility of the event not occurring. In Logistic Regression, there is no condition about the linearity of the relationship. The probability of the variable is represented by

$$\log it\,(P_1)=\ln\left(\frac{P_1}{1-P_1}\right)=\beta_0+\beta 1Z_1+...+\beta_q Z_q \quad (9)$$

Where β_0 represents interrupt and the $Z_1, Z_{2,....}Zq$ denotes the explanatory variables and the $\beta_{0,}\beta_{1,}\beta_{2,}...\beta_q$ defines the explanatory coefficients [9, 11].

3. RESULTS AND DISCUSSION

The performances of the classifiers are analyzed using the following metrics. As per binary classification, the output of a prediction for a specific class can be one of four options. They are True positive (TP-normal class exactly identified as normal) , True Negative (TN-abnormal class exactly identified as abnormal), False Positive (FP-abnormal class identified as normal), False Negative (FN-normal class exactly identified as abnormal).

Table 2.Classifiers Outputs for FCM Clustering Features

S.No	Classifiers	TP	TN	FP	FN	MSE
1	EM	24	24	18	18	8.51E-05
2	GMM	28	34	8	14	3E-05
3	LDA	29	35	7	13	2.22E-05
4	BLDC	28	31	11	14	3E-05
5	Logistic Regression	33	35	7	9	1.29E-05

The Mean Square Error (MSE) of FCM clustering based feature for the classifiers are shown in table 3. The stopping criterion for the training of the classifier is MSE. The MSE for all the classifiers are good except EM.

The database consists of 84 images, in that 80% of the images are used for training purpose and 20% of the images are used for testing. The target fixed for Abnormal class is 0.85 and for normal case 0.1.

Table 3. Performance Analysis of Classifiers for FCM Features

S.No	Parameters (%)	Classifiers				
		EM	GMM	LDA	BLDC	Logistic Regression
1	Sensitivity	57.14286	66.66667	69.04762	66.66667	78.57143
2	Specificity	57.14286	80.95238	83.33333	73.80952	83.33333
3	Accuracy	57.14286	73.80952	76.19048	70.2381	80.95238
4	Precision	57.14286	77.77778	80.55556	71.79487	82.5
5	Error Rate	42.85714	26.19048	23.80952	29.7619	19.04762
6	Mathew Correlation coefficient (MCC)	0.142857	0.481125	0.529238	0.405798	0.619751
7	Classifier Success Index (CSI)	14.28571	44.44444	49.60317	38.46154	61.07143

The performance of the classifiers based on the standard benchmark parameters are shown in Table 3. The Logistic Regression classifier achieved good accuracy of 80.95% with error rate of 19.04% and CSI of 61.07%. The LDA classifier is the next best performer, with an accuracy of 76.19% and an error rate of 23.80 % and 49.60% of CSI. The EM attained a least accuracy of 57.14 % among the 5 classifiers.

4. CONCLUSION

In this paper the abnormal liver image is classified using five different classifiers. Ultrasonic Liver images from the Signal Processing Lab Database are used in this analysis. The FCM clustering based features are utilized for classification. The classifier performances are investigated using standard parameters, in that the Logistic regression attain the average accuracy of 80.95%, whereas the average accuracy of EM, GMM,LDA,BLDC are 57.14 %,73.80%,76.19%,70.23% as respectively. The performance of the Logistic regression classifier is superior to that of other classifiers. The future research will be in the direction of other machine learning algorithms to detect the liver abnormalities.

5. REFERENCES

[1] Ozyurt, Fatih, T. Tuncer, E. Avci, M. Koç and I. Serhatlioglu. "A Novel Liver Image Classification Method Using Perceptual Hash-Based Convolutional Neural Network." Arabian Journal for Science and Engineering 44 (2019): 3173-3182.).

[2] Marina Galicia-Moreno, Jorge A Silva-Gomez, Silvia Lucano-Landeros, Arturo Santos, Hugo C Monroy-Ramirez, Juan Armendariz-Borunda, "Liver Cancer: Therapeutic Challenges and the Importance of Experimental Models", Canadian Journal of Gastroenterology and Hepatology, vol. 2021, Article ID 8837811, 10 pages, 2021. https://doi.org/10.1155/2021/8837811

[3] Ahmed S.Bafaraj, "Performance Analysis of Best Speckle Filters for Noise Reduction in Ultrasound Medical Images "International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 6 (2019)

[4] R. Xu, D. Wunsch II, ―Survey of clustering algorithms‖, IEEE Transcations on Neural Networks, 16, 2005, 645-678

[5] A. Jain, M. Murty, P. Flynn, ―Data Clustering: A review‖, ACM Computing Surveys, 31, 1999, 264-323.

[6] Mahipal Singh Choudhry, Rajiv Kapoor, Performance Analysis of Fuzzy C-Means Clustering Methods for MRI Image Segmentation, Procedia Computer Science,Volume 89, 2016, Pages 749-758, ISSN 1877-0509,https://doi.org/10.1016/j.procs.2016.06.052.

[7] HarikumarRajaguru, VijayakumarThangavel, Wavelets and Morphological Operators Based Classification of Epilepsy Risk Levels, Mathematical Problems in Engineering Volume 2014, Article ID 813197, 13 pages http://dx.doi.org/10.1155/2014/813197 Publication Year 2014

[8] F. Ghasemian and M. M. Homayounpour, "Towards better GMM-based acoustic modeling for spoken language identification," 2011 19th Iranian Conference on Electrical Engineering, 2011, pp. 1-1.

[9] S. K. Prabhakar, H. Rajaguru and S. Lee, "A Comprehensive Analysis of Alcoholic EEG Signals with Detrend Fluctuation Analysis and Post Classifiers," 2019 7th International Winter Conference on Brain-Computer Interface (BCI), 2019, pp. 1-6, doi: 10.1109/IWW-BCI.2019.8737328.

[10] HarikumarRajaguru, Sunil Kumar Prabhakar, Bayesian Linear Discriminant Analysis with Hybrid ABC-PSO Classifier for Classifying Epilepsy from EEG Signals, Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication 2017.

[11] O. Gevaert, F. De Smet, E. Kirk et al., “Predicting the outcome of pregnancies of unknown location: Bayesian networks with expert prior information compared to logistic regression,” Human Reproduction, vol. 21, no. 7, pp. 1824–1831, 2006

[12] K. Yasoda, R. Ponmagal, K. Bhuvaneshwari, and K. Venkatachalam, "Automatic detection and classification of EEG artifacts using fuzzy kernel SVM and wavelet ICA (WICA)," *Soft Computing*, vol. 24, no. 21, pp. 16011-16019, 2020.

[13] C. Viji, N. Rajkumar, S. Suganthi, K. Venkatachalam, and S. Pandiyan, "An improved approach for automatic spine canal segmentation using probabilistic boosting tree (PBT) with fuzzy support vector machine," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-10, 2020.