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

Analysis Of K Means Clustering And Classifiers In Diagnosing Abnormality Of The Ultrasonic Liver Images

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

[1]*Assistant Professor, Electronics And Instrumentation Engineering, Bannari Amman Institute Of Technology, Sathyamangalam, India*
[2]*Professor, Department Of Electronics And Communication Engineering, Bannari Amman Institute Of Technology, Sathyamangalam, India*

Karthikamani@Bitsathy.Ac.In[1], Harikumarr@Bitsathy.Ac.In[2]

Abstract.This Paper Investigates The Diagnosis Of Liver Abnormalities From Ultrasonic Images Using K Means Clustering Algorithm And Five Classifiers. Nowadays, Liver Cancer Is One Of The Most Serious Health Problems. Medical Imaging Is Powerful Tool For Diagnosing The Abnormalities In The Earlier Stage. The Features Are Extracted Using K Means Clustering And Principal Component Analysis (PCA),Expectation Maximization (EM),EM PCA, Kernel PCA,Gaussian Mixture Model (GMM) Classifier Are Used To Detect The Liver Image As Normal Or Abnormal. The Parameters Such As Sensitivity, Specificity, Accuracy, Precision, Error Rate, Mathew Correlation Coefficient (MCC), And Classifier Success Index (CSI) Are Analyzed And Compared. When Compared With All The Classifier The Logistic Regression Attained A Higher Accuracy Of 80.95%.

Keywords:PCA, GMM, EM, CAD

.

1. INTRODUCTION

Computer Aided Diagnosis (CAD) Is Commonly Utilized In The Healthcare Industry Due To Advancements In Computer Technology And Pattern Reorganization. Various Imaging Modalities Are Used For Disease Diagnosis But In Many Predictive And Therapeutic Applications Ultrasound Is Commonly Preferred Because It Is Cost Effective, More Efficient And Does Not Produce Any Radiation Effect [1]. The Detection Of Liver Abnormality Is One Of The Applications Of Ultrasonic Imaging Techniques. Liver Cirrhosis Is The 4 Thforemostcause Of Cancer Associated Demise Globally. The World Health Organization (WHO) Almost One Million Individual Die From It By 2030. Therefore Earlier Diagnosis Is Important For Proper Treatment And Cure Liver Cancer. Clinician's Interpretations Of Medical Images Only Bring 75 % Of Accuracy, But Advances In Computer Vision And Artificial Intelligence Techniques Contributed To The Development Of A CAD System Provide Better Diagnosis Of The Disease [2]. Figure 1 Depicts The Work Flow Of The CAD System For Diagnosis Of Abnormal Liver Ultrasonic Images From The Dataset.The Ultrasonic Images From Signal Processing Lab Database Were Used For Analysis. The Database Comprises Of 42 Normal And 42 Abnormal Liver Images. The Collected Images Were Preprocessed In Next Stage; Then K Means Clustering Based Features Are Extracted. The Detection Of Liver

Abnormality Is Detected Using Five Classifiers. Analysis Was Done With The Use Of Standard Bench Mark Parameter

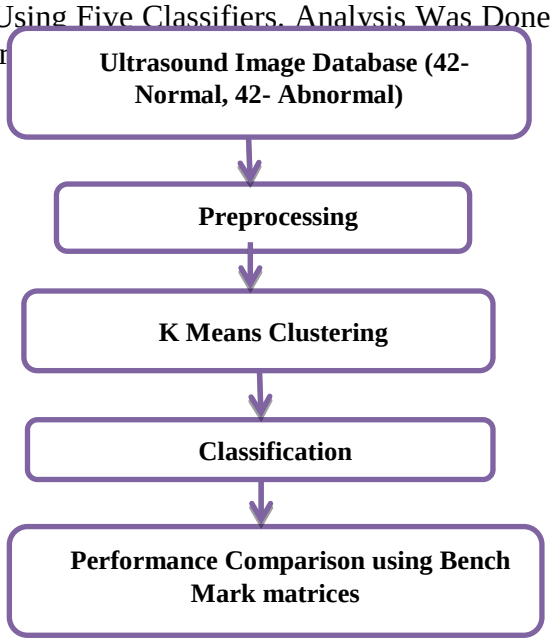

Fig. 1.Work Flow Of The Cad Model

2. MATERIALS AND METHODS

The Dataset Consists Of Normal And Abnormal Liver Images. The Images Are Affected By Noise While Capturing The Images. The Noises Are Reduced Using Filters In Preprocessing. The Useful Information Of The Images Is Extracted Using K Means Clustering Algorithm. Mean, Variance, Skewness, Kurtosis, Pearson Correlation Coefficient Are The Features Used In This Work. These Features Are Inputted To The Classifier For Abnormality Detection.

Preprocessing

Unwanted Signals Have An Impact On Ultrasonic Images. Speckle Noise Is The Most Common Type Of Noise That Affects Ultrasonic Images [3]As A Result; The Proper Filters Are Used To Eliminate The Noise. In This Work, The Noise Will Be Removedusing Fuzzy Median Filter.

K Means Clustering

K-Means Is A Distance-Based Or Clustering Approach That Uses The Distances Between Points To Assign A Point To A Cluster. Every Cluster In K-Means Is Linked With A Centroid. The K-Means Clustering Technique Calculates The Centroids And Iterates Until The Optimum Centroid Is Discovered. Mac Queen In 1967 Introduced This Algorithm And It Comes Under Portioning Technique [4]. The Pseudo Code For K Means Clustering Process Is As Follows

1. Initialize Cluster Centroids
2. Based On Euclidian Distance Each y_i Is Assigned To Nearest Cluster Centroid c_k

$$KM\,(Y,C)=\sum_{i=1}^{m}\left\|Y_i-C_j\right\|^2,\ j\in\{1,2,...k\} \qquad (1)$$

$$d(p,q)=d(q,p)=\sqrt{\sum_{i=1}^{m}\left(q_i-p_i\right)^2} \qquad (2)$$

3. Make Every Cluster Centerc_kEqual To The Average Of Entirelyy_iIn The Cluster.
4. Steps 2-4 Should Be Continual Till The Cluster Centers Are Constant.
Here K Denotes Number Of Clusters And Number Of Cases Given By M. K-Means Is An Efficient Method. But The Number Of Clusters Must Be Specified Beforehand, And The End Results Are Dependent To Initialization, Typically Ending At A Local Optimum [5].The Mean, Variance, Skewness, Kurtosis, Pearson Correlation Coefficient (PCC) Is Shown In Table 1 And These Are Statistical Parameters. For K Means Clustering Based Features Thecanonical Correlation Coefficient Is 0.765648, Which Infers That The Features For Normal And Abnormal Classes Are Highly Correlated And They Are Nonlinear.

Table 1 Statistical Parameters For K Means Cluster Features In Abnormal And Normal Liver Images

Sl.No	Statistical Parameters	Abnormal	Normal
1	Mean	0.118654	0.122705
2	Variance	0.000568	0.000644
3	Skewness	-0.75284	-0.82041
4	Kurtosis	1.086325	1.463358
5	Pearson Correlation Coefficient	0.524039	0.45945
6	CCA	**0.765648**	

The Scatter Plot And Histogram Plot Are Used To Analyze The Characteristics K Means Cluster Features. Figure 2 Shows The Scatter Plot For K Means Cluster Feature Of Normal And Abnormal Classes And It Indicates The Overlapping Nature Of K Means Features Among The Two Classes.Therefore The Features For Normal And Abnormal Classes Are Non-Separable And Figure 3 Indicates The Non-Linearity Of The Extracted Features.

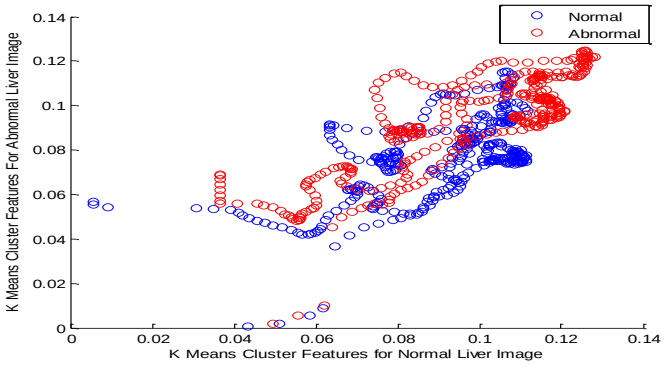

Fig 2.Scatter Plot For K Means Cluster Features For Normal And Abnormal Liver Images

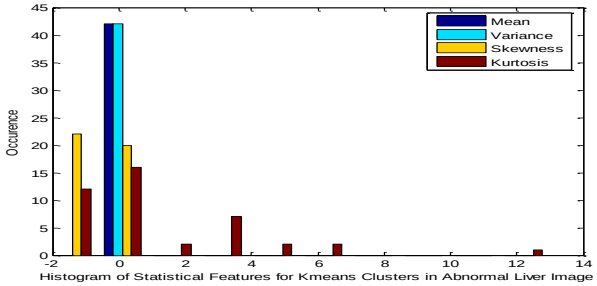

Fig .3.Histogram Of K Means Clusters Features For Abnormal Liver Images

1 Implementation Of Liver Classification Using Classifiers

1.1 Principal Component Analysis(PCA)

PCA Is A Statistical Method For Transforming Highly Correlated Features Into A Set Of Linearly Non - Overlapping Features Known As Principle Components, Which Is Based On Orthogonal Transformation. PCA Is Works On The Basis Of Karhunen-Loeve Transform [6].PCA Also Gives A Strategy For Converting A Complicated Data Set To A Lower-Dimensional Sample In Order To Discover Hidden And Simpler Dynamics Within It.PCA Always Makes A Strong Assumption (I.E., It Must Adopt The Linearity Principle).PCA Can Now Be Limited To Describing Data As A Linear Combination Of Its Basis Vectors Based On This Assumption. It Can Be Represented As $\boldsymbol{X}$ And $(m\times n)$ Matric Represented As Y . The Bias Change Is Given As

$$PX = Y \tag{3}$$

Here P Denotes The Linear Representation [7].

1.2 Expectation Maximization (EM)

The EM Algorithm Is A Repetitive Method That Alternates Between Two Modes. The Estimation-Step Or E-Step Is The First Phase, Which Tries To Estimate Missing Or Latent Variables. The Second Phase, Known As The Maximization-Step Or M-Step, Focuses On Improving The Model's Parameters To Best Quantify The Problem.

1. E-Step. Determine The Dataset's Missing Variables.
2. M-Step. In The Presence Of Data, Enhance The Model's Variables.

$$f_i\left(x/\theta_i\right)=\frac{1}{\left(2\pi\right)^{d/2}\left|\sum_i\right|^{1/2}}\exp\left[-\frac{1}{2}\left(x=\mu_i\right)^i\sum_i^{-1}\left(x-\mu_i\right)\right] \tag{4}$$

Where $\theta_i=\left(\mu_i,\sum_i\right)$.X Is A D States That Dimensional Feature Vector. The Mean Vector States μ_i And $\sum_i$, $\left|\sum_i\right|$, $\sum_i^{-1}$ Are Stated As D-By-D Covariance Matrix, Its Element And Transposed Correspondingly [8].

1.3 EMPCA

PCA Has Some Constraints. Training The Principle Components With A Huge Number Of Data Sets Or High Dimensional Data Is Difficult; Hence The EM Algorithm Is Employed In Conjunction With PCA [9]. The EM Learning Technique Used In PCA Is A Periodic Iterative Technique For Determining The Subspace Covered And Controlled By The I Leading Eigen Vectors.

1.4 Kernel PCA

On The Basis Of PCA, The Kernel Principal Component Analysis Approach Is Developed.PCA Is Used For Linear Datasets, If We Apply It To Non-Linear Datasets, The Outcomes Will Not Be Efficient Because It Is Developed For Linearly Separable Data. The Non-linear Data Is Transformed Into Linear Dataset In Kernel PCA. The Covariance Matric In Feature Space Can Be Written As

$$C_f=\frac{1}{N}\sum_{i=1}^{N}\phi(x_i)\phi(x_i)^T \tag{5}$$

The Principal Component Is Obtained By Solving The Below Equation

$$C_f\, v = \lambda v \qquad (6)$$

C_f Is The Covariance Matric In Feature Space, Λ Is Eigen Value And v Represent Eigen Vector. The $N \times N$ Kernel Can Be Created Using

$$K = k(x, y) = \Phi(x)^T\, \Phi(y) \qquad (7)$$

The Kernel PCA Is Mainly Used For Classification Problem [10].

1.5 Gaussian Mixture Model (GMM)

The GMM Is Initialized Using Probability Density Function As Given As

$$P(b) = \sum_{c=1}^{N} v_c\; p_c(b) = \sum_{c=1}^{N} v_c\, Q\left(b | \mu_c, \sum_c \right) \qquad (8)$$

Where N Represents Fixed Number Of A Model, The Weighting Coefficient Is Denoted v_c . The Parameters Are Evaluated In Order To Obtain The Probability Density Function's Maximum Likelihood Estimate. The Maximum Expected Valued Approach Is Used To Analyze GMM.

3. RESULTS AND DISCUSSION

Table 3 Depicts The Mean Square Error (MSE) Of K Means Clustering Based Feature For The Classifiers. The Mean Square Error Is The Stopping Criteria For The Classifiers. The MSE Represented As

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (O_i - P_i)^2 \qquad (9)$$

Where O_i Represents The Observed Values And The Predicted Value Is Represented By P_i And N Denotes The Total Number Of Images. In This Analysis 84 Images Are Utilized. The Classifiers Are Trained By 80% Of The Images And 20% Of The Images Are Used For Testing Purpose. The Target For Normal Class Is 0.1 And Abnormal Class It Is 0.85.

Table 2 .Classifiers Outputs For K Means Clustering Features

S.No	Classifiers	TP	TN	FP	FN	MSE
1	PCA	26	24	18	16	6.51E-05
2	EM	32	27	15	10	3.56E-05
3	EMPCA	24	30	12	18	5.22E-05
4	Kernel PCA	29	27	15	13	4.3E-05
5	GMM	**36**	**32**	**10**	**6**	**1.17E-05**

The K Means Clustering Features Are Extracted And It Is Classified With Principal Component Analysis, Expectation Maximization, EM PCA, Kernel PCA, GMM And Standard Parameters Like Sensitivity. Specificity, Accuracy, Precision, Error Rate, Mathew Correlation Coefficient(MCC), Classifier Success Index Are Estimated, Analyzed And The Results Are Depicted In Table 3

Table 3. Performance Analysis Of Classifiers For K Means Features

S.No	Parameters (%)	Classifiers				
		PCA	EM	EM PCA	Kernel PCA	GMM
1	Sensitivity	61.90476	76.19048	57.14286	69.04762	85.71429
2	Specificity	57.14286	64.28571	71.42857	64.28571	76.19048
3	Accuracy	59.52381	70.2381	64.28571	66.66667	80.95238
4	Precision	59.09091	68.08511	66.66667	65.90909	78.26087
5	Error Rate	40.47619	29.7619	35.71429	33.33333	19.04762
6	Mathew Correlation Coefficient (MCC)	0.190693	0.407661	0.288675	0.333712	0.621874
7	Classifier Success Index (CSI)	20.99567	44.27558	23.80952	34.95671	63.97516

The Performance Of The Classifiers Based On The Standard Benchmark Parameters Are Shown In Table 3. The GMM Classifier Achieved Good Accuracy Of 80.9523% With Error Rate Of 19.04% And CSI Of 63.97%. The EM Classifier Is The Next Best Performer, With An Accuracy Of 70.23% And An Error Rate Of 29.76 % And 44.27 % Of CSI. The PCA Reached A Least Accuracy Of 59.52 % Among The Five Classifiers.

4. CONCLUSION

In This Work Five Classifiers Are Employed To Classify The Abnormalities In The Liver Ultrasonic Images. The Evaluation Is Performed Using Ultrasonic Images From Signal Processing Labdatabase. The K Means Clustering Features Are Extracted And Given As Input Of The Classifier. The Standard Parameters Are Used For The Analysis And Compare The Classifier Performance, In That The GMM Attain The Average Accuracy Of 80.95%, Whereas The Average Accuracy Of PCA, EM,EM PCA,Kernel PCA Are 59.52 %,70.23%, 64.28 %, 66.66 % Respectively. The Performance Of The GMM Classifier Is Superior To That Of Other Classifiers. In Future Work The Bio Inspired Classifiers With Different Feature Extraction Techniques Will Be Used To Detect The Liver Abnormalities Using Same Liver Image Database.

5. REFERENCES

[1] Alalehalivar, Habibollahdaniali, Mohammad Sadeghhelfroush,' Classification Of Liver Diseases Using Ultrasound Images Based On Feature Combination'ieee Conference 2014.

[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. Harikumar, B. Vinothkumar , G. Karthick, Performance Analysis For Quality Measures Using K Means Clustering And EM Models In Segmentation Of Medical Images,International Journal Of Soft Computing And Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-6, January 2012.

[5] Stephen J Redmond And Conorheneghan,‖ A Method For Initializing The K-Means Clustering Algorithm Using Kd-Trees‖, Pattern Recognition 28(2007), 965-973.

[6] A. Rehman, A. Khan, M. A. Ali, M. U. Khan, S. U. Khan And L. Ali, "Performance Analysis Of PCA, Sparse PCA, Kernel PCA And Incremental PCA Algorithms For Heart Failure Prediction," 2020 International Conferenceon Electrical, Communication, And Computer Engineering (ICECCE), 2020, Pp. 1-5, Doi: 10.1109/ICECCE49384.2020.9179199

[7] S. K. Prabhakar And H. Rajaguru, "Morphological Operator Based Feature Extraction Technique Along With Suitable Post Classifiers For Epilepsy Risk Level Classification," 2015 International Conference On Intelligent Informatics And Biomedical Sciences (ICIIBMS), 2015, Pp. 446-451, Doi: 10.1109/ICIIBMS.2015.7439471.

[8] 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

[9] H. Rajaguru And S. K. Prabhakar, "Multilayer Autoencoders And EM - PCA With Genetic Algorithm For Epilepsy Classification From EEG," 2018 Second International Conference On Electronics, Communication And Aerospace Technology (ICECA), 2018, Pp. 353-358, Doi: 10.1109/ICECA.2018.8474658.

[10] C. Liu, T. Zhang, D. Ding And C. Lv, "Design And Application Of Compound Kernel-PCA Algorithm In Face Recognition," 2016 35th Chinese Control Conference (CCC), 2016, Pp. 4122-4126, Doi: 10.1109/Chicc.2016.7553997.