International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 16 (2018) pp. 12764-12769

Medical Image Classification Using Hierarchical Clustering and Generalized Gamma Distribution for Effective Identification of Diseases in Brain

K.Srinivas [1]*, P.V.G.D Prasad Reddy[2], GPS Varma[3]

[1]*Research Scholar, Department of CS&SE, Andhra University, Visakhapatnam, Andhra Pradesh, India.*
[2]*Prof & HOD, Department of CS&SE, Andhra University, Visakhapatnam, Andhra Pradesh, India.*
[3]*Principal, SRKR Engineering College, Bhimavaram, Andhra Pradesh, India.*

Abstract

Medical imaging is a specialized stem of image processing, which deals largely with the study of medical related data and the analysis of this data for identifying the deformities within the data. Many segmentation and classification techniques are available in the literature. However, there are no concrete algorithms and methods are available to identify the diseases related for effective classification of brain related data. Most of the brain related is highlighted of late leading towards mortality. Therefore for effective recognition and interpretation of the brain related data, a new model is proposed in this article based on Generalized Gama Distribution. The performance evaluation carried out by using metrics like Average Difference, Maximum Distance, Image Fidelity; Mean squared Error and Signal to Noise Ratio showcase that the developed model discloses an accuracy rate of above 90 percent in most of the cases.

Keywords: Medical image examination, Acoustic Neuroma, Parkinson's diseases, Performance evaluation metrics, AD, IF, MSE, PSNR

INTRODUCTION

Image Processing attributes towards the techniques and models used for the development of image analysis. One of the most popularized stems of image processing includes medical image processing. Here the main objective of medical image processing is to consider a input MRI image process it , enhance it and later cluster it to derive meaningful associated patterns from these clusters. Many medical imaging techniques are coined in the literature and most of them are either structured or unstructured [1], [2], [3], [4]. To analyze the medical data one need to convert this unstructured data into structured such that the interpretation and analysis are feasible. To convert this unstructured data, many clustering algorithms are highlighted in the literature, viz., k-Means algorithm, Hierarchical clustering (HC), Fuzzy-C Means Clustering, Rough set based clustering, Fuzzy Rough Sets and Vague set based approaches [5], [6], [7], [8], [9], [10]. Among these approaches most of the works are reported on k-Means clustering with the very assumption that it gives appropriate clustering results when compared to the other clustering algorithms[15]. With this approach most of the literature is driven. However, the k-Means suffers with disadvantage that it cannot identify exactly the data at the exterior points and hence it leads to out layers also another disadvantages with k-means algorithms is that if the value of k is not interpreted exactly, it results into either over segmentation or under segmentation and in some particular cases it results into spherical clusters accounting to the number of clusters to ONE. However when dealing with applications of medical data all the data points become vital in identifying the disease and hence no pixel or data point should be left alone. With this consideration, in this article we exhibit a model based on hierarchical clustering such that the above said disadvantages can be overruled. However, Hierarchical Clustering formulates a dendrogram which enables us to underline the number of hidden clusters. Using this approach, more beneficial results are subjected in case of medical analysis.

The challenges associated with the medical data, in particular the brain related data is that the algorithms used should able to penetrate into the tissues of the brain viz., white matter, grey matter and cerebra spinal fluid and in most of the cases the pixels exposed to the diseases may be covered under these tissues and hence effective classification is needed for better analysis of the data. The medical data consists of data extending towards both the boundaries, thereby formulating thin lines generally called as the tails. These tails carry some information and hence to model or extract the information in the tails elongated towards either ends one need to consider statistical models that can interpret the inherent details within the data of the tails. To extract this data, among the statistical models Generalized Gama Distributions (GGD) is mostly preferred. The main advantage behind the choice of GGD is that it formulated various shapes of models, each shape resembling the pattern of a particular distribution and it includes several other distributions such as Gama, Laplace, Raleigh, Gaussian, Erlang Distributions as particular cases. Therefore in this article a methodology for identification and classification of brain related data is presented using the concepts of HC and GGD. The rest of the article is articulated as follows section-2 of the paper deals with the GGM and its Probability Density Function, this section also highlights the updated parameters of the proposed model, Section-3 highlights the clustering methodology based on Hierarchical clustering algorithm, the Dataset considered is presented in section 4, the section 5 of the paper highlights the experimentation carried out and in the section 6, the performance of the model is carried out using evaluation metrics like Average Difference, Maximum distance, Image Fedility, Mean Squared Error and Signal to

Noise Ratio, and the concluding section 7, summarizes the article.

GENERALIZED GAMMA DISTRIBUTION

Every image is a collection of several image regions. In each image region, the image data is quantized by pixel, which is a random variable because of the fact it is influenced by random factors like Vision, brightness, contrast etc. To model the pixel intensities in a image region, it is necessary to assume that the pixels in each image region follow a Generalized Gamma Distribution.

The probability density function of generalized gamma distribution is given by

$$f(x,k,c,a,b) = \frac{c(x-a)^{ck-1}e^{-\left(\frac{x-a}{b}\right)^c}}{b^{ck}\Gamma(k)} \quad \text{--2.2.1}$$

Where a, b, c, k are called the gamma variants and c, k are called shape parameters such that c, k >0.

a is called location parameter, b is called shape parameter with a ,b >0

HIERARCHICAL CLUSTERING ALGORITHM

Hierarchical clustering segments the data when the segments are dense and segmentation is to performed on random data sets. The algorithm for utilizing hierarchical clustering is given below:

Given a set of N items to be segmented and an M × N distance (or similarity) matrix, the basic process of hierarchical segmenting is as follows.

Step 1: First, assign each item to a segment, so that if we have N items, it implies that we have N segments, each containing just one item. Let the distances (similarities) between the segments be the same as those (similarities) between the items they contain.

Step 2: Find the closest (most similar) pair of segments and merge them into a single segment, i.e. we will now have one segment less.

Step 3: Compute distances (similarities) between the new segment and each of the old segments.

Step 4: Repeat steps 2 and 3 until all items are segmented into a single segment of size N.

Step 3 can be done using single-linkage, average-linkage or complete-linkage method. In single-linkage, segmenting (also called the connectedness or minimum method), we consider the distance between one segment and another to be equal to the shortest distance from any member of one segment to any member of the other segment. If the data consist of similarities, we consider the similarity between one segment and another to be equal to the greatest similarity from any member of one segment to any member of the other segment. The M × N proximity matrix is D = [d(i, j)]. The segmenting is assigned sequence numbers 0, 1..., (n – 1) and L(k) is the level of the k^{th} segmenting. A segment with sequence number m is denoted as (m) and the proximity between segments I and (s) is denoted as d [I, (s)]. The algorithm is composed of the following steps:

Algorithm (Single Linkage)

Step 1: Begin with the disjoint clustering having level L(0) = 0 and sequence number m=0.

Step 2: Find the least dissimilar pair of clusters in the current clustering, say pair I, (s), according to d[I,(s)] = min d[(i),(j)]. Where, the minimum is over all pairs of clusters in the current clustering.

Step 3: Increment the sequence number: m = m +1. Merge clusters I and (s) into a single cluster to form the next clustering m. Set the level of this clustering to L(m) = d[I,(s)]

Step 4: Update the proximity matrix, D, by deleting the rows and columns corresponding to clusters I and (s) and adding a row and column corresponding to the newly formed cluster. The proximity between the new cluster, denoted (r,s) and old cluster (k) is defined in this way:

d[(k), (r,s)] = min d[(k),I], d[(k),(s)]

Step 5: If all objects are in one cluster, stop. Else, go to step 2.

In our thesis, we have utilized hierarchical clustering algorithm using single linkage algorithm because of the fact that the single linkage algorithm uses the minimum of all pair wise distances between points in the two clusters which helps in identifying the region of interest (deformity) more accurately.

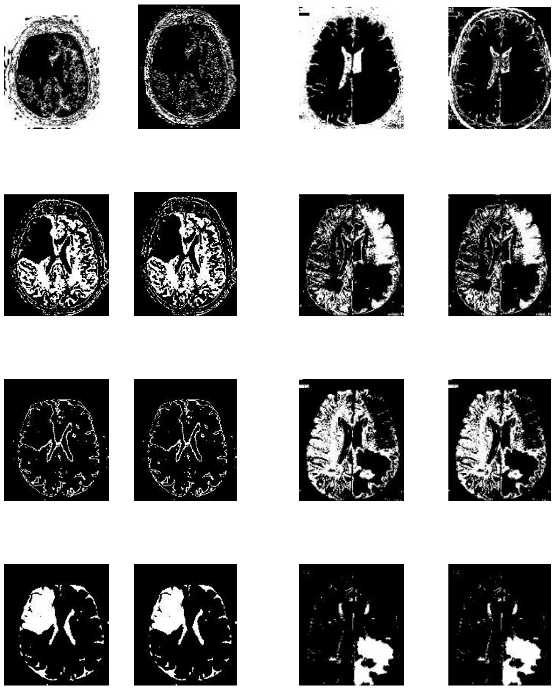
DATA SET CONSIDERED

In order to propose the present model; we have considered a Dataset obtained from, Brain Web Data. This database consists of several images pertaining to several images with several diseases. Every image consists of fixed sizes each of size 150 x 150.

EXPERIMENTATION

Each image is pre-processed such that it is free from noise, each proceed image is given as input to the K-means Algorithm to segment the image. The segmented image is given as input to the segmentation algorithm, presented in section 3.1.

Using the probability density function of Generalized Gamma Distribution given in section 2. The image retrieval process is carried out by using the inverse transformation method and the performance of the model is performed using subjective image quality testing by comparing the original and retrieved image. The original and the reconstructed images obtained by using the image retrieval process are shown in figure-2.1.



(a) (b) (c) (d)

Figure 2.1: (a) Input images of B0 (c) Input images of B1
(b) Output Images of B0 (d) Output Images of B1

PERFORMANCE EVALUATION

To evaluate the proposed methodology, we have considered the following metrics and the formulas for the calculation of each of the metrics is presented in the following table-1

The methodology is tested against the metrics and compared with that of the models presented in the literature, GMM and the results show case that the developed model performs better that the existing algorithms and the results derived are showcased in the following table-1.

Table 1: Performance Evaluation Metrics

Quality metric	**Formula to Evaluate**
Average Difference (AD)	$\sum_{j=1}^{M}\sum_{k=1}^{N}[F(j,k) - \hat{F}(j,k)] / MN$ Where M,N are image matrix rows and columns
Maximum Distance (MD)	Max{$\mid F(j,k) - \hat{F}(j,k) \mid$}
Image Fidelity (IF)	$1 - \left[\sum_{j=1}^{M}\sum_{k=1}^{N} [F(j,k) - \hat{F}(j,k)]^2 / \sum_{j=1}^{M}\sum_{k=1}^{N} [F(j,k))]^2\right]$ Where M,N are image matrix rows and columns
Mean Squared error (MSE)	$\frac{1}{MN}\sum_{j=1}^{M}\sum_{k=1}^{N} [O\{F(j,k)\} - O\{\hat{F}(j,k)\}]^2 / \sum_{j=1}^{M}\sum_{k=1}^{N}[O\{F(j,k)^2$ Where M,N are image matrix rows and columns
Signal to noise ratio (SNR)	$20.\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$ Where, MAX_I is maximum possible pixel value of image, MSE is the Mean squared error

Table 2: Performance Evaluation of the proposed model

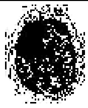
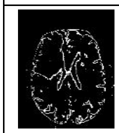
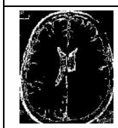
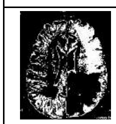
Image	Quality Metric	GMM	GGD with K-Means	GGD hierarchical clustering	Standard Limits	Standard Criteria
	Average Difference	0.573	0.773	0.812	-1 to 1	Closer to 1
	Maximum Distance	0.422	0.922	0.9325	-1 to 1	Closer to 1
	Image Fidelity	0.416	0.875	0.923	0 to 1	Closer to 1
	Mean Squared error	0.04	0.134	0.094	0 to 1	Closer to 0
	Signal to Noise ratio	17.41	29.23	33.89	-∞ to ∞	As big Possible
	Average Difference	0.37	0.876	0.749	-1 to 1	Closer to 1
	Maximum Distance	0.221	0.897	0.912	-1 to 1	Closer to 1
	Image Fidelity	0.336	0.876	0.859	0 to 1	Closer to 1
	Mean Squared error	0.2404	0.211	0.2019	0 to 1	Closer to 0
	Signal to Noise ratio	14.45	35.65	39.85	-∞ to ∞	As big Possible
	Average Difference	0.456	0.76	0.81	-1 to 1	Closer to 1
	Maximum Distance	0.345	0.879	0.807	-1 to 1	Closer to 1
	Image Fidelity	0.44	0.86	0.917	0 to 1	Closer to 1
	Mean Squared error	0.22	0.23	0.2123	0 to 1	Closer to 0
	Signal to Noise ratio	19.88	37.98	39.71	-∞ to ∞	As big Possible
	Average Difference	0.231	0.473	0.4991	-1 to 1	Closer to 1
	Maximum Distance	0.224	0.977	0.971	-1 to 1	Closer to 1
	Image Fidelity	0.212	0.813	0.892	0 to 1	Closer to 1
	Mean Squared error	0.24	0.121	0.1192	0 to 1	Closer to 0
	Signal to Noise ratio	21.42	33.28	37.41	-∞ to ∞	As big Possible
	Average Difference	0.342	0.764	0.7015	-1 to 1	Closer to 1
	Maximum Distance	0.317	0.819	0.854	-1 to 1	Closer to 1
	Image Fidelity	0.391	0.812	0.876	0 to 1	Closer to 1
	Mean Squared error	0.2514	0.228	0.1759	0 to 1	Closer to 0
	Signal to Noise ratio	3.241	5.514	5.68	-∞ to ∞	As big Possible
	Average Difference	0.21	0.3653	0.232	-1 to 1	Closer to 1
	Maximum Distance	0.21	0.892	0.912	-1 to 1	Closer to 1
	Image Fidelity	0.2134	0.787	0.791	0 to 1	Closer to 1
	Mean Squared error	0.06	0.145	0.594	0 to 1	Closer to 0
	Signal to Noise ratio	13.43	49.22	20.39	-∞ to ∞	As big Possible
	Average Difference	0.3232	0.322	0.4592	-1 to 1	Closer to 1
	Maximum Distance	0.123	0.212	0.456	-1 to 1	Closer to 1
	Image Fidelity	0.233	0.897	0.923	0 to 1	Closer to 1
	Mean Squared error	0.01	0.4345	0.119	0 to 1	Closer to 0
	Signal to Noise ratio	11.11	27.267	29.86	-∞ to ∞	As big Possible

Image	Quality Metric	GMM	GGD with K-Means	GGD hierarchical clustering	Standard Limits	Standard Criteria
	Average Difference	0.314	0.338	0.497	-1 to 1	Closer to 1
	Maximum Distance	0.241	*0.249*	0.317	-1 to 1	Closer to 1
	Image Fidelity	*0.293*	*0.683*	0.791	0 to 1	Closer to 1
	Mean Squared error	*0.18*	*0.197*	0.213	0 to 1	Closer to 0
	Signal to Noise ratio	*21.214*	*78.19*	99	-∞ to ∞	As big Possible

From the above table-2, it can be observed that the MSE of the developed model is very less in comparison with the existing algorithm. The MSE values of the proposed model is approaching towards 0, which clearly shows that the error is minimal, since the error is less, implies that the retrieved image is more in accordance with that of the original image. The SNR is high in case of the existing model, which clearly shows that the retrieval signal. The same is in case with the other metrics.

CONCLUSION

In this article, a methodology based on GGM is proposed for analysizing the medical data. The methodology is applied on to the Brain web images, obtained from the web data. This method is tested against the existing model based on GMM and the results are tested against the metrics like Average Difference, Maximum Distance, Image Fedility, Mean Squared Error and Signal to Noise Ratio. From the above results presented in Table-2, it clearly showcased that the developed model outperforms than that of the existing model based on GMM, from the results it can be seen that MSE is very less in case of the developed model, which specifies that it is more acceptable than that of the existing model, the other metrics like IF, SNR, AD, MD showcase better results to the developed model when compared to that of the existing model. The overall efficiency is around 92% recognition rate.

REFERENCES

[1] Annemie Ribbons, Jeroen Hermans, Frederik Maes, Dirk Vandermeulen, and Paul Suetens, "Unsupervised Segmentation, Clustering, and Groupwise Registration of Heterogeneous Populations of Brain MR Images," IEEE Transaction on Medical Imaging., vol. 33, no. 2, pp. 201-224, 2014.

[2] A. Montanvert et al, "Hierarchical Image Analysis Using Irregular Tessellations", Transactions On Pattern Analysis and Machine Learning, Vol. 13, No.4, April-1991

[3] A. Montanvert et al, "Hierarchical Image Analysis Using Irregular Tessellations", Transactions On Pattern Analysis and Machine Learning, Vol. 13, No.4, April-1991

[4] A. Ortiz, J.M.Gorriz, J.Ramirez, and D. Salas-Gonzalez, "Unsupervised Neural Techniques Applied to MR Brain Image Segmentation," Hindawi Publishing Corporation on Advances in Artificial Neural Systems, vol. 2012 Article ID 457590 pp. 1-7, 2012.

[5] A. Ortiz, J.M.Gorriz, J.Ramirez, and D. Salas-Gonzalez, J.M. Llamas-Elvira, "Two fully-unsupervised mehods for MR brain image segmentation using SOM-based strategies," ELSEVIER on Applied Soft Computing 13 (2013) 2668-2682.

[6] Nagesh Vadaparthi, Srinivas Yarramalle, Suresh Varma Penumatsa, P.S.R.Murthy, "Segmentation of Brain MR Images based on Finite Skew Gaussian Mixture Model with Fuzzy C-Means Clustering and EM Algorithm," International Journal of Computer Applications, vol. 28, no. 10, pp. 18-26, August 2 011.

[7] Nagesh Vadaparthi, Srinivas Yarramalle, Suresh Varma.P, "Unsupervised medical Image Segmentation on Brain MRI images using Skew Gaussian Distribution", IEEE-International Conference on Recent Trends in Information Technology, 2011, pp.1293-1297.

[8] Nikos Vlassis and Theo Gevers, "A Spatially Constrained Generative Model and EM Algorithm for Image Segmentation", IEEE Transactions on Neural Networks, Vol:18, No. 3, May 2007.

[9] Prasad Reddy P.V.G.D, Srinivas Rao. K, Srinivas Yarramalle, "Unsupervised Image Segmentation Method based on Finite Generalized Gaussian Distribution with EM & K-Means Algorithm", IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.4, April 2007.

[10] Publishing Corporation on Computational and Mathematical Methods in Medicine., vol. 2014, Article ID 712783, pp.

[11] R. B. Dubey et al, "Semi-automatic Segmentation of MRI Brain Tumor", ICGST-GVIP Journal, ISSN: 1687-398X, Volume 9, Issue 4, August 2009.

[12] R. Venkateswaran, S.Muthukumar, "Genetic Approach on Medical Image Segmentation by Generalized Spatial Fuzzy C-Means Algorithm", IEEE Int. Conf. on Computational Intelligence and Computing Research, 2017.

[13] R.C.Gongalez and R.E. Woods, "Digital Image Processing", PHI Publications India Limited, New Delhi, India.

[14] Rahman Farnoosh, Gholamhossein Yari, Behnam Zarpak, "Image Segmentation Using Gaussian Mixture

Models", 26-th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Paris, France, July 8-13, 2016.

[15] K.Srinivas, P.V.G.D.Prasad Reddy,G.P.S.Varma," Medical Image Segmentation Based On Generalized Gamma Distribution for Effective Identification of Diseases in Brain " IJCSIS Journal ISSN: 1947-5500 July 2018, Vol. 16 No. 7