

**International Journal of Innovative Research in Science,
Engineering and Technology**

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

K-MEANS CLUSTERING TO IDENTIFY HIGH ACTIVE NEURON ANALYSIS FOR LSD

Dr.D. Gladis¹, S.Rani²

Department of Computer Science, Presidency College Chepauk, Chennai-51, India¹

Department of Computer Science (Female campus) , King Khalid University ,Abha KSA²

Abstract: Computing process and mathematical modelling are used to solve many complex issues and are applied in all the major developments of Scientific, Engineering, Medical and Social applications. Computational approaches are attempted to identify the root cause of the disease in the medical domain. Medical Diagnosis, medical image processing, Bio informatics analysis techniques are widely used for the detection of the medication factors and its related causes. The internal complexity of the process is challenge in the medical field. Medical applications are build based on mathematical modelling and algorithms. The mathematical model is integrated on medical applications are attempted to solve different medical problems. This study identifies the functional relationship of Lysosomal storage disorder (LSD) Patients' brain neuron high density and its active status.

Keywords: Neuron Pattern Analysis, Functional Association of LSD, Data Mining, K-Means Clustering

I.INTRODUCTION

The computing techniques are used to resolve the complex issues in all domain applications [5]. The computing processes are build upon mathematical model and provides a possible and optimum solution based on the existing data[7]. It is used to predict the features using Data mining techniques such as association [6], clustering[10] [19] , and classification [23]. It determines the optimum solution using suitable algorithms in accordance to the domain and nature of the problems[4]. Computational model and the algorithms are applied in all the domains according to the needs. The association identifies the relationship according to the existing values of data.

Data mining techniques are applied into the intelligent analysis, business, scientific, mathematical and the medical applications [9] [1]. This research work is attempted to apply the clustering analysis on medical images. This research application is focused to identify the brain activeness of LSD challenged person. The process is attained through clustering brain image and by identifying high active neuron.

II. SCOPE

Data mining techniques are used to predict the relationship between the different attributes using associative relationship. This work is to determine the associative relationship of LSD patient and their neuron activeness. The neuron image analysis processes are implemented with clustering to determine high active neurons. This paper covers the analysis of the LSD patients' neuron image analysis using the K-means clustering and identifies the potential and active neuron. Section III provides orientation on LSD followed by the K-Means clustering algorithm[13] along with implementation results

III. LYSOSOMAL STORAGE DISORDERS (LSDS)

Lysosomal Storage Disorders (LSD) is a deficiency of metabolism that will affect human process and external activities of the human[24]. A genetic (inherited) problem causes a deficiency or malfunction of a particular enzyme in the body, so the enzyme cannot properly rid cells of waste material[22]. Waste products accumulate in the lysosomes of cells, which leads to disruption of cell function. However, while all LSDs have a similar underlying cause, there is great variety of signs and symptoms of the individual diseases. As a group, the LSD affect nearly every part of the body in people of all ages and races.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

In LSD, the classical acute neuronopathic (type 2) form; neuronopathic symptoms are characterised by an onset initial stage. In addition to the systemic manifestations, prominent brainstem findings are present, including spasticity, persistent retroflexion of the head, strabismus, oculomotor apraxia, trismus, dysphagia, laryngeal stridor and exaggerated deep tendon reflexes. Chronic neuronopathic (type 3) disease comprises a variety of disease variants. Patients may have an onset during childhood with hepatomegaly and skeletal lesions, and in addition, early development of a slowing of conjugate horizontal saccadic eye movements. Later on, treatment-resistant generalised tonic-clonic and myoclonic seizures, dementia, progressive spasticity and ataxia may develop and intellectual function may be lost which can be identified from the MRI brain analysis. The following part provides the concept of K-means clustering[20].

IV. K-MEANS ALGORITHM CLUSTERING ALGORITHM

K-Means is a partitioning (P) method. It takes n number of objects as input along with the parameter k, that means the algorithm has to create k number of clusters. Each cluster has to contain at least one object, and each object has to belong to exactly one cluster[12]. In this manner the k-means algorithm does not handle noise, thus in our model applying the noise model is an empty step (P1 = NULL, D0 = D1). K-means algorithm creates an initial partitioning where randomly selected objects are considered as the center of a cluster, and those objects belong to a given cluster, that are the nearest to it considering the distance measure, that is also an input for the algorithm. Thus, in this manner P2 is applied on the input object set and creates the initial clusters C0. This process iterates until the criterion function converges. This reallocation means that the k-means algorithm tries to minimize the root mean square of the clustering. In this paper we are used Squared Euclidean to determine the distance for clustering.

Squared Euclidean Distance

The General Squared Euclidean Distance Coefficient compares two cases i and j, and is defined as follows:

$$d_{ij}^2 = \frac{\sum_k w_{ijk}(x_{ik} - x_{jk})^2}{\sum_k w_{ijk}}$$

Where x_{ik} is the value of variable k in case i, and w_{ijk} is a weight of 1 or depending upon whether or not the comparison is valid for the kth variable; if differential variable weights are specified. It is the weight of the kth variable, or 0 if the comparison is not valid. It should be noted that the effect of the denominator $\sum_k w_{ijk}$ is to divide the sum of the distance scores by the number of variables; or if variable weights have been specified, by the sum of the weights.

Euclidean Distance: Euclidean Distance d_{ij} is obtained by taking the square root of Squared Euclidean Distance Od_{ij}^2 as computed above.

V. METHODOLOGY

The following steps are executable in pre-processed MRI using K-MEANS algorithm.

Step 1: Fetch the pre-processed MRI in the .nrrd format from LSD person

Step 2: Convert the fetched MRI .nrrd format file into sequence of frames in a 2D model.

Step 3: Fixation of frame number for the sequenced frames.

Step 4: Convert the frames into corresponding digital numbers.

Step 5: The converted digital numbers into triplet array sequence.(X,Y,Value)

Step 6: Cluster the triplet converted values using K-MEANS algorithm

Step 7: Fix a index value (ranking) of clustered object

Step 8: Sequential integration of same clusters and determine the pattern

Step 9: Determines of active and inactive neuron set based on pattern

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

VI. MRI analysis for LSD

The biological cell into a structure that is capable of receiving and integrating input, making a decision based on that input, and signalling other cells depending on the outcome of that decision is a remarkable feat of evolution[18]. The clustering process[2] is based on the digital values of the obtained MRI[11].

Source of the data : A MRI converted nrrd image is captured from the slicer public database. The data set is presented in the nrrd format and it is converted into the cubical data format using matlab. The converted cubical data set is presented into 256x256x120 representation. The each layer is presentable in a two dimensional format. The three dimensional axes points of the data are fetched and the changes between the 139 pixels are computed and graph is generated.

Magnetic resonance imaging (MRI) into cubical database: Magnetic resonance imaging (MRI) is one among the familiar and famous three-dimensional viewing of the brain and structures precise spatial relationships [3] . The image resolution is somewhat limited. Stained sections, on the other hand, offer excellent resolution and the ability to see individual nuclei (cell stain) or fiber tracts (myelin stain), however, there are often spatial distortions inherent in the staining process[8]. In this work ,nrrd file is fetched from slicer 3D download data base. Nrrd stands for "nearly raw raster data".Nrrd is a library file format designed to support scientific visualization and image processing involving N-dimensional raster data.

The network path way analysis is made by Modha, et.al., identified the movement and the distance of the neuron via analysing the MRI three dimensional coordinated image[14][21] [15] [16]. A similar approach is made to attain the signal communication analysis to increase the speed of the neuron process[17]. The images which is fetched at the time of MRI scanning is processed using matlab and converted to the two dimensional image and converted into the corresponding digital values.

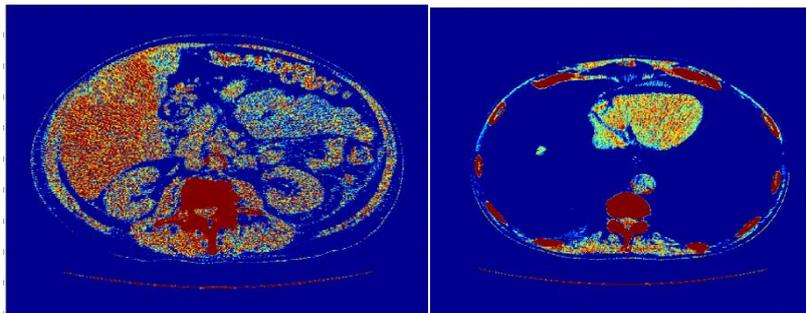


Figure 1 : Converted MRI

Figure 2: MRI layer Image

The image is sequenced into 1: 120 based on the nature of the MRI file. The files corresponding digital values are converted and presented to compute the frequency of the changes. The changes on the digital number consider as an activation of neuron.

VII. CLUSTER ANALYSIS

The neuron processes are observed on the processed MRI. The major two patterns active and inactive neurons are identified. In frequency process evaluated from 262144 generated frequency images. Using K-means on the Digital values neuron are clustered in 8,16,32,64 and 128.The obtained results are summarised and discussed below.

**International Journal of Innovative Research in Science,
Engineering and Technology**

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

Table 1: Eight cluster neuron count

Cluster no	1	2	3	4	5	6	7	8
Neuron Count	0	0	0	0	72025	80057	51237	3053

The first 4 clusters are empty clusters and does not have elements. However, elements are not active. The cluster images below

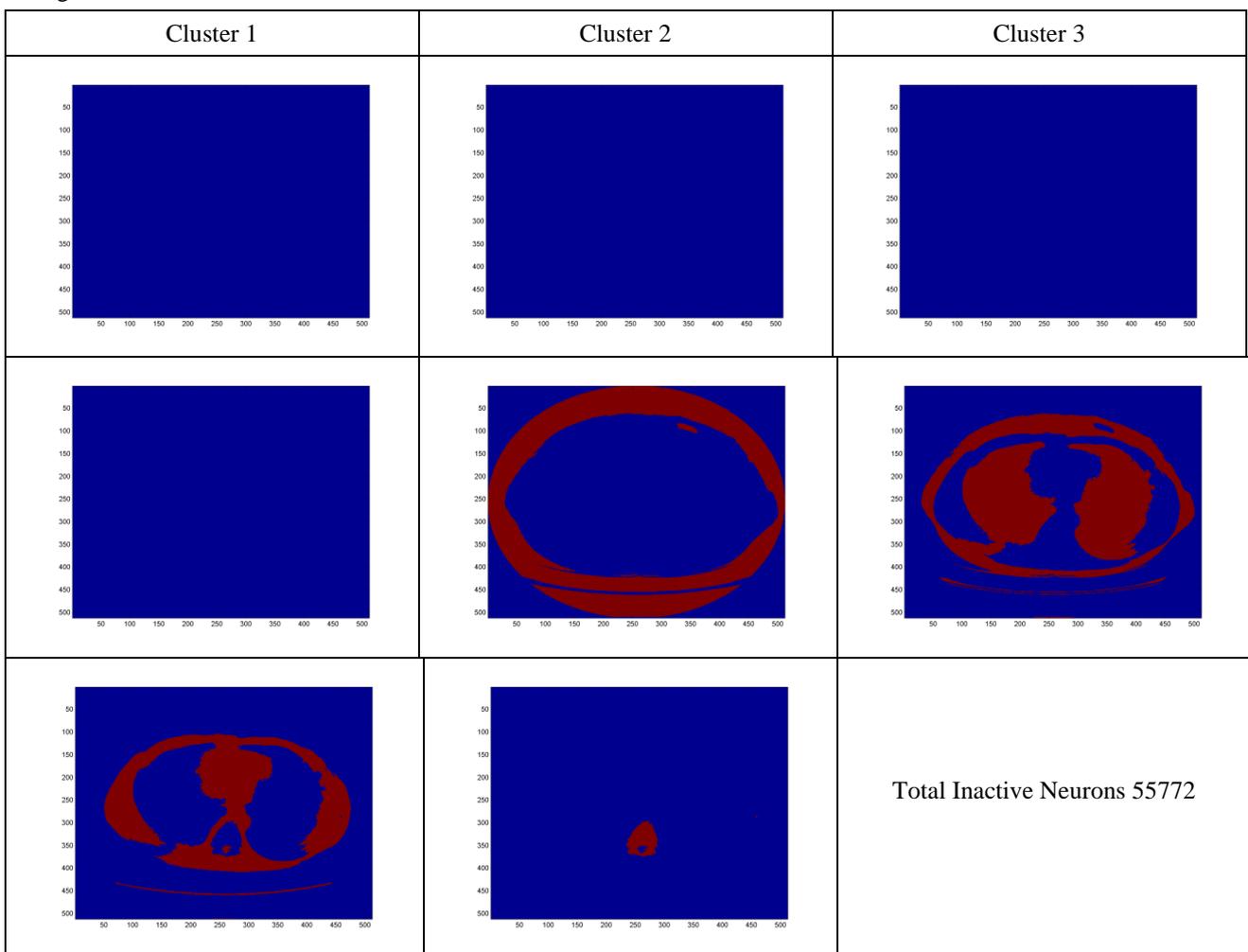


Figure 3 : Cluster images for 8 clusters

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

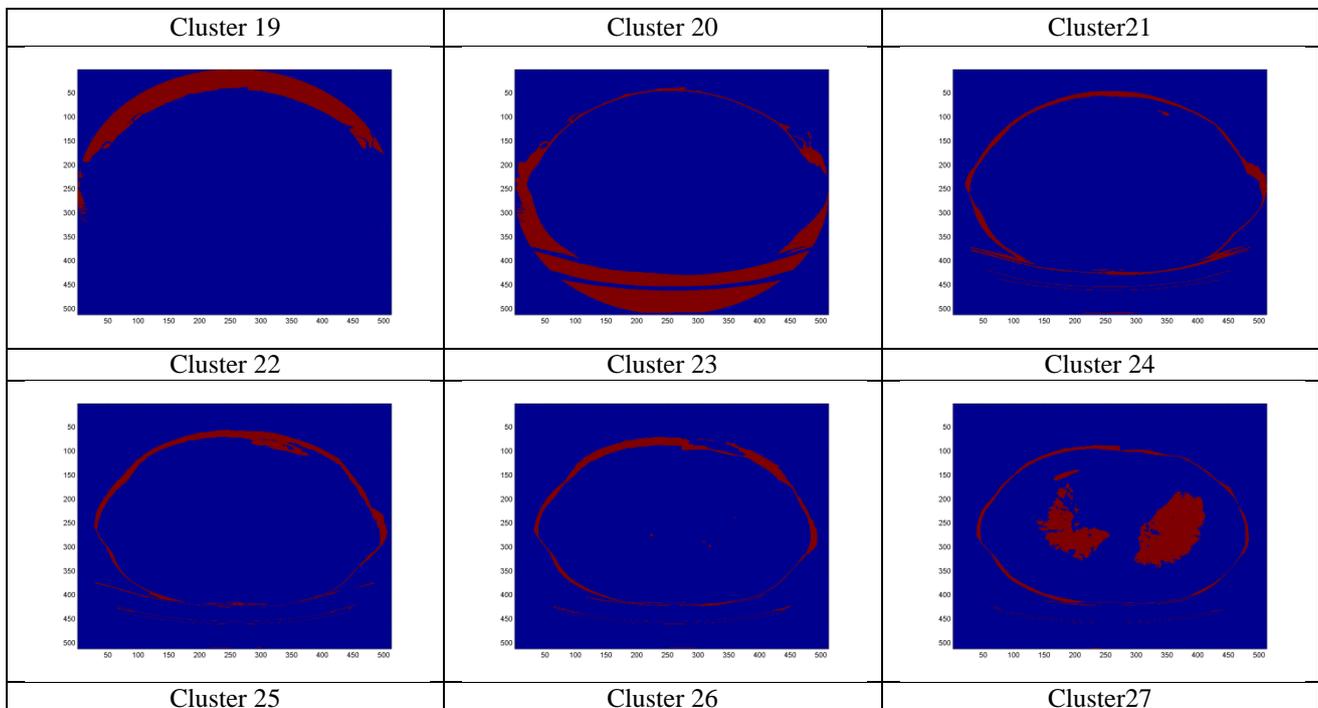
Vol. 2, Issue 9, September 2013

Similarly the neurons are clustered into 32 and the following results are obtained

Table 2: 32 cluster neuron count

Cluster no	1	2	3	4	5	6	7	8
Active Neuron	0	0	0	0	0	0	0	0
Cluster no	9	10	11	12	13	14	15	16
Active Neuron	0	0	0	0	0	0	0	0
Cluster no	17	18	19	20	21	22	23	24
Active Neuron	0	0	21447	36458	10098	11187	10077	20642
Cluster no	25	26	27	28	29	30	31	32
Active Neuron	29623	24055	19923	11109	7464	2132	1901	256

As per the observation of 32 clustering the non active neurons are 55772 and active neurons are 206372. The first 1-18 clusters are inactive. The remaining 19 to 32 are active clustering. The minimum is 56 and the maximum is 36458 in the neuron clusters. The obtained active neuron cluster images are.



**International Journal of Innovative Research in Science,
Engineering and Technology**

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

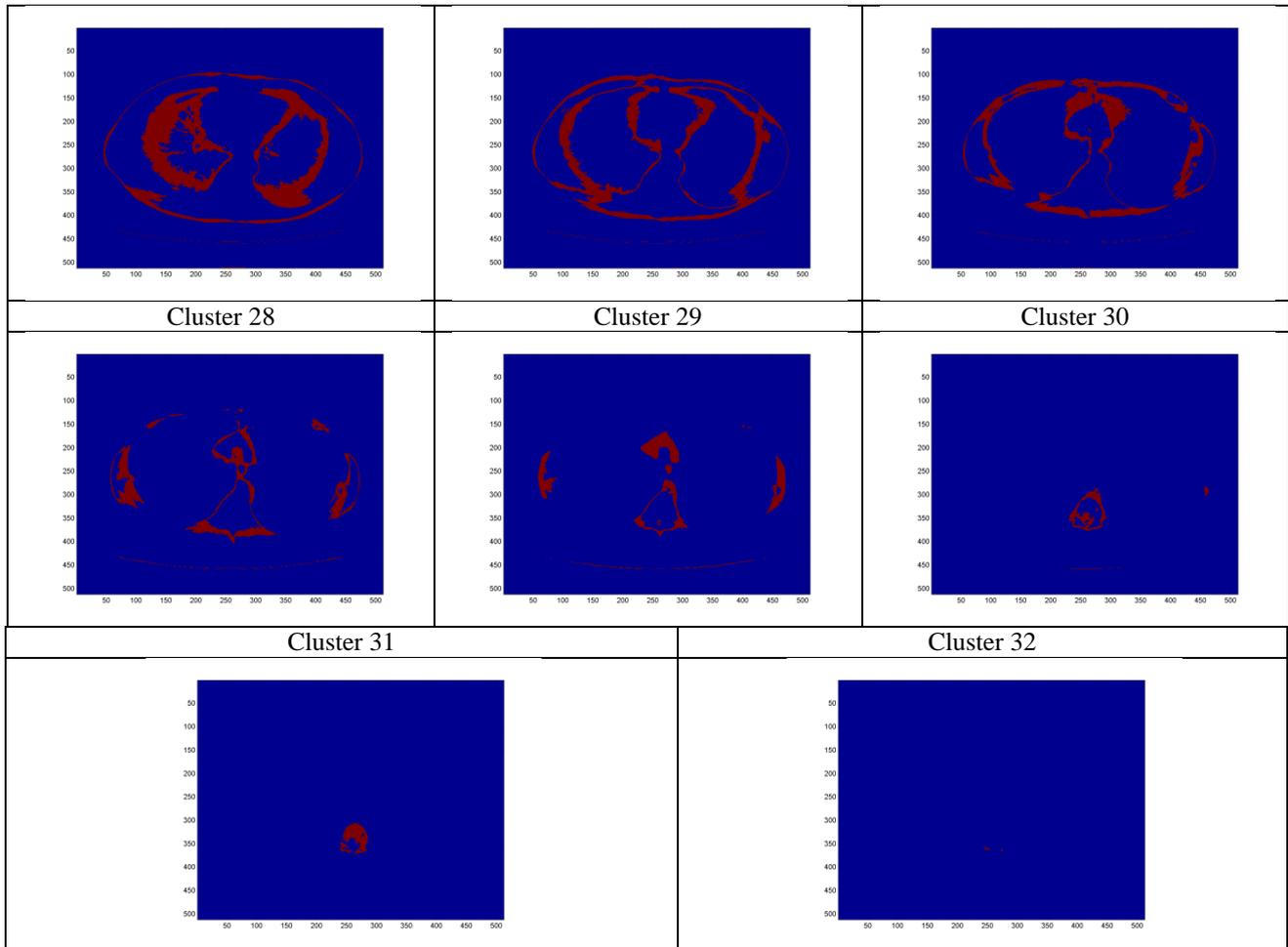


Figure 4: Cluster images for 32 clusters

While clustering into the high level of clustering the communication neuron are reduced. Comparison of Communication strength of Neuron

Table 3: comparison of Active neurons

Cluster	Non Active Cluster	Active Cluster	Min	Max	Average	Active Neuron	Non Active Neuron
8	4	4	2853	206572	82549	206372	55772
16	9	7	671	61742	29482	206372	55772
32	18	14	56	36458	14741	206372	55772
64	38	26	68	30888	7937	206372	55772
128	77	51	37	23543	4046	206372	55772

When the number of cluster are increased the neuron are communicated quickly with the clusters. When numbers of neurons in the cluster are less the communication speed is increased.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

VIII. CONCLUSION

The MRI cluster analysis of the communication of neuron process will increase the speed then the LSD impact will be decreased. The MRI image is converted nrrd format and presented into corresponding cubical data set for the DN generation and clustered according to the frequency. The active neurons are decreased then communication process increased while analyzing the LSD. The low levels of neuron areas are identified. The density based as well as hierarchal clustering algorithms to be compared for high active neurons for the further mapping process.

ACKNOWLEDGMENT

The author is acknowledging to Integrated Intelligent Research (IIR) for the support of computations and programming.

REFERENCES

1. Antonie, M. L., Zaiane, O. R., Coman, A., (2001) "Application of Data Mining Techniques for Medical Image Classification", Proceedings of the Second International Workshop on Multimedia Data Mining MDM/KDD 2001 in conjunction with ACM SIGKDD conference, San Francisco, August 26, 2001
2. B. Andreopoulos, A. An, X. Wang and M. Schroeder "A roadmap of clustering algorithms: Finding a match for a biomedical application", Briefings Bioinformatics, vol. 10, no. 3, pp.297-314 2009
3. Barnathan M, Megalooikonomou V, Faloutsos C, Faro S, Mohamed FB. TWave: high-order analysis of functional MRI. Int. J. Neuroimage. 2011 Sep 15;58(2):537-48.
4. Cios KJ, Moore GW. Medical data mining and knowledge discovery: an overview. In: Cios KJ, editor. Medical data mining and knowledge discovery. Heidelberg: Springer, 2000. p. 1-16 [chapter 1].
5. Dunham, M. H., Sridhar S., (2006) "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006
6. Fadi Thabtah, A review of associative classification mining. The Knowledge Engineering Review, Volume 22, Issue 1 (March 2007), Pages 37-65, 2007.
7. George, D. and Hawkins, J. (Oct. 2009) Towards a mathematical theory of cortical microcircuits. PLoS Computational Biology 5, 10, e1000532.
8. Gruber O, Tost H, Henseler I et al. Pathological amygdala activation during working memory performance: evidence for a pathophysiological trait marker in bipolar affective disorder. Hum Brain Mapp 2010; 31: 115-125.
9. Harleen Kaur, Siri Krishan Wasan and Vasudha Bhatnagar, THE IMPACT OF DATA MINING TECHNIQUES ON MEDICAL DIAGNOSTICS, Data Science Journal, Volume 5, 19 October 2006 pp 119-126.
10. A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. ACM Computing Surveys, 31(3):264-323, 1999
11. M. C. Jobin Christ, R. M. S. Parvathi, "Segmentation of Medical Image using K-Means Clustering and Marker Controlled Watershed Algorithm European Journal of Scientific Research ISSN 1450-216X Vol.71 No.2 (2012), pp. 190-194
12. Jiawei Han and Micheline Kamber, "Data Mining Concepts and techniques", 2nd ed., Morgan Kaufmann Publishers, San Francisco, CA, 2007.
13. M.C. Su and C. H. Chou, "A Modified Version of the K - Means Algorithm with a Distance Based on Cluster Symmetry," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.23, no.6, pp. 674 - 680, June. 2001
14. D. S. Modha, (2009) "A conceptual cortical surface atlas," PLoS ONE, vol. 4, no. 6, p. e5693.
15. Modha, D.S. and Singh, R. (June 2010) Network architecture of the long-distance pathways in the macaque brain. Proceedings of the National Academy of Sciences of the USA 107, 30, 13485-13490.
16. Dharmendra S Modha's, (2012) A scalable simulator for an architecture for Cognitive Computing IBM and LBNL presented the next milestone towards fulfilling the vision of DARPA SyNAPSE program at Supercomputing 2012.
17. Moore GW, Berman JJ. Anatomic pathology data mining. In: Cios KJ, editor. Medical data mining and knowledge discovery. Heidelberg: Springer, 2000. p. 61-108 [chapter 4].
18. Onkamo, P. and Toivonen, H., "A survey of data mining methods for linkage disequilibrium mapping", Henry Stewart Publications 1473 - 9542. Human Genomics. VOL 2, NO 5, age No. 336-340, MARCH 2006.
19. D. Pelleg and A. Moore. X-means: Extending K-means with efficient estimation of the number of clusters. In Pat Langley, editor, Proceedings of 17th International Conference on Machine Learning, pages 727-734. Morgan Kaufmann Publishers San Francisco, USA, 2000.
20. R. Xu and D. Wunsch II "Survey of clustering algorithms", IEEE Trans. Neural Networks, vol. 16, no. 3, pp.645-678 2005
21. J. Sherbondy, R. Ananthanayanan, R. F. Dougherty, D. S. Modha, and B. A. Wandell, (2009) "Think global, act local; projectome estimation with bluematter," in Proceedings of MICCAI 2009. Lecture Notes in Computer Science.
22. Strakowski SM, Adler CM, Holland SK, Mills NP, DelBello MP, Eliassen JC. Abnormal fMRI brain activation in euthymic bipolar disorder patients during a counting Stroop interference task. Am J Psychiatry 2005; 162: 1697-1705.
23. Sunita Soni, O.P. Vyas, Using Associative Classifiers for Predictive Analysis in Health Care Data Mining, International Journal of Computer Application (IJCA, 0975-8887) Volume 4- No.5, July 2010, pages 33-34.

**International Journal of Innovative Research in Science,
Engineering and Technology**

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 9, September 2013

24. Wenger DA, Coppola S, Liu SL. Insights into the diagnosis and treatment of lysosomal storage diseases. Arch Neurol 2003;60(3):322-8.

BIOGRAPHY

Dr.D. Gladis received her MCA degree from Bharathidasan University in 1990, M.Phil degree in Computer Science from Periyar University and the Ph.D in Computer Science from the University of Madras. She is working as an Associate Professor in Presidency College, Chennai. Her area of interest includes Neural Networks and Application, Adhoc Network, Artificial Intelligence, Pattern Recognition .

S. Rani working as a Lecturer , Department of Computer Science , (Female campus), King Khalid University , Kingdom of Saudi Arabia since Sep 2012.She did her M.Phil in computer Science from periyar University , M.Sc(IT) from Bhrathidasan University and BSc Computer Science from University of Madras.Her area of interest includes datamining , Medical Image Analysis and converstion of algorithms into mathematical model.