

Analysis of Feature Extraction Methods for the Classification of Brain Tumor Detection

¹A.Harshavardhan, ²Dr. Suresh Babu,, ³Dr. T. Venugopal

¹Research Scholar, Department of Computer Engineering,

¹JNTUH & Asst. Prof. Dept of CSE, S R Engineering CollegeC, Wgl, India

²Associate Professor, Department of Computer, KDC, Warangal, India

³Associate Professor, Department of Computer Engineering, JNTU, Sulthanpur, Medak, India

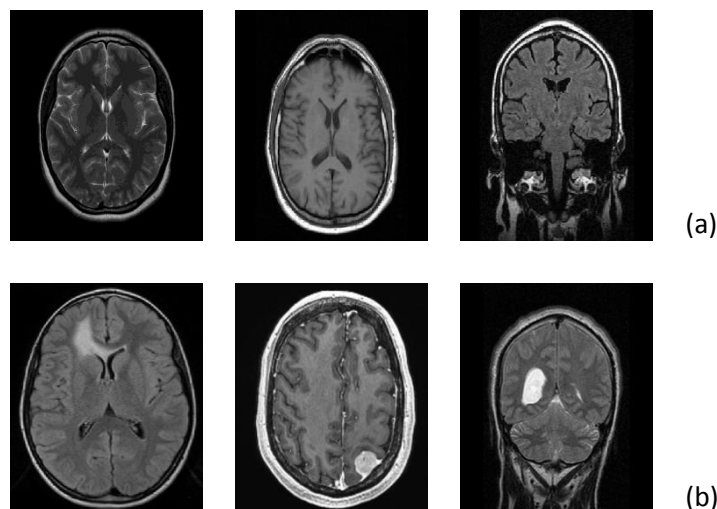
ABSTRACT

Diagnosing Brain tumor cancer is a difficult task for Urologists, Radiologists, and Oncologists. Ultrasound imaging is one of the hopeful techniques used for early detection of Brain tumor cancer. The aim of this article is to focus on current and presented brain tumor detection and classification methods from MRI brain images. A brain tumor is a collection, or mass, of irregular cells in brain. Skull, which covers brain, is very stiff. Any development inside such a restricted space can cause troubles. Brain tumors can be cancerous (malignant) or noncancerous (benign). When benign or malignant tumors develop, they can cause the force inside skull to increase. In this paper, the assessment of texture features is significant for several image handing out applications. The performance of the features pull out from the various texture methods such as histogram, Gray Level Cooccurrence Matrix (GLCM), Gray-Level Run-Length Matrix (GRLM), are analyzed separately. In this paper, it is proposed to combine histogram, GLRLM and GLCM in order to study the performance. The Support Vector Machine (SVM) is used to classify the extracted features into benign or malignant. The performance of texture methods are assessed using a variety of statistical performance measurements such as sensitivity, specificity and accuracy. The comparative analysis has been performed over 5500 digitized images of Brain tumor.

Keywords: Brain tumor, Histogram, GLCM, GLRLM, SVM, Segmentation

1. INTRODUCTION

In present scenario most of the population affecting with brain tumor. A tumor is irregular tissue that grows by unrestrained cell distribution. Two types of brain tumors are a primary tumor and secondary or metastatic tumor [1]. Usually, the primary brain tumor outsets in the brain and tends to stay during its growth tenure. Whereas, the secondary brain tumor commences elsewhere as cancer in the body and later spreads to the brain region. Further, the primary brain tumor has two sub-division namely, (i) Benign tumor and (ii) Malignant tumor. Identifying the tumor size and type of tumor is difficult task for physicians. Hence the automated brain tumor segmentation methods are very demanded in detection of accurate size, location and type of tumor. However there are many methods for this process of segmentation



of brain tumor, there is scope to maximize accuracy.

Figure 1. MRI brain images (a) Typical MRI brain images (b) MRI brain images with the tumor

Grade I tumor	Grade II tumor	Grade III tumor	Grade IV tumor
Slow growth	Relatively slow growth	Active growth	Abnormal growth
Almost appears to be normal	Slightly seems to be abnormal	Appears to be abnormal	Seems to be very normal
Least preferred malignant	Possibly invade adjacent tissues	Penetrate to adjacent tissues	Maintains rapid growth
Less Harm	Possibilities of recurring as higher-grade tumors	Chances to recur as higher-grade tumor	Extremely harmful

Table 1: Tumor Grades and its Salient Features

Basic Steps for brain tumor classification is as shown in figure 1.it consists of 4 stages.

1. Image preprocessing stage
2. Feature extraction stage
3. Classification stage

Image preprocessing stage

In image preprocessing stage the image smoothness, skull stripping and filtering, enhancement and segmentation and defining ROI will be done .filtering process is used to remove the noise from MRI image why because the MRI images are noisy. Skull stripping process is used to remove skull from brain tissue. Segmentation is process of sorting out an image into many pieces and object region. For segmentation we have so many methods like Region growing, watershed algorithm, clustering, K -means Clustering and fuzzy means Clustering. In this paper for the segmentation process used is fuzzy c means algorithm.

Feature extraction stage

Feature extraction is a process used to extract the significant features of the images, which are used to comprehend the image easier. This input image is transformed into the squashed form is called feature extraction. It extract the features like contrast, homogeneity, energy, entropy, sum of average, sum of variance, auto correlation, standard deviation etc...It will be very helpful in classification stage. The outputs of the feature- extraction stage will be given as input to classifier. Here the author is uses GLRLM feature extraction technique is used

Classification stage

Supervised learning method (SVM) tool used to analyze and classify the data.SVM is very effective even for large data. SVM is used to classify two or more classes. It works based on the decision plane; it separates the items with different class attributes. The Brain tumor detection and classification done by support vector machine.SVM is used to identify tumor class present in the image.

2. EXISTING METHODS

The most important one is Texture analysis of an image. By using this to illustrate the spatial variation in an image intensity which is associated to image property such as coarseness, and regularity. It can be done using arithmetic manipulation of digital image to acquire quantitative measurement. Textures parameters are used to increase visual skills of the expert user eye by retrieving image features that are closed for problem diagnosis and not compulsorily visual retrievable. To get the spatial dependent gray level values, the observation of the texture, a 2 dimension texture parameters matrix is considered for texture analysis. Each pixel and pixel values represent the texture characteristics. There are 4 types texture analysis: statistical-based, transform-based, Structural-based and model-

based. Statistical analysis describes pixel intensity values of image with pure numerical methods. Transform approaches usually carry out some sort of changes to the image, obtain a new “reply” image, is then analyzed as agent proxy for the ac actual image [16]. Structural method used to understand the image hierarchical structure. Model-based methods are dependence on the idea of expecting pixel values dependence on a arithmetical model

This paper focused on statistical Methods; represent texture features based on relation among gray levels of an image. Every individual tissue is having unique textures [13]. Benign tumors are as regular masses with similar in-house echo, while carcinomas are masses with fuzzy borders and varied in-house echoes. Different Texture features are built using statistical features with the ROI of Brain Scan image. Texture features are calculated using statistical delivery of practical grouping of intensities at specific points relative to each other in the image. Based on no. of pixels in each group, statics are categorized into first-order, second-order and higher-order statistics. A typical Brain scan image of Brain tumor includes a vast amount of various information that depicts different pieces. Using available information of brain scan image diagnostic system can constructively classify diagnose between normal and abnormal tissue [14, 15].

In this paper, segmented image (ROI) is utilized to construct the feature sets using Histogram method, Gray-Level Run-Length Method (GLRLM), and Grey- Level Co-occurrence Matrix (GLCM). In this paper we analyze Brain scan images using three different texture extraction methods. Performance of the combinations of the above three methods are also analyzed

a) Intensity Histogram Features

Intensity Histogram analysis been widely investigated in the early phases of improvement of this algorithm. The traits of the histogram have close association with the trait of image such as intensity and contrast. The trait of a histogram. And hence, the trait of an image can be represented using the following dimensions [17]. Mean retrieves the intensity of an image. Bright image have elevated mean while dim image have short mean, and as well mean values exemplify individual classifications. The contrast of an image retrieved by Standard deviation or variance retries. Image with high-quality contrast have elevated variance. Standard Deviations (SD) also exemplify the cluster. Skew measures is how irregularity (unbalance) the delivery of the gray level. Image with bimodal histogram delivery (object in contrast background) have elevated variance but short skew delivery (one peak at each side of mean). Energy measurement is closely related to skew. Highly skew distribution usually gives high-energy dimension. Entropy dealings the average number of bits to code each gray level. It has opposite association with skew and energy dimensions. Highly skew delivery tends to yield short Entropy. These are recapitulating in Table 1. Within ROI (i.e. segmented Brain tumor region) a histogram delivery of the image is calculated. Then six features are calculated for classification.

S.No	Features	Formulae
1	Mean	$\sum_{i=1}^N ih(i)$
2	Variance	$\sum_{i=1}^N (i - \mu)^2 h(i)$
3	Skewness	$\sum_{i=1}^N (i - \mu)^3 h(i)$
4	Kurtosis	$\sum_{i=1}^N [i - \mu]^4 h(i)] - 3$
5	Entropy	$-\sum_{g=0}^{L-1} P(g) \log_2 [P(g)]$
6	Energy	$\sum_{g=0}^{L-1} [P(g)]^2$

Table 2: Histogram Features

b) Gray Level Co occurrence Matrix

The Gray Level Cooccurrence Matrix (GLCM) procedure is a way of pull out second order statistical texture features [15, 18]. It models the relationships among pixels inside the region by building Gray Level Co-occurrence Matrix. The GLCM is based on an inference of the second-order mutual limited probability density functions $p(i, j | d, \theta)$ for a diversity of direction $\theta = 0,$

45, 90, 135°, etc., and unlike distances, $d = 1, 2, 3, 4,$ and 5 . The function $p(i, j | d, \theta)$ is the probability that 2 pixels, which are placed with an intersample distance d and a direction θ , have a gray level i and j . The spatial relationship is defined in terms of distance d and angle θ . If the texture is coarse, and distance d is small, the pair of pixels at distance d should have same gray values. On the other hand, for a fine texture, the couple of pixels at distance d should often be quite unlike, so that the value in the GLCM should be stretching out moderately uniformly [19, 20]. Similarly, if the texture is coarser in one direction than another, then the degree of spread of the values about the main diagonal in the GLCM should vary with the direction θ [21]. The figure 2 represents the formation of the GLCM of the grey-level (4 levels) image at the distance $d = 1$ and the direction $\theta = 0^\circ$.

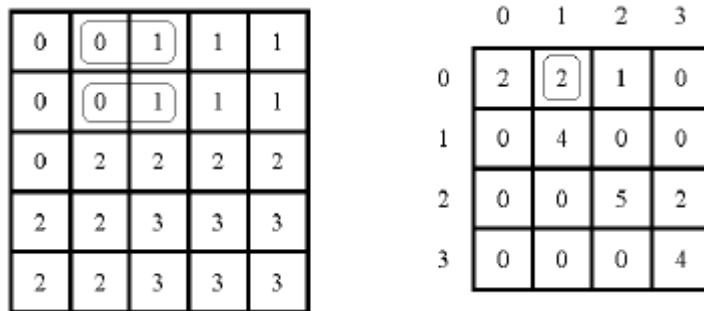


Figure 2(a): image with 4 grey level 2(b): GLCM for $d = 1$ and $\theta = 0^\circ$

The thin box in figure 2(a) represent pixel-intensity 0 with pixel intensity 1 as its neighbor in the direction $\theta = 0^\circ$. There are two occurrences of such pair of pixels. Therefore, the GLCM matrix created with value 2 in row 0 and column 1. This procedure is repetitive for additional pair of intensity values. As a effect, the pixel matrix represented in Figure 2(a) can be transformed into GLCM as shown in Figure 2(b). In addition to the direction (0°), GLCM can also be created for the other directions $45^\circ, 90^\circ$ and 135° as shown in Figure 3.

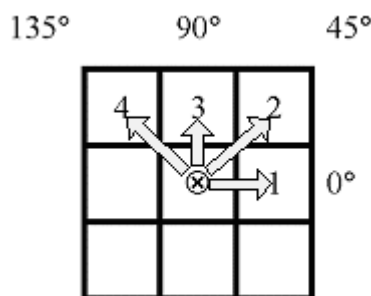


Fig 3: directions $45^\circ, 90^\circ$ and 135°

The pixels 1, 2, 3 and 4 are representing the directions (θ) $0^\circ, 45^\circ, 90^\circ$ and 135° respectively for distance $d = 1$ from the pixel x .

a) Gray-Level Run-Length Matrix

Texture is understood as a blueprint of grey intensity pixel in a particular direction from the reference pixels. Grey-Level Run-Length Matrix (GRLM) is a matrix from which the texture features can be pull out for texture analysis [22]. It is a way of searching the image, always across a given direction, for runs of pixels having the similar gray level value. Run length is the number of neighboring pixels that have the similar grey intensity in a particular direction. Gray-level run-length matrix is a 2-dimensional matrix where each element is the number of elements j with the intensity i , in the direction θ . Thus, given a direction, the run-length matrix measures for each acceptable gray level value how many times there are runs of, for example, 2 consecutive pixels with the same value. Next it does the same for 3 consecutive pixels, then for 4, 5 and so on. Note that many different run-length matrices may be computed for a single image, one for each Chosen direction. The GLRLM is based on computing the number of gray level runs of a variety of lengths [23]. A gray level run is a set of consecutive and collinear pixel points having the same gray level value. The length of the run is the number of pixel points in the run. The gray level run length matrix

is as follows' $(\theta) = (g(i, j) | \theta)$, $0 \leq i \leq Ng$, $0 \leq j \leq Rmax$; Where Ng is the utmost gray level and $Rmax$ is the maximum length. Figure 3 shows the sub image with 4gray levels for constructing the GLRLM. Figure 5 shows that the GLRLM in the direction of 0 of the sub image in Figure 4.

1	2	3	4
1	3	4	4
3	2	2	2
4	1	4	1

Figure 4: Matrix of Image

Gray Level	Run Length(j)			
	1	2	3	4
1	4	0	0	0
2	1	0	1	0
3	3	0	0	0
4	3	1	0	0

Figure 5: GLRL Matrix

In addition to the 0° direction, GLRLM can also be formed in the other direction, i.e. 45°, 90° or 135°.

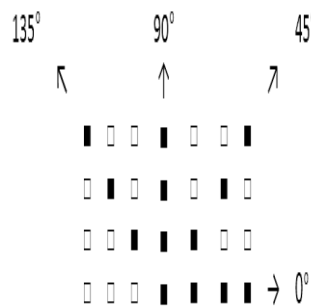


Figure 6: Run Direction

<i>S.No</i>	<i>Features</i>	<i>Formulae</i>
1	Short Run Emphasis (SRE)	$\frac{1}{n} \sum_{i,j} \frac{p(i, j)}{j^2}$
2	Long Run Emphasis (LRE)	$\frac{1}{n} \sum_{i,j} j^2 p(i, j)$
3	Grey Level Non-uniformity (GLN)	$\frac{1}{n} \sum_i \left(\sum_j p(i, j) \right)^2$
4	Run Length Non-uniformity	$\frac{1}{n} \sum_i \left(\sum_j p(i, j) \right)^2$
5	Run percentage (RP)	$\sum_{i,j} \frac{n}{p(i, j) j}$
6	Low Grey Level Run Emphasis (LGRE)	$\frac{1}{n} \sum_{i,j} \frac{p(i, j)}{i^2}$
7	High Grey Level Run Emphasis (HGRE)	$\frac{1}{n} \sum_{i,j} i^2 p(i, j)$

Table 3: GRLM Features

Seven texture features can be extracted from the GLRLM. These features use grey level of pixel in sequence and are intended to distinguish the texture that has the same value of SRE and LRE but have differences in the distribution of gray levels. Once features sets are constructed using Histogram features, GLCM, GRLM, and their combination. Then the next section explicates SVM classifier for the classification of extracted features

3. Discussion on Experimental Analysis

Classification, the sensitivity, specificity and accuracy were calculated using below formulas:

- True Positive (TP): Abnormal brain correctly identified as abnormal.
- True Negative (TN): Normal brain correctly identified as normal.
- False Positive (FP): Normal brain incorrectly identified as abnormal.
- False Negative (FN): Abnormal brain incorrectly identified as normal.

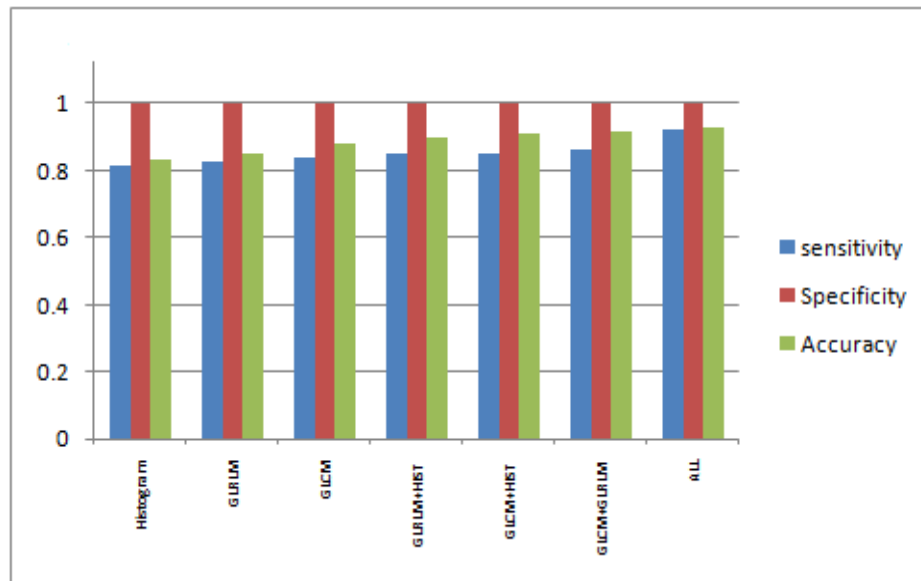
- 1) Sensitivity = TP / (TP+FN) *100%
- 2) Specificity = TN / (TN+FP) * 100%
- 3) Accuracy = (TP+ TN) / (TP+ TN+FP+FN)* 100 %

All these three parameters are used to check the classifiers performance

<i>Methods</i>	<i>sensitivity</i>	<i>Specificity</i>	<i>Accuracy</i>
Histogram	0.81090909	0.9979591	0.82978333
GLRLM	0.82495575	1.0000000	0.85000000
GLCM	0.83790900	1.0000000	0.87895500
GLRLM+HIST	0.84615384	0.9992960	0.89683333
GLCM+HIST	0.84745762	1.0000000	0.90700000
GLCM+GLRLM	0.86206896	1.0000000	0.91333333
ALL	0.91743119	1.0000000	0.92833333

Table 4. performance measurements

From the table 4, we observed that the maximum and minimum classification accuracies are 93% and 83% with SVM classifier. The histogram features discriminate between malignant masses and benign masses on Brain tumor images with 83% accuracy, 81% sensitivity and 99% specificity levels that are relatively poorer compare to others. GLRLM features yielded an accuracy of 85% for distinguishing malignant and benign masses on Brain tumor images. It is 2% higher than features based on histogram. The GLCM features achieved an accuracy of 88% where 84% sensitivity and 100% specificity. The accuracy of GLCM features 3% higher than GLRLM feature and 5% higher than Histogram features. The combination of Histogram features and GLRLM features is achieved 90% of the accuracy, where as 91% of accuracy is produced when combined the Histogram features with GLRLM features. The accuracy difference between these two methods is only 1% even while the sensitivity and specificity of these two are almost same 85% and 100% respectively. The accuracy of 91 % is arrived by the combination of GLCM features and GLRLM features, whilst 86% sensitivity and 100% specificity. The predicted accuracy is 5%, which is 3% higher than GLRLM, GLCM respectively. The combination of Histogram features, GLCM features and GLRLM features produces the highest accuracy of 93%, 91% sensitivity, and 100% specificity followed by GLCM features combined with GLRLM features with the accuracy of 91%, 86% sensitivity, and 100% specificity.



Measures with respect to proposed methods. By considering the different texture methods independently, it is not able to confirm that there is a universal method for best classification. However, usually the statistical cooccurrence (GLCM) features are used. The combination of various methods features produces a significant increase in the accuracy levels. It is interesting to note that using combined features produces relatively good classification results. The percentage of accuracy of combined features is higher than the values obtained from others. These analyses conclude that the combination of Histogram, GLCM and GLRLM texture features achieves best classification accuracy for distinguishing between malignant masses and benign masses on Brain tumor images.

4. Conclusion

Texture analysis is a potentially valuable and versatile in imaging for Brain tumor cancer interpretation. In some cases, radiologists face difficulties in directing the tumors. In this work, feature extraction methods for the Brain tumor cancer classification problem and an innovative approach (combined features) of finding the malignant and benign masses from the Brain tumor medical images are proposed and analyzed. The performances of classifiers for the texture-analysis methods are evaluated using various statistical parameters such as sensitivity, specificity and accuracy. The experiment results show that there is considerable performance variability among the various texture methods. The histogram features and GLRLM features performances are considerably poor. The combination histogram features, GLRLM features and GLCM features outperformed well in discriminating between or among Brain tumor cancer. Using proper feature selection method accuracy may be improved efficiently in future.

References

- [1] Vidhi Rawat ,Alok jain, and Vibhakar shrimali: "Investigation and Assessment of Disorder of Ultrasound Bmode Images" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 2, February2010, PP.289 – 293.
- [2] Ransohoff DF, McNaughton Collins M, Fowler FJ. Why is Brain tumor cancer screening so common when the

- evidence is so uncertain? A system without negative feedback. *Am J Med.* 2002 Dec 1;113(8):663-7. Review
- [3] A. Harshavardhan, Dr. Suresh Babu, Dr. T. Venugopal "A Survey on Brain Tumor Segmentation Methods with the Remedial Approach" (2017)
- [4] Cookson MM. prostate cancer: screening and early detection. *Cancer Control.* 2001 Mar-Apr;8(2):133-40. Review
- [5] Ferdinand Frauscher: "Contrast-enhanced Ultrasound in Brain tumor Cancer" *Imaging and Radiotherapy*, © T O U C H B R I E F I N G S 2 0 0 7. Pp 107-108.
- [6] Ethan J Halpern, MD: "Contrast-Enhanced Ultrasound Imaging of Brain tumor Cancer",
- [7] Thangavel, K., Manavalan, R. and Laurence Aroquiaraj . I.(2009), Removal of Speckle Noise from Ultrasound Medical Image based on Special Filters: Comparative Study, *ICGSTGVIP Journal*, ISSN 1687-398X, Volume (9), Issue (III), pp.25-32.
- [8], A. Harshavardhan, Dr. T. Venugopal, Dr. Suresh Babu "Brain Tumor Segmentation Methods-A survey" (2017)
- [9] Anil K. Jain. (1989) *Fundamentals of Digital Image Processing*: Prentice Hall.
- [10] Gonzalez, R. and Woods, R. (2002). *Digital Image Processing*. 3rd Edn., Prentice Hall Publications, pp. 50-51.
- [11] Ester, M., Kriegel, H.P., Sander, J., and Xu, X. (1996) 'Adensity-based algorithm for discovering clusters in Large spatial databases with noise'. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*, Portland: AAAI Press. pp. 226-231.
- [12] Sinan Kockara, Mutlu Mete, Bernard Chen, and Kemal Aydin: (2010). 'Analysis of density based and fuzzy c-Means clustering methods on lesion border extraction in dermoscopy images', From Seventh Annual MCBIOS Conference: Bioinformatics Systems, Biology, Informatics and Computation Jonesboro, AR, USA. pp. 19 - 20.
- [13] Amadasun, M. and King, R., (1989) 'Textural features corresponding to textural properties', *IEEE Transactions On Systems, Man, and Cybernetics*, vol. 19, no. 5, pp. 1264 -1274.
- [14] Haralick, R.M. , Shanmugan, K.. and Dinstein, I. (1973) 'Textural Features for Image Classification', *IEEE Tr. On Systems, Man, and Cybernetics*, Vol SMC-3, No. 6, pp. 610- 621.
- [15] Haralick, R.M. (1979) 'Statistical and Structural Approaches to Texture', *Proceedings of the IEEE*, Vol. 67, No. 5, pp. 786-804. [16] Mohamed, S.S. and Salama M.M. (2005) 'Computer Aided diagnosis for Brain tumor cancer using Support Vector Machine' Publication: Proc., medical imaging conference, California, SPIE Vol. 5744, pp. 899-907.
- [17] Chitrakala, S. Shamini, P. and Manjula, D: "Multi-class Enhanced Image Mining of Heterogeneous Textual Images Using Multiple Image Features" , *Advance Computing Conference, 2009. IACC 2009. IEEE International* , 2009 ,Page(s): 496 – 501.
- Rajesh, M., and J. M. Gnanasekar. "Congestion control in heterogeneous WANET using FRCC." *Journal of Chemical and Pharmaceutical Sciences* ISSN 974 (2015): 2115.
- Rajesh, M., and J. M. Gnanasekar. "A systematic review of congestion control in ad hoc network." *International Journal of Engineering Inventions* 3.11 (2014): 52-56.
- Rajesh, M., and J. M. Gnanasekar. " Annoyed Realm Outlook Taxonomy Using Twin Transfer Learning." *International Journal of Pure and Applied Mathematics* 116.21 (2017) 547-558.
- Rajesh, M., and J. M. Gnanasekar. " Get-Up-And-Go Efficient memetic Algorithm Based Amalgam Routing Protocol." *International Journal of Pure and Applied Mathematics* 116.21 (2017) 537-547.
- Rajesh, M., and J. M. Gnanasekar. " Congestion Control Scheme for Heterogeneous Wireless Ad Hoc Networks Using Self-Adjust Hybrid Model." *International Journal of Pure and Applied Mathematics* 116.21 (2017) 537-547.
- [18] Haralick R. M., Shanmugam K., Dinstein I. Textural Features of Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*. 1973. Vol. 3(6). P. 610-621.
- [19] Soh L., Tsatsoulis C. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices. *IEEE Transactions on Geoscience and Remote Sensing*. 1999. Vol 37(2). P. 780 – 795.
- [20] Claudi D A. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*. 2002. Vol. 28(1). P. 45-62.

