Brain Tumor Detection and its Area Measurement using K-Means Clustering baesd on Genetic Algorithm

Arnavi A. Patil[1], S. S. Badhe[2], A. A. Mulajkar[3]

[1]Dr.D.Y.Patil Educational Academy's, D.Y.Patil School of Engineering Academy, Ambi, Pune 410506,
Savitribai Phule Pune University, Maharashtra, India

[2]Professor, D. Y. Patil College of Engineering, Ambi, Pune 410506, Savitribai Phule Pune University, India

[3]Professor, D. Y. School of Engineering Academy, Ambi, Pune 410506, Savitribai Phule Pune University, India

Abstract: ***Brain tumor detection and its analysis are tough tasks in medical image processing because brain image and its structure is complex that can be inspected only by expert radiologists. Segmentation plays an important role in the processing of medical images. MRI has become a particularly useful medical indicative tool for diagnosis of brain and other medical images. This paper presents a study of segmentation method implemented for tumor detection. The methods include optimized k-means clustering with genetic algorithm. Traditional k-means algorithm is sensitive to the initial cluster centers. Genetic k-means clustering techniques are used to detect tumor in MRI of brain images. At the end of process the tumor is diagnosed from the MRI image and its actual position and the shape are determined. The experimental results indicate that traditional k-means not only eliminate the over segmentation problem, but also provide fast and efficient clustering results.***

Keywords: brain tumor, MRI image, segmentation, k-means clustering, Genetic algorithm

1. Introduction

Brain is most important part of human nervous system. It controls all the body functions, there for care must be taken than brain will not be affected by any abnormality. Brain tumor is the most challenging task for medical field. Brain tumor causes stroke which can over the life of patient. So physician give treatment for stroke instead of tumor. Before treat the tumor it is necessary that doctor can analyses presence of tumor, its location and size so they can treat it easily and correctly. What is tumor and which are there types?

A. Tumor

Tumor is nothing but abnormal growth of tissues. Depending on the growth of tissues there are three types of tumor.
Benign tumor, pre-malignant tumor, malignant tumor.

a) Benign Tumor: It is the tumor which does not affected on neighboring area and healthy tissues.
b) *Pre-Malignant Tumor*: It is the stage before cancer. If proper care is not taken it may be conclude in serious diseases.
c) *Malignant Tumor*: it leads to cancer. It spreads time to time and result into death of patient.[8]

B. Magnetic resonance imaging (MRI)

MRI is usually used in biomedical field to find the fault in far internal body structure. Find differences in tissues. It is more efficient than computed tomography.

C. Image Segmentation

Segmentation is vital term in medical field. It plays an important role in recognition and treatment of diseases, find tumor in body, its size and then decide the way to treat it. Segmenting an image refers to the technique of segmenting the space of an image into meaningful homogeneity areas in which no overlaps present which are the same in some features like, intensity or tissues . The segmenting techniques are classified into two categories:

1. Edge-based techniques
2. Area-based techniques[14]

In edge based approach pixels are categorized as edged or non-edged pixel based on filter output. In area based aproch grouping pixel on similarities and splitting dissimilar pixels.

2. Proposed Method

As before various segmentation techniques used for MRI image segmentation in biomedical field. Each technique having some unfulfilled requirements. Next technique is better than previous one which complete the requirements with better result. In this project some disadvantages from previous one are covered. K means clustering is the methad used for segmentation from far time.

K-means algorithm is the most well-known and widely-used unsupervised clustering technique in splited clustering algorithms. Purpose of this algorithm is to minimize the distances of all the elements to their cluster centres. Many of the algorithms in this field are developed by inspiring or improving k-means. The algorithm pramotes the clusters iteratively and runs in a loop until it reaches to optimal solution.[13] Pseudo-code of the K-means clustering algorithm is shown. Performance of K-means algorithm depends on initial values of cluster centers. Therefore the algorithm should be examined for different outcomes with different initial cluster centers by multi-running.

Not all the techniques are suitable for segmentation because of complexity and inaccuracy. There is no standard image segmentation technique that can produce adequate outcomes for all imaging applications like brain MRI, brain cancer diagnosis etc. Problems and challenges of brain image segmentation are:

- Traditional K-means algorithm is sensitive to the initial cluster centres; cluster results variable with different initial input and are easy to fall into local optimum.
- Over segmentation of Image due to limitations of the conservative watershed algorithm.

In this project we develop advanced k means clustering in which the result get from k means clustering is again treated using genetic algorithm.

Block diagram

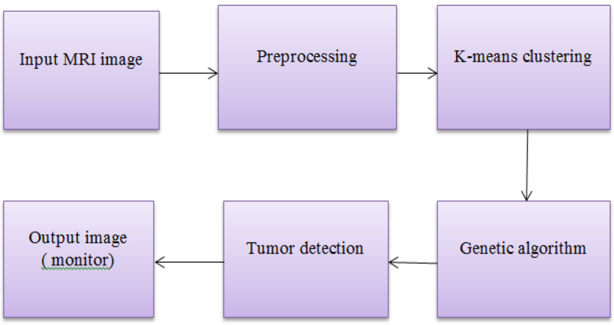


Figure 1: basic block diagram

Input image is nothing but the MRI image.

Pre-Processing:

Preprocessing and improvement techniques are used to improve the detection of the suspicious region from Magnetic Resonance Image (MRI).This section presents the inclined-based image enhancement method for brain MR images which is based on the first derivative and local census. The preprocessing and enhancement method consists of two steps; first the removal of film vestiges such as labels and X-ray marks are removed from the MRI using tracking algorithm. [12]Second, the removal of high frequency components using glutted median filtering technique. It gives high resolution MRI compare than median filter, Adaptive filter and spatial filter.

K-Means Clustering Technique:

First k means clustering technique is applied on image. A cluster is a collection of objects which are similar between them and are dissimilar to the objects relating to other clusters. Clustering is an unsupervised learning method which deals with finding a structure in a collection of unnamed data. A loose description of clustering could be the process of organizing objects into groups whose members are identical in some way.

K-means clustering is an algorithm to group objects based on vartue/features into k number of groups where k is a positive integer. The grouping (clustering) is done by minimizing the Euclidean distance between the data and the corresponding cluster centroid. Thus the function of k-means clustering is to cluster the data.

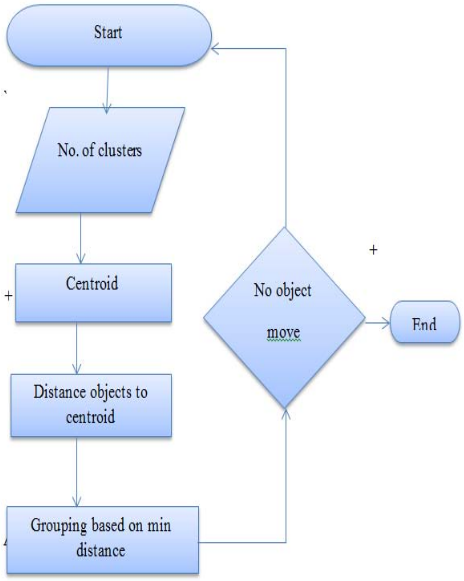


Figure 2: K means clustering algoritm

Genetic algorithm:

Genetic algorithm is a natural motivated meta heuristic algorithm. In GA each solution is represented as chromosome and each chromosome is make up from genes. The best generated solutions will be added to the next iteration while the bad results will be rejected. While the algorithm iterates its solutions, these solutions are improved up to a point where it near to optimal solution [9].

Generally, a GA has five stages: initialization of population, evaluation of fitness function, selection, crossover, mutation and termination. Initial population is created randomly, which can be done by applying genes to random values. After the initialization process, fitness function of each chromosome is evaluated. In the selection process, the fittest members in the present population are selected for reproducing the new solutions in crossover process two chromosomes are collected and enterchange genes by some point. In mutation process a gene is selected frequently and its value is changed. In last termination of the iteration is done when a certain criteria is met. Generally termination is done by number of repetitions. Fig.1 shows the methodology for Genetic algorithm.[9]

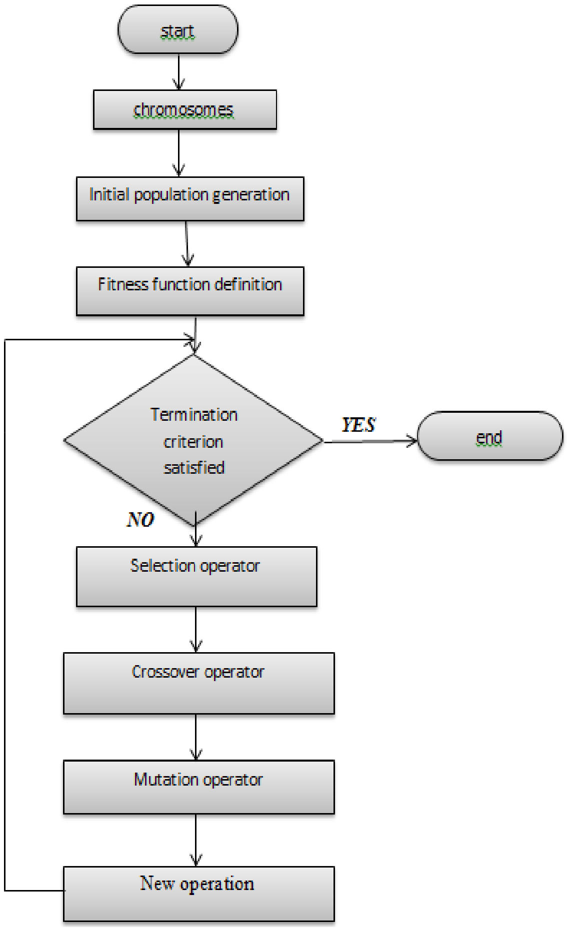


Figure 3: Genetic algorithm

3. Various Steps involved in GA Procedure

i) Representation

In a binary coded genetic algorithm, each variable is first coded in a fixed-length binary string. For example, the following is a string, representing N problem variables:

$$\underbrace{11010}\ \underbrace{1001001}\ \underbrace{010}\ \ldots\ldots\underbrace{0010}$$

The ith problem variable is coded in a binary substring of length li , so that total number of alternatives permitted in that variable is 2^li . The lower bound solution X_i^{min} is represented by(0,0,0,…) solution and the upper level solution X_i^{max} is represented by (1,1,..1)the solution . Any other substring S_i decodes to a solution x_i as follows:

$$x_i = x_i^{min} + \frac{x_i^{max} - x_i^{min}}{2^{l_i} - 1} DV(s_i), \quad (1)$$

Where $DV(S_i)$ is decoded value of string s_i .The decoded value of a binary substring $S=(S_{l-1}\ s_{l-2}\ S_2\ S_1\ S_0)$calculated as $\sum_{j=0}^{l-1} 2^j s_j$where S_j €{0,1} . The length of substring is usually decided by precision required in a variable. For example if three decimal places of accuracy are needed in the i th variable, whole number of alternatives in the variable must be $(X_i^{max} - X_i^{min})/0.001$, which can be set equal to and can be computed as follows:

$l_i=\log_2 (X_i^{max} - X_i^{min} /€_i)$

Here, the parameter desired precision in i th variable. The total string length of a N-variable solution is then

$$l = \sum_{i=1}^{N} l_i .$$

In the population, -bit strings are created at random (at each of positions, there is a equal probability of creating a 0 or 1). Once such string is generated, the first bits can be extracted from the complete string and respective value of the variable can be calculated using Equation (1) and using the chosen lower and upper limits of variable . This process is repeated until all N-variables are obtained from complete string. Thus, an bit string represents a entire solution indicating all N variables uniquely. Once these values are known, the objective function $f(X_1\ X_2\ldots.X_N)$ can be computed.

In a GA, each string created either in the first population or in the respective generations must be assigned a fitness value which is relevant to objective function value. For incremented problem, a string's fitness can be same to string's objective function value. However, for minimization problems, the aim is to find a solution having least objective function value. Thus, the fitness can be evaluated as the negative of the objective function so that solutions with equal objective function value get high fitness.

There are many of advantages of using a string indication to code variables. First, this allows a bonding between working of GA and actual problem. The same GA code can be used for variable problems by only changing meaning of coding a string. This allows a GA to have worldwide applicability. Second, a GA can exploit the similarities in string coding to make its search rapid, a matter which is important in working of a GA.

ii) Reproduction

Reproduction is normally the first operator applied to a population. Reproduction selects good strings in a population and design a mating pool. The imperative idea is that above-average strings are picked from the current population and duplicates of them are inserted in the mating pool. The commonly used reproduction operator is the proportionate selection operator, where a string in the present population is elected with probability proportional to the string's fitness. Thus, the i^{th} string in the population is selected with probability proportional to . Since the population size is usually kept established in a normal GA, the aggregate probability for all string in the population must be one. Therefore, the probability for electing i^{th} string is,

$$f_j \Big/ \sum_{j=1}^{N} f_j$$

where N is the population size. One way to achieve this comparable selection is to use a roulette-wheel with the circumference marked for each string proportionate to the string's fitness.

iii) Crossover

The crossover operator is tested next to the string of the mating pool. In crossover operator, two strings are took from the mating pool at random and some portion of the strings is enterchanged between the strings. At a single-point crossover operator, are cuts both string at an arbitrary place and right-side portion of both strings are exchanged among themselves to form two new strings, as adorned in the following:

Parent 1 00 000 | 00 111 child 1
$\Longrightarrow$
Parent 2 11 111 | 11 000 child 2

good substrings from either parent string can be combined to form better child string if an appropriate site is chosen. Since the knowledge of an appropriate site is normally not known, a random site is normally chosen. However, it is important to understand that the choice of a random site does not make this search operation random. With a single-point crossover on two l -bit parent strings, the search can only find at most $2^{(i-1)}$ different strings in the search space, whereas there are a total of 2^{i} strings in the search space. With a random site, the child strings generated may or may not have a combination of good substrings from parent strings depending on whether the crossing site falls in the appropriate site or not. if good strings are created by crossover, there will be more copies of them in the next mating pool produced by the reproduction operator. But good strings are not created by crossover, they will not bear beyond next generation, because reproduction will not select bad strings for the next mating pool. In a two-point crossover operator, two random sites are taken. This idea can be extended to create multi-point crossover operator and the extreme of this extension is known as a uniform crossover operator. In a uniform crossover for binary strings, each bit from either parent is collected with a probability of 0.5. The main purpose of the crossover operator is to search the parameter space. Other condition is that the search need to be performed in a way to preserve the information gathered in the parent string maximally, because these parent strings are instances of good strings selected using the selection operator. In the single-point crossover operator search is not extensive, but the maximum information is maintained from parent to children. On the other hand, in the uniform crossover, the search is very extensive but minimum information is maintained between parent and children strings. If a crossover probability of P_c is used then 100 P_c% strings in the population are recycled in the crossover operation and 100(1-P_c)%of the population are simply copied to the new population.

iv)Mutation

Crossover operator is mainly answerable for the search aspect of genetic algorithms, even though the mutation operator is also used for this purpose casually. The mutation operator changes a 1 to a 0 and vice versa with a small mutation probability $P_{m:}$

0 0 0 0 0 $\Longrightarrow$ 0 0 0 1 0

In the above example, fourth gene has changed its value, thereby creating a new solution. The need for mutation is to maintain diversity in population. For example, if in a particular position along the string length all strings in the population have a value 0, and a 1 is required in that position to obtain optimum or a near-optimum solution, then mutation operator described above will be able to create a 1 in that position. The inclution of mutation introduces some probability of turning that 0 into 1. Furthermore, for local improvement of a solution, mutation is useful.

After reproduction, crossover, and mutation are tested to whole population, one generation of GA is completed. These three operators are uncomplecated and straightforward. The reproduction operator selects good strings and the crossover operator recombines good substrings from two good strings together to hopefully form a better substring. The mutation operator alters a string locally to surely create a better string. Even though none of these claims guaranteed and / or tested while creating a new population strings, it is expected that if bad strings are created they will be eliminated by the reproduction operator in next generation and if good strings are created, they will be emphasized. To make a faster convergence of a GA to real-world problems, problem-specific operators are often developed and used, but the above three operators portray fundamental operations of a genetic algorithm and facilitate a comparatively easier mathematical treatment.

4. Results

Whenever we execute the project we observe the result step by step which will be easily understandable for new user also.

Step I: load input MRI image:

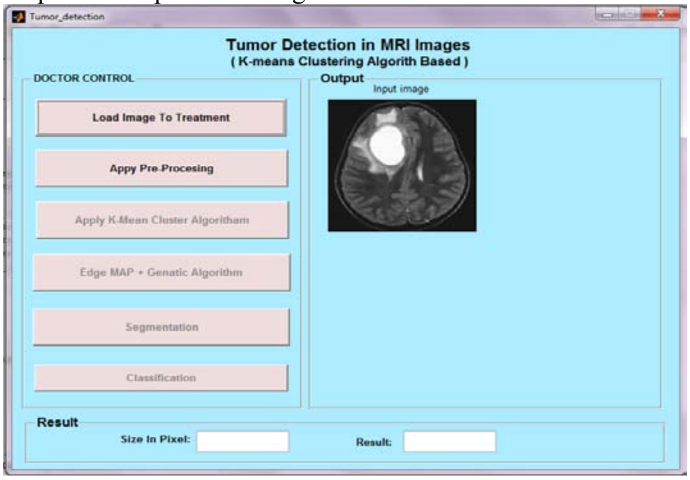


Figure 4: load image

Step II: apply pre-processing:

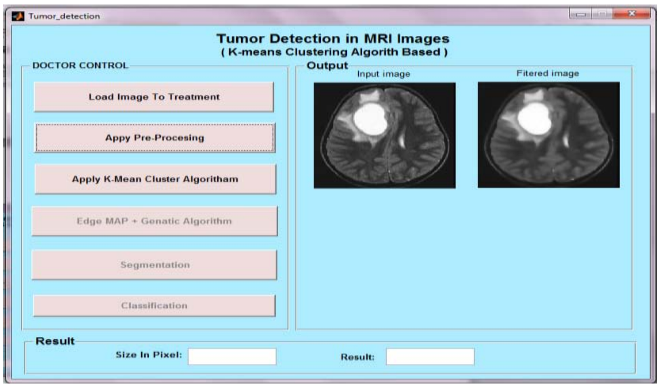
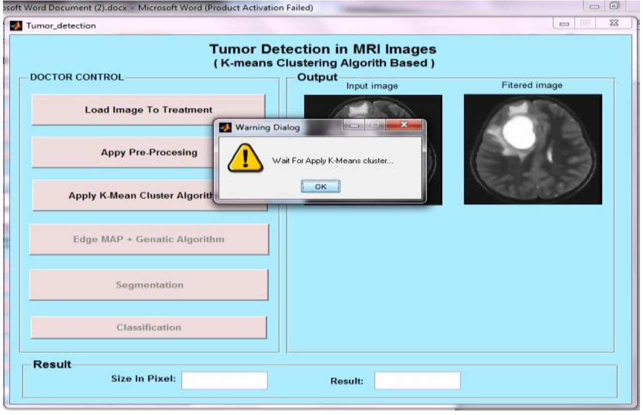


Figure 5: Pre-processing

Step III: apply k means clustering:



Wait For Apply K-Means cluster...

Figure 6: k means clustering

Step IV: Edge map+ genetic algorithm and segmentation

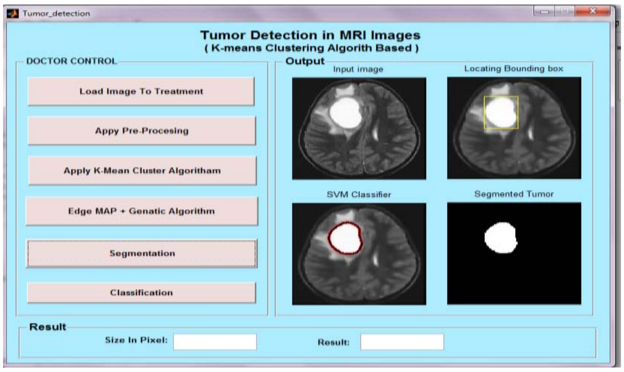
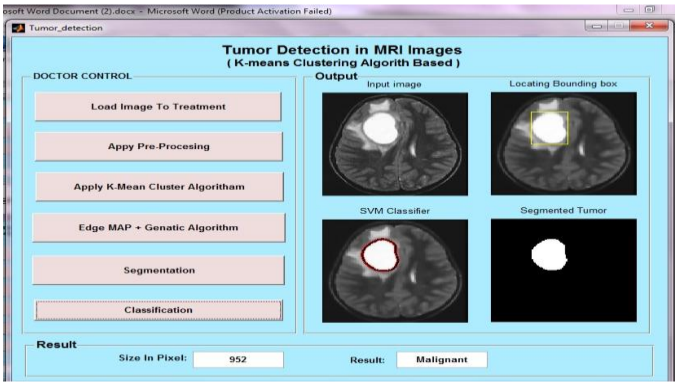


Figure 7: Genetic Algorithm

Step V:Tumor type and size in pixel:



Size In Pixel: 952

Result: Malignant

Figure 8: Tumor Size and Type

5. Conclusion

After observing and comparing the results of tumor detection using segmentation we can conclude that

- Segmentation was achieved for all the proposed methods tumor detection was done.
- The problem of over segmentation was also addressed.
- The genetic algorithm improves the convergence and computing time is reduced.

There are many points which we can improve in future as future work. These points are

- Improving the accuracy, exactness, and computational speed of segmentation.
- Minimizing the amount of manual interaction
- Implement the methods for 3D images of brain tumor MRI
- Volume and growth rate of tumor can be extracted to produce better segmentation results and their performance

References

[1] Kailash Sinha, G.R.Sinha. “Efficient Segmentation Methods for Tumor Detection in MRI Images”. *2014 IEEE Student’s Conference on Electrical, Electronics and Computer Science*.

[2] Kazuki Ichikawa,Shinichi Morishita, "A Simple but Powerful Heuristic Method for Accelerating k-Means Clustering of Large-Scale Data in Life Science", *IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, VOL. 11, NO. 4, JULY/AUGUST 2014*.

[3] Manali Patil,Samata Prabhu,Sonal Patil, Sunilka Patil, Mrs.Prachi Kshirsagar. “Brain Tumor Identification Using K-Means Clustering”. *International Journal of Engineering Trends and Technology- Volume4Issue3- 2013*.

[4] Kshitij Bhagwat, Dhanshri More, Sayali Shinde, Akshay Daga, Assistant Prof Rupali Tornekar “Comparative Study of Brain Tumour Detection Using K means, Fuzzy C Means and Hierarchical Clustering Algorithms ”. *International Journal of Scientific & Engineering Research, Volume 4, Issue 6, June-2013 626 ISSN 2229-5518*.

[5] Amrutha Ravi1, Sreejith S., "A Review on Brain Tumour Detection Using Image ", *International Journal of Emerging Technology andAdvanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 5, Issue 6, June 2015)*

[6] Rohini Paul Joseph1, C. Senthil Singh2, M.Manikandan3. “BRAIN TUMOR MRI IMAGE SEGMENTATION AND DETECTION IN IMAGE PROCESSING”. *IJRET: International Journal of Research in Engineering and Technology Volume: 03 Special Issue: 01 | NC-WiCOMET-2014 | Mar-2014,*.

[7] RAZAFIPAHATELO*D., RAKOTONIAINA S.,RAKOTONDRAOMPIANA S.” AUTOMATIC FLOODS DETECTION WITH A KERNEL K-MEANS APPROACH”, *2014 IEEE Canada International Humanitarian Technology Conference - (IHTC).*

[8] Anam Mustaqeem, Ali Javed, Tehseen Fatima. “ An Efficient Brain Tumor Detection Algorithm Using Watershed & Thresholding Based Segmentation”. *I.J. Image, Graphics and Signal Processing, 2012, 10, 34-39 Published Online September 2012 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijigsp.2012.10.05*

[9] Anubha Lakra, R.B. Dubey. “A Comparative Analysis of MRI Brain Tumor Segmentation Technique”. *International Journal of Computer Applications (0975 – 8887) Volume 125 – No.6, September 2015*

[10] Mr. Vishal Shinde, Miss. Priti Kine, Miss. Suchita Gadge, Mr. Shekhar Khatal. “Brain Tumor Identification using MRI Images*”. International Journal on Recent and Innovation Trends in Computing and Communication Volume: 2 Issue: 10 October 2014.*

[11] TIAN Jinlan ZHU Lin ZHANG Suqin LIU Lu . "Improvement and Parallelism of k>Means Clustering Algorithm". *TSINGHUA SCIENCE AND TECHNOLOGY ISSN 1007-0214 01/21 pp277-281 Volume 10, Number 3, June 2005*

[12] Baris Yuce, *Member, IEEE*, and Yacine Rezgui. "An ANN-GA Semantic Rule-Based System to Reduce the Gap between Predicted and Actual Energy Consumption in Buildings". *IEEE Transactions on Automation Science and Engineering.*

[13] Riddhi.S.Kapse, Dr. S.S. Salankar, Madhuri.Babar, "Literature Survey on Detection of Brain Tumor from MRI Images", *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 10, Issue 1, Ver. II (Jan - Feb. 2015), PP 80-86* .

[14] Akbar Shahrzad Khashandarag, Mirkamal Mirnia and Aidin Sakhavati. "A New Method for Medical Image Clustering Using Genetic Algorithm". *IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 1, January 2013 ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814* .

Author Profile

Arnavi Patil received her B.E. in Electronics and Telecommunication engineering and currently pursuing M.E. signal processing at D.Y.Patil school of engineering, Ambi, Talegoan, Pune.

Mr. Sanjay S. Badhe received his M.E. in Electronics. Currently working as Assistant Professor, Electronics and Telecommunication department at D.Y.Patil college of Engineering Ambi, Talegoan, Pune.

Mr. Ashish Mulajkar received his M.Tech in Embedded System. Currently working as Assistant Professor, Electronics and Telecommunication department at D.Y.Patil School of Engineering Academic Ambi, Talegoan, Pune.