A New Method for Medical Image Clustering Using Genetic Algorithm

Akbar Shahrzad Khashandarag[1], Mirkamal Mirnia[2] and Aidin Sakhavati[3]

[1] **Department of Mechatronic Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran.**

[2] **Department of Mathematics Science, Tabriz University, Tabriz, Iran.**

[3] **Department of Electrical Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran.**

Abstract

Segmentation is applied in medical images when the brightness of the images becomes weaker so that making different in recognizing the tissues borders. Thus, the exact segmentation of medical images is an essential process in recognizing and curing an illness. Thus, it is obvious that the purpose of clustering in medical images is the recognition of damaged areas in tissues. Different techniques have been introduced for clustering in different fields such as engineering, medicine, data mining and so on. However, there is no standard technique of clustering to present ideal results for all of the imaging applications. In this paper, a new method combining genetic algorithm and K-Means algorithm is presented for clustering medical images. In this combined technique, variable string length genetic algorithm (VGA) is used for the determination of the optimal cluster centers. The proposed algorithm has been compared with the K-Means clustering algorithm. The advantage of the proposed method is the accuracy in selecting the optimal cluster centers compared with the above mentioned technique.

Keywords: Medical Image, Clustering, Genetic Algorithm, K-Means.

1. Introduction

Segmentation is vital in the analysis of medical images which can be useful in many applications such as distinguishing the arteries borders from each other in angiography, the size of tumor and its response to treatment, interpretation of operation, the study of brain growth and recognition of tumor and so on. Therefore, it is obvious that segmentation is often used as the first and most important phase in the recognition and treatment of a disease in analyzing medical images like MRI images, and by the intuitive nature of the image, segmentation can be totally different. Segmenting an image refers to the technique of segmenting the space of an image into meaningful homogeneity areas with no overlaps which are the same in some features like, intensity or tissues [1]. The segmenting techniques are classified into two categories:

1. Edge-based techniques
2. Area-based techniques

In the former, both the borders presented in the images and their surroundings are found. While in the latter, the purpose is to initiate from image histogram, and is based on the intensity of pixel intensity whether it is less or more than the given amount of the given value. Clustering as a technique of segmentation, is a segmenting process in which series of information, being usually multidimensional, are divided into groups, so that the members of groups are the same in some criteria, while the members of different groups are different.

Clustering involves finding a structure in a cluster with data having no label. Clearly labels are often observable in medical images like MRI which are analyzed by the physician. When the given labels are not clear, the computer should be applied in labeling. The process of labeling and segmenting can be either synchronously or separately.

There have been different clustering techniques for the image segmentation such as, K-Means [2], Fuzzy C-means [3] and Average link [4] algorithms. These algorithms play important roles in the analysis of imaging in medicine, engineering and etc. In the most techniques mentioned above, the numbers of classes are applied as initiate input, where the clustering is defined as the distribution of N sample, in the space of n dimension in the group K [5][6]. One of the problems in these methods to determine the number of optimal classes for each of the images. Genetic algorithm as an optimal searching technique in the length of searching process is used for the optimal search of cluster center in medical images regarding the high volumes of image pixels [7-9]. One of the well-known clustering methods is K-Means algorithm which is implementable easily. Clustering by K-Means algorithm has high speed convergence; however, its accuracy is not satisfied for abnormal brains (such as tumor, swelling and

etc). Unfortunately, its original version has some objections like its dependence on the initial values of centers and convergence to the local optimal response [10]. Using genetic algorithm, the problems mentioned above have been removed. However, by combining two algorithms, incredible results have been obtained [11][12]. Different clustering algorithms present different results and evaluating these results is very important. Thus, cluster validity is an important challenge. Two main criteria of combination and separation are used for evaluating optimal clusters. Different accuracy criteria have already been proposed such as, Rand Index [13], Dunn's separation measure [14], Davies Bouldin [15], C-index [16], Adjusted Rand Index [17] and etc.

2. The Proposed method

The process of clustering of the proposed model is shown step by step as follows:

Input: Medical image.
Step1: Application of a smoothing filter with a 3×3 neighborhood to reduce the radio frequency noise and small effects of image.
Step2: Computing the number of maximum chromosomes ($k_{\max}$) using noiseless image histogram.
Step3: Application of variable string length genetic algorithm (VGA) for obtaining optimal cluster centers.
Step4: Application of K-Means algorithm clustering on the images using the optimal cluster centers.
Output: Clustered medical image.

2.1 Image pre-processing

MRI image is often the results of noise in the imaging environment of MRI device. One of these noises is the noise of radio frequency (RF) which is reduced by a smoothing filter with the 3×3 neighborhood. In Fig 1, the smoothing filter with the 3×3 neighborhood is shown.

2.2 Evaluation of cluster maximum

In this section, the number of maximum chromosomes ($k_{\max}$) is computed by using noiseless image histogram as one of the features of image. So that $k_{\max} = (rand()\%k^*) + 2$ and k^* is the number of given peaks appearing in image histogram.

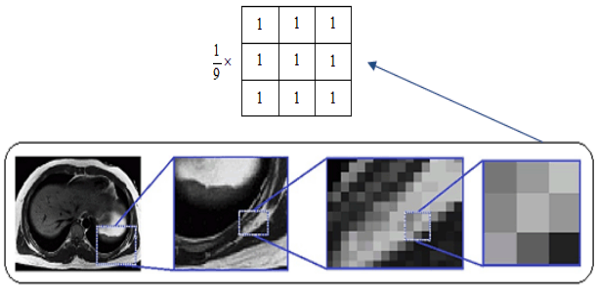

Fig. 1 Smoothing filter with the 3×3 neighborhood.

2.3 Variable string length genetic algorithm

In this paper, a genetic algorithm with a variable string length of chromosome has been presented. In the following, the implementation of the proposed genetic algorithm and the applications used in which will be explained.

2.3.1. Chromosome representation

In the proposed genetic algorithm, chromosomes are indicate the cluster centers. Supposing that the MRI images are two dimensions, l_i as the ith chromosome length which is $2 \times k_i$, and k_i as the number of clusters.

2.3.2. Generating initial population

For each i chromosome, the initial population is produced randomly from the surrounding borders of the image. After generating initial population, the qualification of each chromosome is obtained from the fitness function.

2.3.3. Constructing fitness function

The qualification of each member of community is determined by using fitness function after creating initial community. By using this fitness function, a number is assigned to each chromosome which is the value of that chromosome. This number is used as a merit determining the presence of this chromosome for the next generation.
To compute the fitness, suppose $V = \{v_1, v_2, v_3, ..., v_k\}$ be a chromosome that each of its members is the center of each cluster. Thus, fitness function is computed in equation (1).

$$F(v,x)=\frac{\left|v_i-v_j\right|}{\sum_{t=1}^{n_1}\left|v_i-x_l\right|+\sum_{p=1}^{n_2}\left|v_j-x_h\right|} \quad (1)$$

$$\forall i,j \quad i\neq j,\ 1\le t\le n_1,\ 1\le p\le n_2$$

Where $\left|v_i-v_j\right|$ is the distance between the cluster centers of i and j, $\sum_{t=1}^{n_1}\left|v_i-x_l\right|$ and $\sum_{p=1}^{n_2}\left|v_j-x_h\right|$ are the sum of differences between i and j cluster members and the cluster centers, n_1 and n_2 are the number of i and j cluster members. To obtain the optimal clustering, the purpose is an increase of the fitness is required which is equivalent to clustering with minimum distribution into clustering and minimum separation between cluster centers.

2.3.4. Defining genetic operators

Selection: In this paper, for selecting parental chromosomes, some are selected randomly among the chromosomes with low fitness functions. Then, two chromosomes among them with the highest quality are selected as parental chromosomes.

Crossover: After selecting parental chromosomes, the process of composing must be done. Let P_1 and P_2 be the parental chromosomes and C_1 and C_2 be the chromosomes obtained from the process of composing. From a random number (α), the process of combining will be applied on the parental chromosomes as shown in the equation (2) and (3).

$$C_1=\alpha\times P_1+(1-\alpha)\times P_2 \quad (2)$$

$$C_2=\alpha\times P_2+(1-\alpha)\times P_1 \quad (3)$$

The process of composing is shown in Fig 2.

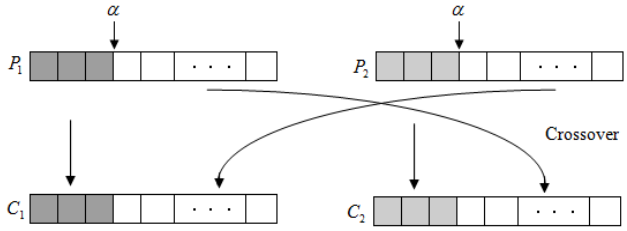

Fig. 2 Crossover operator.

2.3.5. Full genetic algorithm

The different processes of genetic algorithm with variable string length are shown as a follow:

Input: Medical image, maximum generation (G), size of population (P), crossover probability (P_c) and mutation probability (P_m).

Step1: Determining the maximum numbers of clusters ($k_{\max}$) through image histogram to create chromosomes using $k_{\max}$.

Step2: The initial values of chromosomes. Consider the number of generation (g) as 1.

Step3: Computing the fitness function for each chromosome in initial population.

Step4: Rearrange the population and assigning a degree for each chromosome in the population.

Step5: Selection among population chromosomes and then transferring them into mating pool. Apply the process of composing and mutation on the chromosomes which are in the mating pool to create child chromosomes.

Step6: Compute fitness function for each child chromosome.

Step7: Apply elitism by combining child and parent chromosomes, rearrange and finally, selecting the best chromosome to create the next generation. Add one unit to the number of generation ($g=g+1$).

Step8: If the maximum Iterative steps G is not reached ($g\le G$), go to Step 4.

Output: Select the qualified chromosome in the final population.

Being found the qualified chromosome with the optimal cluster centers, the process of clustering must be applies on the medical images using K-Means clustering algorithm.

2.3.6. K-Means clustering algorithm

K-Means algorithm is considered as a technique based on gravity center, which is a basic method for many clustering methods such as fuzzy clustering, because of its simplicity. Different forms have been presented for the K-Means algorithm. But all of them have the same repetitive procedures which try to estimate only a fixed number of clusters as follows:

- Obtaining some points as the cluster centers. In fact, these points are the averages of points in each cluster.
- Assigning each sample data to a cluster in which the given data has the least distance from the center of the cluster.

Alternatively, corresponding to the number of needed clusters, some points are selected randomly. Then, data is assigned to these clusters regarding the amount of similarities. So some new clusters are obtained. By repeating the same procedures, it is possible to compute new centers in each iterative through rearranging the given data and then assigning data to new clusters.
It is possible to explain the stopping in K-Means algorithm as follows:
Let $v_i^{(n)}$ be the center of i in step n and $v_i^{(n-1)}$ be the center of i in step $(n-1)$, in step 4 the equation of $\left|v_i^{(n)} - v_i^{(n-1)}\right| < \delta$ is investigated. If the difference between two quantities of some hierarchical center is less than the value of a pre-determined initial δ, then the algorithm will stop. Otherwise; the process will be repeated as many times as it is needed and finally, the number of repeats reaches to a certain level.

Step1: The point K is selected as cluster centers.
Step2: Each sample data is assigned to the cluster center with least distance from the given data.
Step3: After assigning all data to all clusters, a new point is computed for each cluster as a center one by one (The average of points related to each cluster).
Step4: If there is no change in the cluster centers ($\left|v_i^{(n)} - v_i^{(n-1)}\right| < \delta$), go to step 2.

---

3. Cluster validity measure

In this paper, Dunn index has been used to measure validity of cluster. Dunn index, which is one of the internal criteria for the accuracy of clustering, is obtained from equation (4).

$$D = \min_{i=1\ldots n_c}\left\{\min_{j=1\ldots n_c, j\neq i}\left(\frac{d(K_i,K_j)}{\max_{h=1\ldots n_c}(\Delta(K_h))}\right)\right\} \quad (4)$$

where $d(K_i,K_j)$ and $\Delta(K_i)$ are obtained from equations (5) and (6).

$$d(K_i,K_j) = \min_{x\in K_i, y\in K_j}\{d(x,y)\} \quad (5)$$

$$\Delta(K_i) = \max_{x,y\in K_i}\{d(x,y)\} \quad (6)$$

in which $d(K_i,K_j)$ is the distance from clusters i and j, $\Delta(K_h)$ is the maximum hth cluster diameter, $d(x,y)$ is the distance between the given x and given y and n_c is the number of clusters.
By combining criteria of clusters and separating criteria between clusters, the greater value for Dunn index will be an optimal value.

4. Experimental results

In the simulation of the proposed model, MRI images of brain and heart [18] have been used. For mutation and selection degrees, different values are considered. The maximum numbers of generations (g) is considered as 12 and the minimum numbers, is considered to be 4. The maximum numbers of chromosomes in each generation is considered to be 150 and the minimum numbers considered to be 30. The obtained results are shown in the Table1 and the Table2.

Table 1: The proposed model results for MRI brain image by different parameters.

#	P	G	P_m	P_s	$F(v,x)$		
					worst	***averag***	***best***
1	30	4	0.2	0.2	0.00657	0.00968	0.02486
2	30	8	0.15	0.5	0.31253	0.34973	0.72302
3	80	4	0.2	0.2	0.34522	0.45995	0.85545
4	80	8	0.15	0.5	0.00563	0.00789	0.15649
5	100	4	0.2	0.2	0.43050	0.55914	0.87859
6	100	8	0.15	0.5	0.32957	0.40699	0.91176
7	120	8	0.2	0.2	0.55791	0.65192	1.10000
8	120	4	0.15	0.5	0.29416	0.33681	0.93287
9	150	8	0.2	0.2	0.55396	0.64968	0.99788
10	150	4	0.15	0.5	0.30648	0.35442	0.90580

For example, the fourth test results are shown in Fig 3 and 4.

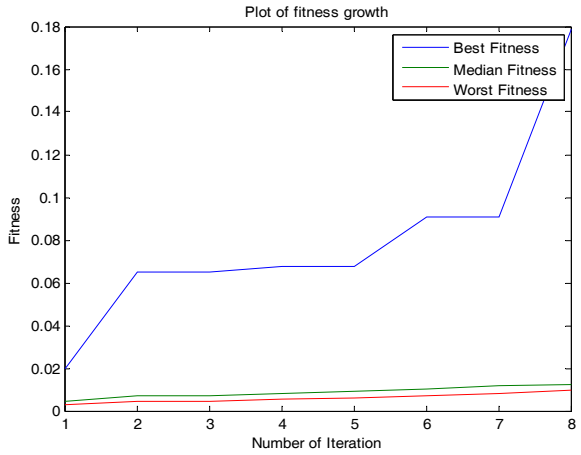

Fig. 3 Growth charts fitness function

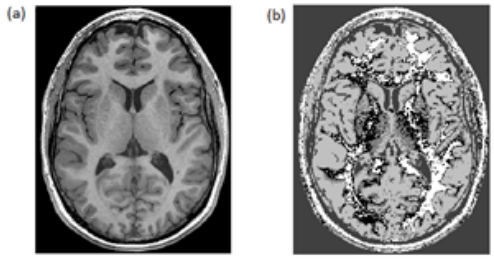

Fig. 4 MRI brain image

Table 2: The proposed model results for MRI heart image by different parameters.

#	P	G	P_m	P_s	$F(v,x)$		
					worst	***averag***	***best***
1	30	4	0.2	0.2	0.088837	0.130050	0.19349
2	30	8	0.15	0.5	0.031170	0.040669	0.23242
3	80	4	0.2	0.2	0.054473	0.071793	0.18693
4	80	8	0.15	0.5	0.048480	0.083582	2.40290
5	100	4	0.2	0.2	0.123110	0.167100	1.76150
6	100	8	0.15	0.5	0.039079	0.055477	0.33636
7	120	12	0.2	0.2	0.160360	0.257860	0.82045
8	120	4	0.15	0.5	0.023779	0.034712	0.28146
9	150	8	0.2	0.2	0.162350	0.252040	1.22620
10	150	12	0.15	0.5	0.083363	0.125290	4.47320

For example, the sixth test results are shown in Fig 5 and 6.

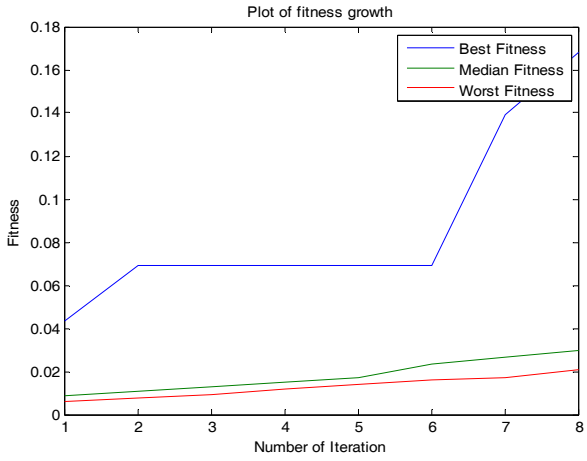

Fig. 5 Growth charts fitness function

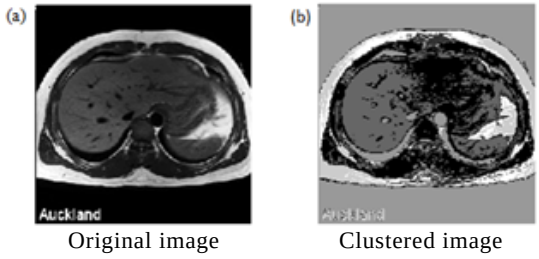

Fig. 6 MRI heart image

Different clustering algorithms show different results, and evaluating those results is very important. Thus, cluster validity has become a very important challenge. In this paper, the proposed model has been compared with the different clustering algorithms such as K-Means clustering algorithm. The comparison is based on the use of criteria and Dunn index, and its results are shown in Table3 and Fig 7. It is important to note that the greater value of Dunn index shows an optimal clustering.

Table 3: Clustering validation results for MRI images.

#	**K-Means**		**Proposed Method**	
	K	**Dunn index**	**K**	**Dunn index**
Image 1	7	0.2381	6	0.70000
Image 2	8	0.42857	6	0.66071
Image 3	7	0.73077	8	0.81818
Image 4	7	0.32203	7	0.40625

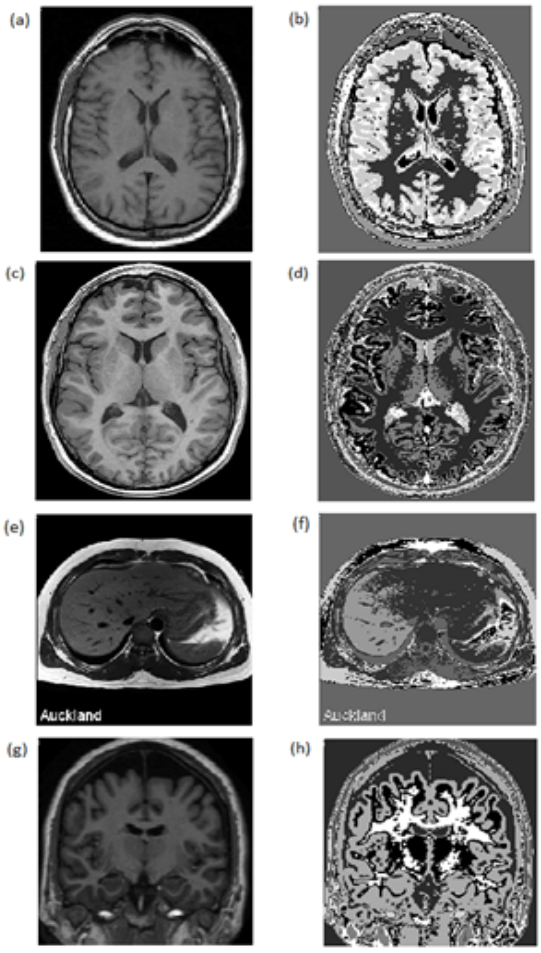

Fig. 7 (a,c,e,g) Original MRI images, (b,d,f,h) Clustered MRI images.

5. Conclusion

The results of investigation on clustering in medical images showed that K-Means algorithm and other similar algorithms are relatively more efficient in clustering normal brain MRI images with no noise and getting the best centers of cluster for clustering purposes must be done by random. Thus, using such methods are not suitable to measure the volume of tumor and its response to treatment, interpretation of operation, the study of brain growth, recognition of tumor and so on. Therefore, to overcome to this problem, a technique is required to be presented. In this paper, a new method has been introduced for clustering medical images, with respect to the importance and application of clustering in the segmentation of images. The advantage of the proposed method is obtaining the optimal cluster centers using variable string length genetic algorithm.

References

[1] R.C. Gonzalez, R.E. Woods, "Digital Image Processing", Addision-Wesley, Massachusette, 1992.

[2] T. Kanungo, D.M. Mount, N.S. Netanyahu, C. Piatko, R. Silverman, A.Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation", IEEE Trans. Patt. Anal. Mach. Intell. 24 (2002).

[3] R. L. Cannon, J. V. Dave, AND J. C. Bezdek, "Efficient Implementation of the Fuzzy c-Means Clustering Algorithms", IEEE Trans. Patt. Anal. Mach. Intell. 8, 2 (1986).

[4] P. Filzmoser, R. Baumgartner, E. Moser, "A hierarchical clustering method for analyzing functional MR images", Magnetic Resonance Imaging, Volume 17, Issue 6, July 1999, Pages 817-826.

[5] E. Forgey, "Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classification", Biometrics, vol. 21, p. 768, 1965.

[6] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. "A local search approximation algorithm for k-means clustering". Comput. Geom., 28(2-3):89–112, 2004.

[7] S. Cagnoni, A.B. Dobrzeniecki, R. Poli, J.C. Yanch, "Genetic algorithm-based interactive segmentation of 3D medical images". Image and Vision Computing, 17(12), 881-895, 1999.

[8] E. Angelié, P. J. H de Koning, H. C van Assen, M. Danilouchkine, G. Koning, R.J van der Geest, J.H.C Reiber, "Automatic tuning of left ventricular segmentation of MR images using genetic algorithms". International Congress Series, Volume 1256, June 2003, Pages 1102-1107.

[9] F. Xie, A. C. Bovik, "Automatic segmentation of dermoscopy images using self-generating neural networks seeded bygenetic algorithm". Pattern Recognition, In Press, Corrected Proof, Available online 29 August 2012.

[10] J. B. MacQueen "Some Methods for classification and Analysis of Multivariate Observations", Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp. 281–297 (1967).

[11] D. X. Chang, X. D. Zhang, C. W. Zheng, "A genetic algorithm with gene rearrangement for k-means clustering", Pattern Recognition, Volume 42, Issue 7, July 2009, Pages 1210-1222.

[12] Y. Liu, X. Wu, Y. Shen, "Automatic clustering using genetic algorithms", Applied Mathematics and Computation, Volume 218, Issue 4, 15 October 2011, Pages 1267-1279.

[13] W. M. Rand, "Objective Criteria for the Evaluation of Clustering Methods". Journal of the American Statistical Association, 1971, 66(336): 846-850.

[14] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". Journal of Cybernetics, 1973, 3(3): 32-57.
[15] D. L. Davies, D. W. Bouldin, "A Cluster Separation Measure", IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1 (2): 224–227.doi:10.1109/TPAMI.1979.4766909 (1979).
[16] L. J. Hubert and J. R. Levin. "A general statistical framework for assessing categorical clustering in free recall". Psychological Bulletin, 1976, 83, 1072-1080.
[17] P. N. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining". 2006.
[18] http://brainweb.bic.mni.mcgill.ca/brainweb/.

Akbar Shahrzad Khashandarag was born in Tabriz, Iran, in 1978. He received the B.Sc. degree in computer engineering from the Islamic Azad University Bonab Branch, Iran in 2009. He is currently M.Sc. student in Mechatronics engineering at the Islamic Azad University Tabriz Branch. His research interests include image processing, signal processing and wireless sensor network.

Dr. Mirkamal Mirnia received the B.Sc. degree in mathematical Physics from Ferdowsi University, Mashhad, Iran in 1967, and M.Sc. degree in pure mathematics from teacher training University of Tehran in 1969 also M.Sc. degree in numerical analysis and computing from Owen university of Manchester, UK in 1975 and PhD. in applied mathematics (optimization) from university of St.Andrews, UK in 1979. He is a member of the Mathematical society of Iran, Iranian operations researches, institute of applied mathematics, UK and member of SIAM.

Dr. Aidin Sakhavati was born in Orumieh, Iran, in 1978. He received his BS, MS and Ph.D. degrees in 2000, 2003 and 2010 from Islamic Azad university-Tabriz branch (IAUT), and Tabriz University, Tabriz, Iran, and Islamic Azad University, science and research branch, Tehran, Iran, respectively. He has been holding the Assistant Professor position at IAUT since 2010. He is the author of more than 30 journal and conference papers. His teaching and research interest include power system and transformers transients and power electronics applications in power systems. His researching interests include power systems dynamic and control application of Quantitative Feedback Theory, Particle Swarm Optimization and Genetic Algorithm in FACTS devices and Load Frequency control.