

Ensemble Semi-supervised Frame-work for Brain Magnetic Resonance Imaging Tissue Segmentation

Reza Azmi, Boshra Pishgoo, Narges Norozi, Samira Yeganeh

Departments of Computer Engineering, Alzahra University, Tehran, Iran

Submission: 14-03-2013 Accepted: 15-04-2013

ABSTRACT

Brain magnetic resonance images (MRIs) tissue segmentation is one of the most important parts of the clinical diagnostic tools. Pixel classification methods have been frequently used in the image segmentation with two supervised and unsupervised approaches up to now. Supervised segmentation methods lead to high accuracy, but they need a large amount of labeled data, which is hard, expensive, and slow to obtain. Moreover, they cannot use unlabeled data to train classifiers. On the other hand, unsupervised segmentation methods have no prior knowledge and lead to low level of performance. However, semi-supervised learning which uses a few labeled data together with a large amount of unlabeled data causes higher accuracy with less trouble. In this paper, we propose an ensemble semi-supervised frame-work for segmenting of brain magnetic resonance imaging (MRI) tissues that it has been used results of several semi-supervised classifiers simultaneously. Selecting appropriate classifiers has a significant role in the performance of this frame-work. Hence, in this paper, we present two semi-supervised algorithms expectation filtering maximization and MCo_Training that are improved versions of semi-supervised methods expectation maximization and Co_Training and increase segmentation accuracy. Afterward, we use these improved classifiers together with graph-based semi-supervised classifier as components of the ensemble frame-work. Experimental results show that performance of segmentation in this approach is higher than both supervised methods and the individual semi-supervised classifiers.

Key words: Brain magnetic resonance image tissue segmentation, ensemble semi-supervised frame-work, expectation filtering maximization classifier, MCo_Training classifier

INTRODUCTION

“Magnetic resonance imaging (MRI) is an imaging technique that used primarily in the clinical diagnosis and biomedical research to produce high resolution and contrast images of the soft-tissue anatomy.”^[1] By increasing size and number of medical images, the use of computers in their processing and analysis is necessary. Segmenting different tissues in MRIs of the human brain is an important pre-processing stage for different tasks.

There are many methods for brain MRIs tissue segmentation. These methods can be categorized into three groups: (i) contour-based segmentation such as active contour algorithm.^[2,3] (ii) Region-based segmentation techniques.^[4] (iii) Classification-based segmentation that includes supervised and unsupervised methods. Supervised segmentation methods such as neural networks^[5] and support vector machines^[6] lead to high accuracy but they require a large amount of labeled data, which is hard, expensive and slow to obtain. Furthermore, they cannot use unlabeled

data to train classifiers. On the other hand, unsupervised learning methods such as Marcov random field (MRF),^[7] Fuzzy C means^[8] and spectral clustering,^[9] remove the costs of labeling and do not use the label of training data. So, these methods have no prior knowledge and will have lower performance respect to supervised methods.

To overcome these problems, in this paper, we propose a semi-supervised approach for segmenting of brain MRIs tissues. Recently, several semi-supervised algorithms such as Co_Training,^[10] Transductive Support Vector Machines (TSVMs),^[11] expectation maximization (EM)^[12] and graph-based methods^[13] have been presented although none of them have been used for MRI segmentation. The proposed method in this paper is an ensemble frame-work that uses the results of several semi-supervised classifiers simultaneously. Figure 1 illustrates an overall view of this frame-work.

In this figure, first, several different semi-supervised classifiers are trained. Then, test data labeling is carried out

Address for correspondence:

Mrs. Boshra Pishgoo, Department of Computer Engineering, Alzahra University, Tehran, Iran. E-mail: boshra.pishgoo@gmail.com

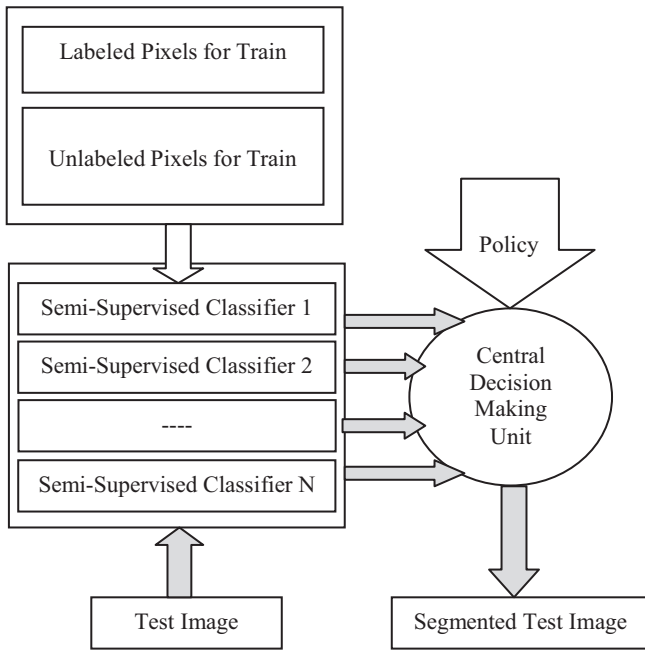


Figure 1: An overall view of ensemble semi-supervised framework. White arrow shows training and gray arrows show test image segmentation process

by using all trained classifiers and a central decision making unit. This unit also uses some policies to make the appropriate decision. Clearly, selecting appropriate classifiers and using efficient decision making policies have important roles in the performance of this framework. For this purpose, in this paper, we present two semi-supervised algorithms that are improved versions of semi-supervised classifiers EM and Co_Training. Then, we use these improved classifiers together with graph-based semi-supervised classifier as the components of the ensemble approach.

This paper is organized as follows: In section 2, we introduce three methods that were used for extracting features from brain MRIs. Semi-supervised classifiers including expectation filtering maximization (EFM), MCo_Training and graph-based method are presented in section 3. The proposed ensemble semi-supervised framework will be explained in section 4. Section 5 investigate the experimental results of our framework and compare them by supervised and individual semi-supervised classifiers. In final, Discussion and conclusion comes in section 6.

FEATURE EXTRACTION

Pixel classification is a common method that models image segmentation task as a labeling problem. In this model, each MRI image with its segmentation can be shown by the pair $\langle X^{MRI}, Y \rangle$ that is defined on 2D lattice $P = \{p_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n\}$, where P is a set of pixels, m and n are image dimensions, $X^{MRI} = \{\vec{x}_k \mid k \in P\}$, \vec{x}_k is the feature vector of pixel k , $y = \{y_k \in \Gamma \mid k \in P\}$, y_k is assigned label of pixel k , $\Gamma = \{\lambda_1, \lambda_2, \dots, \lambda_M\}$ is a set of labels that are associated with

tissues and M is the number of different brain tissues. Using this description, the purpose of image segmentation is to assign appropriate labels to all pixels in a MRI image.

The feature vector \vec{x} , can be obtained using various feature extraction methods. The intensity-based feature extraction method is very straightforward. In this method $x^{MRI} = \{\vec{x}_k \in G \mid k \in P\}$ and $G \in [0, 255]$ is the range of gray level. However, intensity-based method is not sufficient for good segmentation.^[14] Therefore, in this paper we extract three types of features: Stationary wavelet transform (SWT) features, edge features, and fractal features. Then, we combine these features with intensity of each pixel and the horizontal and vertical location of it to create a new feature set with totally 65 features for each pixel (25 SWT features, 25 edge-based features, 12 fractal features, intensity feature and horizontal and vertical locations of pixel).

Combining these features for each pixel may decrease both accuracy and speed. Therefore, selecting an appropriate subset of them has a significant role in improving accuracy and managing complexity. In this paper, we use a forward feature selection method for selecting nine salience features and identify them as final selected feature set. The first three features of this set is belong to edge-based feature set (the 9th, 20th and 21th features of edge-based feature set), the next three features is belong to fractal feature set (the 9th, 11th and 12th features of fractal feature set), the next feature is the 6th feature of SWT feature set and two last features are intensity and horizontal location of each pixel. In next subsections, we describe these methods by details.

Stationary Wavelet Transform Feature

WT^[15] is used frequently in feature extraction of MRIs. The WT produces several representation of image at various resolutions and also captures both frequency and location information. The discrete wavelet transform (DWT) is a powerful implementation of the WT using the dyadic scales and positions. DWT can be expressed as Eq. 1 Where cA_{jk} and cD_{jk} refer to the approximation and the detailed components coefficients respectively and $l(n)$ and $h(n)$ represent the low and high-pass filters.^[15-17]

$$cA_{jk}(n) = DS \left[\sum x(n) l_j^+(n - 2^j k) \right]$$

$$cD_{jk}(n) = DS \left[\sum x(n) h_j^+(n - 2^j k) \right] \quad (1)$$

However, "DWT suffers from time variant property. This means that if a MR image is a little shifted, the features extracted by DWT change notably."^[16,17] Hence, in this paper, we use SWT that is a WT algorithm designed to overcome the lack of translation-invariance of the DWT.^[18] Translation-invariance is achieved by removing the up and down samplers in the DWT and up sampling the filter coefficients by a factor of 2^{j-1} in the j^{th} level of the algorithm.

In this study, we performed a 3-level SWT on MRIs and extracted 25 features for each pixel.

Edge Features

Edge detection operators are based on this idea that edge information in an image are found by looking at the relationship a pixel has with its neighbors.^[19] An edge is defined by a discontinuity in gray-level values. In edge detection methods two masks that is G_x and G_y , are used to approximate the gradient of an input image x^{MRI} to determine edge points. Edge detectors can perform well on uncorrupted images, but are highly sensitive to noise. To reduce the effect of noise, it's attempted to find edges in smoothed images rather than the original ones. Here, first the 3×3 Gaussian filter is used to remove the noisy pixels from the original MRIs. Then, we applied edge detection operators Sobel and Laplacian on smoothed images. We obtain 25 edge-based features using the output of derivative filters $(d_x, d_y, d_{xy}, d_{xx}, d_{yy})$ at five different scales.

Fractal Features

"Fractals are geometric objects that have a non-integer fractal dimension (FD). These objects are found in many places in nature, including mountains and coastlines. Some parts of the human body, such as the lungs, brain tissues, and tumors also appear to grow in the form of fractals or exhibit fractal characteristics."^[20] So, we use FD analysis to extract features for brain MRIs tissue segmentation and calculate FD by Blanket, Piecewise Modified Box-Counting (PMBC) and Piecewise Triangular Prism Surface Area (PTPSA) that are three FD computation algorithms.

First, we divide each image into sub-images with specific dimension. Then, we use one of the three algorithms to compute the FD for each sub-image. In PMBC algorithm,^[20,21] Each sub-image was divided into boxes of several sizes, and the difference between the maximum and minimum gray values was computed for each box. After that, through a final processing, FD is calculated for that sub-image. In PTPSA^[20,22] technique as in PMBC, each sub-image was divided into several boxes. Next, a more accurate procedure using the four corners and center gray values was used to gain FD for sub-image.

Unlike two preceding algorithms, the blanket method^[20] does not divide each sub-image into boxes. Instead, it uses an iterative process for comparing each pixel with its surrounding pixels and finally calculates FD for each sub-image. In all three algorithms, after calculating FD value for each sub-image, this value is assigned to all pixels in it. If we divide each 256×256 image in to 4×4 , 8×8 , 16×16 , and 32×32 sub-images and use three algorithms to compute the FD for each sub-image, then we will get 12

FD values for each pixel.

SEMI SUPERVISED CLASSIFICATION METHODS

As mentioned in previous sections, the purpose of image segmentation is to assign appropriate labels to all pixels in an MRI Image. This can be achieved using supervised and/or unsupervised methods. In supervised method, pixels of images with their labels are used for training learners. Then, these learners are used to label new images. However, in unsupervised methods, there is not any labeled image in training step and image segmentation process is performed by exploiting structure and information existing in unlabeled pixels. Unlike supervised and unsupervised methods, semi-supervised methods train the appropriate classifier using a few labeled data and many unlabeled data as well. In this section, we introduce some semi-supervised methods that are used in our proposed frame-work.

EFM Algorithm

Expectation maximization (EM) is an algorithm that provides a flexible approach for clustering. For the first time, McCallum and Nigam^[23] in 1998, applied this probabilistic learning algorithm for semi-supervised learning. Semi-supervised EM first creates an initial weak classifier based solely on the labeled examples. Then, it repeatedly performs a two-step procedure: At the first step (E), it assigns probabilistically label to all unlabeled examples and at the second step (M), it learns a new maximum a posteriori (MAP) hypothesis based on the examples labeled with high probability in the previous step.^[24]

As mentioned above, EM algorithm in step E, assigns probabilistically label to all unlabeled examples; but in practice, some of these assigned labels are incorrect. This can decrease the accuracy of the classifier that will learn in step M and stop progressing in the next iterations of the algorithm. To solve this problem, a Modified form of EM algorithm called EFM is presented here.

EFM is a semi-supervised multi-learner algorithm that first, creates N initial weak classifiers based solely on the labeled examples by different learning algorithms. Then, it repeatedly performs a three-step procedure: At the first step (E), probabilistically label is assigned to all unlabeled examples by N learners. Therefore, in this step, each unlabeled data has N different labels and N different probabilities of labels. At the second step (F), algorithm selects some reliable unlabeled data for which (i) all N assigned labels are equal, and (ii) each probability label is greater than a specific threshold that is determined by problem information. Unlabeled data that do not satisfy these two constraints are filtered, but the others are called Reliable_Unlabeled data. These data are added to labeled

data examples and removed from unlabeled data in the next iteration. At the third step (M), a new MAP hypothesis is learned based on initial labeled data and Reliable_Unlabeled examples. At the end, N learned classifiers cooperate to label new data.

MCo_training Algorithm

Co_training algorithm was proposed by Blum and Mitchell^[25,11] in 1998. In this algorithm, first, two classifiers are trained with a few labeled training data using two views. Then, each classifier classifies the unlabeled data, chooses a few unlabeled examples whose labels have been predicted most confidently and adds those examples to the labeled training set. The classifiers are retrained and the process repeats. According to,^[25] there are three initial assumptions in the Co_training framework that guarantee that this algorithm works well: (i) Feature sets can be split into two naturally independent views, (ii) each view of the dataset is sufficiently good for training learners (sufficiency assumption) and (iii) the two views are conditionally independent given the class (independence assumption).

However, many datasets in the real world do not satisfy these assumptions. To overcome this problem, first a method was presented based on training classifiers with the same view that ignored the first assumption of Co_training algorithm.^[26] Afterward, Nigam and Ghani^[27] empirically studied the performance of standard Co_training algorithm in this case. Their experimental results suggested that when the feature set is sufficiently large, randomly splitting the features and then conducting standard Co_training may lead to a good performance. However, there are some problems with small feature sets in which random splitting cannot lead to a good performance. To solve this problem, an improved version of Co_training algorithm called MCO_training (Modified Co_training) is presented here.

The main idea of this algorithm is to create two required views of Co_training algorithm via feature selection methods. For this reason, first two subsets of initial features that have the best performance based on respective classifiers are selected by two different feature selection algorithms. Afterward, training classifiers in each step is carried out via the selected feature sets. Remaining steps are carried out in the same manner with Co_training algorithm except that to increase performance we use defined conditions in filtering step of EFM algorithm. At the end, two learned classifiers cooperate to label new data.

Graph-based Algorithm

Graph based semi-supervised methods are composed of two phases, graph construction, and label inference. Graph construction is a crucial step, which affects the performance.

In constructed graph, nodes are labeled and unlabeled examples and edges represent similarities between them. Graph construction involves three selections: (i) Similarity function for computing affinity matrix, (ii) sparsification method, and (iii) reweighting method.^[28]

In label inference phase, known labels diffuse to all the unlabeled nodes in the graph through semi-supervised learning algorithms such as Gaussian random field (GRF) method,^[29] local and global consistency (LGC) method^[30] and the graph transduction via alternating minimization method.^[31] These methods can be viewed as estimating a function f on the graph that should be close to the given labels y_L on the labeled nodes, and be smooth on the whole graph. More details about graph based methods can be found in references.^[32] After labeling the unlabeled data through label inference algorithm, K-nearest neighbors (KNN) classifier is trained through all examples and we use this classifier to label new data.

ENSEMBLE SEMI-SUPERVISED FRAMEWORK

In this paper, an ensemble semi-supervised framework is presented for brain tissues segmentation that uses the results of each three algorithms introduced in section 3. Figure 2 shows, the training process and the segmentation of a test image in our proposed model.

In training step, first three feature sets are extracted for each pixel of training images according to section 2. Then, in each image, some pixels are randomly chosen as training pixels. A few of these pixels are labeled and others remain unlabeled. Afterward, a feature selection process is performed to find salience features. Three semi-supervised classifiers explained in section 3 are trained using labeled and unlabeled samples. Each classifier can determine the label of unseen input pixels and output probability as well. In Figure 2 white arrows show the training process.

For segmenting an image, first salience features are extracted for each pixel of it. Then, these pixels are entered to trained classifiers for labeling. Each classifier produces two matrixes of size $m \times n$ for each $m \times n$ pixels image. These matrixes that are called *ClassifierName_Label* and *ClassifierName_Prob* show allocated label to each pixel and its probability, respectively. Hereby, after labeling step, three labels are assigned to each pixel that may be similar or not. Three label matrixes and three probability matrixes are entered into decision making unit. This unit assigns final labels to each pixel based on input information and decision policies. These policies will be described in the following subsections. In Figure 2, the gray arrows show the image segmentation procedure.

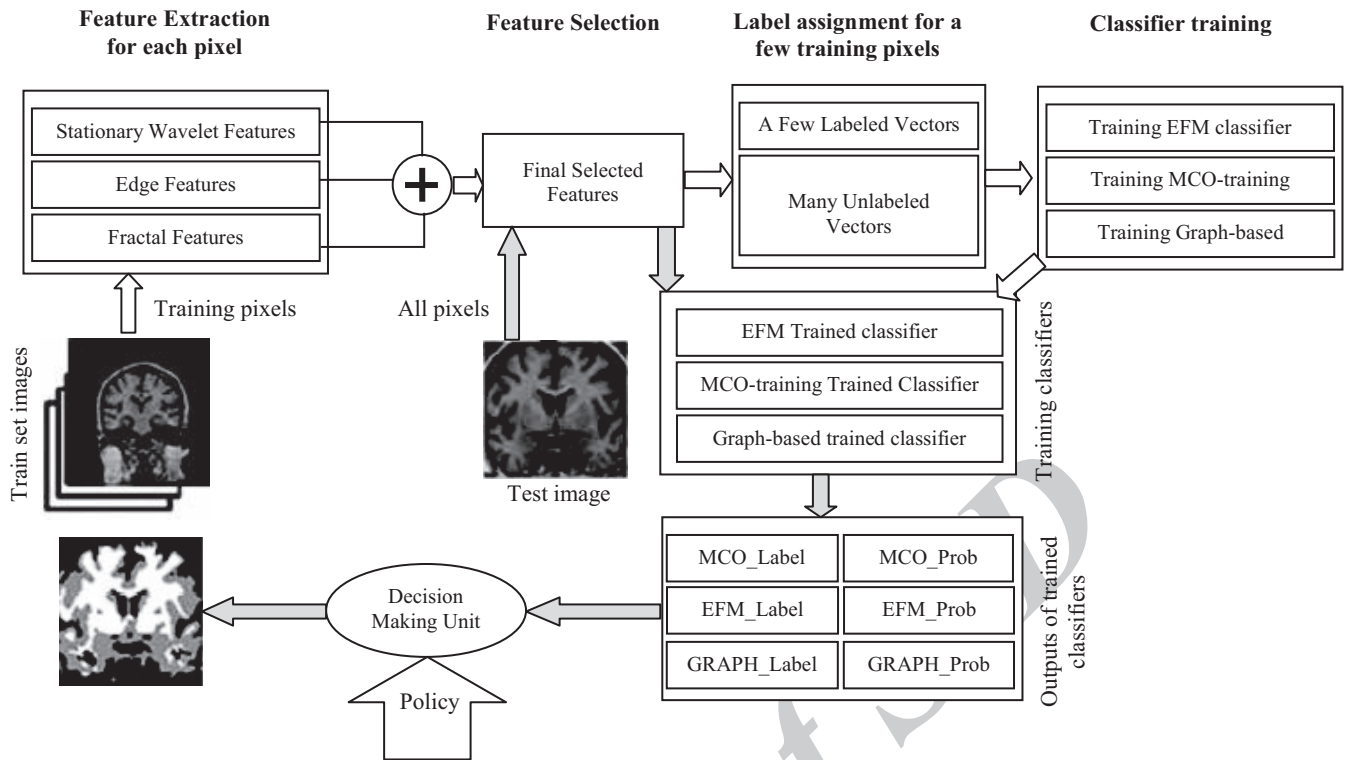


Figure 2: Illustration of ensemble semi-supervised framework. White arrows show training and gray arrows show test image segmentation process

Simple Voting Policy

This policy is carried out in two steps. At the first step, decision making unit performs voting on the results of three classifiers for each pixel and assigns the winning label to that pixel as temporary label. Temporary label for pixel (s, t) can be calculated using Eq. (2)

$$Temp_label(s, t) = \arg \max_{\lambda_q \in \Gamma} \left(\sum_{k=1}^3 \alpha_k(s, t) \right) \quad (2)$$

Where $\Gamma = \{\lambda_1, \lambda_2, \lambda_3\}$ is a set of associated labels to three white, black and gray brain tissues and $\alpha_k(s, t)$ is defined through Eq. 3.

$$\alpha_k(s, t) = \begin{cases} 1 \rightarrow label(s, t, k) = \lambda_q \\ 0 \rightarrow otherwise \end{cases} \quad (3)$$

Where label is a matrix with the size of $m \times n \times 3$. It contains all labels that different classifiers assign to pixels of the image. For example, label can be defined for pixel (m, n) as follow:

According to equations 2 and 3, decision making unit performs voting on the results of three classifiers for each pixel and assigns the winning label to that pixel as temporary label. After assigning temporary labels to all pixels, second step begins. In this step, final label for each pixel will be the one that gets the maximum number of votes between temporary labels of its surrounding neighbors. Eq. 4

presents this process for each pixel in location (i, j) .

$$Final_label(i, j) = \arg \max_{\lambda_m \in \Gamma} \left(\sum_{s=i-1}^{i+1} \sum_{t=j-1}^{j+1} \beta(s, t) \right) \quad (4)$$

Where the $Final_label(i, j)$ is the final label allocated to pixel (i, j) , s and t are variables for defining neighborhood around pixel (i, j) and $\beta(s, t)$ is defined by Eq. 5.

$$\beta(s, t) = \begin{cases} 1 & \text{if } Temp_{label(s, t)} = \lambda_m \\ 0 & \text{other} \end{cases}$$

Probabilistic Voting Policy

As mentioned in the previous section, the simple voting policy only uses allocated labels of each pixel and its neighbors in the label predicting process. However, probabilistic voting policy uses both labels and their probabilities. Like simple voting, this policy is performed in two steps. At the first step, decision making unit performs probabilistic voting on the results of the three classifiers for each pixel and assigns the winning label to that pixel as temporary label. For example, if predicted labels by three classifiers are λ_1, λ_2 and corresponding probability of these labels are 0.3, 0.4 and 0.8, then the scores of the labels $\lambda_1, \lambda_2, \lambda_3$ equal to 0.7, 0.8, and 0 respectively. Therefore, with this policy, λ_2 will be the temporary label of that pixel whereas in this situation, simple voting policy selects λ_1 . Temporary label for pixel (s, t) can be calculated using Eq. (6)

$$Temp_label(s,t) = arg \max_{\lambda_q \in \Gamma} \left(\sum_{k=1}^3 (\alpha_k(s,t) \times prob(s,t,k)) \right) \quad (6)$$

Where $\alpha_k(s,t)$ is defined by Eq. 3 and $prob$ is a matrix with the size of $m \times n \times 3$. This matrix contains all label probabilities that different classifiers assign to pixels of the image. For example, $prob$ can be defined for pixel (m,n) as follow:

$$prob(m,n,1) = MCO_Prob(m,n)$$

$$prob(m,n,2) = MCO_Prob(m,n)$$

$$prob(m,n,3) = MCO_Prob(m,n)$$

After allocating temporary label to all pixels, next step is performed like the second step of the simple voting policy. In probabilistic voting policy, final label allocation can be shown by Eq. 4 in which we can get $\beta(s,t)$ through Eq. 5.

Maximum Probability Policy

This policy is performed in two steps and leads to better results compared to two other policies. In the first step, decision making unit temporarily assigns a label with maximum probability to the pixel by evaluating proposed labels and their probabilities with three classifiers. Temporary label for pixel (s,t) can be calculated using Eq. (7).

$$Temp_label(s,t) = label \left(s,t, arg \max_{k \in \{1,2,3\}} prob(s,t,k) \right)$$

The next step is performed like step 2 in previous policies. Here, the final label prediction for each pixel (i,j) can be done through Eq. 4 and we can get $\beta(s,t)$ using Eq. 5.

EXPERIMENTAL RESULTS

In this section, the performance of our semi-supervised framework is investigated using Internet Brain Segmentation Repository (IBSR) dataset that were provided by the Center for Morphometric Analysis at Massachusetts General Hospital.^[33] This dataset includes brain images and their ground truth (GT) segmentation. GT is used as a reference for performance evaluation of segmentation methods in our experiments.

Here, we used 24 images for the training process that one of them is labeled and the others are unlabeled images. Then, 180 pixels are chosen randomly from each image and three sets of features are extracted for each of them according to section 2. Hereby, we have the total of 180 labeled and 4140 unlabeled pixels for the training classifiers. By applying forward feature selection method on training examples, nine salience features are selected, which lead to higher speed in addition to appropriate accuracy.

In the next step, six brain images that are different in terms of shape and structure of tissues are used as the test images. These images and their GTs have been shown in

two first rows of Figure 3. Then, nine selected features are extracted from all pixels of the test images and these pixels are labeled by the trained classifiers. Finally, these labels are used for performance evaluation of different classifiers.

Evaluation Criteria

To investigate classifiers we use three criteria; accuracy, precision, and energy of images. These criteria will be described in the following subsections.

Accuracy

The accuracy criterion is a known concept in the evaluation of learning methods, which is described in reference^[34] for two classification problem. Here, we only customized this criterion for our tissue segmentation problem, which is a three classification problem. This between assigned labels by classifier C and real labels in GT. Through this description, the accuracy of classifier C is obtained using Eq. 8.

$$Accuracy(C) = \frac{Tr[Eval_Matrix]}{m \times n} \quad (8)$$

Where, m and n are dimensions of the test image and $Tr[\cdot]$ is trace operator. $Eval_Matrix$ is a $M \times M$ matrix that M is the number of brain tissues (here $M = 3$). Each (i,j) element of this matrix shows the number of pixels that classifier C assigns label λ_i to them where their actual label is λ_j according to GT. Through this description, if $i = j$, then, classifier C has a correct prediction. Therefore, by applying trace operator (an operator that sums the diagonal elements of a matrix) on the $Eval_Matrix$, the number of pixels that are correctly predicted by classifier C can be obtained.

Precision

The Precision criterion is a known concept in the evaluation of learning methods, which is described in reference^[34] for two classification problem. Here, we only customized this criterion for our tissue segmentation problem, which is a three classification problem. Unlike accuracy, precision criterion is used to measure reproducibility or repeatability of assigning a label in the same conditions. Therefore, precision of classifier C can be obtained through Eq. 9.

$$Precision(C) = \sum_{j=1}^M \left(\frac{Eval_Matrix(i,j)}{\sum_{i=1}^M Eval_Matrix(i,j)} \right) / M \quad (9)$$

Energy function

Before describing this criterion, first we investigate MRF method. MRF is an unsupervised classification method in image segmentation that models spatial coherence among pixels using a neighboring system. In 1971, a new theory was ascribed to MRF by Hammersley and Clifford. Based on this theory, random field formed by assigned labels to an

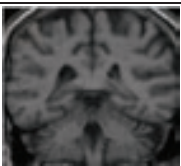
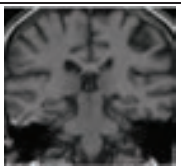
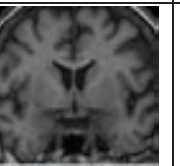
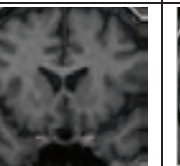
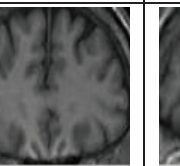
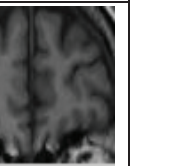

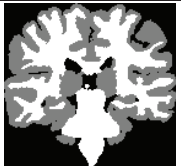

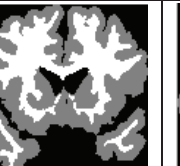
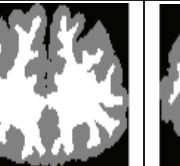
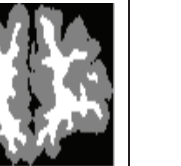

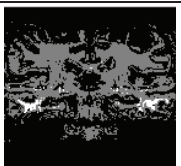
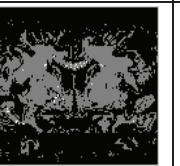
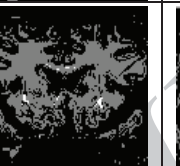
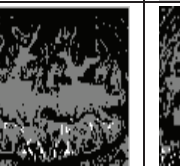

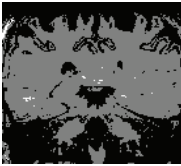
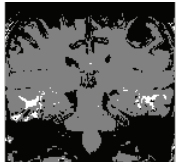
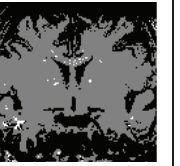
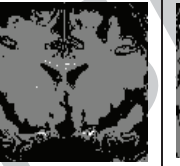
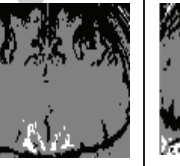

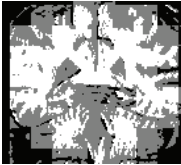



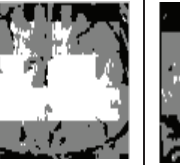

	Segmentation Result					
	Test Image 1	Test Image 2	Test Image 3	Test Image 4	Test Image 5	Test Image 6
Original MR Images						
Ground Truth						
SVM						
KNN						
Bayesian						

Figure 3: Segmentation results for supervised methods: Support vector machine, K-nearest neighbors and Naïve Bayesian

image is a MRF if and only if, its probability distribution be Gibbs distribution according to Eq. 10^[7].

$$P(y) = \frac{1}{Z} \exp\left(\frac{-E(y)}{T}\right) \quad (10)$$

Where y is an array of all assigned labels to an image, Z is normalization constant, T is temperature parameter and $E(y)$ is defined by Eq. 11.^[7]

$$E(y) = \sum_{i=1}^m \sum_{j=1}^n \sum_{k \in N} v(k, y(i, j)) \quad (11)$$

Where $N = \{y(i-1, j), y(i+1, j), y(i, j-1), y(i, j+1)\}$ is a set of 4 immediate neighbors of $y(i, j)$, m and n are image dimensions and $v(k, q)$ is defined according to Eq. 12.^[7]

$$v(k, q) = \begin{cases} -1 \rightarrow \text{if } k = q \\ +1 \rightarrow \text{if } k \neq q \end{cases} \quad (12)$$

Now, we should find the labels of the pixels (means y array) in a way that minimizes energy function of Eq. 13.^[7]

$$\text{Energy_Function} = \sum_{i=1}^m \sum_{j=1}^n \left(\frac{(X(i, j) - \mu_i)^2}{2\pi\sigma_i^2} + \ln\sqrt{2\pi\sigma_i} \right) + \frac{E(y)}{T} \quad (13)$$

Where $l = y(i, j)$, $X(i, j)$ is the gray value of pixel (i, j) . μ_l and

σ_l^2 are the mean and covariance for gray level of pixels with label λ_l respectively.

In this paper, however, we do not use MRF method for image segmentation. Here, we use energy function of Eq. 13 as the criterion for performance evaluation of presented classifiers. It is clear that unlike accuracy and precision criteria, lower values of the energy function in a constant temperature show more appropriate performance of classifiers in brain image segmentation.

Performance Evaluation for Supervised Classifiers

Before evaluating the segmentation performance in semi-supervised methods, first we investigate the performance of supervised methods in MRI image segmentation. For this purpose, three supervised classifiers (KNN with $k = 10$), Support Vector Machine (SVM), and naïve Bayesian are trained separately using 180 labeled data. Then, we compute accuracy and precision of these classifiers for test data using Eq. 8 and 9. Table 1 represents segmentation accuracy and precision of the three supervised classifiers on test images and Figure 3 illustrates their segmentation results.

As can be observed in Figure 3, supervised methods cannot produce acceptable results when we have a few labeled data. For example, KNN and SVM classifiers fail to recognize white tissue of the brain and Bayesian classifier fails in recognition of black tissue. These weak results lead us toward using semi-supervised classifiers.

Performance Evaluation for Semi-supervised Classifiers

In this section, three semi-supervised EM, Co_Training and graph-based methods are trained separately using 180 labeled and 4140 unlabeled data. In this experiment, EM method uses Bayesian classifier as the basic classifier and Co_Training algorithm with random feature split uses Bayesian and SVM for this purpose. In graph-based method, for construct a graph, we use Euclidean distance to compute affinity matrix, KNN algorithm as graph sparsification method, and binary method for reweighting. For label inference, GRF method that is one of the graph Laplacian-based methods is used. After training above semi-supervised classifiers, we compute accuracy and precision of them for test images using Eq. 8 and 9.

Experimental results show that the graph-based classifier has produced appropriate results and can be used as one of the semi-supervised classifiers in presented ensemble frame-work. However, EM and Co_Training classifiers do not have the desired accuracy. Hence, in the next step of the experiment, improved versions of these classifiers called EFM and MCo_Training are evaluated. Three basic classifiers

in EFM algorithm are Bayesian, SVM and KNN ($k = 3$), which the threshold value of all of them is 0.8. On the other hand, MCo_Training algorithm uses Bayesian and SVM algorithms as basic classifiers, which the threshold values of them are 0.8 and 0.5 respectively.

Table 2 represents segmentation accuracy and precision of the three traditional semi-supervised classifiers and two improved semi-supervised classifiers on test images and Figure 4 illustrates their segmentation results. According to the results in this table, MCo_Training and EFM improve segmentation accuracy and precision in all test images respect to Co_training and EM respectively. So, these two improved classifiers can be used in ensemble frame-work together with the graph-based method. It is important to note that each of the three semi-supervised classifiers yields more appropriate results compared to all supervised classifiers. Comparing Figure 4 with Figure 3 confirms this progress.

Performance Evaluation for Ensemble Semi-supervised Framework

After specifying appropriate semi-supervised classifiers to be used in presented ensemble frame-work, we can evaluate the accuracy of semi-supervised frame-work with different decision policies. For this purpose, three policies; simple voting, probabilistic voting and maximum probability, are applied separately to decision making unit and therefore the segmentation accuracy and precision of all test images are evaluated. Table 3 reports these results for ensemble semi-supervised frame-work and individual semi-supervised

Table 1: Segmentation accuracy and precision for supervised methods: SVM, KNN and Naïve Bayesian

Learning algorithms	Accuracy and precision (%)												
	Test image 1		Test image 2		Test image 3		Test image 4		Test image 5		Test image 6		
	Acc	Pre	Acc	Pre	Acc	Pre	Acc	Pre	Acc	Pre	Acc	Pre	
Supervised learning approach													
SVM	38.47	51.20	46.53	60.74	48.96	62.74	45.34	60.92	35.75	52.21	46.92	46.40	
KNN	48.86	40.78	53.99	58.01	57.05	54.37	54.55	41.84	48.13	57.63	57.39	42.78	
Bayesian	67.92	70.15	64.65	65.21	62.94	66.02	64.32	66.32	68.24	70.57	62.92	63.61	

SVM – Support vector machine; KNN – K-nearest neighbors

Table 2: Segmentation accuracy and precision for semi-supervised methods: Co_training, MCo_training, EM, EFM and graph-based method

Algorithm or policy	Accuracy and precision (%)												
	Test image 1		Test image 2		Test image 3		Test image 4		Test image 5		Test image 6		
	Acc	Pre	Acc	Pre	Acc	Pre	Acc	Pre	Acc	Pre	Acc	Pre	
Semi-supervised learning approach													
Co-training	54.72	59.40	55.06	57.79	49.37	54.48	51.18	54.09	58.64	63.72	48.89	51.86	
MCO-training	62.69	70.94	66.94	69.96	68.50	70.59	70.19	74.95	69.88	72.03	68.72	73.84	
EM	66.78	70.09	62.39	64.59	61.87	65.91	67.39	70.25	67.72	70.76	60.10	59.28	
EFM	69.47	70.92	71.45	71.84	71.97	72.83	72.16	75.50	68.47	70.98	67.69	67.32	
Graph-base	75.45	76.52	77.98	77.53	76.65	76.45	79.32	79.61	77.31	76.59	70.78	71.27	

EM – Expectation maximization; EFM – Expectation filtering maximization; MCo_training – Modified Co_training; Co_training

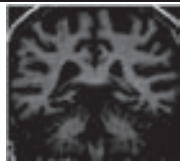
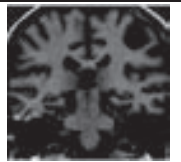
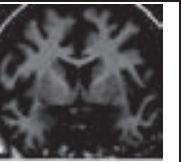
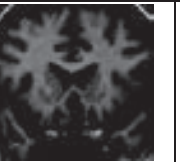
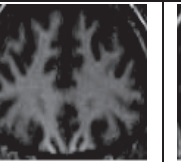
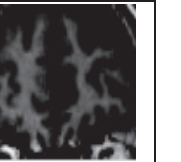
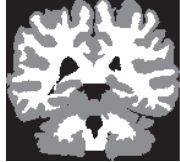
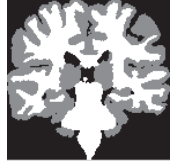




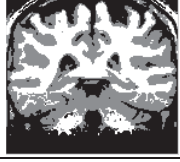
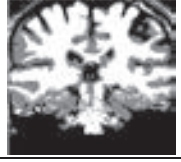
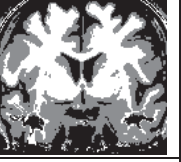
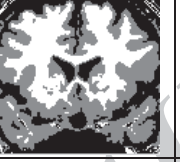
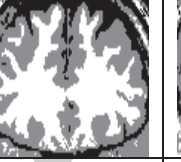
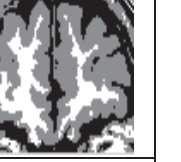










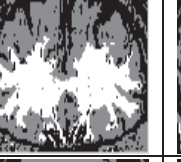










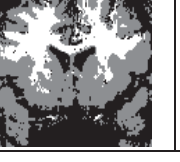
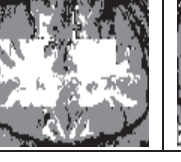

	Segmentation Result					
	Test Image 1	Test Image 2	Test Image 3	Test Image 4	Test Image 5	Test Image 6
Original MR Images						
Ground Truth						
Graph-based Method						
Co-training						
MCO-training						
EM						
EFM						

Figure 4: Segmentation results for semi-supervised methods: Co_training, MCo_Training, expectation maximization, expectation filtering maximization and graph-based method

Table 3: Segmentation accuracy and precision for individual semi-supervised methods and semi-supervised ensemble framework

Algorithm or policy	Accuracy and precision (%)											
	Test image 1		Test image 2		Test image 3		Test image 4		Test image 5		Test image 6	
	Acc	Pre	Acc	Pre	Acc	Pre	Acc	Pre	Acc	Pre	Acc	Pre
Semi-supervised learning approach												
MCO-training	62.69	70.94	66.94	69.96	68.50	70.59	70.19	74.95	69.88	72.03	68.72	73.84
EFM	69.47	70.92	71.45	71.84	71.97	72.83	72.16	75.50	68.47	70.98	67.69	67.32
Graph-base	75.45	76.52	77.98	77.53	76.65	76.45	79.32	79.61	77.31	76.59	70.78	71.27
Semi-supervised ensemble method												
Policy 4.1	71.79	76.89	75.77	77.74	76.54	77.72	79.81	82.66	79.03	80.83	76.37	79.16
Policy 4.2	72.76	77.22	76.58	77.88	77.69	78.54	80.04	82.81	79.76	81.46	75.42	78.59
Policy 4.3	75.63	78.48	79.79	79.86	81.23	81.17	81.69	82.89	81.34	81.69	73.05	76.09

EFM – Expectation filtering maximization; MCo_training – Modified Co_training

classifiers. Figure 5 illustrates segmentation results of these methods.

As Figure 5 shows, by applying each of the three decision policies especially the maximum probability, segmented images have higher clarity and less noise compared to results of individual semi-supervised classifiers and are more similar to GT images. These properties can be investigated through

energy function criterion. Table 4 reports segmentation energy of each image using Eq. 13. According to this table, the energy of segmented images produced by ensemble method is clearly less than individual semi-supervised methods. The diagram in Figure 6 shows this different. According to Table 4, each of the three policies improves segmentation accuracy and precision of the test image compared to all semi-supervised classifiers in average.


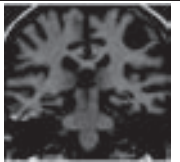
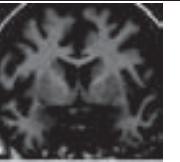

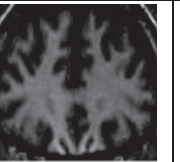
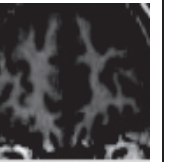

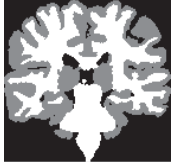








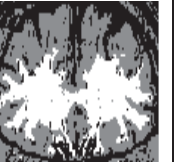


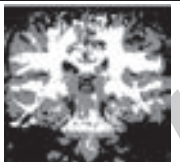

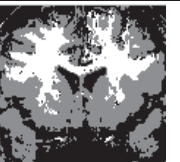


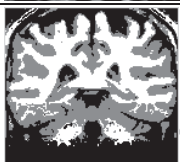

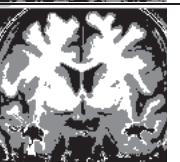
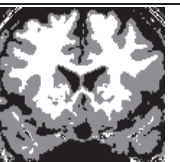

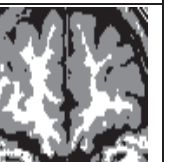
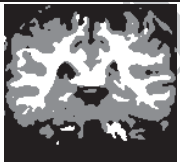




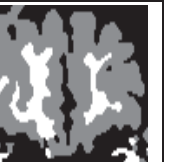

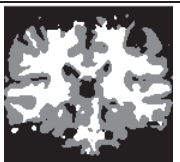




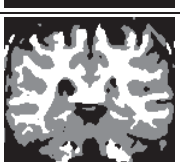
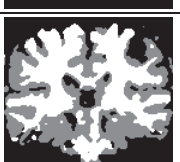

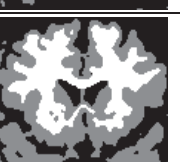


	Segmentation Result					
	Test Image 1	Test Image 2	Test Image 3	Test Image 4	Test Image 5	Test Image 6
Original MR Images						
Ground Truth						
MCO-training						
EFM						
Graph-based Method						
Ensemble Method: Policy 4.1						
Ensemble Method: Policy 4.2						
Ensemble Method: Policy 4.3						

Figure 5: Segmentation results for individual semi-supervised methods and semi-supervised ensemble frame-work

Table 4: Segmentation energy for individual semi-supervised methods and semi-supervised ensemble frame-work

	Algorithm or policy	Energy					
		Test image 1	Test image 2	Test image 3	Test image 4	Test image 5	Test image 6
Semi-supervised learning approach	MCO-training	28875	32859	26531	22087	14351	13571
	EFM	35893	35232	29514	21006	15242	12116
	Graph-base	22887	27661	23063	16711	9488	9783
Semi-supervised Ensemble Method	Policy 4.1	17047	19809	15576	12243	6809	7427
	Policy 4.2	16879	19505	15575	12085	6503	7311
	Policy 4.3	16306	18611	15220	11796	6700	7629

EFM – Expectation filtering maximization; MCo_training – Modified Co_training

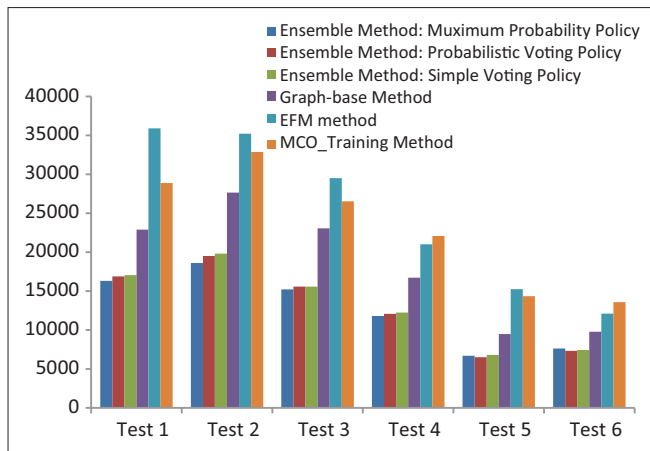


Figure 6: Illustration of energy levels for results of individual semi-supervised methods and semi-supervised ensemble frame-work

Computation Time Evaluation

In previous subsections, we evaluate the performance of supervised methods (Bayesian, KNN, SVM), individual semi-supervised methods (EM, EFM, Co_training, MCo_training, graph-based) and ensemble semi-supervised framework. In this part, we concentrate on computation time of these methods in testing phase and compare them in terms of this criterion.

For simplicity, we assume that the computation time of all mentioned supervised methods (Bayesian, KNN, SVM) is equal to M in testing phase. So, this time is equal to $3 \times M$ for EFM algorithm because it uses Bayesian, KNN and SVM classifiers as its basic classifiers. In the other hand, the computation time of MCo_training will be almost equal to $2 \times M$ because this method uses Bayesian and SVM classifiers as its basic classifiers. Finally, the computation time of graph-based method is M because this method utilizes KNN algorithm as final classifier in test phase.

As mentioned previous, EFM, MCo_training and graph-based methods are three main components of ensemble frame-work. So, it is clear that the computation time of this method will be equal to $(2 \times M) + (3 \times M) + M + K = (6 \times M) + K$ in test phase where k is the computation time of decision making process.

It's noteworthy that all mentioned semi-supervised methods (EFM, MCo_training and Ensemble frame-work) can utilize parallel computation. So in this situation, the running time of test process in EFM and MCo_training methods are almost equal to M , because their basic classifiers can be ran parallel. In the other hand, the running time of our proposed ensemble frame-work is equal to $(M + K)$ because its main components can be ran parallel in time, M , and after that, decision making process can be applied in time k .

CONCLUSION AND DISCUSSION

In this paper, semi-supervised approach was presented as a new approach for brain tissue segmentation in MRI images and evaluated through 3 criteria; accuracy, precision, and energy of images. This approach can produce better results compared to supervised classifiers through using a few number of training labeled data and also information and structure existing in unlabeled data. For this reason, two improved semi-supervised EFM, MCo_Training classifiers and an ensemble semi-supervised frame-work was presented for brain tissue segmentation. The evaluation results of these classifiers on several test images reveal a number of interesting points:

- Supervised classifiers have appropriate accuracy in image segmentation when they are trained with many labeled data. However, in many cases, obtaining labeled data is expensive. This motivates us to use semi-supervised methods. According to the results of experiments, supervised classifiers cannot produce the suitable results when limited labeled data are available
- In conditions that limited labeled data are available, presented semi-supervised classifiers can produce more appropriate results compared to supervised classifiers by exploiting information, which exist in labeled and unlabeled data
- By adding Filtering phase to the EM classifier, EFM classifier trains a confident learner in phase M and improves accuracy and precision of segmented images according to the experimental results
- By using feature selection methods, MCO_training classifier creates two feature sets and improves performance of Co_training classifier that randomly splits features

- v. By using the results of several semi-supervised classifiers simultaneously, ensemble semi-supervised classifier improves accuracy and precision compared to all supervised and semi-supervised methods in average. Moreover, by reducing the energy of images and applying ensemble policies based on neighborhood, these methods increase the quality of segmented images and decreases noise
- vi. Our experiments in the evaluation of unsupervised approaches like MRF for brain image segmentation^[7] show that in this approach, the segmentation of an image requires a lot of time. While semi-supervised classifiers consume a lot of time in training phase but perform the test image segmentation quickly. Therefore, supervised and semi-supervised approaches are better than unsupervised methods in terms of segmentation time.

So, according to points i and vi, when a few labeled data are available semi-supervised approach is better than supervised approach in terms of segmentation accuracy and is better than unsupervised approach in terms of required segmentation time. Furthermore, according to point v, ensemble semi-supervised approach produces more suitable results compared to individual semi-supervised classifiers. It is clear that using appropriate policies in decision making unit has an important role in improving the results. Therefore, defining effective policies and applying the presented framework in other problems like brain tumor segmentation can be investigated in future works.

REFERENCES

1. Limperopoulos C, Clouchoux C. Advancing fetal brain MRI: Targets for the future. *Semin Perinatol* 2009;33:289-98.
2. Duta N, Sonka M. Segmentation and interpretation of MR brain images: An improved active shape model. *IEEE Trans Med Imaging* 1998;17:1049-62.
3. Li H, Yezzi A, Cohen LD. Fast 3D brain segmentation using dual-front active contours with optional user-interaction. *Proc. International Workshop on Computer Vision for Biomedical Image Applications, Lecture Notes in Computer Science*. Vol. 3765. 2005. p. 335-45.
4. Pohle R, Toennies KD. Segmentation of medical images using adaptive region growing. *Proc SPIE Medical Imaging*. Vol. 4322. 2001. p. 1337-46.
5. Sammouda R, Niki N, Nishitani H. A Comparison of hopfield neural network and Boltzmann machine in segmenting MR images of the brain. *IEEE Trans Nucl Sci* 1996;43:3361-9.
6. Zhou J, Chan KL, Chong VF, Krishnan SM. Extraction of brain tumor from MR images using one-class support vector machine. *Conf Proc IEEE Eng Med Biol Soc* 2005;6:6411-4.
7. Yousefi S, Zahedi M, Azmi R. 3D MRI brain segmentation based on MRF and hybrid of SA and IGA. *Biomedical Engineering (ICBME) Proceeding of the 17th Iranian Conference of Biomedical Engineering*; 2010. p. 1-4.
8. Shen S, Sandham W, Granat M, Sterr A. MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. *IEEE Trans Inf Technol Biomed* 2005;9:459-67.
9. Crum WR. Spectral clustering and label fusion for 3D tissue classification: Sensitivity and consistency analysis. *Annals of the BMVA* 2009;2009:1-12.
10. Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. In: *Proceedings of the 11th Annual Conference on Computational Learning Theory, Madison*; 1998. p. 92-100.
11. Nigam K, McCallum A, Thrun S, Mitchell T. Text classification from labeled and unlabeled document using using em. *Proc Int Conf Mach Learn* 2000;39:103-34.
12. Maeireizo B, Litman D, Hwa R. Co-training for predicting emotions with spoken dialogue data. *Companion Proc. of Annual Meeting of the Association for Computational Linguistics*; 2004. p. 202-5.
13. Zhou D, Scholkopf B, Hofmann T. Semisupervised learning on directed graphs. *Adv Neural Inf Process Syst* 2005;17:1633-40.
14. Wu FJ, Poehlman S, Noseworthy MD, Kamath MV. Texture feature based automated seeded region growing in abdominal MRI segmentation. *Biomed Sci Eng* 2009;2:1-8.
15. Chaplot S, Patnaik LM, Jagannathan NR. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomed Signal Process* 2006;1:86-92.
16. El-Dahshan ES, Hosny T, Salem AB. Hybrid intelligent techniques for MRI brain images classification. *Digit Signal Process* 2009; 20:441-3.
17. Briassouli A, Matsiki D, Kompatsiaris I. Continuous wavelet transform for time-varying motion extraction. *IET Image Process* 2010;4:271-82.
18. Zhang Y, Dong Z. Feature extraction of brain MRI by stationary wavelet transforms. *Biomedical Engineering and Computer Science (ICBECS) IEEE*; 2010. p. 1-4.
19. Umbaugh SE. *Computer Vision And Image Processing: A Practical Approach Using CVP Tools*. Published by Prentice Hall PTR; 1997.
20. Reson JM. Fractal-based brain tumor detection algorithms. *Imaging* 2005;23:671-8.
21. Iftexharuddin KM, Jia W, Marsh R. A fractal analysis of tumor in brain MR images. *Mach Vis Appl* 2003;13:352-62.
22. Iftexharuddin KM, Zheng J, Islam MA, Ogg RJ. Fractal-based brain tumor detection in multimodal MRI. *Appl Math Comput* 2009;207:23-41.
23. McCallum A, Nigam K. Employing EM and poolbased active learning for text classification. *Int Conf on Machine Learning*; 1998. p. 359-67.
24. Muslea I, Minton S, Knoblock CA. Active + semi-supervised learning = robust multi-view learning. In: *Proceedings of the 19th International Conference on Machine Learning*; 2002. p. 435-42.
25. Pise N, Kulkarni P. A survey of semi-supervised learning methods. *Int Conf Comput Intell Secur* 2008;2:30-4.
26. Ng V, Cardie C. Bootstrapping coreference classifiers with multiple machine learning algorithms. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Sapporo, Japan; 2003. p. 113-20.
27. Nigam K, Ghani R. Understanding the behavior of co-training. In: *Proceedings of KDD-2000 Workshop on Text Mining*; 2000. p. 15-7.
28. Jebara T, Wang J, Chang S. Graph construction and b-matching for semi-supervised learning. In: *Proceedings of the 26th International Conference on Machine Learning, Montral, Canada*; 2009. p. 441-8.
29. Zhuxj XZ, Ghahramani Z, Lafferty J. Semi-Supervised learning using gaussian fields and harmonic functions. *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington DC, 2003. p. 912-9.
30. Zhou D, Bousquet O, Lal T, Weston J, Scholkopf B. Learning with local and global consistency. In: *Advances in Neural Information Processing Systems*. Vol. 16. 2004. p. 321-8.
31. Wang J, Jebara T, Chang S. Graph transduction via alternating minimization. In: *Proceedings of the 25th International Conference on Machine Learning*; 2008. p. 1144-51.

32. Zhu X. Semi-Supervised learning literature survey. Computer Sciences Technical Report 1530, Univ. of Wisconsin, Madison; 2006.
33. Stan L. Markov Random Field Modeling in Image Analysis. Tokyo: Springer-Verlag, IBSR; 2001. p. 1-40. Available from: <http://www.cma.mgh.harvard.edu/ibsr/>.
34. Available from: http://en.wikipedia.org/wiki/Accuracy_and_precision.

How to cite this article: Azmi R, Pishgoo B, Norozi N, Yeganeh S. Ensemble semi-supervised frame-work for brain magnetic resonance imaging tissue segmentation. *J Med Sign Sens* 2012;3:94-106.

Source of Support: Nil, **Conflict of Interest:** None declared

BIOGRAPHIES



Reza Azmi received his BS degree in Electrical Engineering from Amirkabir university of technology, Tehran, Iran in 1990 and his MS and PhD degrees in Electrical Engineering from Tarbiat Modares university, Tehran, Iran in 1993 and 1999 respectively. Since 2001, he has joined Alzahra university, Tehran, Iran. He was an expert member of Image Processing and Multi-Media working groups in ITRC (From 2003 to 2004), Optical Character Recognition working group in supreme council of information and communication technology (From 2006 to 2007) and Security Information Technology and Systems working groups in ITRC (From 2006 to 2008). He was Project Manager and technical member of many industrial projects.

E-mail: azmi@alzahra.ac.ir



Boshra Pishgoo received the BSc and MSc degrees in computer engineering from Alzahra University, Tehran, Iran in 2009 and 2012 respectively. Currently she is doing some researches in the field of Operating System Security in OSSL Laboratory of Alzahra University. Her research Interests include Medical Image processing, Pattern Recognition,

Artificial Immune Systems, Operating System Security and Smart Card Operating Systems.

E-mail: boshra.pishgoo@gmail.com



Narges Norozi received the BSc degree in computer engineering from Abhar University, Zanjan, Iran in 2008 and his MSc degree in Artificial Intelligence from Alzahra University in 2012, Tehran, Iran. Her research interests include Medical Image processing, Machine Vision and Pattern Recognition.

E-mail: na.norozi@gmail.com



Samira Yeganeh received the BSc degree in computer engineering from Tarbiat Moallem University, Karaj, Iran in 2008 and his MSc degree in Artificial Intelligence from Alzahra University in 2012, Tehran, Iran. Her research interests include Face Recognition, Medica Image Processing, Machine Vision and Pattern Recognition.

E-mail: yeganeh30@gmail.com