Machine Learning Based Analysis of Structural MRI for Epilepsy Diagnosis

Ghazal Sahebzamani, Mansour Saffar
CIPCE, School of Electrical and Computer Engineering
University of Tehran
Tehran, Iran
ghazal.sahebzamani@gmail.com
msaffarm1371@alumni.ut.ac.ir

Hamid Soltanian-Zadeh
CIPCE, School of Electrical and Computer Engineering,
University of Tehran, Tehran, Iran
Radiology Image Analysis Lab, Henry Ford Health System,
Detroit, MI, USA
hszadeh@ut.ac.ir, hsoltan1@hfhs.org

***Abstract*— Epilepsy is a common neurological disorder, characterized by abnormal firing of neurons. Magnetic Resonance Imaging (MRI) techniques can be integrated with machine learning methods to diagnose epileptic patients noninvasively. In this study, we use structural MRI data of 17 subjects (10 epileptic patients and 7 normal control subjects) and segment brain tissues using a Gram-Schmidt orthogonalization method and a unified tissue segmentation approach. We then compute first-order statistical and volumetric gray-level co-occurrence matrix (GLCM) texture features and train SVM classifiers for epilepsy diagnosis based on the features of the whole brain or those of the hippocampus. We achieve an accuracy of 94% using the unified segmentation method and whole-brain analysis approach.**

Keywords- Magnetic Resonance Imaging (MRI), Gram-Schmidt orthogonalization, unified segmentation method, Support Vector Machine (SVM), machine learning.

I. INTRODUCTION

Epilepsy in one of the most common neurological disorders, characterized by abnormal firing of neurons in the brain. Magnetic Resonance Imaging (MRI) techniques play an important role in the diagnosis of epilepsy. Both structural and functional MRI have been used by researchers to analyze epileptic patients. Machine learning methods allow efficient analysis of discriminative features extracted from brain images for epilepsy diagnosis.

Support Vector Machine (SVM) is one the most commonly used machine learning techniques in medical image processing research. Many researchers have applied SVMs on the MRI data for detection and classification of epilepsy. Focke et al. [1] used SVMs on T1-weighted and Diffusion Tensor Imaging (DTI) data of mesial temporal lobe epilepsy (mTLE) patients. Cantor-Reivera et al. [2] used quantitative relaxometry and DTI data with SVM to improve detection of Temporal lobe epilepsy (TLE). In all of these studies, SVM proved to be a robust tool.

In voxel-based MRI analysis, brain tissue segmentation is used to extract white matter, grey matter, and CSF maps from MRI (T1-weighted, T2-weighted, FLAIR, and proton density-weighted images). In this study, we use two techniques for brain tissue segmentation and then compare the results. The first technique is based on Gram-Schmidt orthogonalization [3] and the second method is the unified segmentation method presented in [4]. Our study is the first one to compare these segmentation techniques on MRI images of epileptic patients. We use first-order statistical and volumetric gray-level co-occurrence matrix (GLCM) texture features [5] to extract features from images. Sujitha et al. [6] used GLCM and first-order statistical features in their experiments. We extend the GLCM features to volumetric GLCM features. For feature selection, we apply a three-step filter of Relief-F, k-means clustering, and Sequential Floating Forward Selection (SFFS) methods [7] and use the results to train an SVM classifier. We compare the effectiveness of these features in this work. For a comprehensive analysis, we conduct both whole-brain and hippocampus analysis and compare the results.

The rest of the paper is outlined as follows. In Section II, we describe the dataset used in out experiments. In Section III, we elaborate on the methodologies used for tissue segmentation and feature extraction. In Section IV, the results are presented and in Section V, we analyze the results. We discuss possible approaches and ideas for future work in Section VI.

II. DATA

The data used in this study consist of 17 sets of structural MRI data. Each set contains a sequence of T1-weighted, T2-weighted, FLAIR, and proton density-weighted images. The data was acquired from 7 normal subjects, 5 patients with right temporal lobe epilepsy, and 5 patients with left temporal lobe epilepsy.

III. METHODOLOGY

Images were first skull stripped, i.e., the brain tissue was kept and non-brain tissues such as eyes, skull, and scalp were removed using BET (http://www.fmrib.ox.ac.uk/fsl/) [8]. Next, the images were transferred to a standard coordinate system. This step was carried out by FSL-FLIRT, using the MNI-152 T1 atlas with 1 mm^3 voxels [9] and generated 218x182x182-voxel volumes. The next step was to segment brain tissues into grey

matter, white matter, and CSF. In order to perform this stage, two different methods were applied: the Gram-Schmidt orthogonalization method and the unified segmentation method. These methods are explained briefly in the following subsections.

A. The Gram-Schmidt Orthogonalization Method

In this method, three tissue images are produced from a weighted summation of a sequence of images, each by maximizing the projection of the desired tissue while minimizing the projection of the undesired tissues [3].

If *n* is the number of images in the sequence and *m* represents the number of tissues we wish to segment the images to, we define a pixel vector for each pixel corresponding to the (j,k)-th pixels in the MR images by $\overrightarrow{\mathbf{P}_{jk}} = [P_{jk1}P_{jk2} \dots P_{jkn}]^T$. A desired tissue pattern $\vec{d} = [d_1 d_2 \dots d_n]^T$ is a vector whose i-th element is the average grey level of a selected region of interest (ROI) of the images which covers the desired tissue, while undesired tissue patterns $\vec{u}_i = [u_{1i}u_{2i} \dots u_{ni}]^T,\ 1 \le i \le m$ are similarly defined for interfering tissues. These ROIs may be chosen based on ICBM Tissue Probabilistic Atlas [1]. By generating brain masks for small regions including the intensities in each slice of each tissue, we extract the desired tissues. Using these definitions, the grey level of pixel (j,k) in the desired tissue image are obtained by the inner product of its corresponding pixel vector and a weighting vector which maximizes the signal-to-noise ratio of the desired tissue. It has been proved that these weighting vectors (denoted by $\vec{e}$) are given by

$$\vec{e} = \vec{d} - \vec{d}^P \qquad (1)$$

where $\vec{d}^P$ is the projection of $\vec{d}$ onto the subspace spanned by $\{\vec{u}_i, i = 1, \dots, m\}$ and can be computed using a Gram-Schmidt orthogonalization procedure. In order to guarantee the existence of the solution, the number of unique images in the sequence (n) must be greater than the number of undesired tissues (m). Here, a unique image is one that is not a linear combination of other images in the sequence [10]. An example of a brain slice image segmented by this method is shown in Figure 1.

B. The Unified Segmentation Method

In this method, the intensity distribution of the three brain tissues (grey matter, white matter, CSF) is modeled by a mixture of three Gaussian functions [11], each with mean μ_k, variance σ_k^2, and a mixing proportion γ_k ($\gamma_k \ge 0$, $\sum_{k=1}^{K} \gamma_k = 1$) [4]. To fit this mixture of Gaussians (MOG) model, the probability of observing a grey level intensity of y_k given that it belongs to the *k*-th Gaussian with respective parameters should be maximized. Additionally, the prior probability of each voxel, irrespective of its intensity, belonging to the *k*-th Gaussian can be chosen based on additional information about subjects' brain images. The ICBM probabilistic atlas is used to determine each voxel's prior probability of belonging to the *k*-th Gaussian irrespective of its intensity

$$P(c_i = k|\gamma) = \frac{\gamma_k b_{ik}}{\sum_{j=1}^{K} \gamma_j b_{ij}}, \qquad (2)$$

Assuming that all elements are independent, it has been shown that the maximization of the probability of the entire dataset y with respect to the unknown parameters ($\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\gamma}$) is equivalent to minimization of the following cost function:

$$\varepsilon = -\sum_{i=1}^{I} \log\left(\frac{1}{\sum_{j=1}^{K} \gamma_j b_{ij}} \sum_{k=1}^{K} \frac{\gamma_k b_{ik}}{(2\pi\sigma_k^2)^{\frac{1}{2}}} \times \exp\left(-\frac{(y_i - \mu_k)^2}{2\sigma_k^2}\right)\right) \qquad (3)$$

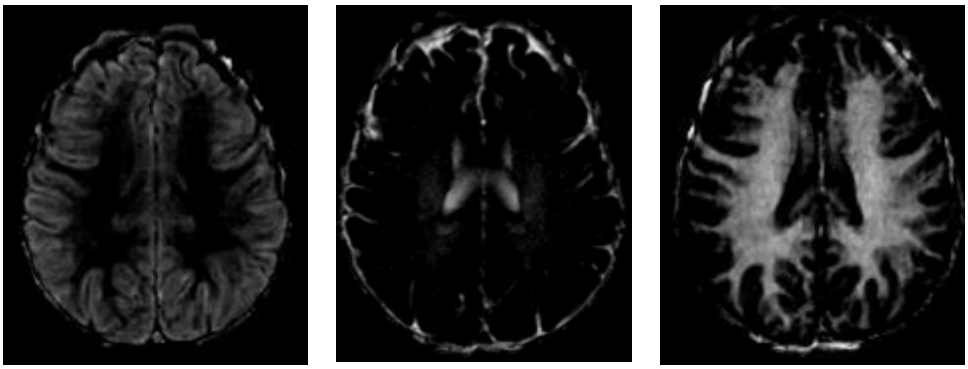

Figure 1. Results of the Gram-Schmidt segmentation method. Left: gray matter, middle: CSF, right: white matter.

[1] Available from http://www.loni.usc.edu/atlases/Atlas_Download.php

The parameters of interest (μ, σ, and γ), summarized by θ, are found by an expectation maximization (EM) algorithm (see [11], [12] for details), which can be considered as using some distribution q_{ik} to minimize the

$$\varepsilon = \varepsilon_{EM} = -\sum_{i=1}^{I}\sum_{k=1}^{K} q_{ik} \log P(y_i, c_i = k|\theta) + \sum_{i=1}^{I}\sum_{k=1}^{K} q_{ik} \log q_{ik} \quad (4)$$

By substituting the above-mentioned values and differentiating the cost function with respect to each parameter, the updates for parameters are computed as:

$$\mu_k^{(n+1)} = \frac{\sum_{i=1}^{I} q_{ik}^{(n)} y_i}{\sum_{i=1}^{I} q_{ik}^{(n)}} \quad (5)$$

$$\sigma_k^{2(n+1)} = \frac{\sum_{i=1}^{I} q_{ik}^{(n)} \left(\mu_k^{(n+1)} - y_i\right)^2}{\sum_{i=1}^{I} q_{ik}^{(n)}} \quad (6)$$

$$\gamma_k^{(n+1)} = \frac{\sum_{i=1}^{I} q_{ik}^{(n)}}{\sum_{i=1}^{I} \frac{b_{ik}}{\sum_{j=1}^{K} \gamma_j^{(n)} b_{ij}}} \quad (7)$$

An example of a brain slice image segmented by this method is shown in Figure 2.

C. Feature Extraction

After segmentation and initial preprocessing, data is ready for extraction of discriminative features. Two classes of features characterizing MRI images are extracted [6], [13]. These features are:

- First order statistical features
- Second order statistics that are computed from spatial gray-level co-occurrence matrix (GLCM), also known as second order texture features.

The first order statistical features are derived using:

$$Mean\ \mu = \sum_{i=0}^{G-1} iP(i) \quad (8)$$

$$Variance\ \sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 P(i) \quad (9)$$

$$Skewness\ m_3 = \sum_{i=0}^{G-1} (i-\mu)^3 P(i) \quad (10)$$

$$Kurtosis\ m_4 = \sum_{i=0}^{G-1} (i-\mu)^4 P(i) \quad (11)$$

where G is the number of gray levels in the image and $P(i)$ denotes the gray level intensity. The GLCMs are constructed by mapping the gray level co-occurrence probabilities based on spatial relations of different angular directions. In Haralick's initial study [6], 23 features were defined, among which 12 features are used in this study: Entropy, Correlation, Contrast, Variance, Sum average, Dissimilarity, Cluster shade, Cluster tendency, Homogeneity, Maximum probability, and Inverse variance. If we consider P as the two-dimensional co-occurrence matrix, N_g as the number of discrete grey levels in the image, for a two-dimensional image, these 12 features are defined as follows

$$energy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} [P(i,j)]^2 \quad (12)$$

$$entropy = --\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P(i,j) \log_2 [P(i,j)] \quad (13)$$

$$correlation = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijP(i,j) - \mu_x(i)\mu_y(j)}{\sigma_x(i)\sigma_y(j)} \quad (14)$$

$$contrast = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} |i-j|^2 P(i,j) \quad (15)$$

$$sum\ average = \sum_{i=2}^{2N_g} \left[iP_{x+y}(i)\right] \quad (16)$$

$$dissimilarity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} |i-j| P(i,j) \quad (17)$$

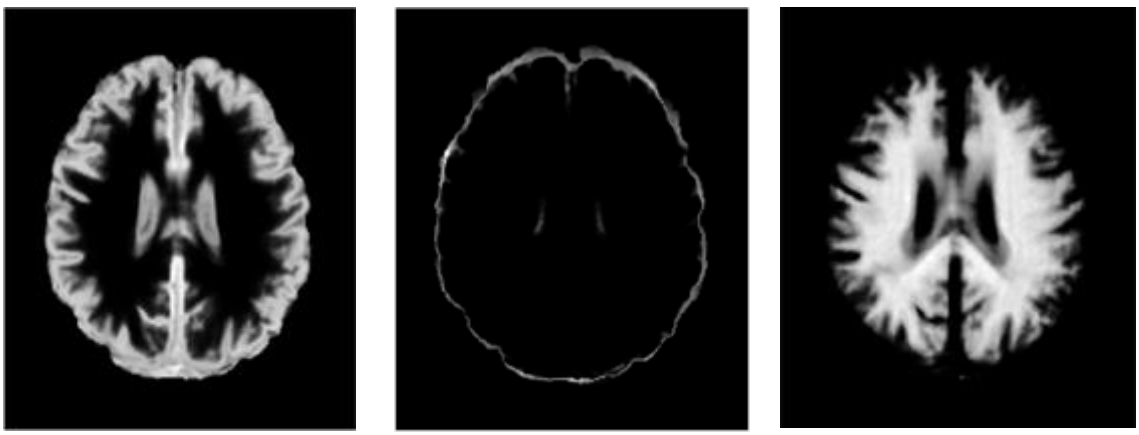

Figure 2. Results of the unified segmentation method. Left: gray matter, middle: CSF, right: white matter.

$$variance = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu_x(i))^2\,P(i,j) + \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(j-\mu_y(j))^2\,P(i,j) \quad (18)$$

$$cluster\ shade = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\Big((i+j-\mu_x(i)-\mu_y(j))^3 \times P(i,j)\Big) \quad (19)$$

$$cluster\ tendency = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\Big((i+j-\mu_x(i)-\mu_y(j))^2 \times P(i,j)\Big) \quad (20)$$

$$Homogeneity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{1+|i-j|} \quad (21)$$

$$maximum\ probability = \max(P(i,j)) \quad (22)$$

$$inverse\ variance = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{P(i,j)}{|i-j|^2} \quad , i \neq j \quad (23)$$

We define these features for a three-dimensional space [5]. In 2D, each slice leads to creation of several co-occurrence matrices and thus, for each image consisting of 182 slices, hundreds of co-occurrence matrices should be computed. When we define a 3D co-occurrence matrix, the number of these matrices, and as a result, the required computations will decrease considerably. Plus, similar to computer graphics that use 3D textures as a more realistic alternative to 2D texture mapping, texture features derived from volumetric data are expected to have more discriminating power than 2D features. 3D co-occurrence matrices are $G \times G$ matrices, where G is the number of grey levels in the image, capable of capturing spatial dependence of grey-level values across multiple slices, instead of capturing the spatial dependence of grey levels within a specific slice. In this approach, the displacement vector is defined as a 3D vector $d = (dx, dy, dz)$. By defining this volumetric co-occurrence matrix, we can generalize the definition of each 2D feature to 3D. In this study, we extract the 3D features in three directions *d1= (0,1,0), d2= (1,1,0), d3= (0,1,−1),* each with a distance of one pixel. Therefore, 36 texture features are obtained.

D. Feature Selection

In this paper, a three-step algorithm is utilized for dimensionality reduction [7]. First, the irrelevant features are removed using the Relief-F algorithm [14], [15]. For each sample, it finds *k* nearest "hits" (other samples of the same class) and "misses" (samples of a different class) and adjusts the relevance value of each feature according to the average square of the feature difference between the sample and the hits and misses. The relevance threshold was chosen such that one half of the data with highest Relief-F scores are considered as relevant features and passed to the next filter. The second step is a redundancy filter that uses the k-means clustering algorithm to cluster features according to how well they correlate to each other [16]. When feature clusters are discovered, only the feature with the highest Relief score is kept and the other features in the cluster are removed from the feature set. The third and final filter is a combinatorial selection algorithm, which uses a Sequential Floating Forward Selection (SFFS) algorithm to find features with the least classification error among the remaining features [17]. Applying this method, we were able to reduce the number of features from hundreds to a few (2 to 4).

IV. Implementation

As some previous studies demonstrate that temporal lobe epilepsy is related to hippocampal disorders [18], by segmenting hippocampus from MR images and running the above analysis on the results, we can limit our search area. Extraction of hippocampus from brain images are carried out by the FSL-FIRST tool. In this work, we run the algorithms both on the whole brain images and the hippocampus separately. First, we implement the feature extraction methods on the whole-brain images, segmented by either the Gram-Schmidt method or the unified segmentation method and compare the results. Next, we repeat these steps on the hippocampal images.

A. Whole-brain

1) Unified Segmentation

Performing the feature selection steps on the features extracted from these images lead to the following four features:

- Contrast of white matter for FLAIR images in direction *d1= (0,1,0)*
- Contrast of white matter for FLAIR images in direction *d3= (0,1,−1)*
- Contrast of white matter for FLAIR images in direction *d2= (1,1,0)*
- Sum average of white matter in FLAIR images in direction *d2= (1,1,0)*

By training an SVM classifier based on these features and cross validating the performance, we get the confusion matrix presented in Table I.

TABLE I. CONFUSION MATRIX FOR EPILEPSY DIAGNOSIS BASED ON WHOLE BRAIN TISSUES SEGMENTED USING UNIFIED SEGMENTATION METHOD.

	Predicted Healthy	**Predicted Patient**
True Healthy	7	0
True Patient	1	9

Accuracy= 94%, specificity= 100%, and sensitivity= 90%.

2) Gram-Schmidt Segmentation

Repeating the forenamed methods on the images segmented by the Gram-Schmidt method, we get the following features:

- Dissimilarity in direction *d1= (0,1,0)* for gray matter
- Dissimilarity in direction *d2= (1,1,0)* for gray matter
- Dissimilarity in direction *d3= (0,1,−1)* for gray matter

Similarly, training the SVM classifier and testing it by the leave-one-out method leads to the confusion matrix shown in Table II.

TABLE II. CONFUSION MATRIX FOR EPILEPSY DIAGNOSIS BASED ON WHOLE BRAIN TISSUES SEGMENTED USING GRAM-SCHMIDT METHOD.

	Predicted Healthy	**Predicted Patient**
True Healthy	6	1
True Patient	2	8

Accuracy = 82.3%, specificity = 85.7%, and sensitivity = 80%.

B. Hippocampus Images

1) Unified Segmentation

By extracting the mentioned features from hippocampus images and reducing the feature vector dimensionality, the following features are selected:

- Cluster tendency of grey matter for T2 images in direction *d1= (0,1,0)*
- Cluster tendency of grey matter for T2 images in direction *d2= (1,1,0)*

The leave-one out cross validation algorithm on these training features generates the confusion matrix shown in Table III.

TABLE III. CONFUSION MATRIX FOR EPILEPSY DIAGNOSIS BASED ON HIPPOCAMPUS AND UNIFIED SEGMENTATION METHOD.

	Predicted Healthy	**Predicted Patient**
True Healthy	5	2
True Patient	1	9

Accuracy = 82.3%, specificity = 71.4%, and sensitivity = 90%.

2) Gram-Schmidt Segmentation

Finally, applying the Gram-Schmidt segmentation on hippocampus, the selected features are:

- Homogeneity of white matter in direction *d1= (0,1,0)*
- Homogeneity of white matter in direction *d2= (1,1,0)*
- Homogeneity of white matter in direction *d3= (0,1,−1)*
- Homogeneity of white matter in direction *d3= (0,1,−1)*

which lead to the confusion matrix shown in Table IV.

TABLE IV. CONFUSION MATRIX FOR EPILEPSY DIAGNOSIS BASED ON HIPPOCAMPUS AND GRAM-SCHMIDT SEGMENTATION METHOD.

	Predicted Healthy	**Predicted Patient**
True Healthy	4	3
True Patient	1	9

Accuracy = 76.4%, specificity = 57.1%, and sensitivity = 90%.

V. CONCLUSION

Unified segmentation led to a more accurate epilepsy diagnosis compared to the Gram-Schmidt method, both in whole-brain and hippocampus images. This could be due to the fact that unlike the Gram-Schmidt method, which generates three tissue images form the MRI data consisting of T1-weighetd, T2-weighted, FLAIR, and proton density-weighted images, the method based on the unified segmentation uses all four of the images. Therefore, the features extracted from the unified segmentation results are more comprehensive and more discriminative than those extracted from the Gram-Schmidt images.

Additionally, we notice that volumetric texture features, unlike first-order statistical features, are capable of training the classifier with high performance.

Moreover, it can be seen that although hippocampus features are discriminative, whole-brain image analysis leads to a superior diagnosis. This indicates that other brain structures are also involved in epilepsy.

Overall, the most distinctive features were contrast and homogeneity of white matter, and dissimilarity and cluster tendency of gray matter. The best accuracy was obtained from contrast and sum average of white matter from whole brain images.

VI. SUMMARY AND FUTURE WORK

In this paper, we studied approaches for segmentation, feature extraction, and feature selection of structural MRI images to train a classifier for diagnosis of epileptic patients, and compared the results. The highest accuracy was about 94%. For future works, we will study physiological interpretations of the features. We will also increase the number of subjects to obtain more reliable results and be able to generalize the methods to a multiclass problem which includes lateralization of the epilepsy patients.

In addition, in the unified segmentation method, the parameters which model intensity nonuniformity, artifacts, and deformation of tissue probability maps were not optimized. Although these parameters may not be inspected visually, they affect the results. It is also worth noting that the severity of

epilepsy could also account for the classification errors. As the patients used in this study were not totally similar in their history of epilepsy, it is possible that the features of the patients with milder levels of epilepsy displayed less distinction abilities that those with severe levels of epilepsy.

REFERENCES

[1] N. K. Focke, et al. "Automated MR image classification in temporal lobe epilepsy." *Neuroimage* 59.1 (2012): 356-362.

[2] D. Cantor-Rivera, Ali R. Khan, Maged Goubran, Seyed M. Mirsattari, and Terry M. Peters. "Detection of temporal lobe epilepsy using support vector machines in multi-parametric quantitative MR imaging." *Computerized Medical Imaging and Graphics* 41 (2015): 14-28.

[3] H. Soltanian-Zadeh, Donald J. Peck, Joe P. Windham, and Tom Mikkelsen. "Brain tumor segmentation and characterization by pattern analysis of multispectral NMR images." *NMR in Biomedicine* 11, no. 4-5 (1998): 201-208.

[4] J. Ashburner, Karl J. Friston. "Unified segmentation." *Neuroimage* 26, no. 3 (2005): 839-851.

[5] A. S. Kurani, Dong-Hui Xu, Jacob Furst, and Daniela Stan Raicu. "Co-occurrence matrices for volumetric data." *Heart* 27 (2004): 25.

[6] V. Sujitha, P. Sivagami, and M. S. Vijaya. "Support vector machine based epilepsy prediction using textural features of MRI." *Procedia Computer Science* 2 (2010): 283-290.

[7] B. A. Draper "Feature selection from huge feature sets." In *null*, p. 159. IEEE, 2001.

[8] S. Smith, "Fast robust automated brain extraction." *Human brain mapping* 17.3 (2002): 143-155.

[9] J.C. Mazziotta, Arthur W. Toga, Alan Evans, Peter Fox, and Jack Lancaster. "A probabilistic atlas of the human brain: Theory and rationale for its development: The international consortium for brain mapping (icbm)." *Neuroimage* 2, no. 2 (1995): 89-101.

[10] H. Soltanian-Zadeh, J. P. Windham and A. E. Yagle, "Optimal transformation for correcting partial volume averaging effects in magnetic resonance imaging," *IEEE Transactions on Nuclear Science*, vol. 40, no. 4, pp. 1204-1212, Aug 1993.

[11] Chris Bishop, Christopher M. Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.

[12] A. P. Dempster, Nan M. Laird, and Donald B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the royal statistical society. Series B (methodological)* (1977): 1-38.

[13] R. M. Haralick, K. Shanmugam, and Its'Hak Dinstein. "Textural features for image classification." *IEEE Transactions on systems, man, and cybernetics* 3, no. 6 (1973): 610-621.

[14] K. Kira, and Larry A. Rendell. "The feature selection problem: Traditional methods and a new algorithm." In *Aaai*, vol. 2, pp. 129-134. 1992.

[15] I. Kononenko, "Estimating attributes: analysis and extensions of RELIEF." In *European conference on machine learning*, pp. 171-182. Springer, Berlin, Heidelberg, 1994.

[16] J. MacQueen, "Some methods for classification and analysis of multivariate observations." In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14, pp. 281-297. 1967.

[17] P. Pudil, Jana Novovičová, and Josef Kittler. "Floating search methods in feature selection." *Pattern recognition letters* 15, no. 11 (1994): 1119-1125.

[18] S. R. Das, et al. "Structure specific analysis of the hippocampus in temporal lobe epilepsy." *Hippocampus* 19.6 (2009): 517-525.