Classification of Hippocampal Region using Extreme Learning Machine

Muhammad Hafiz Md Zaini, Mohd Ibrahim Shapiai
Electronic System Engineering, MJIIT, Center of Artificial Intelligence and Robotics (CAIRO), UTM,
54100 Kuala Lumpur, Malaysia

Ahmad Rithauddin Mohamed
Institut Pediatrik, Hospital Kuala Lumpur,
Jalan Pahang Kuala Lumpur, 50586 Kuala Lumpur, Malaysia

Norrima Mokhtar
Faculty of Engineering, University of Malaya,
50603 Kuala Lumpur, Malaysia

Zuwairie Ibrahim
Universiti Malaysia Pahang,
26600 Pekan, Pahang, Malaysia

E-mail: muhdhafizzaini@gmail.com, md_ibrahim83@utm.my, *drahmadrm@gmail.com, norrimamokhtar@um.edu.my,*
zuwairie@ump.edu.my

Abstract

Important brain parts like hippocampal usually being manually segmented by doctors. But with the introduction of hybrid between machine learning along with neuroimaging technique, it has proved to shows some promising results regarding on segmenting subcortical structures. However, it is known that Extreme Learning Machine (ELM) is to be superior machine learning technique. This study will investigate on the usage of ELM to segment hippocampal by using various hidden nodes configuration. This study also will address on the usage of full image and region of interest (ROI) using ELM. Bag of features is used as a feature extractor where it will segment the hippocampal of the MRI in order to get its visual words. ELM will used it to learn its feature. Results shows that with suitable hidden nodes, it could achieve up to 100% performance on both cases for full image and ROI in hippocampal segmentation.

Keywords: Extreme learning machine, Hippocampal segmentation, Magnetic resonance imaging, Neuroimaging

1. Introduction

With the introduction of biomedical imaging modality, it has brings great help especially for doctors and medical expert community. Medical imaging can be defined as a modality that deliver information on the subject of the volume underneath the skin [1]. There are other biomedical imaging modalities that has been introduced such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), radiography and many more to detect stroke, cancer, Alzheimer and epilepsy. However the problem arises when doctors need to make further diagnosis as it only provide raw images for them

Muhammad Hafiz Md Zaini, Mohd Ibrahim Shapiai, Ahmad Rithauddin Mohamed, Norrima Mokhtar, Zuwairie Ibrahim

to diagnose [2]. This can be seen in the cases where doctors usually detect hippocampal manually which can consume a lot of their time and also lead to error at some point [2, 3].

Computer Aided Diagnosis (CAD) is introduced to solve this problem, whereas it can be describe by the means of the usage of machine learning and image processing to process image and interpret it in order to assist doctors and experts [4]. The process to segment subcortical structures is difficult and important task at the same time too especially in brain analysis, this is because, it is where the hippocampal is located [5]. Many research has shown that a patient's hippocampal who is effected by Epilepsy or Alzheimer tends to become smaller in size compare to a healthy person [6]. The likes of Atlas-based Segmentation, Statistical Model and Machine Learning is the result of a research that has been done in this past years in order to segment hippocampal [6-8]. Machine learning, technique such as Structure Vector Machine (SVM) [5], radial basis [9], Artificial Neural Network (ANN) [10] has been widely used as the learning method in order to learn the structure of the hippocampal. Although it is proven that the aforementioned machine learning could achieve good result, however ELM is known to be superior in machine learning technique point of view as it offer a unified framework where the generalized decision from training data can be used to solve binary classification and predicting a scalar number for regression problem [11]. Therefore, this study will investigate on the usage of ELM to segment hippocampal by using various hidden nodes configuration with normal control from ADNI dataset as its data. Furthermore, this study also will address on the usage of full image and region of interest (ROI) using ELM. Bag of features [12] will be used as a feature extractor where it will segment the hippocampal of the MRI in order to gets its visual words. Thus, it could be feed into ELM to learn its feature.

2. Related Works

Lately, a lot of research has been done particularly in the area of segmenting subcortical structure. Technique such as Atlas selection, Statistical model and Machine learning which can be seen in Figure 1 are some of the result achieved by extensive research that have been done [5-7].
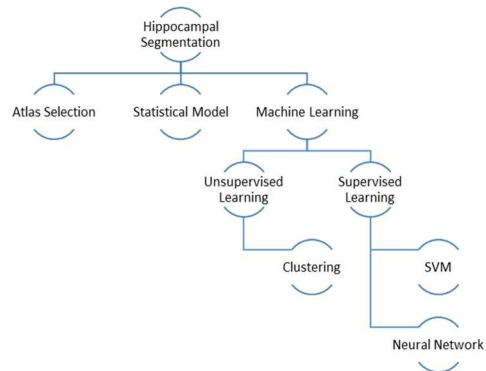

Fig. 1. Hippocampal segmentation technique

2.1. Atlas Selection, Statistical Model and Machine Learning

One of the technique that is popular among researcher is Atlas-based segmentation, where it will register atlas image to the subject image. And the result of the registered atlas will be used to map the coordinates of the structure of interest from the atlas image to the subject image [7]. Since image registration is basically the essence of an extensive variety of medical applications together with visualisation, and image guided surgery and voxel-based morphometry [6, 7]. Hence allowing Atlas-based segmentation to gain benefits from methodological advances driven by a wide range of application areas. However, it has been noted that its performance depends on the image registration accuracy and anatomical differences between the target image [5].

Another technique is statistical models. The idea of statistical model is that a priori shape information will be used as its learning set so that it could learn the variation from it and limit the search space to only acceptable instances defined by the trained model [6]. This technique is widely used among researchers as expert knowledge can be captured in the forms of training examples [13].

Machine learning is one of the method that vastly become favourite among researcher [8, 14, 15], it is used not only to segment subcortical structure but also to detect tumour [6]. Besides, it also can handle large amount of data especially in segmenting or classifying brain anatomical as it needs a lot of raw data in order to get better result [8]. However, machine learning has several limitations that could hinder its accuracy, it needs a lot of data in

order to be more accurate [6], and thus a lot of experiments need to be done in order to get its optimal parameter.

2.2. SVM and neural network

The idea of SVM is that it separates the data in the future space by looking for support vectors is one of the reason of the growing interest in applying SVM in neuroimaging as it has good generalization ability [14]. Besides, it also has the capability to classify non-linearly separable data too [6]. Nevertheless, SVM also has several disadvantages that might deter the performance during classification as it is difficult to determine the optimal parameters and to understand its structure algorithm too. Apart from that, SVM also takes a lot of time for training compare to other machine learning approach [6]. Besides that, SVM also is not a sparse model since support vector tends to grows along the size of training samples [9].

Whereas ANN is a data processing technique that can be categorised by the form of architecture between connections of the neurons, its method to determine the weights on the connections which is the training or learning algorithm and its activation function [16]. Its performance depends entirely on the structure of the training set as if it is not well design it might hinder the classification result and might be resulting to overfitting too [6]. But on the hindsight, researcher still opt for ANN as it is capable for classification and regression and it is tolerant towards noise because of the ANN structure which it keeps on improving until it finds the optimal parameter. Plus, ANN can handle and classified more than one output too [6].

3. Employed Techniques

This section will briefly explain the techniques that has been employed in this study.

3.1. Spatial normalization

Spatial normalization is a step where it will transform a brain subject by establish a one-to-one correspondence between the brains by matching it to standard brain form called "template" [12]. This technique is developed mainly to simplify inter-subject comparisons by placing all subjects into a standardized stereotactic template space [17]. Spatial normalization is an important process in this research as it involves large inter-individual variability of human brain, thus with spatial normalization it helps to decrease the number of inter-individual variability and also to ease subject comparison.

This permit an exploratory approach looking for group effects across the entire brain, or a hypothesis-driven approach whereby common ROIs may be utilised across all subjects, keeping away from the necessity for ROI tracing for each subjects [18].

3.2. Bag of Feature

Bag of Feature (BoF) is an approach where it represent the whole image or an ROI as a histogram occurrence quantized visual features that also known as "visual signature" of the image [12]. It is also one of the most popular method for content-based visual information retrieval (CBVIR) especially in medical field. One of the advantages of BoF is that it represent a direct identification of the features instead of their quantization, thus this approach is able to classify the intended classification whether the image represent hippocampal or not- hippocampal [12].

3.3. ELM

ELM is a supervised learning technique which is based on single-hidden layer feedforward neural networks (SLFN) [19]. The main idea of ELM is that the hidden node parameters don't need to adjust as they can be assigned with random values. Unlike conventional neural network, ELM also does not need much parameter selection which make it appropriate to be applied in neuroimaging, and ELM also fast in training with good generalisation performance too [20].

ELM can be formulated by:

$$f(x) = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x) \quad (1)$$

From the formula above, a_i are the input nodes and b_i is the bias of the ith, which they are defined as the learning parameters of the ith hidden nodes. Whereas x is defined as the input vector with d dimensions and β_i is the output weight from the ith hidden node. As for $G(a_i, b_i, x)$ is the output of ith hidden node with respect to input x and G activation function. The activation function can be nonlinear continuous functions such as sigmoid, Gaussian and many more.

During training phase, hidden nodes and output nodes parameter must be determined. According to ELM theory, the hidden node parameters a_i and b_i are assigned with values randomly regardless of its nature and will

remained fixed after that. But in this scenario, β_i is the only parameter that needs to be determined based on training data. In ELM, given a training data $(x_i, t_i), i = 1, \ldots, N$ where $x_i \in R^d$ and $t_i \in \{-1, +1\}$, thus in order to minimize the training error in the cost function that is formed in least square sense and given can be seen in equation (2).

$$E = \min \sum_{j=1}^{N} \left(\sum_{i=1}^{L} \beta_i G(a_i . x_j + b_i) - t_j \right)^2 \quad (2)$$

Equation 2 can be further simplified as shown in formula (3).

$$H\beta = T \quad (3)$$

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix},$$

$$H = \begin{bmatrix} G(a_1, b_1, x_1) & \cdots & G(a_L, b_L, x_1) \\ \vdots & \cdots & \vdots \\ G(a_1, b_1, x_N) & \cdots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_L \end{bmatrix}_{L \times 1} \text{ and } T = \begin{bmatrix} t_1 \\ \vdots \\ t_N \end{bmatrix}_{N \times 1}$$

Where H is the hidden layer output matrix of SLFN. While the ith column of is the ith hidden node output with respect to input $x_1, x_2, \ldots, x_N . h(x)$ is defined as $h(x) = [G(a, b, x), \ldots, G(a_L, b_L, x)]$ and called the hidden layer feature mapping. The ith row H is the hidden layer feature mapping with i^{th} input x_i. In this case, solving the linear system in equation for β is equivalent to training the network. If the number of training samples equal to the number of hidden nodes, $L = N$ then H is a square matrix and β can be found by calculating the inverse of H where a zero training error is obtained. Thus, if L < N then it is not a square matrix and a solution can be found using Moore–Penrose generalized inverse of matrix H as given in Equation (4)

$$\beta = H^\dagger T \quad (4)$$

In binary classification problems, the decision function for ELM with one output node can be written in a vector form from Equation (1) and is given in Equation (5).

$$f(x) = sign\left(\sum_{i-1}^{L} \beta_i G(a_i, b_i, x) \right) = sign(\beta . h(x)) \quad (5)$$

where, β is the estimated output weight vector in equation (4) and $h(x)$ is a vector that maps the d-dimensional input space to L-dimensional hidden layer feature space.

4. Proposed Methodology

Since the objective of this study is to investigate on the usage of ELM to segment hippocampal by using various hidden nodes configuration. And also, to investigate on the usage of full image and region of interest (ROI), the overall process is being highlighted on flowchart in Figure 2.
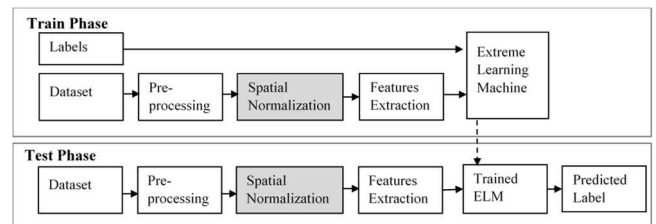

Fig. 2. Experiment methodology

The process first started with acquiring dataset from ADNI database, it can be downloaded directly from adni.loni.usc.edu website. A total of 68 MRI from Normal Control (NC) data from the ADNI1: Annual 2 Year 1.5T dataset has been downloaded from http://ida.loni.usc.edu details can be seen in table 1. Next, is the pre-process step, all of them has undergoes through spatial normalisation process that will register the MRI image into template's image. Then, the normalised image will be fed into BoF in order to get its visual words where it will be later on used as the training and testing feature for the ELM.

Table 1. Dataset descriptions

Specification	Description
Age	61 – 91 years' old
Gender	Male (11), Female (9)

4.1. Spatial Normalisation

The first step is to correct the origin of the raw MRI data, the main purpose is to reposition the crosshair position so that it would provide better normalisation result of the subject's structural MRI to the template's structural MRI which in this case, it is the MNI152 space.

Next, is to realign and reslice. Fundamentally, the function of realign and reslice is to registers all images of a subject to generate parameter files so that it can be used later on to correct for head motion. By realigning the images of the subject, it will match the image by transforming it to manipulate the scan. The realign function only allows translations by moving the image in X, Y and Z direction and rotation. After the realignment process the next step will be reslice. Reslice is a function where it will refine which images are needed as sometimes the MRI might not have the same thickness as others thus corrects its motion.

The last step of spatial normalisation is to normalise the subject's MRI to the template's MRI, so that it could put the subject's MRI into standardised MNI space. The function of this step is mainly to determine the transformation that minimises the differences between two scans by minimising the sum of squares of intensity differences. The result of the normalised MRI can be seen in Figure 3
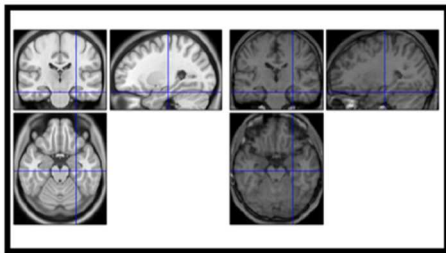

Fig. 3. Spatial normalisation result

4.2. Feature extraction using BoF

The next step is to segment hippocampal and non-hippocampal so that the segmented image can be feed into BoF to get its Visual Words. There are several steps involved in order to extract the Visual Words which are to extract its features, learn "Visual Vocabulary" and quantize features using Visual Vocabulary and lastly to represent each image by frequencies of Visual Words which can be seen in Figure 4.
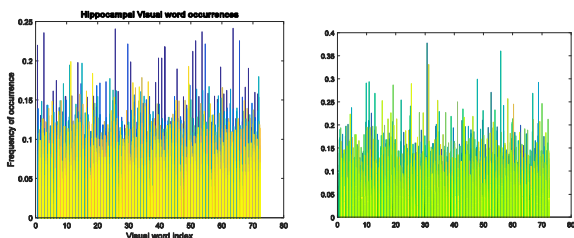

Fig. 4. Hippocampal and non-hippocampal visual word

5. Experiment, Result, and Discussion

This part will discuss on the result of the experiment that have been conducted for hippocampal classification in two experiments 1) full MRI image and 2) ROI of MRI image. 68 MRI from NC has been pre-processed using spatial normalisation technique. From the pre-processed MRI, 2-4 hippocampal from sagittal, axial and coronal has been extracted resulting to 140 extracted images of hippocampal and non-hippocampal. All of them has been fed into bag of feature in order to extract its feature vector. Both of the experiment will used 60% from the extracted images as the training sample and another 40% as testing set. Table 2 below shows the clarification on how the experiment has been setup. Whereas the parameter for both of the experiment can be seen on table 3

Table 2. Distribution dataset for training and testing

Model	Training Set	Testing Set
Full image	Training 60%: 84 feature vector	
ROI	Test 40%: 56 feature vector	

Table 3. Experiment's parameter setup

Parameter	Experiment 1	Experiment 2
Hidden Nodes	10, 30, 50, 100, 300, 500, 800, 1000, 1200, 1500	
Activation Function	Sigmoid function	
Experiment Runs	30	

As mentioned previously, two set of experiment have been conducted where the first experiment is to test on different configuration of hidden neurons for full image of MRI. The investigated hidden nodes are 10, 30, 50, 100, 300, 500, 800, 1000, 1200 and 1500 with 30 set of experiment where all of the image will be sorted by using pseudorandom method. The results for the first experiments are tabulated in Table 4.

Table 4. Experiment 1 result

Hidden	Training			
	AVG	MIN	MAX	STD
10	0.684325	0.619048	0.755952	0.028971
30	0.905357	0.863095	0.940476	0.022126
50	0.969048	0.952381	0.982143	0.009431
100	0.999405	0.994048	1	0.001816
300	1	1	1	0
500	1	1	1	0
800	1	1	1	0
1000	1	1	1	0
1200	1	1	1	0
1500	1	1	1	0
Hidden	Testing			
	AVG	MIN	MAX	STD
10	0.63869	0.544643	0.75	0.044719
30	0.82381	0.741071	0.883929	0.03632
50	0.892262	0.839286	0.946429	0.025463
100	0.88244	0.803571	0.928571	0.029106
300	0.971131	0.919643	0.991071	0.01718
500	0.982143	0.955357	1	0.013055
800	0.98869	0.964286	1	0.010469
1000	**0.993452**	**0.964286**	**1**	**0.008432**
1200	0.988393	0.955357	1	0.011523
1500	0.996131	0.991071	1	0.0045

From the table result above, it shows that ELM could clearly distinguish between hippocampal and non-hippocampal image where both training and testing result could achieve as high as 100%. This could be seen in bar chart below that shows the average of training and testing accuracy. When the more the number of hidden neuron is applied, the higher the accuracy will be. In this experiment 1000 hidden neurons are the optimum hidden neuron parameter for this experiment set, because the average of the testing accuracy will start to decrease a little bit when the number of hidden neuron applied is increased to 1200 but when 1500 hidden neuron is applied it started to increased back.

As for the second experiment, all of the segmented image of both hippocampal and Non-hippocampal will be classified using ROI of MRI image using similar parameters setup as given in Table 3. The results for the second experiments are tabulated in Table 5. Experiment will be setup with the same configuration as experiment 1 which can be seen in table 1 and table 2. The result can be seen on table below.

Table 5. Experiment 2 result

Hidden	Training ROI			
	AVG	MIN	MAX	STD
10	0.803968	0.744048	0.845238	0.026247
30	0.940476	0.89881	0.964286	0.014327
50	0.984325	0.952381	1	0.011107
100	1	1	1	0
300	1	1	1	0
500	1	1	1	0
800	1	1	1	0
1000	1	1	1	0
1200	1	1	1	0
1500	1	1	1	0
Hidden	Test ROI			
	AVG	MIN	MAX	STD
10	0.763988	0.696429	0.839286	0.034911
30	0.891071	0.848214	0.9375	0.024072
50	0.905952	0.821429	0.955357	0.035462
100	0.88006	0.794643	0.9375	0.036976
300	0.982738	0.964286	1	0.009649
500	**0.993155**	**0.973214**	**1**	**0.006062**
800	0.998214	0.991071	1	0.003632
1000	0.999405	0.991071	1	0.002265
1200	0.999702	0.991071	1	0.00163
1500	0.999702	0.991071	1	0.00163

The results of the second experiment show that it could offer a better result compare to experiment 1. This is because, the average testing accuracy in experiment 2 is higher than experiment as tabulated in Table 5. The comparison performance between experiment 1 and experiment 2 are shown in figure 5

The optimum hidden neurons for experiment 2 is 500 because after that, the testing accuracy remains the same when the number of hidden neurons applied is increased.

However, some things worth to be mention, where ELM could not give a good classification result when the parameter between hippocampal and non-hippocampal is almost the same. This could be seen in the result of Experiment 2 on hidden neurons 500 to 1500. Which in this case the testing result shows 1 of the testing is classified wrong. One of the reason that could be deducted is its visual word's is almost the same with the non-hippocampal's visual words, figure 6 shows the comparison between the classified wrong image's visual words and non-hippocampal's visual words. Aside from that, this study proves that ELM could be applied in Neuroimaging problems especially on segmenting brain subcortical problem.
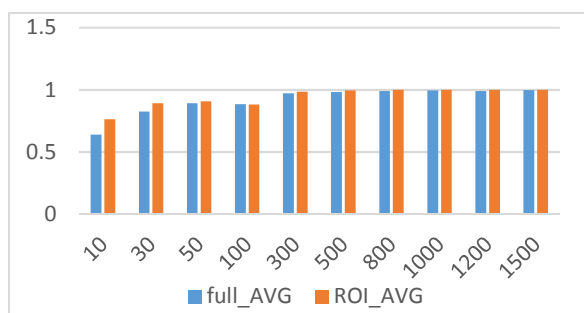

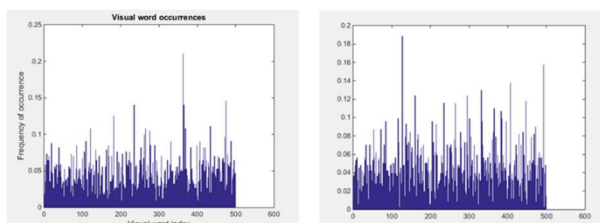Fig. 5. Comparison test average between full image and ROI

Fig. 6. Visual word occurrences between wrongly segment hippocampal and non-hippocampal

6. Conclusion

In general, ELM has been tested to two different type of hippocampal segmentation which are full image and ROI. Besides that, several hidden nodes configuration for ELM has been investigated too. From the result of experiment 1 and 2, this study could deduct that ELM could easily classify between hippocampal and non-hippocampal image this is because the result for both experiment 1 and 2 could achieve as high as 100%! Besides that, from the experiment conducted also, a hypothesis could be made, where the accuracy result is

influenced by the number of hidden neurons applied. This clearly shows the ability of ELM's where it has good generalisation performance, as its only need to adjust the hidden neuron's parameter in order to increase its accuracy. For future works, ELM could be further improved into structured-ELM so that it could be experimented in more complicated problem especially regarding hippocampal segmentation problems.

References

1. I. Bankman, *Handbook of medical image processing and analysis*: academic press, 2008.
2. M. Mizotin, J. Benois-Pineau, M. Allard, and G. Catheline, "Feature-based brain MRI retrieval for Alzheimer disease diagnosis," in *2012 19th IEEE International Conference on Image Processing*, 2012, pp. 1241-1244.
3. P. Suppa, H. Hampel, L. Spies, J. B. Fiebach, B. Dubois, and R. Buchert, "Fully automated atlas-based hippocampus volumetry for clinical routine: Validation in subjects with mild cognitive impairment from the ADNI cohort," *Journal of Alzheimer's Disease,* vol. 46, pp. 199-209, 2015.
4. C. L. Leggett and K. K. Wang, "Computer-aided diagnosis in GI endoscopy: looking into the future," *Gastrointestinal Endoscopy,* vol. 84, pp. 842-844, 2016.
5. Y. Hao, T. Wang, X. Zhang, Y. Duan, C. Yu, T. Jiang, *et al.*, "Local label learning (LLL) for subcortical structure segmentation: application to hippocampus segmentation," *Human brain mapping,* vol. 35, pp. 2674-2697, 2014.
6. J. Dolz, L. Massoptier, and M. Vermandel, "Segmentation algorithms of subcortical brain structures on MRI for radiotherapy and radiosurgery: a survey," *IRBM,* vol. 36, pp. 200-212, 2015.
7. O. T. Carmichael, H. A. Aizenstein, S. W. Davis, J. T. Becker, P. M. Thompson, C. C. Meltzer, *et al.*, "Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment," *Neuroimage,* vol. 27, pp. 979-990, 2005.
8. N. Oktar and Y. Oktar, "Machine Learning and Neuroimaging," *Journal of Neurological Sciences (Turkish),* vol. 32, pp. 001-004, 2015.
9. Y. Wang, Y. Fan, P. Bhatt, and C. Davatzikos, "High-dimensional pattern regression using machine

learning: from medical images to continuous clinical variables," *Neuroimage,* vol. 50, pp. 1519-1535, 2010.

10. M. Lai, "Deep learning for medical image segmentation," *arXiv preprint arXiv:1505.02000,* 2015.

11. Z. Bai, G.-B. Huang, D. Wang, H. Wang, and M. B. Westover, "Sparse extreme learning machine for classification," *IEEE transactions on cybernetics,* vol. 44, pp. 1858-1870, 2014.

12. O. B. Ahmed, J. Benois-Pineau, C. B. Amar, M. Allard, and G. Catheline, "Early Alzheimer disease detection with bag-of-visual-words and hybrid fusion on structural MRI," in *Content-Based Multimedia Indexing (CBMI), 2013 11th International Workshop on*, 2013, pp. 79-83.

13. T. F. Cootes and C. J. Taylor, "Statistical models of appearance for computer vision," ed: Technical report, University of Manchester, 2004.

14. J. Morra, Z. Tu, A. Toga, and P. Thompson, "Machine learning for brain image segmentation," *Biomedical Image Analysis and Machine Learning Technologies: Applications and Techniques: Applications and Techniques,* p. 102, 2009.

15. J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson, "Comparison of AdaBoost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation," *IEEE transactions on medical imaging,* vol. 29, p. 30, 2010.

16. L. Fausett, *Fundamentals of neural networks: architectures, algorithms, and applications*: Prentice-Hall, Inc., 1994.

17. P. T. Fox, "Spatial normalization origins: Objectives, applications, and alternatives," *Human brain mapping,* vol. 3, pp. 161-164, 1995.

18. S. Krishnan, M. J. Slavin, T.-T. T. Tran, P. M. Doraiswamy, and J. R. Petrella, "Accuracy of spatial normalization of the hippocampus: implications for fMRI research in memory disorders," *Neuroimage,* vol. 31, pp. 560-571, 2006.

19. A. Annema, K. Hoen, and H. Wallinga, "Precision requirements for single-layer feedforward neural networks," 1994.

20. N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Transactions on Neural networks,* vol. 17, pp. 1411-1423, 2006.