MRI based Techniques for Detection of Alzheimer: A Survey

Ruaa Adeeb Abdulmunem Al-falluji
University of Babylon,
Babylon , Iraq

ABSTRACT

Alzheimer's disease(AD) is a neurological disease. It affects memory of the patient. The livelihood of the people that are diagnosed with AD. Magnetic resonance imaging (MRI) is one of the most commonly used imaging modality for the diagnosis of Alzheimer's. Different features and classifiers that are used Computer Aided Diagnosis (CAD) for diagnosis of Alzheimer's are presented.

Keywords

Alzheimer Magnetic Resonance Imaging (MRI) , Independent Component Analysis(ICA), Principal Component Analysis(PCA), Gray Level CoOccurance Matrix(GLCM), Gabor Filter.

1. INTRODUCTION

Dementia is a syndrome due to disease of the brain, usually chronic, characterized by a progressive, global deterioration in intellect including memory, learning, orientation, language, comprehension and judgment.

Alzheimer's disease (AD) is a one of the most important example of dementia which mostly affects people over 65 years old and whose incidence rate grows exponentially with age, almost doubling in every 5 years. Still, apart from a few exceptions, the factors that trigger the onset of AD remain unknown. It is a progressive disease, this means that it worsens over time, and for which there is currently no cure, leading eventually to death. The very early stages are often mistakenly confused with the normal process of ageing or linked to stress and it is often characterized by episodic losses of short term memory and difficulty to grasp new ideas.

The early detection of Alzheimer's disease still a challenge because of the estimation of the scans depends on manual directing and visual reading. This preclinical stage is also known as Mild Cognitive Impairment (MCI). As the brain damage progresses, other cognitive deterioration appear and the disease becomes obvious. In the late stages, persons are completely dependent on caregivers even for the most basic daily tasks such as eating, bathing or dressing [1][2][3][4]. Till now, there is no cure for Alzheimer's disease, but its early detection is important to a successful treatment, slacken the progression of symptoms. So, the development of automatic diagnostic tools, which is reliant on information obtained from 3D images of the brain, has attracted great attention in last years. Computer Aided Diagnosis (CAD) allows early detection of the disease at early stages, and structural brain images are useful in this task.

Structural and Functional neuroimaging allow studying of brain pathology at macro and micro molecular level like Magnetic Resonance Imaging (MRI), single photon emission computed tomography (SPECT), and Positron emission tomography (PET).
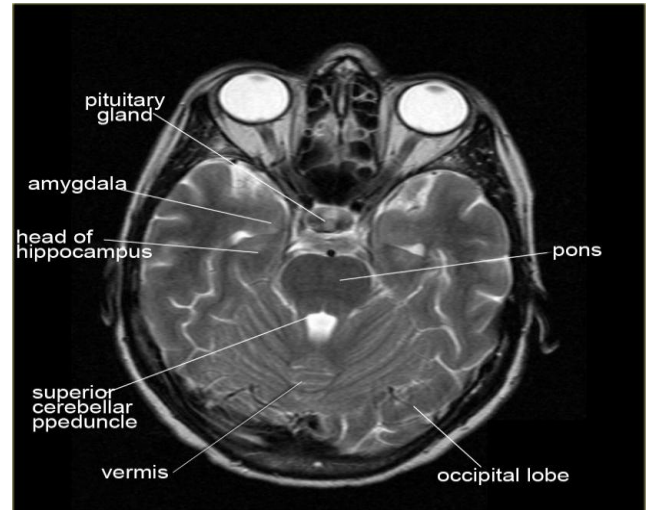

Fig. 1 MRI of the Brain

The most commonly used imaging technique by doctors is Magnetic resonance imaging (MRI) [5]which uses a magnetic field and pulses of radio wave energy to make pictures of organs and structures inside the body. Figure 1 shows the MRI of the brain. In many cases, MRI gives different information about structures in the body than can be seen with an X-ray, ultrasound, or computed tomography (CT) scan.Magnetic resonance imaging (MRI) is considered the preferred neuroimaging examination for Alzheimer disease because it allows for accurate measurement of the 3-dimensional (3D) volume of brain structures, especially the size of the hippocampus and related regions.Figure 2 shows the main atrophied regions in brain that usually diagnosed as AD. As the nerve cells die, they are replaced with fluid. Fluid appears dark on MRI. V indicates the ventricles and frontal horn which are the fluid-filled spaces within the brain. H is in the position of the hippocampus and intraparietal sulci.
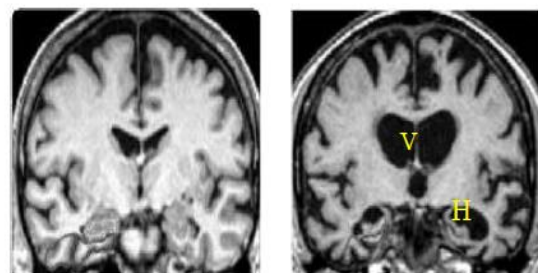

(a)   **(b)**

Fig. 2 MRI (a) Cognitive Normal Brain (b) Atrophied Brain

In this paper we present the different techniques used to detect Alzheimer disease in MRI scans.

2. FEATURE EXTRACTION TECHNIQUES

A. Independent Component Analysis (ICA)

The ICA transformation is used for capturing group-differences from high order voxel relations,generating from the original average images sources characteristic vectors (pk) a new set of

statistically independent components. For achieving the ICA transformation, we make use ofFastICA [6] for iterative fixed-point algorithm and it has the contrast function chosen to be acubic polynomial. Once the basis vector {pk} of the subspace S is obtained, a set of features can be extracted to train a SVM with labeled data[7].

B. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is used when the goal is to reduce the number of features, and convert them into space having high variance [8][9].

Method for calculating PCA involves the following steps:

Step 1: Get some data.

Step 2: Subtract the mean.

Step 3: Calculate the covariance matrix.

Step 4: Calculate the eigenvectors and eigenvalues of the covariance matrix.

Step 5: Choose components and form a feature vector.

Step 6: Derive the new data set.

The use of PCA [10] within this work is justified not as a dimensional reduction technique but it is applied to the average images vector. Therefore, this paper [11] introduce PCA as a procedure for extracting a decorrelated eigenvector basis from the previous defined average image vectors. Similar frameworks in which a reduced number of sources are obtained from a linear combination of the mixtures have also been used in other fields as face recognition problem [12], or functional magnetic resonance imaging (fMRI) [13].The number of vectors in this new basis is the same as the number of classes used to define the average images. Therefore, none of the eigenbrains are neglected, independently from its variance eigenvalue.

C. Wavelets

Wavelets are mathematical functions that decompose data into different frequency components and then study each component with a resolution matched to its scale. While the Fourier Transform only provides representation of an image based on its frequency content, so it loses time information of the signal, the Wavelet Transform provides both time and frequency information. Therefore, the WaveletTransform is a better tool for feature extraction from images.

The Discrete Wavelet Transform (DWT) is a linear transformation that operates on a data vector whose length is an integer power of two, transforming it into a numerically different frequency components, and then studies each component with resolution matched to its scale.Wavelet coefficients were used as features for detection of alzhimer's.[14]

D. Gray Level Co-Occurrence Matrix

Statistical texture features based on the Gray Level Co-occurrence Matrix (GLCM) were extracted from Hippocampus region of the brain. This matrix characterizes the spatial distribution of gray levels in an image or a region.

An element at location (i, j) of the co-occurrence matrix expresses the joint probability density of the occurrence of gray levels i and j in a specific orientation θ and distance d from each other.The choice of different orientation and distance generates different matrices. Here, four directions (θ=0O, θ=45O, θ=90O, θ=135O) and one distance (d=1pixel) were used to generate GLCMs. Texture measures can be extracted from each GLCM and the mean value was used as one feature. Hence, 56 texture features were computed for each closed regions according to Haralick's definition [15][16].

E. HOG Descriptors

The MRI images were divide into 93 ROIs based on a Jacob template, by non-rigid brain registration Given the MRI images separated by 93 ROIs, HOG features were extracted from each ROI. Specifically, first down-sample the original MRI images, i.e., from 256 × 256 × 256 to 64 × 64 × 64, followed by partitioning the whole brain into 93 ROIs. Dilate each ROI with 3 voxels to achieve a soft boundary among ROIs. Following the method in [6], set the number of orientation bins to 9, with each bin with 8 orientations to describe a descriptor by a 72-dimensional feature vector. And also set the size (in voxels) of the spatial bins and the size of the blocks, respectively, as 5 and 2, to extract 1728 descriptors from each ROI. Note that descriptor information was divided into overlapping blocks, each of which contained 2×2×2 3-dimensional cells. We further clustered descriptors of each ROI of all MRI images to form a 50-dimensional bag-of-words for each ROI.

Finally use a 4650 (= 93 × 50) dimensional HOG feature vector to obtain a multi-view representation of an MRI image[17].

F. Gabor Filter

Gabor filter is a linear filter. Two dimensional Gabor filter [18] was extended by using the concept of one-dimensional Gabor filter. Gabor filter is modulation of complex sinusoidal wave with Gaussian envelope.

Gabor filter at scale m and orientation n is represented by Gmn. For an 2d input image I(x,y), of dimension X and Y, convolving input image with Gabor filter in spatial domain, gives Gabor response GTmn(x,y). GTmnis a complex coefficient contains two parts: Re(GTmn) real part as well as imaginary part Im(GTmn). The magnitude and phase of Gabor response gmn is calculated for each slice as follows

$$g_{mn}(GT_{mn}) = \sqrt{\mathrm{Re}^2(GT_{mn}) + \mathrm{Im}^2(GT_{mn})}$$

$$\phi_{mn}(GT_{mn}) = \arctan(\frac{\mathrm{Im}(GT_{mn})}{\mathrm{Re}(GT_{mn})})$$

Gabor filter on different frequency behave as scaled version of each other. Higher frequency obtains finer detail. Thus, applying Gabor filter at different scale results in computation of finer-to-course details of MRI images. But the calculated Gabor features is of high dimension. Thus, there is a need for the method for reduction of dimension. For this the mean and standard deviation statistics has been used. Feature vector was created by using mean and standard deviation of magnitude of Gabor filter response. Mostly this technique is used to extract features from the hippocampus region of the brain.[19]

Table 1 Advantages and disadvantages of features Extraction Techniques

METHOD	ADVANTAGES	DISADVANTAGES
ICA	The ICAtransformation is used for capturing groupdifferences from highorder voxel relations, generating from the original average images sources.	Don't exist a criteria for determining how many components represent the dynamic of the data.
PCA	Reduce the redundant features and large dimensionality of the data.	PCA only takes into account pair-wise relationships between voxels of the brain images.
Wavelet	Its advantage is that discrete set of scales and shifts are used which provides sufficient information and offers high reduction in computation time.	Wavelets are shift sensitive and does not provide any information about the phase.
GLCM	The advantage of the co-occurrence matrix calculations is that the co-occurring pairs of pixels can be spatially related in various orientations with reference to distance and angular spatial relationships, as on considering the relationship between two pixels at a time.	It ignores the spatial relationship between the texture patterns and sensitive towards image noise.
HoG	It captures edge or gradient structure that is very characteristic of local shape, and it does so in a local representation with an easily controllable degree of invariance to local geometric and photometric transformations.	Disadvantage of HOG are orientations that are densely sampled and are therefore likely to capture additional structures, which are irrelevant for object representation, thus leading to a poor object detection.
Gabor	Gabor filters provide means for better spatial localization	The outputs of Gabor filter banks are not mutually orthogonal; then a significant correlation between texture features may occur. And , Gabor filter banks come with high computational costs.

3. CLASSIFICATION TECHNIQUES

To classify the constructed feature vectors different types of classifiers are used. Some of the common classifiers used are listed below

G. K – Nearest Neighbor(KNN) Classifier

The K-nearest neighbor (KNN) classifier acts as a benchmark and workhorse classifier [20][[21][22][23][24]. Given a test feature vector x0 and a set of labeled instances of feature vectors {xi,yi}M, where xi are the feature vectors and yi

their corresponding labels and M total number of instances, the classifier predicts the class of the test feature vector x0 from the pre-defined classes. The K – Nearest Neighbor(KNN) algorithm performs classification based on the majority vote of the N nearest neighbors i.e. the most occurring class in the N nearest neighbor is assigned to the test vector. The criterion used for finding nearest neighbors is distance. Commonly types of distance measure used includes Euclidean distance, City block distance, Chebychev distance, Hamming distance, Minkowski distance and Mahalanobis distance.

If the N is set one , the KNN classifier is known as Nearest Neighbor Classifier and the test vector is assigned the same label as that of the nearest neighbor.

H. Naïve Bayes Classifier

Naïve Bayes Classifier is a probability classifier based on the Bayes theorem[25]. This classifier considers every feature to contribute independently to the probability, irrespective of the presence and absence of other features. As naïve classifier follows the bayes theorem is assign the most likely classes to a given sample mentioned in its feature set. Aside from this unrealistic assumption, naïve bayes is remarkably successful as compared to other sophisticated classifiers [26].

$$P\left(C|X_i\right) = \frac{P\left(C\right).P\left(X_i|C\right)}{P(X)}$$

Where $P(C|X_i)$ is known as the posterior probability, $P(X_i|C)$ is the class conditional probability,$P(C)$ is prior probability and $P(X)$ is the total probability of the system.

I. Support Vector Machine(SVM) Classifier

Support Vector Machine (SVM) is one of the best supervised learning algorithms. SVM was originally designed as a binary classifier. Here a brief description of the classifier is provided for complete details refer to [27][28]. Consider a data set (x1,y1)(x2,y2)…(xn,yn),where x's are the features and y's are the class labels, which is to be classified into two classes [+1,-1]. Assume a hyper plane $w.x + b = 0$ in some space H separate both the classes and that no information about the data distribution is know , then the hyper plane with the maximum margin is the optimal hyper plane. Lagrange multiplier $\alpha_i (i = 1....m)$ can be used to calculate the values of w and b by solving constrained minimization problem.

$$F(x) = \text{sgn} \sum_{i=1}^{m} \alpha_i y_i K(x_i, x) + b$$

Where SVC learning algorithm is used to find α_i and b. Support vectors are those xi that have α_i .For k(x,y)=x.y, this corresponds to constructing an optimal separating hyper plane in the input space.

J. Artificial Neural Network (ANN)

Elementary principle of human neural system was the initial basis of artificial neural network. A lot of different kinds of networks have been developed, but all are composed of neurons and the connection between them which together determine the behavior of the system. The type of network used is dependent on the application it is used for. One of the most commonly used is back propagation gradient. A

network can have any number of hidden neurons layers, but to decrease calculation time mostly one neuron is used, if the result obtained is satisfactory. All the neurons of each layer, except that of last layer, are connected to each neuron of the next layer by means of an axon[29].

K. Binary Classification Trees

Binary Classification tree as with SVM were primarily used for two class classifications. Improvements have been made to the algorithm to use for multi-class classification. In multi-class classification tree a class group is partitioned into two distinct subgroups at each node, and each node has a binary classifier that in turn assigns the sample pattern to one of the two classes. [30]

Table 2 Advantages and disadvantages Of Classifiers

METHOD	ADVANTAGES	DISADVANTAGES
KNN	It's the simplest technique that provides good classification accuracy.	It has a slow running time.
Naïve Bayes	Extremly simple.	Naive Bayes classifier makes a very strong assumption on the shape of your data distribution.
SVM	Produce very accurate classifiers and very low overfitting.	Its biggest limitation is its speed and size.
ANN	It can handle large amount of data sets.	The images go through many stages. Programming is difficult. Again to run the program, experts are not needed. ANNs are black-box modeling.
Binary Tress	Easy to interpret and explain.	Don't support online learning, so you have to rebuild your tree when new examples come on and they easily overfit.

4. CONCLUSION

In this paper a survey of different features extraction algorithms and classifiers which are used for designing of Alzheimer's CAD system is presented. Different features from different part of the MRI are calculated by various authors using various techniques to either detect or calculate the progression of the Alzheimer's disease. In future these should be compared with other imaging modalities to check which provide better accuracy.

5. REFERENCES

[1] P. Morgado, "Automated Diagnosis of Alzheimer's Disease using PET Images", MSc thesis at Electrical and Computer Engineering Dep., Higher technical institute, Technical University of Lisbon, September 2012.

[2] C. P. Ferri, R. Sousa, E. Albanense, W. s. Ribeiro, and M. Honyashiki, "World Alzheimer Report 2009," 2009.

[3] A. Wimo and M. Prince, "World Alzheimer Report 2010: The global economic impact of dementia," September 2010.

[4] A. Association, "2012 Alzheimer's disease facts and figures," Alzheimer's and Dementia: The Jthenal of the Alzheimer's Association, vol. 8, no. 2, pp. 131–168, 2012.

[5] http://emedicine.medscape.com/article/336281-overview

[6] A .Hyvarinen "Fast and robust fixed-point algorithms for independent component analysis". Ieee Transactions on Neural Networks. 1999; 10:626–634.

[7] Yang, Wenlu, et al. "Independent component analysis-based classification of Alzheimer's disease MRI data." Journal of Alzheimer's disease 24.4 (2011): 775-783.

[8] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, "Overview of the Face Recognition Grand Challenge," in Computer vision and pattern recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005, pp. 947-954.

[9] D. Srinivasulu Asadi, Ch. DV Subba Rao and V. Saikrishna "A Comparative Study of Face Recognition with Principal Component Analysis and Cross-Correlation Technique," International Journal of Computer Applica-tions Vol. 10, 2010.

[10] Fermı´n Segovia, Christine Bastin, Eric Salmon, Juan Manuel Go´ rriz, Javier Ramı´rez, Christophe Phillips "Combining PET Images and Neuropsychological Test Data for Automatic Diagnosis of Alzheimer's Disease" PLOS ONE, www.plosone.org, February 2014, Volume 9, Issue 2, e88687.

[11] 18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis" I.A. Illán, J.M. Górriz,J. Ramírez, D. Salas-Gonzalez, M.M. López, F. Segovia, R. Chaves, M. Gómez-Rio c, C.G. Puntonet, Information Sciences 181 903–916, 2011.

[12] M. Bartlett, J. Movellan, T. Sejnowski, "Face recognition by independent component analysis", IEEETransactions on Neural Networks 13 (6) (2002) 1450–1464.

[13] "Independent component analysis of electroencephalographic data in: Advances in Neural Information Processing Systems" S. Makeig, A.J. Bell, T. ping Jung, T.J. Sejnowski, vol. 8, MIT, 1996, pp. 145–151.

[14] Herrera, Luis Javier, et al. "Classification of MRI Images for Alzheimer's Disease Detection." Social Computing (SocialCom), 2013 International Conference on. IEEE, 2013.

[15] Haralick RM, Shanmugam K, Dinstein I. Textural features for.image classification. IEEE Trans. Syst., Man, Cybern. 1973; 6:610-21.

[16] Rajeesh, Jayapathy. "Discrimination of Alzheimer's disease using hippocampus texture features from MRI." Asian Biomedicine (Research Reviews and News) 6.1 (2012).

[17] Zhu, Xiaofeng, et al. "Multi-view classification for identification of Alzheimer's disease." International Workshop on Machine Learning in Medical Imaging. Springer International Publishing, 2015.

[18] J. Llonen, J. K. Kamarainen and H. Kalviainen, "Efficient computation of Gabor features," Research Report 100, Lappeenranta University of Technology, Department of Information Technology, 2005.

[19] Ahmed, Olfa Ben, et al. "Classification of Alzheimer's disease subjects from MRI using hippocampal visual features." Multimedia Tools and Applications 74.4 (2015): 1249-1266.

[20] Athitsos, V., Alon, J., Sclaroff, S.: Efficient nearest neighbor classification using a cascade of approximate similarity measures. In: CVPR '05, pp. 486–493. IEEE Computer Society, Washington, DC, USA (2005)

[21] Athitsos, V., Sclaroff, S.: Boosting nearest neighbor classifiers for multiclass recognition. In: CVPR '05, IEEE Computer Society, Washington, DC, USA (2005)

[22] Cover, T., Hart, P.: Nearest neighbor pattern classification. IEEE Transactions on Information Theory 13(1), 21–27 (1967)

[23] Peng, J., Heisterkamp, D.R., Dai, H.K.: LDA/SVM driven nearest neighbor classification. In: CVPR '01, p.

58. IEEE Computer Society, Los Alamitos, CA, USA (2001)

[24] Zhang, H., Berg, A.C., Maire, M., Svm-knn, J.M.: Discriminative nearest neighbor classification for visual category recognition. In: CVPR '06, pp. 2126–2136. IEEE Computer Society, Los Alamitos, CA, USA (2006)

[25] Tom M. Mitchell. Machine Learning. McGraw-Hill, 1997.

[26] Rish, Irina. "An empirical study of the naive Bayes classifier," IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 2001.

[27] P. Doll´ar, V. Rabaud, G. Cottrell, and S. Belongie, ―Be-havior recognition via sparse spatio-temporal features‖, In VS-PETS, 2005.

[28] J. C. Niebles, H. Wang, and L. Fei-Fei, ―Unsupervised learning of human action categories using spatial-temporal words‖, In BMVC, 2006.

[29] Gallant, S.I., 1993. Neural Network Learning and Expert Systems. M.I.T. Press, London, pp 365.

[30] B.-B. Chae, T. Huang, X. Zhuang, Y. Zhao, J. Sklansky, "Piecewise linear classifiers using binary tree structure and genetic algorithm," Pattern Recognition, Vol.29, No.11, pp.1905-1917, 1996.