Analysis of Alzheimer's Disease on MRI Image using Transform and Feature Selection Method

B.G. Sharan Gopan[1], P Shruthi[2], M Yokhasri[3], Mr.A Pandian[4]

[1,2,3]Student, Department of ECE, Valliammai Engineering College, Tamilnadu, India.
[4]Asst. Professor, Department ECE, Valliammai Engineering College, Tamilnadu, India.

---***---

Abstract:- Alzheimer's disease is a neurological disorder in which the death of brain cells is the source of amnesia and cognitive decline. It is a neurodegenerative type of dementia, the disease starts mild and gets progressively worse. An important area under medical research is Brain image analysis which is used to detect brain diseases. The main causes for Alzheimer's diseases is low brain activity and circulation of blood. The aim of this project is to identify the region affected in the brain from brain MRI image using HARR transform algorithm based on different feature combinations such as color ,edge ,orientation and texture . Alzheimer's extraction and is analysis are challenging tasks in medical image processing because brain image is complicated .To diagnose the disease at the earliest stage using the MRI scanned image by using Digital image processing .MRI has been considered because it provides accurate visualization of anatomical structure of tissues .To recognize Alzheimer's medical image processing and clustering technique are used. This memoranda use combinations of segmentation, filters and other image processing algorithms to achieve the best results.

Keywords—median filter, k means cluster, HAAR, svm

1. INTRODUCTION

Alzheimer's disease is a type of dementia that gradually eradicates brain cells, affecting a person's memory and also affects their ability to learn, make judgments, communicate and carry out basic daily activities. Alzheimer's disease is characterized by a gradual relapse that generally progresses through three stages: early, middle and late stage disease. These three stages are distinguished by their general features, which tend to progress steadily throughout the course of the disease. Alzheimer's disease is not unavoidable in people with Down syndrome. While all people with Down syndrome are at risk, many adults with Down syndrome will not manifest the variance of Alzheimer's disease in their lifetime. Although the danger increases with each decade of life, at no point does it come close to reaching 100%. This is why it is especially important to be careful and thoughtful about assigning this diagnosis before looking at all other possible causes for why changes are taking place with aging. Evaluation shows that Alzheimer's disease affects about 30% of people with Down syndrome in their 50s.When the reach their 60s, this number comes closer to 50%. The Analysis of Hippocampal Structure helps for the diagnosis of Alzheimer's disease. The manual segmentation is a time consuming procedure and it is necessary for a segmentation method .The hippocampus location is shown in the following MRI scan of brain.

2. LITERATURE SURVEY

In the earlier studies, very few works have considered SFMRI for analysis of Alzheimer's to understand its dysfunctions between various regions. Literature has explored functional connectivity of Alzheimer's affected people and contradicting results are achieved in the same resting state network. For example, both hyper-connectivity and hypoconnectivity within the DMN are shown in the literature corresponding to Alzheimer's. Few investigations performed in Alzheimer's are described in this section. Shen etal studied RS-fMRI to find reduced feature based on low level embedding and defined unsupervised. Classifier to discriminate between Alzheimer's and normal subjects and achieved accuracy of 93.5% for Alzheimer's and 75% for normal. The highest discriminative power is exhibited by functional connectives between some cerebellar regions and frontal cortex is determined. Du et al proposed a work which involves double feature selection techniques namely., kernel principal component analysis (KPCA) and Fisher's linear discriminant analysis (FLD) in the fMRI images acquired during resting state activity and auditory oddball task .They are able to retrieve highly discriminating feature set majority of them regions from frontal, temporal and visual regions and acquired overall 90% classification accuracy in discrimination of Alzheimer's from normal(93% in rest and 98% during task). Venkatraman etal analysed whole brain connectivity and determined important features based on the literature survey and reviews. Then they applied random forest for classification and they acquired 75% accuracy. The supra-normal connectivity between frontal and parietal regions (associated with positive symptoms of Alzheimer's) and sub-normal connectivity between parietotemporal regions (associated with negative symptoms of Alzheimer's).

3. MATERIALS AND METHODS

This section provides detail view on participants involved, MRI acquisition technique and the preprocessing methodology followed to derive temporal correlations between various ROIs

A. Flow diagram
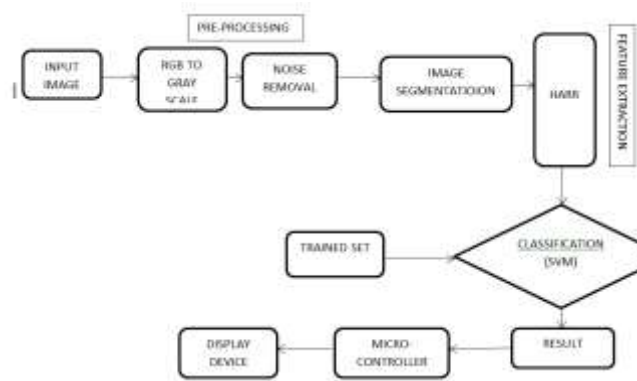
First an image for Alzheimer's is selected

Then it is converted from RGB to gray scale image.

Then it is filtered by median filter to remove the noises

The filtered image is segmented using K-means Clustering algorithm by this method the affected part of theregion affected in the brain is segmented and will be shown.

The features of the brain image are extracted using HAAR Wavelet Transform (DWT)

The extracted features of an image compared with the training dataset by using Support Vector Machine (SVM).

Finally we show how much region affected at the result

B. Pre-processing

This preprocessing section is main thing to identify the region affected in the disease .this involve two section they are RGB to Gray conversion and Filtering

1) RGB to Gray scale

The modification of a color image to gray scale is not unique; different weighting of the color channels effectively represent the effect of shooting black-and-white film with different-colored photographic filters on the cameras.

In order to convert a color from a color space based on an RGB color model to a gray scale representation of its luminance, weighted sums must be calculated in a linear RGB space, which is, after the gamma compression function has been removed first via gamma expansion.

I = rgb2gray (RGB) helps to converts the true color image RGB to the gray scale intensity image I. The rgb2gray function converts RGB images to gray scale by eradicating the hue and saturation information while retaining the luminance.

2) Median Filter

Median filtering is a nonlinear process which is useful in reducing impulsive, or salt-and- pepper noise. This is also useful in preserving edges in an image while reducing random noise. The output of the median filter and the current pixel location is 10, which is the median of the five values. The median filter acknowledges each pixel in the image and in turn looks at its nearby neighbors to conclude whether or not it is representative of its surroundings. Rather than simply replacing the pixel value with the mean of neighboring pixel values, it replaces it with the median of those values. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel in consideration to the middle pixel value.

C. K-means clustering

K-Means algorithm is an unsupervised clustering algorithm which classifies the input data points into various classes based on their inherent distance from each other. Here the algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The points are clustered around the centroids $\mu_i \forall i = 1 \ldots k$ which can be obtained by minimizing the objective

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2$$

In which there are k clusters S_i, i = 1, 2, . . . , k and μ_i is the centroid or mean point of all the points $x_j \in S_i$ respectively.

The objective of *K-means* clustering to partition *n* observations into 'K' clusters in which each observation belongs to the cluster with the nearest mean.
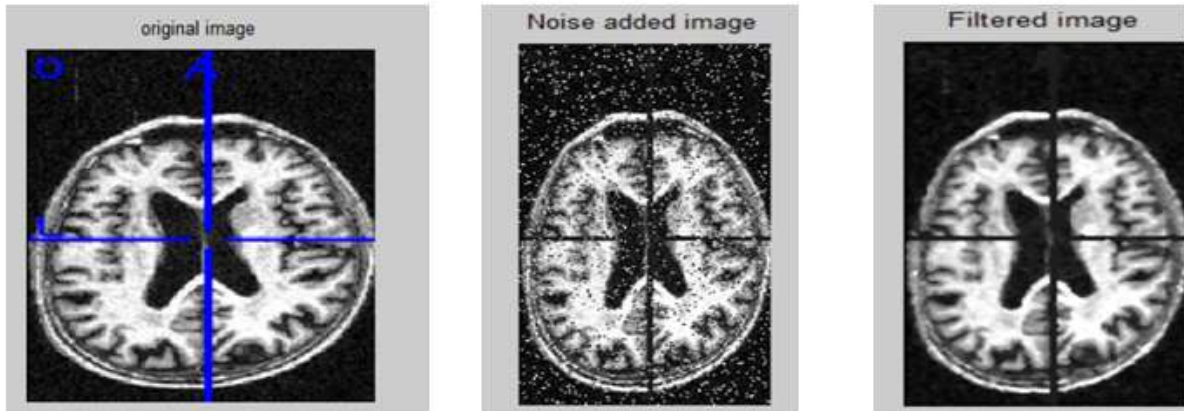
(a)**original image**(b) **expert selection**(c) **K-means selection**

D. Classification

The machine learning approach we propose to identify components associated to Alzheimer's is based on a linear classifier, which is provided with N labeled training pairs of data f(xi; yi)gNi=1. Here, GMC data of subject i is represented by the its S loading coefficients (S: number of sources or components) such that xi = [xi1; : : : ; xiS]>, and yi indicates the subject class (-1 for Ct, +1 for Sz). The estimated class of an unseen example x by a linear classifier is defined by

$$\hat{y} = f(\mathbf{x}) = \mathrm{sgn}\left(\mathbf{w}^T \mathbf{x} + b\right) = \mathrm{sgn}\left(\sum_{d=1}^{S} w_d x_d + b\right), \quad (1)$$

Where w = [w1; : : : ;ws] is the weight vector of the classifier, each element of w being associated to one component. Here we provide a brief insight of the interpretation given to the components weights. For the sake of simplicity, let us assume that all of the components spatial maps are predominantly positive. If the loading coefficients of component d were greater in Sz than in Ct, then GMC would be greater for Sz for that component. Given the class representation of Ct (-1) and Sz (+1), wd would take a positive value. On the other hand, if loading coefficients of the same component were lower in Sz, then GMC would be lower for Sz too and wd would take a negative value. This gives a clear interpretation of the role of each component in the discriminant function defined on . We use a support vector machine (SVM) to train the linear classifier. The SVM, which was coded in MATLAB (http://www.mathworks.com) and solved using the MOSEK optimization toolbox (http://www.mosek.com), finds w and the bias term b in by solving the following optimization

Problem:

$$\begin{aligned}
\min_{\mathbf{w},b,\xi} \quad & \|\mathbf{w}\|_p + C \sum_{i=1}^{N} \xi_i \\
s.t. \quad & y_i \left(\mathbf{w}^T \mathbf{x}_i + b\right) \geq 1 - \xi_i \quad \forall i \in [1, N] \\
& \xi_i \geq 0 \quad \forall i \in [1, N].
\end{aligned}$$

Slack variables _i allow errors in training data, kwkp is an lp-norm regularization term that controls over fitting and C is a parameter that controls the tradeoff between the regularization and the training error terms. The choice of norm p depends on the nature of the classification task. L1-norm regularization enforces sparsity in w, making the weights associated to non-informative features equal to zero. Since the data dimensionality is significantly reduced from hundreds of thousands voxels to S components through SBM, an l1-norm regularization term is well-suited to find informative components. Nonetheless, a selection of informative components cannot rely on an l1-norm SVM approach only, as its sparse selection can vary significantly if some subjects are set aside of the analysis due to the inherent inter- subject

variability of the data. This problem can be solved to a certain extent by using a pool of diverse linear classifiers, as suggested by the star plots method proposed

E. Experimental setup

The classification accuracy rate obtained by bagged SVM was estimated using a leave-one-out cross-validation (loo-CV) procedure. More specifically, 234 classifiers were trained by using the data from all subjects except one, each of them being tested on that left -out example. The fraction of correctly classified examples using this experimental setting defined the method's accuracy rate. There are also several parameters that need to be either fixed to a given value or selected using some validation procedure. These parameters are the number of classifiers of the ensemble (P), the number of examples drawn to train each of these classifiers (M), the sign consistency threshold of the components (r) and the SVM parameter C. The proposed approach did not seem to be sensitive to parameters P and M, which were fixed to 250 and 80% of the available data, respectively. However, the sign consistency threshold defines the number of selected components, so it needs to be validated. Furthermore, C also plays an important role in determining the number of components to be selected, so it needs to be properly estimated. To do so, an inner loo-CV was ran inside each round of the loo-CV used to estimate the classification accuracy. In other words, at each round of the main loo-CV, a nested loo-CV selected the appropriate values of C and r for that specific round. C was picked from 3 logarithmically spaced points in the range [0:1; 1], while was selected from the range [50%; 100%] in steps of 1%.

4. RESULT

A. Classification Accuracy

Table shows the classification accuracy, sensitivity and specificity rates achieved by Bagged SVM. This table also shows the results obtained by a linear SVM with l1- and l2norm regularization terms, as well as for partial least squares (PLS) and random forests (RF) . PLS and RF were also implemented using MATLAB.

Approach	Accuracy	Specification
RF	0.67	0.74
SVM 1	0.68	0.73
SVM 2	0.7	0.73
BAGGED SVM	0.73	0.8

B. Relevant component statistics

The multivariate association of the components to Alzheimer's was evaluated in terms of their weights mean and selection frequencies across CV rounds. Table shows the components that achieved a selection frequency greater than or equal to 0.5 and the p-values associated to a two sample t-test across Sz and Ct on their loading coefficients

Component number	Mean Weight	Frequency	t-test
1	-0.64	0.98	1.50E-08
2	-0.44	0.79	4.42E-08
3	-0.23	0.97	0.01
4	-0.26	0.5	9.09E-6

C. Percentage of affected

From the obtained data set it will be compared with the Trained data set .By using support vector machine classifier both the obtained and trained data sets are being compared And we are getting the result as the percentage of affected areas.

D. Displaying

All the parameters like mean, variance and entropy are displayed by using an LCD display interfaced with arduino micro controller and the affected percentage ofregion affected also showed in the LCD display

REFERENCES

1. Wehenkel M, Bastin C, Phillips C, Gwerts P. Tree ensemble technique and parceling to find brain areas related to Alzheimer's disease. InPattern Recognition in Neuroimaging (PRNI), 2017 International Workshop on 2017 Jun 21 (pp. 1-4). IEEE

2. Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, L Whitwell J, Ward C, Dale AM. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of magnetic resonance imaging. 2008 Apr 1;27(4):685-91

3. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H, Cummings JL. Mild cognitive impairment. The Lancet. 2006 Apr 21;367(9518):1262-70

4. Xiaoxiao Lui, Marc Niethammer, Roland Kwitt, Nikhil Singh, Matt McCormikk and Stephen Aylward, "Low-Rank Atlas Image Analyses in the Presence of Pathologies" , IEEE Transactions On Medical Imaging, 34 (12), 2015.

5. Meiyan Huang,Wei Yang, YaoWu, Jun Jiang , Wufan Chen and Qianjin Feng, "Brain Tumor Segmentation Based on Local Independent Projection-Based Classification", IEEE Transactions On Biomedical Engineering, 61 (10), 2014.

6. H.Cao, J.Duan, D. Lin and Y.P. Wang, "Sparse Representation Based Clustering for Integrated Analysis of Gene Copy Number Variat ion and Gene Expression Data," IJCA, vol. 19, No. 2, June 2012

7. 7. Hongbao Cao, Hong-Wen Deng, Marilyn Li and, Yu-Ping Wang, "Classificat ion of Mult icolor Fluorescence In-situ Hybridizat ion (MFISH) Images with Sparse Representat ion," IEEE Tans. Nanobioscience, Feb. 2012, In press.

8. 8. S.A.Meda, et al., "An fMRI study of working memory in first degree unaffected relat ives of Alzheimer's pat ients," Alzheimer's Research",vol. 104, pp. 85-95, Sep 2008

9. 9. J. H. Callicott, et al., "Variat ion in DISC1 affects hippocampal st ructure and funct ion and increases risk for Alzheimer's," Proc Nat l Acad Sci U S A, vol. 102, pp. 8627-32, Jun. 2005. 10. S. R. Sut rala, et al., "Gene copy number variat ion in Alzheimer's Res, vol. 96, pp. 93-9, Nov 2007.