Classification of Brain MRI for Alzheimer's Disease Based on Linear Discriminate Analysis

Heba Mohsen[1], El-Sayed A. El-Dahshan[2,3], El-Sayed M. El-Horbaty[4], and Abdel-Badeeh M. Salem[4]

[1*] Faculty of Computers and Information Technology, Future University, Cairo, Egypt
[2]Faculty of Science, Ain Shams University, Cairo, Egypt
[3] Egyptian E-Learning University, Giza, Egypt
[4]Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

hmohsen@fue.edu.eg, e_eldahshan@yahoo.com, shorbaty@cis.asu.edu.eg, absalem@asunet.shams.edu.eg

Abstract

Alzheimer's disease (AD) is known to be the most common cause of neurodegenerative dementia that affects people over 65 years. It's an irreversible, progressive brain disorder that slowly destroys memory and thinking skills, and eventually the ability to carry out the simplest tasks and it has no treatment till now expect slowing down its symptoms if it was diagnosed in early stages. The diagnosis of AD includes mental status, physical exam and neurological exam, which is analyzing different imaging techniques such as magnetic resonance images (MRI). And accordingly, AD become a challenging wide area of research in the medical images application that aims to find a reliable methodology that can early detect and differential diagnosis of cognitive normal (CN), mild cognitive impairment (MCI) and AD by examining the brain MRIs. In this work, we proposed a methodology based on Discrete Wavelet Transform (DWT) feature extraction technique and Principal Component Analysis (PCA) for feature vector reduction then these features are entered to linear discriminant analysis (LDA) classifier.

The performance of the proposed methodology was evaluated using two datasets obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database and Harvard Medical School website. Our methodology achieved 94.59%average classification rate with AUC of ROC = 0.963over Harvard medical school dataset and 77.78% average classification rate with area under the ROC curve (AUC) = 0.809 over ADNI dataset using a 6-fold cross validation.

Keywords: *Machine Learning, Alzheimer's disease, Discrete Wavelet Transform, Principle Component Analysis, Linear Discriminant Analysis, Magnetic Resonance Images*

1. Introduction

Alzheimer's disease (AD) is the degeneration of the capability to memorize things and other cognitive abilities decline [1,2]. AD is known to be the most common cause of neurodegenerative dementia which affects the elder people over 65 years worldwide and it is expected that the pervasiveness of AD will double within the next 2 decades and that one out of every 85 people will be afflicted with the disease by 2050[3, 4, 5].

This disease was named after Dr. Alois Alzheimer in 1906 after he inspected a female patient who had changes in the brain tissue and died at age 51 of an unusual mental illness. Her symptoms included memory loss, language problems, and unpredictable behavior. Alzheimer reported two common abnormalities in the brain of this patient when he examined her brain after she died. First, dense layers of protein deposited exterior and stuck between the nerve cells and the second was regions of broken nerve fibres, within the nerve cells, which rather than being straight had become twisted. Till now, these plaques and tangles have been employed in order to assist diagnosing of AD and considered as main features [6].

AD is an irreversible, progressive disease which in its early stages, memory loss is mild, but in the late-stages, patients lose the ability to carry on a conversation or respond to their environment as a result of the degeneration in the normal brain tissues [3]. An accurate and early diagnosis of AD and identification of the risk of progression with awareness of the condition's severity allow the patients to take preventative measures, such as making lifestyle changes and taking medications that can slow down the symptoms of the disease. However, early detection of AD is still challenging process [5].The diagnosis of AD includes mental status, physical exam and neurological exam through analysing the different imaging techniques such as magnetic resonance images (MRI) which are very useful in evaluating the anatomical degeneration using volumetric measures and analysing the structural changes. Therefore, brain MRI analysis is a very important factor in AD diagnosing and identify its progression overtime [2,4].

According to the Alzheimer's Association there are3 phases of AD: preclinical, mild cognitive impairment due to AD, and dementia due to AD. Preclinical AD includes "Measureable transformations in biomarkers (for instance brain imaging and spinal fluid chemistry) that points out the extremely earliest indications of disease, prior to outward symptoms are noticeable". Mild cognitive impairment (MCI) due to AD also includes "mild changes in memory and thinking abilities that are evident enough to be noticed and measured but are not accompanied with injury that compromises day by day activities and execution". Dementia due to AD involves "cognitive and behavioural symptoms that are exist and are of adequate severity to harm the patient's ability to function in daily life" [6].

Researches proposed several methods to identify the imaging biomarkers in order to differential diagnosis of cognitive normal (CN), mild cognitive impairment (MCI) and AD by examining the brain MRIs of the given subjects and focusing on certain regions of interest (ROI) that are more affected in the disease progression [1, 7 ,8]. The MRI scan is preferred due to its clarity in soft tissue differentiation and high spatial resolution with better contrast so that it can provide identification of tiny irregularities in the brain and structural changes [5, 8].

In this study we aim to present a robust methodology for automatic differential diagnosis of CN and AD subjects using brain MRIs based on learning the most discriminate features for each class. The proposed methodology is based on extracting and selecting the discriminate features using Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA) and classification using linear discriminant analysis (LDA) classifier and was tested using two different datasets to evaluate its performance. The paper is planned as follows, in section 2 the general framework of the proposed methodology is analysed. In section 3 the experimental results and evaluation of the presented work performance is carried and finally in section 4 there's the conclusion and future works.

2. Methodology

This work considered proposing a methodology for classifying brain MRIs into CN and AD subjects using two different datasets of brain MRIs. The classification methodology involves DWT based feature extraction, PCA based feature selection and then the selected features are entered to train the LDA classifier. Finally, the performance of this methodology is evaluated using the receiver operating characteristic (ROC) analysis.Figure 1 represents the architecture for the proposed methodology.

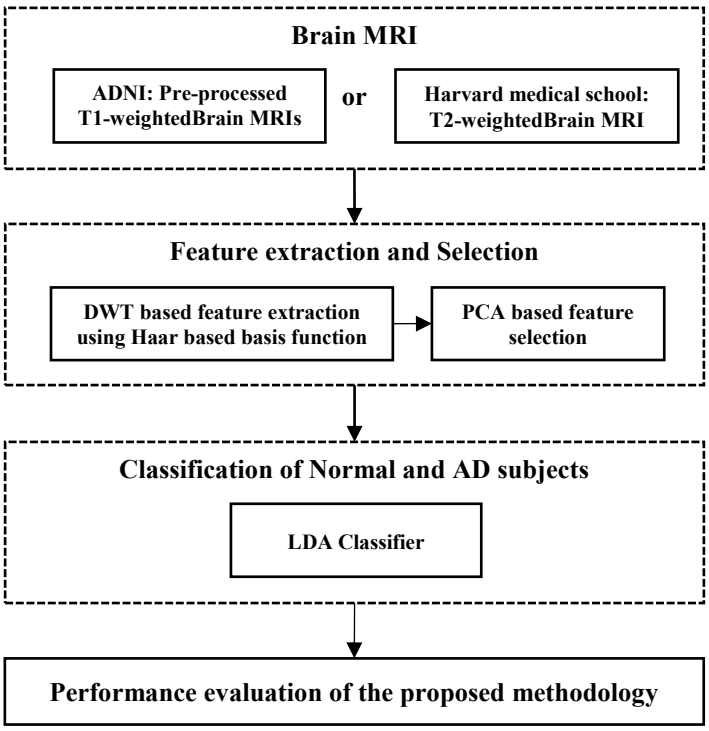


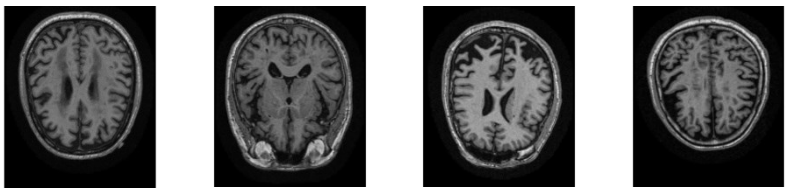
Figure 1.Block diagram of the proposed methodology architecture

2.1 Dataset

Two dataset were used in evaluating the performance of the proposed methodology. The first dataset is obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database [9].ADNI database includes more than 900 subjects of age 50 years to 90+ years with an annual follow-up of 3 years. All ADNI subjects are scanned at screening, three months from the screening MRI, and within two weeks before or after Month 6 and subsequent annual visits. In this work, we used 1.5T axial T1-weighted (MPRAGE) MRI scans acquired from participants according to ADNI acquisition protocol. A total of 126 brain MRIs from different subjects of 63 CN and 63 AD brain MRIs (24Male and 18 Female) with 3 MRI scans from each subject selected from the central section of the brain which is more discriminative in classifying AD [3]. We only considered the brain MRI scans on baseline and screening visits as the same concept could be applied further on the other visits with no difference. Also, for the standardization purpose of the used images, the brain MRIs selected in the dataset are all pre-processed images undergone several correction steps: Grad warp correction to correct image geometry distortion due to gradient non-linearity, B1 non-uniformity correction to

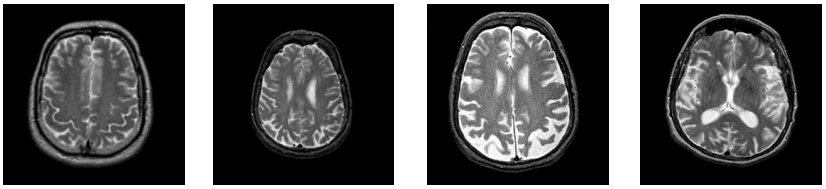
correct the image intensity non-uniformity and N3 histogram peak sharpening algorithm to reduce residual intensity non-uniformity.Figure2 shows a sample of brain MRIs in the dataset.

The second dataset is collected from Harvard Medical School website [10]. We obtained a dataset of 37 brain MRIs of 22 CN and 15 AD. All the brain MRIs are axial T2-weighted MRI scans which provides higher contrast than T1-weighted. A sample of this dataset is shown in Figure3.



(a) (b) (c) (d)

Figure2.Sample of ADNI dataset: (a) and (b) represents the CN sample of MRIs, (c) and (d) represents the AD sample of MRIs



(a) (b) (c) (d)

Figure 3.Sample of Harvard Medical School: (a) and (b) represents the CN sample of MRIs, (c) and (d) represents the AD sample of MRIs

2.2 Feature Extraction and Selection

DWT is an image processing technique that provides a multi-scale representation of a given image and known to be one of the most powerful and frequently used feature extraction techniques in many texture classification methods for medical image analysis problems [8, 11]. DWT gains its popularity from being useful for classification problems that have a lot of details such as in brain MRI since it extracts the basic structural information from an image by giving the localized information at various levels of resolution. DWT concept is separating the data into different frequency components and studying each component with resolution matched to its scale [11, 12, 13]. Practically in 2D images, it is employed to each dimension separately to linearly transform the input data vector whose length is an integer power of two into more meaningful numerically different vector of the same length [11, 13]. DWT is implemented using cascaded filter banks (low-pass and high-pass filters) and the decomposition schema is demonstrated in Figure 4 where the functions *h(n)* and *g(n)* represent the coefficients of the high-pass and low-pass filters, respectively. As a result, at each level of the image decomposition four sub-band imagesare obtained (LL, LH, HH and HL). LL sub-band is considered as the approximation component of the image, while the LH, HL, HH sub-bands considered as the vertical, horizontal, and diagonal detailed components of the image [12, 13].

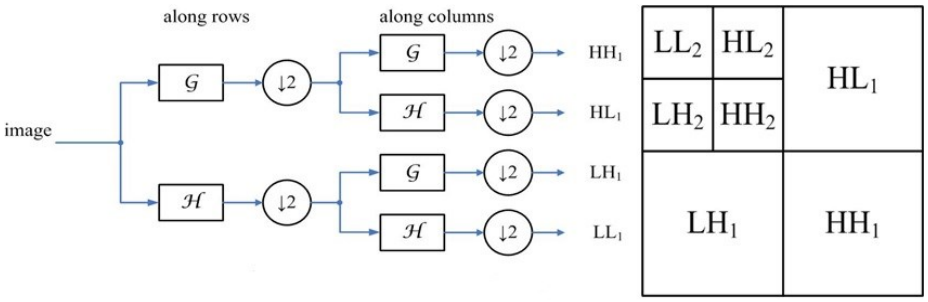


Figure 4. DWT decomposition scheme

There are various types of basis wavelet functions, for example, Haar (haar), Daubechies (db), symlets (sym), coiflets (coif), and biorthogonal (bior) [13, 14]. In this work, we computed a3-levels of decomposition for the approximation coefficient based on Haar basis function wavelet. A demonstration of the 3-levels of wavelet decomposition on a sample brain MRI is shown in Figure5

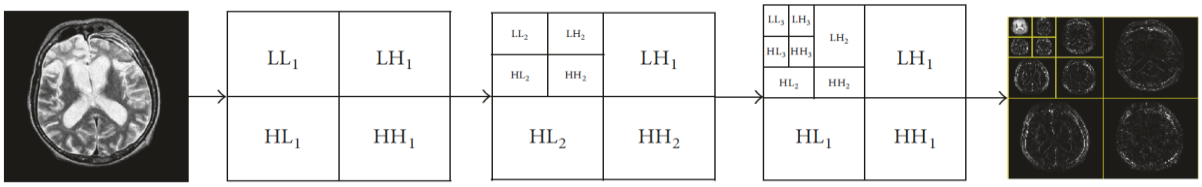


Figure 5. A sample of brain MRI with 3-levels of wavelet decomposition

The extracted features from the 3rd level of approximation only for each image of original size 256 * 256 was 32 * 32 =1024 features which is a large number of features to be used results in more classification complexity. Hence, the feature vector dimensionality should be reduced using an efficient technique as the performance of the classifiers on high-dimensional data depends on the selection of features as an initial step.

Statistical models such as PCA technique are successfully used for this process. PCA is a famous dimensionality reduction technique that is applied to map features into lower dimensional space using linear or nonlinear transformation [8]. However the linear transformation is the most frequently used where an orthogonal transformation is used on the original correlated features to produce a linearly uncorrelated variables which are the principle components. The number of principal components is lower than or equal to the number of original features. This technique is unsupervised one that searches for directions in the data which have largest variance and subsequently project the data onto it resulting into a lower dimensional representation with maximizing the variance of the projected data and the "noisy" directions are moved as well [12, 15].

The PCA technique can be described as following steps [8]:

a) Calculating the mean of the data and zero mean data
b) Constructing the covariance matrix
c) Acquiring the eigenvalue and the eigenvector
d) Projecting the data matrix with eigenvectors corresponding to the highest to lowest eigenvalues.

In this work we employed the PCA for dimensionality reduction to facilitate LDA classification. Both DWT and PCA techniques was implemented in Matlab R2015a using the wavelet and machine learning toolboxes.

2.3 Classifying CN and AD subjects

LDA is a statistical technique that has been widely used to predict class membership of observations, based onlinear projection of features and finding a coordinate axes retains the class separablility while reducing the variation within each class, and it showed high classification accuracy in many applications [8, 16, 17].This technique can easily handle the case when the within-class frequencies are unequal. It is a transform-based technique which maximizes the ratio of between-class variance to the within-class variance in any dataset thereby guaranteeing maximal separability using the following objective function [15, 18]:

$$J(w) = \frac{w^T S_B w}{w^T S_W w} = \frac{\sigma^2{}_{between}}{\sigma^2{}_{within}} \quad (1)$$

Where S_B is the "between classes scatter matrix" and S_W is the "within classes scatter matrix" and defined using the following:

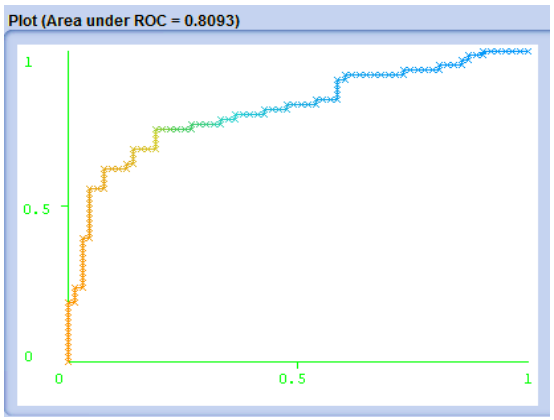
$$S_B = \sum_c (\mu_c - \bar{x})\,(\mu_c - \bar{x})^T \quad (2)$$

$$S_W = \sum_c \sum_{i \in c} (x_i - \mu_c)(x_i - \mu_c)^T \quad (3)$$

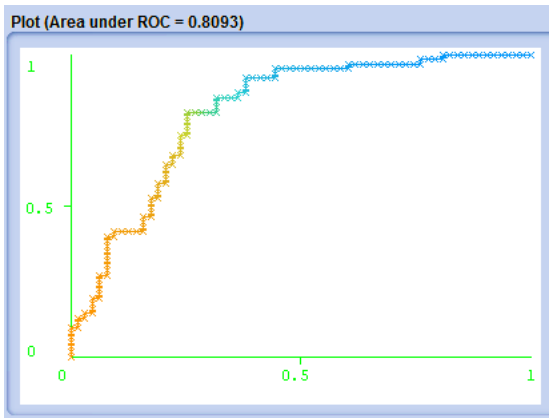
In fact, LDA has a number of highly desirable properties as a classifier. First, it is based on very simple geometric reasoning using the information in both the means and the variances of the data. Second, due to its simplicity, LDA can be applied to multiclass problems. Finally, precisely because of its simplicity, algorithms for implementing it are highly efficient [19].However, it suffers from the singularity of scatter matrices when dealing with a small number of training samples and a high dimensional feature vectors. This problem is resolved by using the combination of PCA and LDA as the dimension of feature vectors is reduced by the PCA. In this work, we used the LDA classifier available in Waikato Environment for Knowledge Analysis (WEKA) package [20].

3. Experimental evaluation

In order to have an unbiased estimate for the performance of the proposed methodology, we applied a 6-fold cross validation over each of the two datasets used where each dataset was randomly partitioned into 6 subsets then use 5 sets out of 6 for training and the remaining one for testing and this process is repeated for 6 times reciprocally. Then, we evaluated the performance with respect to the average classification rate which was 77.78% with area under the ROC curve (AUC) = 0.809 over ADNI dataset and 94.59% over Harvard medical school dataset with AUC of ROC = 0.963. The AUC of ROC over the two datasets for each class (CN and AD) is shown in Figure 6 and Figure 7 also the performance measures for the two datasets are shown in table 1 and table 2.

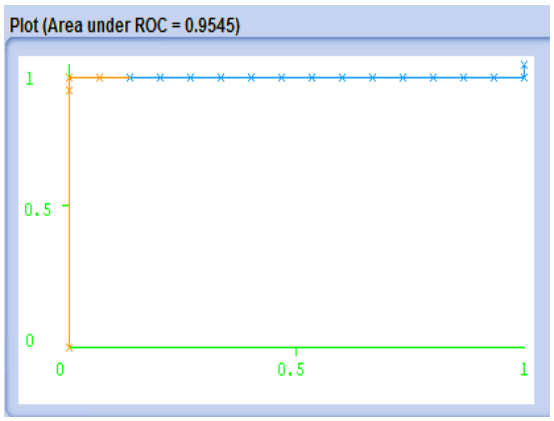


(a) AUC of CN class = 0.8093

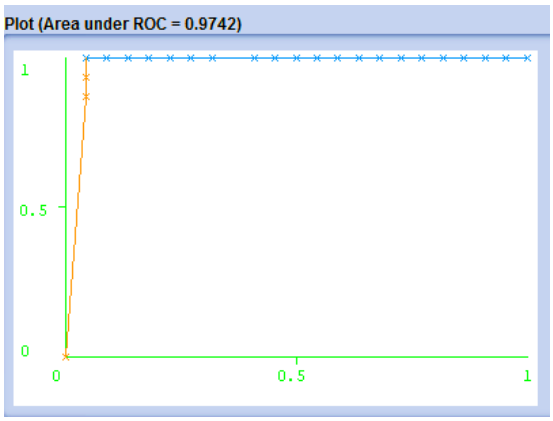


(b) AUC of AD class = 0.8093

Figure 6. AUC for CN and AD classes over ADNI dataset



(a) AUC of CN class = 0.955



(b) AUC of AD class = 0.974

Figure 7. AUC for CN and AD classes over Harvard medical school dataset

Table 1. The performance measures over ADNI dataset

Class	Recall	Precision	F-Measure	AUC (ROC)
CN	0.746	0.797	0.77	0.809
AD	0.810	0.761	0.785	0.809
Average	**0.778**	**0.779**	**0.778**	**0.809**

Table 2. The performance measures over Harvard medical school dataset

Class	Recall	Precision	F-Measure	AUC (ROC)
CN	0.955	0.955	0.955	0.955
AD	0.933	0.933	0.933	0.974
Average	**0.946**	**0.946**	**0.946**	**0.963**

As seen from the tables, the performance of the proposed methodology was also measured in terms of recall, precision and F-Measure besides the classification rate and the AUC of ROC for more accurate evaluation. Our methodology achieved an average recall of 0.778, an average precision of 0.779 and an average F-Measure of 0.78 over ADNI dataset. While it achieved an average recall of 0.778, an average precision of 0.779 and an average F-Measure of 0.78 over Harvard medical school dataset. It was noticed that the difference in contrast of the Brain MRIs in the two datasets was a factor that leaded to higher performance results using T2-weighted MRI scans from Harvard medical school dataset although all the MRI scans from ADNI was pre-processed. Other factors that could affected the methodology performance is the selection of the MRI scans from each subject.

4. Conclusion and future works

The proposed methodology was evaluated using two different Brain MRI datasets and the overall performance considered the differential diagnosis of CN and AD subjects using LDA based classifier with a classification accuracy = 77.78% and average AUC of ROC = 0.809 over ADNI dataset and 94.59% and average AUC of ROC = 0.963over Harvard medical school dataset. The methodology included feature extraction using DWT, feature selection using PCA and finally the classification using LDA classifier.

In future works, more classification techniques could be considered and also image segmentation using various common techniques like fuzzy c-means and k-means or using the common used packages as SPM5 before feature extraction that could be a factor affecting classification performance and leads to a better results of accuracy and robustness over ADNI dataset.

References

[1]. Cai, K., Xu, H., Guan, H., Zhu, W., Jiang, J., Cui, Y., Zhang, J., Liu, T. & Wen, W. (2017). Identification of Early-Stage Alzheimer's Disease Using Sulcal Morphology and Other Common Neuroimaging Indices. PloS one, 12(1), e0170875.

[2]. Amulya, E. R., Varma, S., & Paul, V. (2016, December). Classification of brain images for Alzheimer's disease detection. In Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on (pp. 1-4). IEEE.

[3]. Previtali, F., Bertolazzi, P., Felici, G., &Weitschek, E. (2017). A novel method and software for automatically classifying Alzheimer's disease patients by magnetic resonance imaging analysis. Computer Methods and Programs in Biomedicine, 143, 89-95.

[4]. Kruthika, K. R., Pai, A., Maheshappa, H. D., &Alzheimer's Disease Neuroimaging Initiative. (2017). Classification of Alzheimer and MCI Phenotypes on MRI Data Using SVM. In International Symposium on Signal Processing and Intelligent Recognition Systems, Springer, Cham, September 2017, pp. 263-275.

[5]. Jha, D., Kim, J. I., & Kwon, G. R. (2017). Diagnosis of Alzheimer's disease using dual-tree complex wavelet transform, PCA, and feed-forward neural network. Journal of Healthcare Engineering, 2017.

[6]. Geetha, C., &Pugazhenthi, D. (2017). Classification of alzheimer's disease subjects from MRI using fuzzy neural network with feature extraction using discrete wavelet transform. Biomedical Research, 1-1.

[7]. Suk, H. I., & Shen, D. (2013, September). Deep learning-based feature representation for AD/MCI classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 583-590). Springer, Berlin, Heidelberg.

[8]. Alam, S., Kwon, G. R., Kim, J. I., & Park, C. S. (2017). Twin SVM-Based Classification of Alzheimer's Disease Using Complex Dual-Tree Wavelet Principal Coefficients and LDA. Journal of Healthcare Engineering, 2017.

[9]. http://adni.loni.usc.edu/

[10]. http:// med.harvard.edu/AANLIB/

[11]. A.B. AL-Khafaji (2015) Classification of MRI Brain Images using Discrete Wavelet Transform And K-NN, International Journal Of Engineering Sciences & Research Technology, Vol. 4, No. 11, November 2015.

[12]. Kalbkhani, H., Shayesteh, M. G., &Zali-Vargahan, B. (2013). Robust algorithm for brain magnetic resonance image (MRI) classification based on GARCH variances series. Biomedical Signal Processing and Control, 8(6), 909-919.

[13]. Jha, D., Kim, J. I., Choi, M. R., & Kwon, G. R. (2017). Pathological Brain Detection Using Weiner Filtering, 2D-Discrete Wavelet Transform, Probabilistic PCA, and Random Subspace Ensemble Classifier. Computational Intelligence and Neuroscience, 2017.

[14]. Tzalavra, A., Dalakleidi, K., Zacharaki, E. I., Tsiaparas, N., Constantinidis, F., Paragios, N., & Nikita, K. S. (2016, October). Comparison of Multi-resolution Analysis Patterns for Texture Classification of Breast Tumors Based on DCE-MRI. In International Workshop on Machine Learning in Medical Imaging (pp. 296-304). Springer International Publishing.

[15]. Welling, M. (2005). Fisher linear discriminant analysis. Department of Computer Science, University of Toronto, 3(1).

[16]. Usha.B.L, Supreeth.H.S.G (2017) Brain Tumor detection and identification in brain MRI using supervised learning: A LDA based classification method, International Research Journal of Engineering and Technology (IRJET), Vol. 4, No. 6, June 2017

[17]. Zhang, Y., Zhou, X., Witt, R. M., Sabatini, B. L., Adjeroh, D., & Wong, S. T. (2007). Dendritic spine detection using curvilinear structure detector and LDA classifier. Neuroimage, 36(2), 346-360.

[18]. Balakrishnama, S., &Ganapathiraju, A. (1998). Linear discriminant analysis-a brief tutorial. Institute for Signal and information Processing, 18.

[19]. Vogelstein, J. T., Tang, M., Zheng, D., Burns, R., &Maggioni, M. (2017). Linear Optimal Low Rank Projection for High-Dimensional Multi-class Data. arXiv preprint arXiv:1709.01233.

[20]. https://www.cs.waikato.ac.nz/