International Journal of Latest Research in Engineering and Technology (IJLRET)
ISSN: 2454-5031
www.ijlret.com || Volume 03 - Issue 04 || April 2017 || PP. 90-96

A Comparison between Linear SVM and Logistic Regression Classifiers towards the Early Diagnosis of Alzheimer's disease

Abigail Copiaco[1], Nidhal Abdulaziz[2]
[1,2]*(Faculty of Engineering and Information Sciences, University of Wollongong in Dubai, United Arab Emirates)*

Abstract: Alzheimer's disease (AD) is a tenacious neurodegenerative brain disorder that adversely influences its victims' memory. Due to its progressive property, early detection of the disorder is crucial for regular observation of the patient's condition. This research concentrates towards determining an optimum classification algorithm for an early and rapid diagnosis of AD. The proposed method constitutes for high efficiency through a Computer visionimage enhanced MRI brain scan where brain volume features are extracted via the regional MRI measurementtechnique, fed separately to a Linear Support Vector Machineand a Logistic Regression classifier for result verification and method comparison. Effectiveness of the suggested method is then computed through the accuracy, sensitivity, specificity, and precision. For 100 MRI brain scans, 30 being the test set, results indicate 73% accuracy using the Linear Support Vector Machines method, while 50% accuracy is achieved with Logistic Regression. However, improvement is expected upon the application of the same technique in a larger sample set of brain scans.

Keywords: Alzheimer's Disease, Linear SVM, Logistic Regression, Regional MRI Measurements, Computer Visionimage Enhancement

I. INTRODUCTION

Alzheimer's Disease (AD) is the most customary type of dementia [*1*,*2*]. It is a progressive neurodegenerative brain condition that brings about detrimental effects on the victim's memory and coherence, eventually interceding with daily activities [*1*,*2*]. The distinctive attribute of AD is the progressive accumulation of tangles and plaques within the brain neurons [*3*]. Studies had declared the hippocampus and the entorhinal cortex to be the main regions of AD development, which are mostly affected by tangles and cell loss, causing anexplicit change in the volume of these regions [*1*,*3*,*4*].

Although several AD detection techniquescurrently exist, majority of such tests, such as neuropsychological tests like the mini-mental state exam (MMSE) [*5*], are subjective and cannot be scientifically proven. The conventionality of AD worldwide also manifestsnotable research importance [*6*,*7*,*8*]. Moreover, diagnosis research within the medical field opens opportunities for further innovation towards medical image processing, such as extending the scope towards differential detection of dementia [*9*].

With early detection of AD being the main objective, this research aims for a comparison between two major classification algorithms, the Linear Support Vector Machines (SVM)[*3*,*10*,*11*]and the Logistic Regression[*12*,*13*],in order to achieve optimum efficiency in categorizing extracted features from the regional MRI measurements from a computer visionimage enhanced MRI brain scan[*14*,*15*,*16*,*17*].It is aimed to alleviate the difficulty of coping with the ailment through progress supervision for familiarized mitigation practices, and to alleviate progress done in the field of AD diagnosis through a comparison and identification of the optimum method of feature classification. However, this research solely intends to focus on the early and rapid identification of AD, and not the other forms of dementia.

II. RELATED WORK AND METHOD JUSTIFICATION

Disease diagnosis through medical images normally branches into three categories, namely: Pre-processing, Feature extraction, and Feature classification. Although various techniques exist for each category, each corresponds with particular benefits and drawbacks. In the caseof early diagnosis of AD via MRI images, advantages in terms of time duration and accuracy are considered upon the selection of methodologies.

Pre-processing is a crucial step needed for preparing the MRI image for feature extraction. The desired outcome of this procedure is an enhanced, feature-amplified version of the original image, making it easier to detect and extract the desired features.Putting forward the need for feature amplification and image intensity uniformity, pre-processing via Computer Visionimage Enhancement poses the maximum advantage [*14*,*15*,*16*,*17*]. Such technique results in a contrast and clarity enhanced, scaled intensity image [*14*]. Although other several non-biased techniques such as the Spatial Normalization and the Discrete Wavelet Transform are studied, such methods commence disadvantages in terms of complexity, possibility of information loss, and prolonged completion duration[*18*,*19*].

Several methods of feature extraction were also discussed. However, techniques such as the Skull stripping, Detrended Fluctuation analysis, and Weibull distribution for Edge detection are computationally extensive and are vulnerable to noise[*20*,*21*,*22*]. However, the method of Regional MRI Measurements takes advantage of the change in volume and thickness regions affected by AD [*1*,*3*,*4*], coupled with findings from neuropsychological tests[*5*]. Hence, it supports the theoretical information the best. The approach is purely mathematical, accounting for fast computation time, and uses imaging software called Freesurfer Pipeline for completion [*23*,*24*]. However, a technique to apply the methodology without the Freesurfer software is to be established.

Finally, a variety of feature classification techniques were taken into consideration. Nonetheless, common drawbacks on majority of the classifier algorithms such as the Random forest classifier, Naïve bayes classifier, and Adaboost include accuracy inferiority, biasing, and excessive inclusions of assumptions [*25*,*26*,*27*]. Support Vector Machines (SVM), however, is proven to achieve high accuracies [*3*,*10*,*11*], while Logistic Regression is well known for its flexibility and robustness [*12*,*13*]For these reasons, Linear SVM and Logistic Regression classifiers are going to be compared within this paper.

III. METHODOLOGY

3.1 Pre-processing

The MRI images used for this research work are sets of 3-plane localized axial MRI brain scans of 10 mm slice thickness because of the high perceptibility of the prominent regions of AD [*4*,*28*]. The set of images are also segregated into AD and non-AD patients, through the results of the MMSE (Mini Mental State Exam) given upon downloading the images.Such divergence among data sets is an important component for the feature extraction technique, and is also an important requirement for testing the efficiency of the proposed method.

The MRI images are to be utilized in a DICOM (Digital Imaging and Communications in Medicine) format, which conserves theimage quality, a mandatory factor for medical purposes [*29*]. This format, however, is to be translated into matrix form in order to be processed in MATLAB.

In order to pre-process the image for the preparation of feature extraction, the method of Computer Visionimage Enhancement is utilized [*14*,*15*,*16*,*17*]. This method branches out into Colour translation into grayscale, which is achieved via adjusting every intensity of the image to a scale of 0 to 1 based on the maximum value[*15*]; and a contrast and clarity enhancement, which is achieved via contrast stretch limit rather than the original method of Histogram equalization for a faster result[*30*,*31*]. A summary of the pre-processing method is observed in Figure 1.

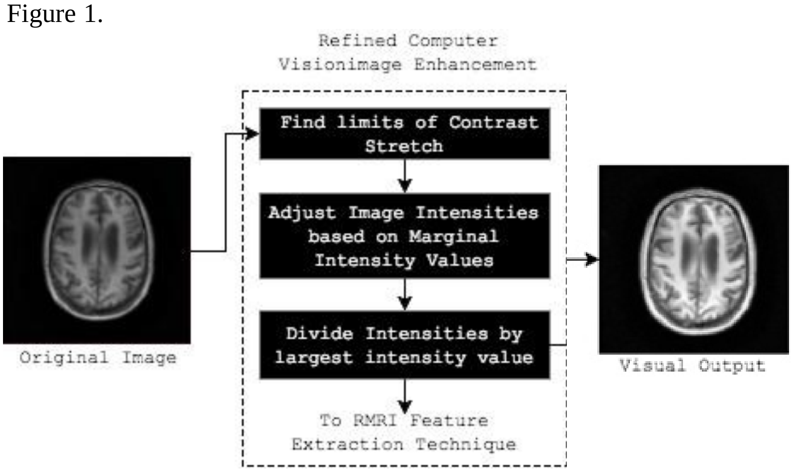

Figure 1: Refined Computer Visionimage Enhancement

As observed, the visual output is a brighter, more clarified version of the original image, with intensities scaled to 0 to 1, making calculation of the brain voxels and volume for feature extraction easier.

3.2 Feature Extraction

For the reason that the main priorities of early Alzheimer's detection are accuracy and time duration, the Regional MRI measurements feature extraction method was selected to conduct feature extraction. This method concentrates on brain volume, which supports the theoretical information provided about AD, which states that the change in the brain upon AD can be perceived from its volume [4,28]. However, instead of the norm methodology of extracting Regional MRI measurements via Freesurfer [23,24],this research aims to conduct this through MATLAB, with the use of image processing and mathematical calculations. A summary of the methodology can be observed in Figure 2.

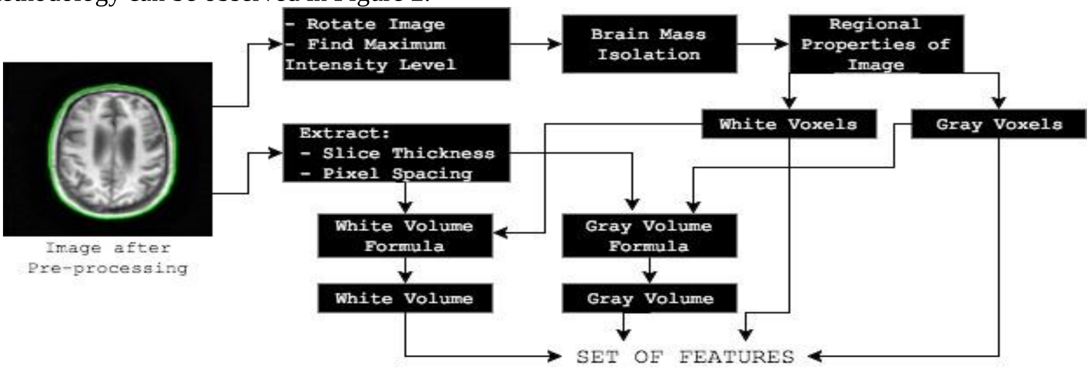

Figure 2: Regional MRI Measurements via MATLAB

The set of features to be extracted composes of four parameters, according to the brain image's white matter and gray matter. The following features are: the white voxel, gray voxel, white volume, and gray volume. The volumes are obtained via the formulas defined in Eq. 1 and 2 [32].

$$Grey\ Volume = Gray\ voxels * \frac{Product\ \ of\ Voxel\ Sizes\ of\ Images}{e^6} \quad \text{(Eq. 1)}$$

$$White\ Volume = White\ voxels * \frac{Product\ \ of\ Voxel\ Sizes\ of\ Images}{e^6} \quad \text{(Eq. 2)}$$

3.3 Feature Classification

3.3.1 Linear Support Vector Machines

The Linear Support Vector Machine works by ensuring that the classification error is kept as low as possible, while keeping the distance of two vectors belonging in different classes at a maximum [3,10,11], Operation starts from generating a linear kernel matrix via the multiplication of MRI voxels for the interpretation of the matrix in proportion to the resemblance measures of every feature [33]. The most favourable choice for the linear hyper plane between two classes is the provision of the utmost margin from both of the classes [34,35].

In MATLAB, Linear SVM is performed by creating an SVM model by training the classifier according to the 60 MRI data set's four features defined by the Regional MRI measurements, initially categorized by the corresponding MMSE score. This model is then utilized to categorize any set of new data having the same four features.

3.3.2 Logistic Regression Classifier

The Logistic Regression Classifier, known for its robustness and flexibility, fundamentally functions by extracting a set of features, followed by the application of the exponential function, and linear integration[12,13]. This therefore implieseach feature to be multiplied to acertain weight, while beinggathered accordingly. Note that the weights are extracted via gradient ascents, which begins with zero and then advances towards the positioning of the partial derivative of the gradient and the objective function[12,13].

The Logistic Regression Classifier is implemented in MATLAB via the Classification Learner Application, where data is uploaded defining both the features and the initial classification and selecting the predictors and response columns. A function must then be generated accordingly, which can be retrained and utilized for categorization.

3.4 Testing

Four performance parameters are calculated upon gathering the results of the experiment in order to ensure the efficiency of the designed methodology. Accuracy, or precision, considers the correctly identified classes against the entire sample set. Sensitivity, regards to theamenability of the methodeven to lower levels of AD. Finally, the specificity and the positive predictive value measure the methodology's efficiency in detecting affected patients[36]. The formulas for these parameters are observed in Eq. 3 to 6 [36]:

$acc = \frac{TP+TN}{TP+TN+FN+FP}$ (Eq. 3) $sens = \frac{TP}{TP+FN}$ (Eq. 4)

$ppv = \frac{TP}{TP+FP}$ (Eq. 5) $spc = \frac{TN}{TN+FP}$ (Eq. 6)

Such that:

TP – True Positive (correctly detected AD carrier)
TN – True Negative (correctly detected non-AD carrier)
FP – False Positive (incorrectly detected AD carrier)
FN – False Negative (incorrectly detected non-AD carrier)

IV. RESULTS

A total of 100 MRI brain scans were subjected to the proposed methodology, with 70 being the training dataset, and 30 being the test set, 15 of which are AD-positive, and the remaining 15 being AD-negative. The MRI data is to be collected from ADNI (Alzheimer's Disease Neuroimaging Initiative), an online database of medical related images available for researchers [*37*]. Following the Computer visionimage enhancement pre-processing, feature extraction through Regional MRI measurements, and the two feature classification techniques, results of the performance measurements for both the Linear SVM classifier and the Logistic Regression classifier is reported in summary in the form of Table 1.

Table 1: Performance Measurements Results

Performance Measurement	Linear SVM Classifier	Logistic Regression Classifier
Accuracy	73%	50%
Sensitivity	67%	0%
Specificity	89%	50%
Positive Predictive Value	93%	0%

As observed, for a 70 MRI training data set, the Linear SVM classifier performs significantly better in all four of the performance measurements. Such difference in values are mostly accustomed to the findings that the Logistic Regression classifier requires a large number of training data in order to expect meaningful results[*38*]. Since the training data used is merely 70, the high accuracy property of the Linear SVM is mostly observed. The 0% result for the sensitivity and positive predictive value of the Linear Regression Classifier implies that it fails to correctly identify positive AD carriers among the 30 test subjects. Testing to a larger test set may be able to improve percentage on this matter.

Aiming to increase the accuracy of the Linear SVM classifier, it is aimed to improve that an increase in training data set would improve the efficiency of the overall technique. To prove this, the pattern of percentage decrease was observed upon the decrease of the training set, as observed in Figure 3. Hence, similarly, it can be deduced that an increase in the number of training data would also increase the percentage of performance measurements.

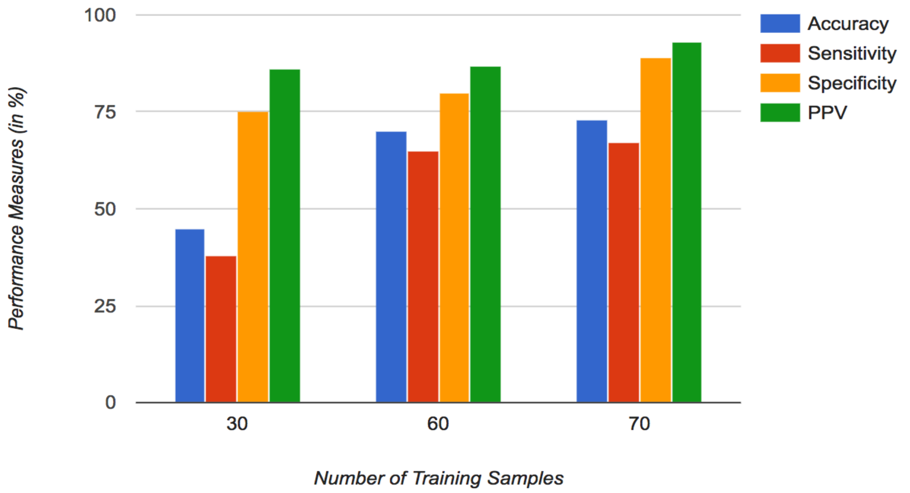

Figure 3: Linear SVM Efficiency Chart

V. CONCLUSION

In conclusion, it was extrapolated that fabricating on the upheaval AD detection permits constant monitoring of the disorder’s progress. Taking into consideration the pros and cons of every technique of pre-processing, feature extraction, and feature classification, it has been finalized that both the Linear SVM classifier [*3*,*10*,*11*], and the Logistic Regression classifier[*12*,*13*]make reasonable choiceswhen it comes to feature classification.Thus, a comparison between these two methods is put forward, concluding with a better method of feature classification. For pre-processing, Computer visionimage enhancement is selected due to its clarity enhancement and intensity adjustment characteristics minus biasing and information loss[*14*,*15*,*16*,*17*]. While for feature extraction, Regional MRI measurementwas chosen due to its alignment to the theoretical information of AD and its minimal time consumption[*23*,*24*].

For a sample set of 100, 70 being the training set, and 30 being the test set, experimental results state 73% accuracy for the Linear SVM classifier, while 50% is achieved via Logistic Regression. Linear SVM classifier also significantly leads the latter in all four of the performance parameters. However, a continual increase in the training data set is proven to be directly proportional to the performance measurements, which therefore increases them accordingly. Once optimum performance is achieved, innovation of the proposed methodology can be deviatedto the advancement of the technique’s rapidness and complexity reduction, as well as fathoming outothertechniques to be implementedalong with the current algorithms. Moreover, the proposed methodology can be developed into differential diagnosis of Dementia through further research[*9*].

REFERENCES

[1]. Alzheimer's Association. (2016) 10 Early Signs and Symptoms of Alzheimer's. [Online]. www.alz.org/10-signs-symptoms-alzheimers-dementia.asp

[2]. M. Archana and S. Ramakrishnan, "Detection of Alzheimer disease in MR Images using structure tensor," in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, 2014, pp. 1043-1046.

[3]. M. Rangini and G.W. Jiji, "Detection of Alzheimer's disease through automated hippocampal segmentation," in Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on, Kottayam, 2013, pp. 144-149.

[4]. R. Alattas and B.D. Barkana, "A comparative study of brain volume changes in Alzheimer's disease using MRI scans," in Systems, Applications, and Technology Conference (LISAT), 2015 IEEE Long Island, Farmingdale, NY, 2015, pp. 1-6.

[5]. M. Wallace L. Kurlowicz, "The Mini Mental State Examination (MMSE)," The Hartford Institute for Geriatric Nursing, New York, Questionnaire Issue Number 3, 1999.

[6]. Alzheimer's Association, "2016 Alzheimer's Disease Facts and Figures," Alzheimer's and Dementia, Includes a Special Report on the Personal Financial Impact of Alzheimer's on Families 12(4), 2016.

[7]. I.O. Korolev, "Alzheimer's Disease: A Clinical and Basic Science Review," Medical Student Research Journal, vol. 4, pp. 24-33, September 2014.

[8]. Alzheimer's Association, "Basics of Alzheimer's Disease," Alzheimer's Association, 770-10-0003, 2011.

[9]. Khan Academy Medicine. (2015, September) Diagnosis of Dementia and Alzheimer's Disease: Video. [Online]. www.youtube.com/watch?v=LieVEfl4luw&list=PLbKSbFnKYVY3_PviSE5ANWtMPXAXpMj5-&index=6

[10]. Z. Tu, L.G. Apostolova, A.E. Green, A.W. Toga, and P.M. Thompson J.H. Morra, "Comparison of Adaboost and Support Vector Machines for Detecting Alzheimer's Disease Through Automated Hippocampal Segmentation," IEEE Transactions on Medical Imaging, vol. 29, no. 1, pp. 30-43, January 2010.

[11]. Z. Pan, H. Lu and Adni L. Huang, "Automated Diagnosis of Alzheimer's Disease with Degenerate SVM-based Adaboost," in Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2013 5th International Conference on, Hangzhou, 2013, pp. 298-301.

[12]. D. Jurafsky and J.H. Martin, "Chapter 7. Logistic Regression," in Speech and Language Processing., 2016.

[13]. S.J. Phillips, R.E. Schapire M. Dudik, "Maximum Entropy Density Estimation with Generalized Regularization and an application to species distribution modeling," Journal of Machine Learning Research, vol. 8, no. 6, 2007.

[14]. K. Padmavathi K. Thangadurai, "Computer Visionimage Enhancement for Plant Leaves Disease Detection," in World Congress on Computing and Communication Technologies, 2014, pp. 173-175.

[15]. K. Verma T. Kumar, "A Theory Based on Conversion of RGB image to Gray Image," International Journal of Computer Applications, vol. 7, no. 2, pp. 0975-8887, September 2010.

[16]. D. Tenorio R. Krutsch, "Histogram Equalization," Microcontroller Solutions Group, Freescale Semiconsudctor Document No.: AN4318, June 2011.

[17]. Y. Singh K. Vij, "Enhancement of Images using Histogram Processing Techniques," Comp. Tech. Appl., vol. 2.

[18]. Y. Zhang and L. Wu, "An MR Brain Images Classifier via Principal Component Analysis and Kernel Support Vector Machine," Progress in Electromagnetics Research, vol. 130, pp. 369-388, June 2012.

[19]. W.A. Serdijn, and I.W. Selesnick A.N. Akansu, "Wavelet Transforms in Signal Processing: A Review of Emerging Applications," Physical Communication, Elsevier, vol. 3, no. 1, pp. 1-18, March 2010.

[20]. A. El-Zaart W.K. Al-Jibory, "Edge Detection for diagnosis early Alzheimer's disease by using Weibull distribution," in 2013 5th International Conference on Microelectronics (ICM), Beirut, 2013, pp. 1-5.

[21]. D. Ray, "Edge Detection in Digital Image Processing," Department of Mathematics, University of Washington, Washington, 2013.

[22]. F. Segonne et al., "A hybrid approach to the skull stripping problem in MRI," Elsevier Journals, vol. NeuroImage, no. 22, pp. 1060-1075, 2004.

[23]. Q. Zhou et al., "Regional MRI measures and neuropsychological test for multi-dimensional analysis in Alzheimer's disease," in Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on, , San Diego, CA, 2013, pp. 533-536.

[24]. M. Jalbrizikowski et al., "Structural abnormalities in cortical volume, thickness, and surface area in 22q11.2 microdeletion syndrome," Elsevier Journals, vol. NeuroImage, no. Clinical 3, pp. 405-415, 2013.

[25]. D. Bhalla. (2014, November) Listen Data: Random Forest Explained in Simple Terms. [Online]. www.listendata.com/2014/11/random-forest-with-r.html

[26]. I.C. Mihaescu, "Naive-Bayes Classification Algorithm," Univesity of Craiova, Department of Engineering Software, Romania, Laboratory Module.

[27]. L. Wang, and E. Sung X. II, "AdaBoost with SVM-based component classifiers," Engineering Applications of Artificial Intelligence, vol. 21, no. 5, pp. 785-795, 2008.

[28]. Alzheimer's Association. (2016) Types of Dementia. [Online]. www.alz.org/dementia/types-of-dementia.asp

[29]. Digital Imaging and Communications in Medicine. About DICOM. [Online]. dicom.nema.org/Dicom/about-DICOM.html

[30]. Image Analyst. (2013, January) Histogram of ROI of an Image : Mathworks, MATLAB Answers, Website. MATLAB.

[31]. Image Analyst. (2009, July) Specific Area Inside of an Image : Mathworks, MATLAB Newsgroup. Website. MATLAB.

[32]. R. Bemis. (2004, May) MRI Brain Segmentation : Mathworks File Exchange. Website.

[33]. S. Kloppel et al., "Automatic classification of MR scans in Alzheimer's disease," in Brain 131, 2008, pp. 681-689.

[34]. F. de la Torre M.H. Nguyen, "Optimal feature selection for support vector machines," Elsevier: Pattern Recognition, vol. 43, pp. 584-591, February 2010.
[35]. T.S. Korting. (2014, January) How SVM (Support Vector Machines) Algorithm Works. Video.
[36]. A. Popov, J. Ramirez, and J.M. Gorriz I. Krashenyi, "Fuzzy computer-aided diagnosis of Alzheimer's disease using MRI and PET statistical features," in 2016 IEEE 36th International Conference on Electronics and Nanotechnology (ELNANO), Kiev, 2016, pp. 187-191.
[37]. Alzheimer's Disease Neuroimaging Initiative. (2016) ADNI, University of Southern California: Mark and Mary Stevens Neuroimaging and Informatics Institute. [Online]. adni.loni.usc.edu
[38]. V. Fang. (2011, May) Advantages and Disadvantages of Logistic Regression. Website.