International Journal of Computer Engineering and Applications,
Volume XII, Issue III, March 18, www.ijcea.com ISSN 2321-3469

EVALUATION OF CLASSIFICATION OF BRAIN IMAGES INTO ALZHEIMER DISEASE AND NORMAL

SUMANTH S[1] , A SURESH [2]

[1] *Research Scholar, Periyar University, Salem, Tamilnadu, India*

[2] *Principal, Maisurii Women's College of Arts and Science, Kakapalayam, Salem, Tamilnadu, India*

ABSTRACT:

Alzheimer's disease (AD) is the most common form of dementia in elderly people worldwide. Most existing pattern classification methods just use one individual modality of biomarkers for diagnosis of AD or MCI, which may affect the overall classification performance. Correlation based Feature Selection (CFS) is a simple filter algorithm that ranks feature subsets and discovers the merit of feature or subset of features according to a correlation based function. This work classifies the brain image as Alzheimer or normal. To classify whether the image is Alzheimer or normal Multi-Layered Perceptron Neural Network (MLPNN) and fuzzy classifiers are used. MLP have been evolved over the years as a very powerful technique for solving a wide variety of problems. Much progress has been made in improving performance of MLP and in understanding how these neural networks gets operate.

Keywords: Alzheimer's disease (AD), Correlation based Feature Selection (CFS), Multi-Layered Perceptron Neural Network (MLPNN) and fuzzy classifiers.

[1] INTRODUCTION

Alzheimer's disease (AD) is the most popular dementia in elderly people worldwide. Its expectation is 1 in 85 people will be affected by 2050 and the number of affected people is double in the next 20 years. Alzheimer reported two common abnormalities in the brain of this patient, "1. Dense layers of protein deposited outside and between the nerve cells. 2. Areas of

damaged nerve fibers, inside the nerve cells, which instead of being straight had become tangled". Moreover, these plaques and tangles have been used to help diagnose AD [1].

There are three phases of AD: preclinical, mild cognitive impairment, and dementia. Preclinical means the starting stage of AD. Mild Cognitive Impairment (MCI) includes "mild changes in memory. Dementia means severity of the disease. The symptoms of AD different between patients. The following are common Symptoms of Alzheimer's: Memory loss that disrupts daily life. Challenges in planning or solving problems. Confusion with time or place. Trouble understanding visual images and spatial relationships. Decreased or poor judgment. Withdrawal from work or social activities.

In image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction [2].

If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data, one of the major problems stems from the number of variables is involved.

Feature selection is the technique of selecting a subset of relevant features for building robust learning models. By removing most irrelevant and redundant features from the data, feature selection helps improve the performance of learning models by: Alleviating the effect of the curse of dimensionality. Enhancing generalization capability. Speeding up learning process. Improving model interpretability. Feature selection also helps people to acquire better understanding about their data by telling them which are the important features and how they are related with each other [3].

A typical feature selection process is divided into 3 steps: Subset Generation: this is the process to conditionally extract the feature subset from the training feature vector matrix prepared for being analyzed by the classifier or a single vector according to some certain criteria. Subset Evaluation: feature selection can be fulfilled in two directions: one is to begin with a feature subset which contains just one element and to increase the capacity of the subset element by element; the other is to begin with a universal set which contains all the elements of the feature subset and to decrease the capacity of the subset element by element. The determination of the stop criterion: each feature subset after assessing is needed to be compared with the stopping criterion to verify if the characteristics of the current subset have attained a pre-set standard. If so, feature selection will stop automatically and the current subset will be considered as the final output; otherwise this process will continue to repeat again and again until a feature subset which meets the stopping criterion appears [4].

Feature selection algorithms are either filter or wrapper models. The former relies on general characteristics of training data to select features without involving learning algorithms. It does not inherit any learning algorithm bias and are cheap computationally as they do not involve induction algorithm. Nevertheless, they risk selecting features subsets which fail to match chosen induction algorithm. The wrapper model needs a predetermined learning algorithm in feature selection where its performance evaluates and determines selected features. For new subset features, a wrapper model learns a hypothesis (or classifier). It ensures superior performance as it locates features suited for a predetermined learning algorithm. On the other hand, it is expensive computationally.

Wrapper model uses classification to measure a features set's importance and so the selected feature depends on classifier used. Wrapper methods lead to better performance than filter methods as feature selection is optimized for classification algorithm used. But, wrapper methods are expensive for large dimensional database regarding computational complexity and time as every feature set is evaluated by a classifier algorithm [5].

Classification is one of the most frequently encountered decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. There are many industrial problems identified as classification problems. For examples, Stock market prediction, Weather forecasting, Bankruptcy prediction, Medical diagnosis, Speech recognition, Character recognitions [6]. These classification problems can be solved both mathematically and in a non-linear fashion. The difficulty of solving such problem mathematically lies in the accuracy and distribution of data properties and model capabilities

Classification is also called supervised learning, as the instances are given with known labels, contrast to unsupervised learning in which labels are not known. Each instance in the dataset used by supervised or unsupervised learning method is represented by set of features or attributes which may be categorical or continuous [7]. Classification is the process of building the model from the training set made up of database instances and associated class label. The resulting model is then used to predict the class label of the testing instances where the values of the predictor features are known. Supervised classification is one of the tasks most frequently carried out by intelligent techniques. The large number of techniques has been developed.

[2] RELATED WORKS

Roman & Pascual [8] analyzed recent findings relevant to the contribution of neuroimaging to the diagnosis of Alzheimer's Dementia (AD) and Vascular Dementia (VaD). Computerized Tomography (CT) or Magnetic Resonance Imaging (MRI) has been provided accurate demonstration of the location and rate of progression of atrophic changes affecting the brain in AD and the different types of vascular lesions observed in mixed dementias and in pure VaD. Quantification of cortical thickness allowed early diagnosis and rate of progression from MCI to dementia. White Matter (WM) involvement can also be quantified with Diffusion Tensor Imaging (DTI) and functional MRI (fMRI), functional connectivity, and Magnetic Resonance Spectroscopy (MRS).

Akbarpour et al., [9] proposed a new method for extraction of regions affected by AD from multispectral medical images. In this method, first two models of MRIs are fused to be achieved an image with high information content. Statistical features of fused image are extracted and then are grouped into three clusters with the help of an unsupervised algorithm to perform initial segmentation. Labeling members of clusters and rearranging image yields final image. Results of quantitative analysis proved combination of fusion and segmentation to result in an image with higher values of quantitative metrics and better visual outcome.

Zheng et al., [10] presented a summary of existing automatic dementia detection protocols in literature from the perspective of patterns classification. Because mostly these protocols comprise features extraction as well as classification, they offer a review on the three groups of features extraction techniques which are voxel-, vertex- as well as RoI-based ones as well as four groups of classifiers which are the Linear Discriminant Analysis (LDA), Bayes classifier, Support Vector Machine (SVM) as well as Artificial Neural Networks (ANN). The performance of the classifiers are contrasted and the comparison reveals that several protocols are capable of distinguishing AD from Head Circumference (HC)s with excellent accuracies although differentiating HCs from those suffering from MCI is still a difficult task.

Chu et al., [11] proposed four common feature selection methods. (1) Pre-selected Region of Interests (ROIs) that are based on prior knowledge. (2) Univariate t-test filtering. (3) Recursive Feature Elimination (RFE), and 4) t-test filtering constrained by ROIs. The predictive accuracies achieved from different sample sizes, with and without feature selection, were compared statistically. To demonstrate the effect, it used Grey Matter (GM) segmented from the T1-weighted anatomical scans collected by the Alzheimer's Disease Neuroimaging Initiative (ADNI) as the input features to a linear SVM classifier. The objective was to be characterized the patterns of difference between AD patients and Cognitively Normal (CN) subjects, and also to be characterized the difference between MCI patients and normal subjects. Therefore, feature selection does improved the classification accuracies, but it depends on the method adopted.

Liu et al., [12] proposed a novel multi-task feature selection method to be preserved the complementary inter-modality information. Specifically, it can be treated feature selection from each modality as a separate task and further impose a constraint for preserving the inter-modality relationship, besides separately enforcing the sparseness of the selected features from each modality. After feature selection, a multi-kernel SVM was further used to be integrated the selected features from each modality for classification. The method was evaluated using the baseline Positron Emission Tomography (PET) and MRI images of subjects obtained from the ADNI database.

Liu et al., [13] proposed a linear sparse SVM to build classifiers for distinguishing AD and MCI subjects from CN subjects based on different combinations of regional measures extracted from imaging data, including perfusion and amyloid deposition information extracted from early and late frames of 11C-PIB separately, and GM volumetric information extracted from structural MRI (sMRI) data. The experimental results demonstrated that the classifier built

upon the combination of imaging measures extracted from early and late frames of 11C-PIB as well as sMRI achieved the highest classification accuracy in both classification studies of AD (100%) and MCI (85%), indicating that multimodality information could aid in the diagnosis of AD and MCI.

Dyrba et al., [14] proposed a SVM classifier to DTI and volumetric MRI data from 35 amyloid-ß42 negative MCI subjects (MCI-Aß42-), 35 positive MCI subjects (MCI-Aß42+), and 25 HCs retrieved from the European DTI Study on Dementia. The SVM was applied to DTI-derived fractional anisotropy, Mean Diffusivity (MD), and Mode of Anisotropy (MOA) maps. For comparison, the studied classification based on GM and WM volume. The results suggest that DTI data provide better prediction accuracy than GM volume in predementia AD.

Salvatore et al., [15] reviewed the SVM for the early and differential diagnosis of AD-related pathologies by means of MRI data, starting from preliminary steps such as image pre-processing, feature extraction and feature selection, and ending with classification, validation strategies and extraction of MRI-related biomarkers. The main advantages and drawbacks of the different techniques were explored. Results obtained by the reviewed studies were reported in terms of classification performance and biomarker outcomes, in order to shed light on the parameters that accompany normal and pathological aging. Unresolved issues and possible future directions were finally pointed out.

[3] METHODOLOGY

Alzheimer's disease Neuroimaging Initiative (ADNI) dataset is used. SPM12 is used to extract the features and CFS and Chi-Square test is used to select the features and fuzzy classifiers and MLPNN are used to classify the images.

3.1 Dataset

The data used in the preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5- year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials [16].

The ADNI general eligibility criteria are described at www.adni-info.org. Briefly, subjects are between 55-90 years of age, having a study partner able to provide an independent evaluation of functioning. Specific psychoactive medications will be excluded. General inclusion/exclusion criteria are as follows: 1) healthy subjects: Mini-Mental State Examination (MMSE) scores between 24-30, a Clinical Dementia Rating (CDR) of 0, non depressed, non

MCI, and non demented; 2) MCI subjects: MMSE scores between 24-30, a memory complaint, having objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a CDR of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia; and 3) Mild AD: MMSE scores between 20-26, CDR of 0.5 or 1.0, and meets the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable AD.

3.2 Feature Extraction

SPM12 [17] was used for tissue segmentation and normalization (http://www.fil.ion.ucl.ac.uk/spm). The SPM5 implementation includes image normalization, segmentation as well as bias correction. Images were normalized to the MNI template and segmented into Grey Matter (GM), White Matter (WM) and CSF using the unified segmentation model in SPM5. Average probability maps of GM, WM and CSF were created from all 380 GM, WM and CSF probability maps. Let $x_{\mathrm{GM},k}$ be the vector of remaining GM voxels for the *k*-th subject (*k*=1,2,3....*m*). Similarly vectors for WM, $x_{WM,k}$ and CSF, $x_{\mathrm{CSF},k}$, were created. The total number of retained tissue densities values for each patient was of the order 10^4.

3.3 Feature selection

Feature selection leads to minimizing the complexity and computational time. Feature selection is the process of detecting the relevant features and discarding the irrelevant ones. A correct selection of the features can lead to an improvement of the inductive learner, either in terms of learning speed, generalization capacity or simplicity of the induced model. Moreover, there are some other benefits associated with a smaller number of features: a reduced measurement cost and hopefully a better understanding of the domain.

3.3.1 Correlation-based Feature Selection (CFS)

CFS is a simple multivariate filter algorithm that ranks feature subsets according to a correlation-based heuristic evaluation function [18].The bias of the evaluation function is toward subsets that contain features that are highly correlated with the class and uncorrelated with each other. Irrelevant features should be ignored because they will have low correlation with the class. Redundant features should be screened out as they will be highly correlated with one or more of the remaining features. The acceptance of a feature will depend on the extent to which it predicts classes in areas of the instance space not already predicted by other features.CFS is a classical filtered algorithm of attribute selection; in this algorithm, the heuristic evaluation for a single feature corresponding to each category label is used to obtain the final feature subset, and the assessment method of CFS is as follows:

$$M_s = \frac{k\overline{r}_{cf}}{\sqrt{k + k(k-1)\overline{r}_{ff}}} \qquad (1)$$

In (1), M_s is the evaluation for an attribute subset *s* including *k* attribute items, $\overline{r}_{cf}$ is the mean correlation degree between attributes and the category label, and $\overline{r}_{ff}$ is the mean

correlation degree among attributes. And the evaluation of CFS is a method of correlation based on attribute subsets. A bigger $\overline{r}_{cf}$ or smaller $\overline{r}_{ff}$ in acquired subsets by the method produce a higher evaluation value, and in CFS, the correlation degree among attributes is calculated by information gain, and the formula of information gain is shown below. *Y* is the category attribute, *y* is any possible value of *Y*, the entropy of *Y* is shown in (2), and for an attribute *X*, entropy of category attribute *Y* under the condition of *X* is in (3). One has

$$H(\mathrm{Y}) = -\sum_{y \in Y} p(y) \log_2 (p(y))$$
$$H(\mathrm{Y} \mid \mathrm{X}) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y \mid x) \log_2 (p(y \mid x)) \quad (2)$$

The difference of $H(Y)-H(Y|X)$ (i.e., the entropy reduction of attribute *Y*) can reflect the information amount provided by attribute *X* to attribute *Y*, and a bigger difference means a higher correlation degree between *X* and *Y*. Information gain is a symmetrical evaluation method; it tends to select the attributes with more values. Therefore, it is necessary to normalize information gain to [0, 1] for keeping equivalent comparison effect among attributes, and (4), below, shows the calculating formula. One has

$$U_{XY} = 2.0 \times \frac{H(Y) - H(Y \mid X)}{H(Y) + H(\mathrm{X})} \quad (3)$$

As a filtering algorithm, CFS evaluates the correlation between attributes and category label, and the redundancy degree among attributes [19].

3.3.2 Chi Square based Feature Selection

The chi-square test is a statistical test of independence to determine the dependency of two variables. It shares similarities with coefficient of determination. However, chi-square test is only applicable to categorical or nominal data while is only applicable to numeric data. From the definition, of chi-square it can be assume the application of chi-square technique in feature selection. If a target variable (i.e., the class label) are presented and some other features (feature variables) that describes each sample of the data, then calculate chi-square statistics between every feature variable and the target variable and observe the existence of a relationship between the variables and the target. If the target variable is independent of the feature variable, then discard the feature variable. If they are dependent, the feature variable is very important.

3.4 Classifier

3.4.1 Multi-Layered Perceptron Neural Network (MLPNN)

MLPs [20] are feed-forward neural networks trained with the standard back-propagation algorithm. It is shown that a network having a single layer of threshold units could classify a set of points perfectly if they were linearly separable. It is shown that for a set of N data points, a two-layer network of threshold units with $N-1$ units in the hidden layer could exactly separate an arbitrary dichotomy. Since it is very likely that one ends up in a “bad” local minimum, the network should be trained a couple of times (typically at least five times), starting from different initial weights.

A MLPNN structure is used as another network model to predict and evaluate the tram and environmental noise characteristics. MLP includes some main layers in it such as input layer, hidden layer, and output layer. MLP can able to solve a more difficult task in a better way if the number of layer or neuron gets increased. MLP maps an input data samples onto the appropriate number of outputs. An MLP employs a supervised learning technique called as back propagation in order to train the network. Each layer has got linear or nonlinear neurons and each individual neuron sums its weighted inputs and gives an output by means of a nonlinear activation function with a bias.

Connections between the each and every layers are typically formed by connecting each of the nodes from a given layer to all neurons in the next layer. During the training phase each connection's scalar weight is adjusted. The outputs are got from the output nodes. The feature vector says x is input at the input layer and the output represents a discriminator between its class and all of the other classes. In training, the training examples are fed and the predicted outputs are computed. The output is compared with the target output and error measured is propagated back through the network and the weights are adjusted. The training set of size m can be represented as $T_M = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ where $x_i \in R^a$ are the input vectors of dimension a and $Y_i \in R^b$ are the output vectors of dimension b and R represents the set of real numbers. Let fx represent the function with w for the neural network. Supervised learning adjusts the weight such that

$$F_w(X_i) = y; \forall (x_i, y_i) \in T_M \quad (4)$$

After the Neural network is trained with all images feature vectors, and is tested on new samples its output will be correct to a certain extent.

3.4.2 Fuzzy Classifier

Fuzzy classifiers consist of interpretable if-then rules representing the input features and the output class of the form shown in equation (5):

$$R_j : \textit{if } x_{p1} \textit{ is } A_{j1} \textit{ and} \ldots \textit{and } x_{pn} \textit{ is } A_{jn} \textit{ then class } C_j \quad (5)$$

where $A_{j1}, \ldots, A_{jn}$ are antecedent fuzzy sets of the input variable $x_{p1}, \ldots, x_{pn}$ and C_j is one of the output class label. Collections of such rules are used as knowledge base of the fuzzy classifier. With input-output relationship expressed as a collection of fuzzy if-then rules, in which the "if" part uses linguistic variables of each fuzzy set and the "then" part have class labels, qualitative reasoning is performed to infer the results. Here the set of input variable is matched against the "if part" of each if-then rule, and the response of each rule is obtained through fuzzy implication operation [21].

The response of each rule is weighted according to the extent to which each rule fires. The responses of all the fuzzy rules for a particular output class are combined to obtain the confidence with which the input is classified to the corresponding output class. Generally the rules and the membership functions used by the fuzzy logic for solving the classification problem are formed from the experience of the human experts. With an increasing number of variables, the possible number of rules for the system increases exponentially, which makes it difficult for experts to define a complete rule set for good system performance. Also the system performance can be improved by tuning the membership functions.

[5] RESULTS AND DISCUSSION

Table 1 and figure 1 to 3 shows the results of accuracy, sensitivity and specificity respectively. Figure 4 shows the average MLP error.

Table 1 Summary of Results

Classification test	**Chi Square-Fuzzy Classifier**	**Chi Square-MLP NN**	**CFS-Fuzzy Classifier**	**CFS- MLP NN**
Accuracy	82.41	83.39	84.36	86.32
Sensitivity	0.7874	0.8056	0.8177	0.83595
Specificity	0.7874	0.8056	0.8177	0.83595

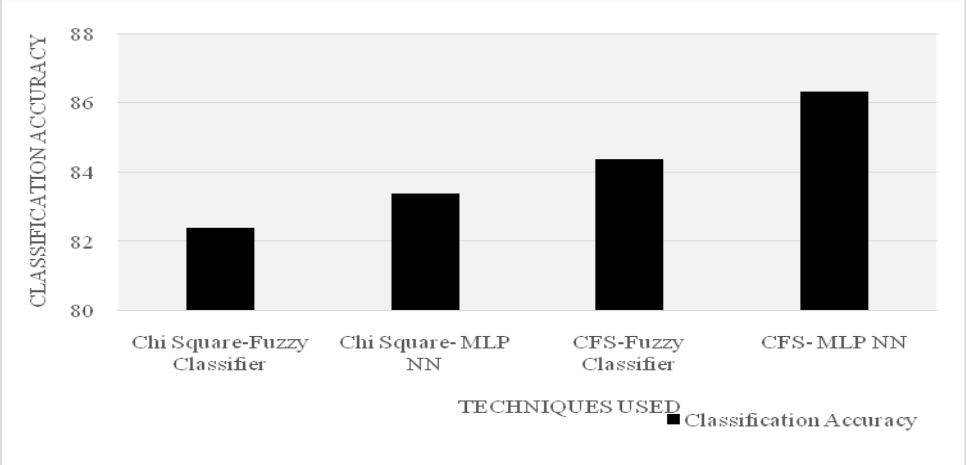

Figure 1 Classification Accuracy for CFS-MLPNN

From table 1 and figure 1 it is observed that the classification accuracy for CFS-MLPNN performs better by 4.63%, by 3.45% and by 2.29% than Chi-Square Fuzzy Classifier, Chi-Square-MLPNN and CFS-Fuzzy Classifier respectively.

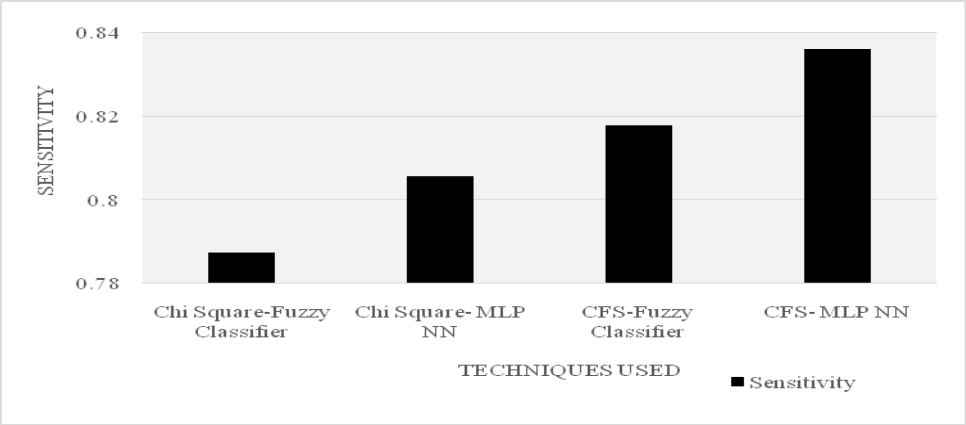

Figure 2 Sensitivity for CFS-MLPNN

From table 1 and figure 2 it is observed that the sensitivity for CFS-MLPNN performs better by 5.98%, by 3.69% and by 2.21% than Chi-Square Fuzzy Classifier, Chi-Square-MLPNN and CFS-Fuzzy Classifier respectively.

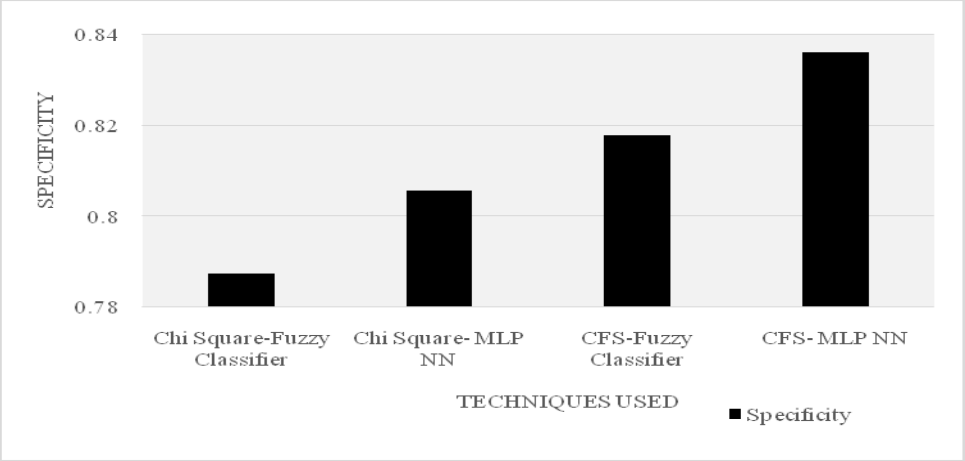

Figure 3 Specificity for CFS-MLPNN

From table 1 and figure 3 it is observed that the specificity for CFS-MLPNN performs better by 5.98%, by 3.69% and by 2.21% than Chi-Square Fuzzy Classifier, Chi-Square-MLPNN and CFS-Fuzzy Classifier respectively.

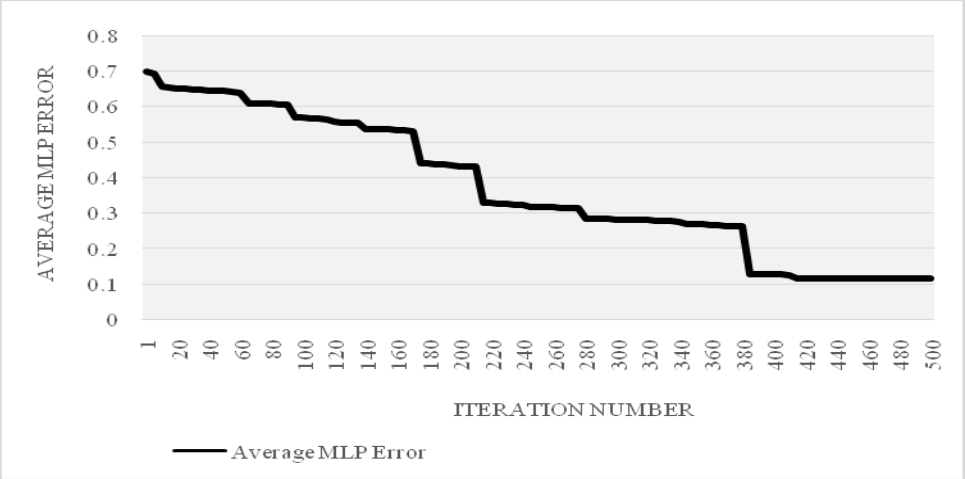

Figure 4 Average MLP Error

Figure 4 shows the occurrence of average MLP error for iteration number 1 to 500. In that convergence occurs at iteration number 425.

[6] CONCLUSION

Diagnostic criteria for Alzheimer's disease are currently based on clinical and psychometric assessment. As the search for effective therapies to arrest or slow the progression of Alzheimer's disease intensifies, there is a need to develop better diagnostic tools. The extracted brain images were selected by ChiSquare and CFS feature selection. Then it is classified by Fuzzy classifier and MLPNN. Results show that the classification accuracy for CFS-MLPNN

performs better by 4.63%, by 3.45% and by 2.29% than Chi-Square Fuzzy Classifier, Chi-Square-MLPNN and CFS-Fuzzy Classifier respectively.

REFERENCES

[1] Mareeswari, S., &Jiji, D. G. W. (2015). A survey: Early detection of alzheimer's disease using different techniques. International Journal on Computational Sciences & Applications (IJCSA) Vol, 5.

[2] Rathi, V. P., & Palani, S. (2012). Brain tumor MRI image classification with feature selection and extraction using linear discriminant analysis. arXiv preprint arXiv:1208.2128.

[3] Rathi, V. G. P., & Palani, D. S. (2012). A novel approach for feature extraction and selection on MRI images for brain tumor classification. In IntConf Comp Sci EngAppl (pp. 225-234).

[4] Zhang, N. (2011). Feature selection based segmentation of multi-source images: application to brain tumor segmentation in multi-sequence MRI (Doctoral dissertation, INSA de Lyon).

[5] Chitra, D., &Nasira, G. M. (2015). WRAPPER BASED FEATURE SELECTION FOR CT IMAGE. ICTACT Journal on Image & Video Processing, 6(1).

[6] Sathya, R., & Abraham, A. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. Int J Adv Res Artificial Intell, 2(2), 34-38.

[7] Bhavsar, H., &Ganatra, A. (2012). A comparative study of training algorithms for supervised machine learning. International Journal of Soft Computing and Engineering (IJSCE), 2(4), 2231-2307.

[8] Roman, G., & Pascual, B. (2012). Contribution of neuroimaging to the diagnosis of Alzheimer's disease and vascular dementia. Archives of medical research, 43(8), 671-676.

[9] Akbarpour, T., Shamsi, M., &Daneshvar, S. (2015, May). Extraction of brain regions affected by Alzheimer disease via fusion of brain multispectral MR images. In Information and Knowledge Technology (IKT), 2015 7th Conference on (pp. 1-6). IEEE.

[10] Zheng, C., Xia, Y., Pan, Y., & Chen, J. (2016). Automated identification of dementia using medical imaging: a survey from a pattern classification perspective. Brain Informatics, 3(1), 17-27.

[11] Chu, C., Hsu, A. L., Chou, K. H., Bandettini, P., Lin, C., & Alzheimer's Disease Neuroimaging Initiative. (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage, 60(1), 59-70.

[12] Liu, F., Wee, C. Y., Chen, H., & Shen, D. (2014). Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification. NeuroImage, 84, 466-475.

[13] Liu, L., Fu, L., Zhang, X., Zhang, J., Zhang, X., Xu, B., ..& Fan, Y. (2015). Combination of dynamic 11 C-PIB PET and structural MRI improves diagnosis of Alzheimer's disease. Psychiatry Research: Neuroimaging, 233(2), 131-140.

[14] Dyrba, M., Barkhof, F., Fellgiebel, A., Filippi, M., Hausner, L., Hauenstein, K., ... &Teipel, S. J. (2015). Predicting Prodromal Alzheimer's Disease in Subjects with Mild Cognitive Impairment Using Machine Learning Classification of Multimodal Multicenter Diffusion-Tensor and Magnetic Resonance Imaging Data. Journal of Neuroimaging, 25(5), 738-747.

[15] Salvatore, C., Battista, P., & Castiglioni, I. (2016). Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines. Current Alzheimer Research, 13(5), 509-533.

[16] Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., & Alzheimer's Disease Neuroimaging Initiative. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*, *55*(3), 856-867.

[17] Vemuri, P., Gunter, J. L., Senjem, M. L., Whitwell, J. L., Kantarci, K., Knopman, D. S., ... & Jack, C. R. (2008). Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *Neuroimage*, *39*(3), 1186-1197.

[18] Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and information systems*, *34*(3), 483-519.

[19] Chen, X. Y., Ma, L. Z., Chu, N., Zhou, M., & Hu, Y. (2013). Classification and progression based on CFS-GA and C5. 0 boost decision tree of TCM zheng in chronic hepatitis B. *Evidence-Based Complementary and Alternative Medicine*, *2013*.

[20] Mendre, M. W., &Raut, R. D. (2013). Neural Network based Decision Support System for the Diagnosis of Thyroid Diseases. *International Journal Of Computer Science And Applications*, *6*(2).

[21] Devaraj, D., & GANESH KUMAR, P. (2010). Mixed genetic algorithm approach for fuzzy classifier design. *International Journal of Computational Intelligence and Applications*, *9*(01), 49-67.