Review of Alzheimer's Disease Classification Techniques

Aow Yong Li Yew[1] and Ghazali Sulong[1, 2]
[1]Faculty of Computing, Universiti Teknologi Malaysia, 81318 Johor Bahru, Johor, Malaysia.
[2]School of Informatics and Applied Mathematics,
Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.
eutherine@live.com.my

***Abstract*—Alzheimer's disease is a brain degeneration illness. It requires earlier detection in order to improve the quality of life of the patient. Many researchers have studied different computed aided approaches in order to help in the diagnosis of the disease. However, there are few well-known techniques that were used in most of the studies. It is challenging in mild cognitive impairment classification too. Therefore, this paper will review the common used methods and the state-of-art methods in Alzheimer's disease classification. This helps researchers to identify the suitable methodologies in different stages of classification.**

***Index Terms*—Alzheimer's Disease; Feature Extraction; Feature Selection; Principal Component Analysis; Support Vector Machine.**

I. INTRODUCTION

Alzheimer's disease (AD) is a degenerative brain illness that affects every aspects of life of a person and eventually cause of death due to complications [1]. It has no cure but the early detection of AD helps in discovering the root of AD and improving patient's quality of life [2]. Hence, many researches are conducted especially in the atrophy of the brain because it is proved to be one of the neurodegeneration biomarkers in AD diagnosis through the brain imaging [3], [4].

There are two important stages in AD detection, which are feature selection or feature extraction and classification. Feature selection helps in reducing the dimension before proceeds to classification stage because most of the existing classifiers do not work well if there are too many input variables. In this process, the important features will be selected and improper features will be omitted [5]. On the other hand, feature extraction is the technique to use in extracting the raw data in voxel of the selected voxel site [6]. After going through either feature selection or feature extraction, or both, classification will be applied.

There are many studies focus on the classification of healthy control (HC), mild cognitive impairment (MCI) and AD [7]–[9]. Recently, the focus of the researches and organization pay more attention to MCI [10], [11]. MCI is a state that an individual has mild cognitive problem. An individual in MCI state has the chance to revert to normal cognition or remain in the stage. But in the meta-analysis of 41 studies, an average of 38 percent of the individuals suffered MCI developed dementia [1]. It is also affirmed to be a process before AD in later study [12], [13]. Therefore, the diagnosis of stable MCI (sMCI) and progressive MCI (pMCI) is getting more significant. SMCI refers to the individual with MCI who remains in the stage while pMCI refers to the individual with MCI who converts to AD after few years.

Every step of AD classification contributes to the classification result. The selection of the features will influence the classifier to identify the stage of the patient. The power of classifier in differentiating different stages of AD also will give impact in the result. Therefore, the review of current techniques is important in order to find the suitable techniques in different stages of AD classification. Ten studies are reviewed in this paper, the result of each study are showed in appendix while the techniques are reviewed in section II and section III.

II. FEATURE SELECTION AND FEATURE EXTRACTION TECHNIQUES

Principal component analysis (PCA) and linear discriminate analysis (LDA) are the commonly used feature reduction techniques [14], [15]. PCA is claimed to be a powerful tool in extracting the important features from a dataset [8], [15], [16]. Even, the traditional PCA is found that it outperforms than more sophisticated dimensionality reduction in real-world cases [17]. Nevertheless, LDA is better in extracting discriminative information compared to PCA. This is because PCA works well when the data has general variance but not to estimate the relationship between the data. But the lower inter-class discrimination of PCA prevents overfitting of the input data. By feeding with the prior knowledge of the disease, LDA becomes the parameter in the pattern categories process to construct feature space. Hence, this will be the crucial point for LDA to perform well in extracting accurate features. Besides that, both PCA and LDA require high computational cost due to the transformation of image matrices into vectors. They also suffer overfitting when the numbers of features greater than the numbers of subjects. Despite of this, both techniques are still widely used in current researches especially PCA in AD diagnosis [9], [18].

Dessouky et al. [15] proposed a new feature reduction technique to deal with the high computational cost of PCA and LDA. The technique removed the pixels have same intensity values in the same place as the first step. After this, the images were divided to two classes, which were demented subjects and non-demented subjects. The significant value of each feature was calculated by dividing the subtraction result of the means of each feature in two classes with the multiplication result of standard deviations of each feature in two classes. In later study, Dessouky et al. [19] combined his proposed technique in 2013 with different existing feature extraction approaches which were discrete cosine transform (DCT), discrete sine transform (DST), discrete wavelet transform (DWT) and mel-frequency cepstral coefficients (MFCC). The conventional feature extraction approaches were performed after the first step of his proposed technique while the second step of the proposed technique was applied as the third step in this study. The combination of MFCC with proposed feature extraction technique outperforms the other three techniques in terms of achieving 100% accuracy by using only 25 features for HC and AD classification. But, the computation time of overall classification by using MFCC is slightly higher than DST and DWT. On the other hand, the overall processing time is highly influenced by the proposed first step of reduction. The computation time of the proposed first step is higher than other steps because it dealt with pixel-by-pixel in each image.

In contrast to the view of Van Der Maaten et al. [17], the later researches showed PCA gives poor result when comparing with other techniques which was not mentioned in the study. Khedher et al. [8] compared partial least squares (PLS) with PCA. PLS takes into account the relationship between observed variables and class labels. The observed variables are defined through the help of latent variables. The orthogonal weight vectors were created through searching the maximum covariance of different data sets instead of the variance of the samples only as PCA. There were only 8 components used in this study for both techniques. PLS gave better result than PCA with smaller computation time and higher accuracy in classification.

Independent component analysis (ICA) works similar to PCA but it was claimed ICA is better than PCA due to ICA finds the independent components while PCA finds the uncorrelated components [20]. Uncorrelated components will influence each other in certain extent while independent components have no influence to each other. Wilette et al. [21] implemented ICA by using Infomax algorithm. This method decomposes the complex data into a new feature space which consists of 30 discrete data. The discrete data is called ICs. Each IC contributes in forming the complex data, which is the whole brain GM in this study. Therefore, dimension reduction can be done through the elimination of less contribution IC towards different subjects in classification process. The significance of IC is computed through multivariate analysis of covariance (MANCOVA) in SPSS. The advantage of ICA is it retains the variance of source image even it is decomposed to new feature space. In the study, the result showed that ICA works well in differentiating HC, MCI and AD. It is good enough to separate sMCI and pMCI too compared to previous research.

Another dimensional reduction approach is called local linear embedding (LLE) [22]. LLE cares about the nearest neighbors of each point when the data is reconstructed from high dimension to lower dimension. The number of neighbors is decided by the researchers and the nearest neighbors of each point are found through Euclidean distance. Then the contribution of each neighbors are calculated to reconstruct the data point. As the result showed in this study, LLE improves the performance of classification especially in sMCI and pMCI without limit to any classifiers. It even helps in learning the nonlinear feature structures of AD while embedding the data into linear coordinates.

Demirhan et al. [23] applied reliefF algorithm in the feature selection. This algorithm was first proposed by Kira & Rendell [24] in year of 1992. The algorithm is fed with a training set which has separated the AD classes. Then, the selected features of testing set are compared with the labeled features to determine the relevancy of the feature towards the class with a predefined threshold. This method can deal with feature interaction which is useful in AD classification because there will be similar features for MCI with other classes. Nevertheless, this algorithm only functions well when more relevant features compared to irrelevant features. It is crucial to define the threshold because it is the key to affect the feature selection. The classification result was compared according different numbers of features. It showed that larger number of features does not give higher accuracy even it have different effects towards different classes.

A texture feature extraction called advanced local binary pattern sign magnitude from three orthogonal planes (ALBPSM-TOP) was proposed by Sarwinda & Bustamam [25] before the implementation of factor analysis (FA) and PCA. The concept of Local Binary Pattern (LBP) was calculating the value assigned to the centre of the circle by multiplying the binary value of the neighbor pixels with 2p where P is the number of neighbors. The difference of ALBPSM-TOP from LBP is it considered the magnitude of the image in obtaining the LBP value. ALBPSM-TOP also dealt with 3D data by combining the LBP values of axial, sagittal and coronal planes into a single histogram. The combination of different planes increased the accuracy in classification because more significant features can be extracted from the image. As a result, ALBPSM-TOP succeeded to classify all classes of AD perfectly by using FA in feature selection in this study even though it only achieved 84.75% accuracy by using PCA. This shows that there is no standalone methods can solve an issue. Besides this, the radius of the circle, r and P are not mentioned in the paper. Both parameters are manually set by the user. It was proved that different value of the parameters will influence the classification performance [26]. So, it is crucial to find the suitable values for both parameters according to different dataset.

A combination of t-test feature ranking and fisher criterion techniques was also developed to compare with PCA by Beheshti & Demerel [27]. T-test was used to find the features relevant to classify AD from HC by calculating the significance difference between means of two classes. The optimal number of features for classification was obtained through fisher criterion approach. The author claimed that the proposed method helps in increasing the discriminative value while reducing the dimension. The study also showed the classification result without dimensionality reduction technique. It showed that there is no much difference by using raw data in classification when using PCA in feature extraction while there is 8.95% increase in the AUC after applying the proposed method.

In conclusion, the existing feature extraction and feature selection techniques improve AD classification by extracting the important features and determining the suitable number of features. Cuingnet et al. [28] also stated that sMCI and pMCI classification will have higher accuracy when only few brain regions are selected. This might be the reason that ICA achieves higher accuracy in the classification of sMCI and pMCI compared to LLE and PCA. SMCI and pMCI consist of similar features which will cause the classifier to have wrong interpretation towards both classes. However, the method which only selects certain brain regions may cause to high computation time, even up to hours per subject just to process the feature selection algorithm. So, it is crucial to achieve high accuracy in classification while maintain the computation time in acceptable range in order to fit the need in real-world. The methods that give good result, even 100 percent in HC, MCI and AD classification shall be explored in sMCI and pMCI classification to ensure the power in selecting the distinct features between different classes. At last, the summary of the strengths and weaknesses of the methods are given in Table 1.

Table 1
Strengths and Weaknesses of the Feature Extraction and Feature Selection Techniques

Feature Selection or Feature Extraction Techniques	Strength	Weakness
PCA	• Perform well when data has general variance	• Overfitting might occur when there are many features
PLS	• The combination of regression, dimension reduction and classification tasks • The variables interrelate to each other is identified	• Prediction oriented which will cause the difficulty in variables interpretation
ICA	• Consider the relationship between data • Find distinct feature for different classes	• Require prior knowledge towards data • Overfitting might occur when there are many features
LLE	• Preserve local properties by taking into consideration of nearest neighbors	• Global features might be neglected
ReliefF algorithm	• Deal with feature interaction between different data	• Threshold is predefined to determine the relevancy of the feature
ALBSM-TOP	• Deal with 3D data • Considering neighbor pixels which gives more precise result in identifying significant features	• Parameters are manually set by user
T-test feature ranking + fisher criterion	• Fast computation • Optimum number of features are defined with automatic approach	• T-test will miss the features that only useful when working with others

III. CLASSIFICATION TECHNIQUES

Most of the studies were based on supervised machine learning approach in AD classification [11]. Support vector machine (SVM) is one of the widely used methods in AD diagnosis. It is a supervised learning model which requires training set to find the hyperplane to separate different group of subjects and implement it to the testing set [29]. The illustration of SVM concept is shown in Figure 1. Hyperplane is the line that maximizes the distance between the classes in higher dimensional space. Hence, this session will compare different SVM approaches. The other approaches also will be discussed.

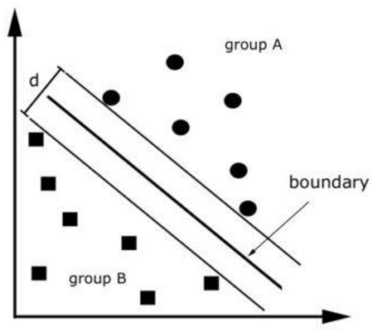

Figure 1: Illustration of support vector machine's concept. The boundary, which is hyperplane is the maximum distance(d) between group A and group B [29].

Linear SVM is sensitive to the dataset. The classification result will be different by using different dataset [29]. The selection of training set and testing set also influence the classification result. But, linear SVM works well no matter involve or not involve the feature selection step [28]. Linear SVM is often compared with radial basis function (RBF) kernel SVM [8], [27]. RBF SVM involves the tuning of two parameters, which are C in regularization and γ for controling the width of the kernel using training set. Khedher et al. [8] concluded that linear SVM gave better result than RBF SVM. The result was supported by the study of Beheshti & Demirel [27] in 2016 after applying data fusion among atrophy cluster. Data fusion by using majority voting increases the classification result. Data fusion combines two or more distinct data resources into single one to provide more accurate description towards the data. In the study, mojority voting integrates the initial classification result of different volume of interests (VOIs). However, RBF SVM gave better result without data fusion. This can prove that the result of RBF SVM is much depends on the features used, but linear SVM gives a good result in general.

SVM is designed for two-class classification. In order to deal with multiclass classification, kernel support vector machine decision tree (kSVM-DT) was developed [9]. RBF kernel was chosen in this study and the parameter used in kSVM-DT was obtained through canonical quadratic programming (QP) and Particle Swarm Optimization (PSO). The concept of PSO is based on the bird flocking where each particle is evaluated with the previous best position. The best position is updated along the iteration in order to find the the best parameters. By obtaining the parameters used in SVM, overfitting and underfitting are prevented and it reduces the computation burden. The performance of PSO was assessed through comparing the classification result by using random selection parameters. It was proved that PSO indeed helped in increasing the accuracy result.

Similar to [9], Demirhan et al. [23] used Gaussian RBF kernal with sequential minimal optimization (SMO) to train SVM. SMO was first proposed in 1998 to solve the large QP optimization issue of the SVM training algorithm [30]. SMO breaks down large QP problem into smaller manageable QP problems. It was claimed that the computation time of SMO is more than 1000 times faster than the standard SVM algorithm. Besides that, SMO can handle very large training sets compared to standard SVM because the memory requirements linear with the size of training set. On the other hand, the tuning parameters of RBF kernel were obtained from grid search with 10 fold cross-validation in 10 times. All the data points of training and test sets were undergone mean-centering to prevent the objective function of SVM was dominated by the features have larger variance. Compared to other studies, the result did not show improvement even in AD versus HC classification. This may because of the combination of SMO and RBF kernel do not suitable for the dataset or the features selected do not distinct enough to classify different classes of AD.

Apart from SVM, there are different techniques can be used in classification, which includes K-mean clustering, fuzzy clustering method (FCM), orthogonal partial least squares to latent structures (OPLS), decision trees (Trees), artificial neural networks (ANN), elastic net (EN) regularized regression and discriminant classification analysis. Farzan et al. [18] compared 4 classification techniques which were K-mean clustering, FCM, linear SVM and RBF-SVM. K-mean clustering and FCM are unsupervised classification method. K-mean grouped the observations to the cluster which has closest mean to them. The mean of the clusters will be recalculated once each observation is grouped to the cluster. The updating process will continue until the mean of cluster remains the same as previous. Then, that cluster is classified to the suitable class. On the contrary, FCM allows each data point belongs to every cluster. Each cluster has the membership value. If the data point has higher membership value, which means the data point is closer to the center of the cluster and vice versa. K-mean and FCM worked well in the specificity but not in sensitivity. This may be because of the initial starting points of the methods were chosen incorrectly since the features may overlap to each class. K-mean and FCM works similarly but FCM needs higher computation time because it is an interative process [31].

OPLS, Trees, ANN and SVM were compared in [32] to classify HC, AD, sMCI and pMCI. All of them gave a good result in AD vs. HC and pMCI vs. AD but not HC vs. sMCI. OPLS is the variant of PLS. The difference between both techniques is OPLS creates a model for class separation. The first component found will be the predictive component and this makes the data interpretation easier [33]. The significant value of the predictive model in group separation will be computed along the process. In this study, the result is matched with the study conducted by Westman et al. [33]. Westman et al. believe that MCI subjects are closer to HC subjects. So, it is hard to differentiate MCI and HC. This explains the reason of bad classification result of HC and sMCI even using different techniques. Trees involves a set of choices which the choices are made based on the attributes contribute to the classes. Greedy algorithm is used to determine the most discriminative attributes for each step. The method is suitable to the classes where they have less interaction among the classification attributes [34]. As a result, this algorithm gives the worst result in sMCI and HC classification among the techniques used in the study. ANN was developed based on the human's neural structure which consists of weighted inputs, hidden layers and outputs. Weka machine learning software was used to implement ANN. It gives similar result to other techniques.

On the other hand, EN regression is a regression analysis method which used in [22] as one of the classifier. EN solves the problems of lasso and ridge regression [35]. Lasso tends to ignore the much-correlated independent variables and only pick one of them while ridge regression shrinks the weighted of correlated independent variables toward each other. EN averages the independent variables into the model which is suitable to be applied when there are many correlated independent variables. It gives the best result among EN, SVM and LDA In different groups' classification. Besides this, EN has less influence compared to other technique with or without using feature reduction techniques. But, the result also showed that three of the techniques are facing the difficulties when the classification involves MCI state. By seeing the classification result of different studies, we can know that SVM is less dependency on the features but itself can classify different groups of AD well. Nevertheless, sMCI and pMCI classification are the main issues left in the AD classification. New classifier will need to be developed in order to distinguish sMCI and pMCI from HC and AD even between them. The summary of the strengths and weaknesses for the classification techniques are shown in Table 2.

Table 2
Strengths and Weaknesses of the Classification Techniques

Classification Techniques	Strength	Weakness
SVM	• Work well without involve feature selection • Give a clear cut for both classes	• Sensitive to dataset
RBF SVM	• Higher distinguishing power of classes when the features are well separated	• Classification result depends much on the features
K-means	• Manage to identify the features in each class	• Initial point determines overall result • Important features might be excluded from some classes when it was assigned to other cluster.
FCM	• Suitable to the classes with close features	• Initial point determines overall result • High computational time
OPLS	• Easier in data interpretation	• First component found determines the overall result
Trees	• High predictive power by giving distinct features	• Suitable to classes have less interaction only.
ANN	• Good in handling large training set with many hidden layers	• Black box learning which disables in the interpretation of relationship between input and output
EN Regression	• Solve the problems of lasso and ridge regression	• Less distinguishing power in pMCI and sMCI classification due to the closeness of the features between both stages

IV. Conclusion

There are pros and cons for each technique. Therefore, the combination of different existing and proposed techniques were used in different studies. The combination of techniques may increase the accuracy of AD classification but it may also raise up another issue such as high computational cost. There is no one solution that can fix all the issues for all studies. Researchers have to tackle the problem according to their significance of studies. In feature selection and feature extraction technique, it is important to extract the significant features which can distinguish different stages of AD. The relationship between the classes has to be studied since it is the process to develop AD. A feature extraction that is suitable for AD classification can be developed with prior knowledge of the disease. After feature extraction or feature selection, the classification approach that can differentiate different classes of AD has to be implemented. The existing techniques proved that it is high accuracy in AD versus HC classification. However, there is room of improvement in MCI classification especially for sMCI and pMCI classification [36]. Last but not least, an automatic classification method is desire but it is more important in achieving high sensitivity and specificity.

REFERENCES

[1] Alzheimer's Association, "2016 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, vol. 12, no. 4, pp. 459–509, Apr. 2016.

[2] Alzheimer's Association, "What is Alzheimer's," 2017. [Online]. Available: http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp. [Accessed: 03-Jan-2017].

[3] G. B. Frisoni, N.C. Fox, J. P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nat. Rev. Neurol*, vol. 6, no. 2, pp. 67–77, 2010.

[4] B. Duthey, "Background Paper 6.11 Alzheimer Disease and other Dementias," 2013.

[5] C. R. Munteanu, C. Fernandez-Lozano, V. Mato Abad, S. Pita Fernández, J. Álvarez-Linera, J. A. Hernández-Tamames, and A. Pazos, "Classification of mild cognitive impairment and Alzheimer's Disease with machine-learning techniques using 1H Magnetic Resonance Spectroscopy data," *Expert Syst. Appl.*, vol. 42, no. 15–16, pp. 6205–6214, 2015.

[6] M. Termenon, M. Graña, A. Besga, J. Echeveste, and A. Gonzalez-Pinto, "Lattice independent component analysis feature selection on diffusion weighted imaging for Alzheimer's disease classification," *Neurocomputing*, vol. 114, pp. 132–141, Aug. 2013.

[7] Y. Zhang, S. Wang, P. Phillips, Z. Dong, and G. Ji, "Detection of Alzheimer ' s disease and mild cognitive impairment based on structural volumetric MR images using 3D-DWT and WTA-KSVM trained by PSOTVAC," *Biomed. Signal Process. Control*, vol. 21, pp. 58–73, 2015.

[8] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, and F. Segovia, "Early diagnosis of Alzheimer â€™ s disease based on partial least squares , principal component analysis and support vector machine using segmented MRI images," *Neurocomputing*, vol. 151, pp. 139–150, 2015.

[9] Y. Zhang, S. Wang, and Z. Dong, "Classification of Alzheimer Disease Based on Structural Magnetic Resonance Imaging by Kernel Support Vector Machine Decision Tree," *Prog. Electromagn. Res.*, vol. 144, pp. 171–184, 2014.

[10] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, and I. Castiglioni, "Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer ' s disease : a machine learning approach," *Front. Neurosci.*, vol. 9, no. September, pp. 1–13, 2015.

[11] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka, "Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects.," *Neuroimage*, vol. 104, pp. 398–412, Jan. 2015.

[12] L. G. Brooks and D. A. Loewenstein, "Assessing the progression of mild cognitive impairment to Alzheimer's disease: current trends and future directions.," *Alzheimers. Res. Ther.*, vol. 2, no. 5, p. 28, 2010.

[13] N. Philippi, V. Noblet, E. Duron, B. Cretin, C. Boully, I. Wisniewski, M. L. Seux, C. Martin-Hunyadi, E. Chaussade, C. Demuynck, S. Kremer, S. Lehéricy, D. Gounot, J. P. Armspach, O. Hanon, and F. Blanc, "Exploring anterograde memory: a volumetric MRI study in patients with mild cognitive impairment," *Alzheimers. Res. Ther.*, vol. 8, no. 1, p. 26, 2016.

[14] Y. Wang, Y. Fan, P. Bhatt, and C. Davatzikos, "High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables.," *Neuroimage*, vol. 50, no. 4, pp. 1519–35, May 2010.

[15] M. M. Dessouky, M. A. Elrashidy, T. E. Taha, and H. M. Abdelkader, "Selecting and extracting effective features for automated diagnosis of Alzheimer's disease," *Int. J. Comput. Appl.*, vol. 81, no. 4, pp. 17–28, 2013.

[16] H. Park, J. ju Yang, J. Seo, and J. min Lee, "Dimensionality reduced cortical features and their use in predicting longitudinal changes in Alzheimer's disease," *Neurosci. Lett.*, vol. 550, pp. 17–22, 2013.

[17] L. J. P. Van Der Maaten, E. O. Postma, and H. J. Van Den Herik, "Dimensionality Reduction: A Comparative Review," *J. Mach. Learn. Res.*, vol. 10, pp. 1–41, 2009.

[18] A. Farzan, S. Mashohor, A. R. Ramli, and R. Mahmud, "Boosting diagnosis accuracy of Alzheimer's disease using High dimensional recognition of longitudinal brain atrophy patterns," *Behav. Brain Res.*, vol. 290, pp. 124–130, Apr. 2015.

[19] M. M. Dessouky, M. A. Elrashidy, T. E. Taha, and H. M. Abdelkader, "Computer-aided diagnosis system for Alzheimer's disease using different discrete transform techniques," *Am J Alzheimers Dis Other Demen*, vol. 31, no. 3, pp. 282–293, 2016.

[20] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. 2001.

[21] A. A. Willette, V. D. Calhoun, J. M. Egan, and D. Kapogiannis, "Prognostic classification of mild cognitive impairment and Alzheimer's disease: MRI independent component analysis.," *Psychiatry Res.*, vol. 224, no. 2, pp. 81–8, Nov. 2014.

[22] X. Liu, D. Tosun, M. W. Weiner, and N. Schuff, "Locally linear embedding (LLE) for MRI based Alzheimer's disease classification," *Neuroimage*, vol. 83, pp. 148–157, Dec. 2013.

[23] A. Demirhan, T. M. Nir, A. Zavaliangos-petropulu, C. R. Jack, M. W. Weiner, M. A. Bernstein, P. M. Thompson, N. Jahanshad, C. Technology, and F. Technology, "Feature selection improves the accuracy of classifying alzheimer disease using diffusion tensor images," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 126–130.

[24] K. Kira and L. Rendell, "The feature selection problem: Traditional methods and a new algorithm," *Aaai*, pp. 129–134, 1992.

[25] D. Sarwinda and A. Bustamam, "Detection of Alzheimer's disease using advanced local binary pattern from hippocampus and whole brain of MR images," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 5051–5056.

[26] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artif. Intell. Med.*, vol. 49, no. 2, pp. 117–125, 2010.

[27] I. Beheshti and H. Demirel, "Feature-ranking-based Alzheimer's disease classification from structural MRI," *Magn. Reson. Imaging*, vol. 34, no. 3, pp. 252–263, 2016.

[28] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M. O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," *Neuroimage*, vol. 56, no. 2, pp. 766–781, 2011.

[29] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, J. Ashburner, and R. S. J. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689, 2008.

[30] J. C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines," *Adv. kernel methods*, pp. 185–208, 1998.

[31] S. Ghosh and S. K. S. Dubey, "Comparative analysis of k-means and fuzzy c-means algorithms," *Ijacsa*, vol. 4, no. 4, pp. 35–38, 2013.

[32] C. Aguilar, E. Westman, J.-S. Muehlboeck, P. Mecocci, B. Vellas, M. Tsolaki, I. Kloszewska, H. Soininen, S. Lovestone, C. Spenger, A. Simmons, and L.-O. Wahlund, "Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment.," *Psychiatry Res.*, vol. 212, no. 2, pp. 89–98, May 2013.

[33] E. Westman, A. Simmons, Y. Zhang, J.-S. Muehlboeck, C. Tunnard, Y. Liu, L. Collins, A. Evans, P. Mecocci, B. Vellas, M. Tsolaki, I. Kłoszewska, H. Soininen, S. Lovestone, C. Spenger, and L.-O. Wahlund, "Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls.," *Neuroimage*, vol. 54, no. 2, pp. 1178–87, Jan. 2011.

[34] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artif. Intell.*, vol. 97, no. 1–2, pp. 245–271, 1997.

[35] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *J. Stat. Softw.*, vol. 33, no. 1, pp. 1–11, 2010.

[36] Y. Cho, J.-K. Seong, Y. Jeong, and S. Y. Shin, "Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data.," *Neuroimage*, vol. 59, no. 3, pp. 2217–30, Feb. 2012.

APPENDIX

Table 3
Comparison of Ten Studies in Alzheimer's Disease Classification

Author, Year	Techniques: Feature Selection and Feature Reduction [Feature]	Techniques: Classification	Dataset	Performances (%)	HC VS. AD	HC VS. MCI	MCI VS. AD	sMCI VS. pMCI	HC VS. sMCI	pMCI VS. AD
Aguilar et al., 2013 [32]	- [Cortical thickness and VOI of whole brain]	OPLS, Trees, ANN, SVM	- AddNeuroMed project - Total 345 subjects, 116 AD, 119 MCI and 110 HC.	*OPLS*						
				ACC	84.5	-	-	-	68.4	81
				SEN	79.3	-	-	-	-	-
				SPE	90	-	-	-	-	-
				Trees						
				ACC	81.9	-	-	-	49	85.7
				SEN	78.5	-	-	-	-	-
				SPE	85.5	-	-	-	-	-
				ANN						
				ACC	84.9	-	-	-	59.2	81
				SEN	80.2	-	-	-	-	-
				SPE	90	-	-	-	-	-
				SVM						
				ACC	83.6	-	-	-	56.1	85.7
				SEN	81	-	-	-	-	-
				SPE	86.4	-	-	-	-	-
Liu et al., 2013 [22]	LLE [volume and thickness of cortical]	EN regression, SVM or LDA	- ADNI database - Total 413 subjects, 138 HC, 93 sMCI, 97 pMCI and 86 AD - T1-weighted MR images	*EN*						
				ACC	90	-	-	68	64	56
				SEN	86	-	-	80	65	51
				SPE	93	-	-	56	63	61
				SVM						
				ACC	90	-	-	66	64	57
				SEN	87	-	-	75	58	48
				SPE	92	-	-	56	67	65
				LDA						
				ACC	89	-	-	68	51	57
				SEN	86	-	-	82	48	51
				SPE	91	-	-	53	49	63
Willette et al., 2014 [21]	ICA [GM voxels, age, sex, education, MMSE, ADAS-Cog]	Discriminant classification of SPSS	- 2 years diagnosis of ADNI database - 93 HC, 162 MCI and 65 AD - T1-weighted MR images	ACC	94.3	83.3	81.4	80	-	-
				SEN	94.9	76.7	86.1	78.3	-	-
				SPE	94	89.1	73	81.5	-	-
Zhang, Wang, & Dong, 2014 [9]	PCA [GM voxels, gender, age, education, socioeconomic status(SES), MMSE, eTIV, nWBV, ASF]	PSO is used to train the parameter of kSVM-DT	- OASIS database - 97 HC, 57 MCI and 14 AD. - Total 178 of T1-weighted MR images	ACC	96	85	88	-	-	-
				SEN	98	87	96	-	-	-
				SPE	89	80	73	-	-	-
Demirhan et al., 2015 [23]	ReliefF algorithm [WM voxels]	SVM which is trained by RBF-kernel and SMO	- ADNI database - 70 HC, 114 MCI and 37 AD diffusion weighted images (DWI) - T2-weightetd images	ACC	87.8	78.48	85.34	-	-	-
				SEN	79.5	86.2	77	-	-	-
				SPE	92.9	65.7	89.4	-	-	-
Farzan et al., 2015 [18]	PCA [whole brain volume atrophy rates]	Comparison among K-means, FCM, linear SVM and RBF-SVM	- ADNI database - 30 AD and 30 HC, which undergone 2 years follow up	*K-means*						
				ACC	83.3	-	-	-	-	-
				SEN	73.3	-	-	-	-	-
				SPE	93.3	-	-	-	-	-
				FCM						
				ACC	83.3	-	-	-	-	-
				SEN	73.3	-	-	-	-	-
				SPE	93.3	-	-	-	-	-
				Linear SVM						
				ACC	90	-	-	-	-	-
				SEN	86.7	-	-	-	-	-
				SPE	93.3	-	-	-	-	-
				RBF-SVM						
				ACC	91.7	-	-	-	-	-
				SEN	90	-	-	-	-	-
				SPE	93.3	-	-	-	-	-
Khedher et al., 2015 [8]	PLS [GM voxels + WM voxels]	Linear kernel SVM	- ADNI database - 229HC, 401 MCI (312 sMCI and 86 pMCI) and 188 AD - Total 1075 of T1-weighted MR images	ACC	88.49	81.89	85.41	-	-	-
				SEN	91.27	82.16	87.03	-	-	-
				SPE	85.11	81.62	83.78	-	-	-
Beheshti & Demirel, 2016 [27]	T-test feature ranking + fisher criterion [GM - VOI of right amygdala, left lateral globuspallidus right inferior parietal lobule region, right anterior cingulated region]	Comparison among linear SVM and RBF SVM. The result of classifiers are integrated through majority voting	- ADNI database - Total 136 subjects, 68 AD and 68 HC - Proportion of training set and test set are 9:1	Linear SVM						
				ACC	96.32	-	-	-	-	-
				SEN	94.11	-	-	-	-	-
				SPE	98.52	-	-	-	-	-
				RBF SVM						
				ACC	95.59					
				SEN	94.11					
				SPE	97.05					

Author, Year	Techniques: Feature Selection and Feature Reduction [Feature]	Techniques: Classification	Dataset	(%)	Performances: HC VS. AD	HC VS. MCI	MCI VS. AD	sMCI VS. pMCI	HC VS. sMCI	pMCI VS. AD
Dessouky et al., 2016 [19]	First, removing the pixels has same intensity values. Then, applying MFCC and at last calculating the maximum differences between the means of two classes and feed it to SVM [GM voxels– 25 features left after feature extraction]	Linear SVM	- OASIS database - 49 very mild to mild AD and 71 HC - T1-weighted MRI images	ACC SEN SPE	100 100 100	- - -	- - -	- - -	- - -	- - -
Sarwinda & Bustamam, 2016 [25]	ALBPSM-TOP + FA [Hippocampus]	Linear and RBF SVM	- ADNI database - 94 AD, 80 MCI and 96 HC - Total 270 T1-weighted 3D MRI images	ACC	Multiclass classification: 100			-	-	-