Data-Driven Approach based on Feature Selection Technique for Early Diagnosis of Alzheimer's Disease

Surendrabikram Thapa[1], Priyanka Singh[2], Deepak Kumar Jain[3], Neha Bharill[4], Akshansh Gupta[5], Mukesh Prasad[2]

[1]Department of Computer Science and Engineering, Delhi Technological University, New Delhi, India
[2]School of Computer Science, FEIT, University of Technology Sydney, Sydney, Australia
[3]Department of Computer Science, Chongqing University of Posts and Telecommunications, China
[4]Department of Computer Science and Engineering, Mahindra Ecole Centrale, Hyderabad, India
[5]Central Electronics Engineering Research Institute, Pilani, Rajasthan, India

Abstract: Alzheimer's disease (AD) is a neurodegenerative disorder resulting in memory loss and cognitive decline caused due to the death of brain cells. It is the most common form of dementia and accounts for 60-80% of all dementia cases. There is no single test for diagnosis of AD, the doctors rely on medical history, neuropsychological assessments, computed tomography (CT) or magnetic resonance imaging (MRI) scan of the brain, etc. to confirm a diagnosis. In terms of the treatment, currently, there is neither a cure nor any way to slow the progression of AD. However, for people with mild or moderate stages of this disease, there are some medications available to temporarily reduce symptoms and help to improve quality of life. Hence, early diagnosis of AD is extremely crucial for overall better management of the disease. The researches have shown some relation between neuropsychological scores and atrophies of the brain. This can be leveraged for the early diagnosis of AD. This paper makes use of feature selection techniques to extract the most important features in the diagnosis of AD. This paper demonstrates the need to combine neuropsychological scores like mini-mental state examination (MMSE) with MRI features to provide better decisional space for early diagnosis of AD. Through the experiments, including MMSE along with other features are found to improve the classification of AD, significantly.

Keywords: Alzheimer's disease, neuropsychological scores, feature selection, mini-mental state examination

I. INTRODUCTION

Alzheimer's disease (AD) was first identified and discussed by a German physicist and neuro-pathologist Alois Alzheimer [1]. According to the World Alzheimer Report [2], there were around 50 million people in the world living with dementia in 2018 and about two-third of them have AD. The number of patients suffering from this disease is likely to rise to around 152 million in 2050 and the cost of this disease is forecasted to be 2 trillion US dollars in 2030 [2]. Currently, the treatment of AD predominantly involves the use of this disease-modifying or delaying drugs rather than drugs that reverse or permanently stop its progression. Thus, there is a necessity for the prediction of AD in the early stage so that it becomes possible to better manage this disease by delaying the occurrence of cognitive decline.

AD is characterized by loss of neurons and synapses in the cerebral cortex and certain subcortical regions which results in gross atrophy of the affected regions, including degeneration in the temporal lobe and parietal lobe, and parts of the frontal cortex and cingulate gyrus [3]. The previous studies have shown that various impairments in AD are correlated with atrophy in various regions like the hippocampus [4], temporal lobe [5], and amygdala [6]. Moreover, the determination of the key atrophied regions across the entire brain could be used as parameters for the delineation of AD patients from cognitively normal (CN) subjects [7].

Earlier, the clinical measures like clinical dementia rating (CDR) were used in the diagnosis of AD [8]. With the rapid advancement in computation and development of imaging techniques, machine learning techniques in early diagnosis of AD have been widely used using neuroimaging images like MRI and positron emission tomography (PET) [9-11]. Among numerous machine learning techniques that have been explored, the widely used technique for AD prediction is Support Vector Machines (SVM) [12]. The various studies have been carried out with single modality or multiple modalities. Using a single modality makes the job easier with less cost and time but is accompanied by more erroneous predictions. The experiments that make use of imaging modalities without neuropsychological scores attain lesser accuracy than the experiments which used imaging modalities with neuropsychological scores. Hinrichs et al. [11] were able to improve the accuracy of classification between AD and Control Normal (CN) by around 5% after incorporating the Mini-Mental State Examination (MMSE) in their experiment. Ewers et al. [13] also combined the primary MRI and Cerebrospinal fluid (CSF) biomarkers with a neuropsychological test for making the classification of AD and CN better. They also used similar combinations to predict the conversion from Mild Cognitive Impairment (MCI) to AD. Similarly, Zhou et al. [7] had an accuracy score of 83.1% using subcortical volume alone whereas the accuracy of classification between AD and CN using subcortical volume with MMSE was 92.3% with a

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

significant increment of more than 9% in accuracy. Some of the neuropsychological tests can be a good determinant of atrophy in various regions of the brain. For example, naming impairment in AD is associated with left anterior temporal lobe atrophy [5]. Due to this reason, the use of neuropsychological test scores can increase accuracy in the classification.

This study discusses the classification of CN, MCI, and AD individuals by combining the primary MRI with neuropsychological test scores. This study uses the MMSE as a neuropsychological test, which concentrates on cognitive aspects of mental function [14]. It is a 30-point questionnaire that is extensively used to examine the individual's orientation to time and place, repeating abilities, attention and calculation, recall, language, and response to simple commands. The proposed study uses MRI images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The structural measures are extracted from raw T1 structural MRI images using Free Surfer. The machine learning algorithms are trained on different CN, MCI, and AD subjects at the screening phase with some feature selection techniques to yield the best possible results. The results obtained in this experiment are expected to portray the role of MMSE along with the volume of important features in the early detection of AD.

II. METHODOLOGY

Figure 1 shows the flow diagram of the proposed approach for the classification of various stages of AD. The process starts with dataset collection and goes through data preprocessing, feature selection, use of learning algorithms, and finally the result obtained from the proposed approach. The framework for the proposed approach is simple and can be replicated with any machine learning algorithms. Each component is described below in detail.

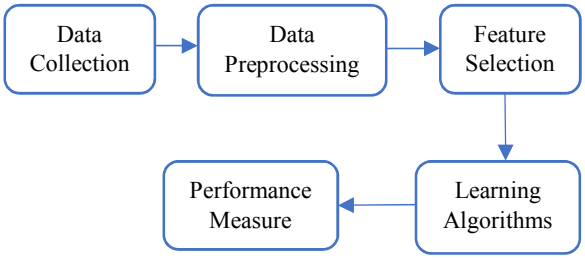

Fig. 1: Flow diagram of the proposed approach

A. Data Collection

The proposed approach uses the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [15]. The ADNI dataset was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of the ADNI dataset was 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.

The ADNI dataset is queried for basic, clinical, and MRI data from CN, MCI, and AD subjects screening. The query results are processed to obtain useful information and to drop the irrelevant information for this paper and only three subsets are considered. The first, second, and third subset contains 119 CN, 155 MCI, and 46 AD subjects, respectively. This paper uses T1 weighted MRI images of 1.5 T. The basic demographic data of the ADNI subjects included in this study are mentioned in Table 1. Data are given as mean ± standard deviation (σ) except for the gender for which frequencies are mentioned and represented by *M* and *F* for males and females, respectively.

B. Data Preprocessing

The attributes of the ADNI dataset are in different ranges such as the range for attribute age and MMSE. The classifiers like SVM work comparatively better with data that is scaled rather than the unscaled data. The main advantage of scaling is to avoid the features in greater numeric ranges dominating those in smaller numeric ranges. Another advantage is to avoid numerical difficulties during the calculation. This is because kernel values depend usually on the inner products of the feature vectors. For the linear kernel and the polynomial kernel, attribute values with a greater numeric range might cause numerical problems. So, for better computation and accuracy, the data are scaled using the *z*-score as shown in Eq. (1).

$$z = \frac{x - \bar{x}}{\sigma} \tag{1}$$

where x is the original feature vector, $\bar{x}$ = is the mean of that feature vector, and σ is the standard deviation.

The number of observations for CN, MCI, and AD is 119, 155, and 46, respectively. Imbalanced data often have poor performance in terms of sensitivity and specificity. So, to nullify this effect, some over-sampling techniques are required. The Synthetic Minority Over-sampling Technique (SMOTE) is used to solve the class imbalance problem by synthesizing new instances from existing minority cases in this paper [16].

C. Feature Selection

The irrelevant features can simply overfit the model, hence there is a need for a feature selection technique to handle the overfitting issue. The feature selection techniques improve the reliability of the model out of

sample if the irrelevant features are dropped out. This paper uses a filter-based feature selection method, which is independent of any classifier [17]. The feature selection method consists of a forward selection (Weka's *Bestfirst*) to look for a combination of attributes with the high individual predictive value of the diagnostic class and low inter-correlation (Wekas's *CfsSubsetEval*) [17]. It evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. The significant features obtained for different classification problems are mentioned in Table 2.

Table 1: Demographic data of ADNI subjects included in this paper

Attributes	CN (N= 119)	MCI (N= 155)	AD (N= 46)
Gender	60 F/ 59 M	49 F/ 106 M	24 F/ 22 M
Age	76.12 ± 5.06	75.33 ± 7.45	76.08 ± 5.92
Years of Education	15.79 ± 3.01	15.74 ± 2.88	24.19 ± 1.49
MMSE	29.11 ± 0.97	27.3 ± 1.75	24.19 ± 1.49

Table 2: Significant features for CN vs AD, MCI vs AD and CN vs MCI

Classification	Rank	Variables
CN *vs* AD	1	MMSE
	2	Hippocampus (Left) Volume
	3	Hippocampus (Right) Volume
MCI *vs* AD	1	MMSE
	2	Hippocampus (Right) Volume
	3	Hippocampus (Left) Volume
CN *vs* MCI	1	MMSE
	2	Hippocampus (Left) Volume
	3	Hippocampus (Right) Volume

D. Machine Learning Algorithms

The best learning algorithm for the given classification problem is very difficult to select by instinct. So, different machine learning algorithms are compared with the proposed approach. Some of the machine learning algorithms used in this paper can be studied under two broad headings viz. Classification and Ensemble Learning algorithms. The ensemble learners are chosen because sometimes ensemble learners outperform the classification algorithms by combining the predictions of multiple base estimators in order to improve the robustness of a given single classifier. Five classification algorithms are used in this paper. Wherever applicable, ten-fold cross-validation is performed to optimize the hyper-parameters of models using a grid-search technique. Ensemble learning algorithms are expected to give better performances than the base learning algorithms. Multiple hypotheses are generated by running the same base learner and are combined to form a better hypothesis. Among some of the popular ensemble learning algorithms, AdaBoost, XGBoost, and Random Forest are used.

Logistic Regression: Logistic regression is a statistical method to model a binary classification problem [18], which fits with the binary classification problems in this paper such as CN vs AD, CN vs MCI, and MCI vs AD. There are certain parameters to be tuned in logistic regression like *C* (regularization) and *penalty*. A grid search method is performed to tune these parameters. The values of parameters *C* and *penalty* for CN vs AD without MMSE are 0.1 and '*L2*', and with MMSE are 1 and '*L2*'. Similarly, the values of parameters *C* and *penalty* for CN vs MCI without MMSE are 0.1 and '*L1*', and with MMSE are 0.01 and '*L2*'. The values of parameters *C* and *penalty* for MCI vs AD without MMSE are 0.1 and '*L2*', and with MMSE are 0.1 and '*L1*'.

Decision Tree: Decision tree is one of the simplest and yet most useful machine learning algorithms. It can be used to solve both classification and regression problems [19]. There are some parameters like *maximum depth* which is the longest path from the tree root to a leaf and *minimum samples* that are required to be at a leaf node. A Grid search method is used to tune

these parameters. The values of parameters *maximum depth* and *minimum samples* for CN vs AD without MMSE are 7 and 1, and with MMSE are 1 and 1. Similarly, the values of parameters *maximum depth* and *minimum samples* for CN vs MCI without MMSE are 1 and 100, and with MMSE are 5 and 5. The values of parameters *maximum depth* and *minimum samples* for MCI vs AD without MMSE are 7 and 1, and with MMSE are and 12 and 1.

Naïve Bayes: Naïve Bayes classifier is a probabilistic machine learning model that is used to solve classification problems. A naïve Bayes classifier assumes that all attributes are independent given the class variable [20].

Factorization Machines: The Factorization machine is a supervised learning algorithm that can be used in solving a wide variety of prediction tasks. It was introduced by Steffen Rendle in 2010 and can be used to solve various prediction tasks like regression, binary classification, and ranking [21]. The values for tuned parameters of rank, order, and learning rates are 10, 3, and 0.001, respectively.

Support Vector Machines: Support Vector Machines are based on the concept of decision planes that define decision boundaries. Support Vector Machine constructs a hyperplane in high-dimensional space which can be used to classify data into two groups [12]. It can also be used for tasks like regression and outlier detection. The performance of SVM is affected by parameters *C* (regularization) and γ (gamma). A Grid search method is performed to tune these parameters. The values of parameters *C* and γ for CN vs AD without MMSE are 32 and 2, and with MMSE are 0.125 and 0.5. Similarly, the values of parameters *C* and γ for CN vs MCI without MMSE are 2048 and 0.0001, and with MMSE are 8 and 2. The values of parameters *C* and γ for MCI vs AD without MMSE are 8 and 8, and with MMSE are 8 and 8.

AdaBoost: AdaBoost, a short form for Adaptive Boosting is a meta-estimator that was developed by Freund and Schapire which has its applications in numerous fields. It can be used in conjunction with many other learning algorithms to improve the accuracy of especially weak learners. This is because AdaBoost combines weak hypotheses by summing their probabilistic predictions [22]. The values of parameters *learning rate*, *loss function* and *maximum number of estimators* for AD vs CN without MMSE are 1, linear and 50, and with MMSE are 0.01, linear, and 100. Similarly, the values of parameters *learning rate, loss function,* and *maximum number of estimators* for CN vs MCI without MMSE are 0.1, exponential and 100, and with MMSE are 0.01, square and 50. The values of parameters *learning rate*, *loss function* and *maximum number of estimators* for MCI vs AD without MMSE are 0.3, linear and 100, and with MMSE are 0.1, square and 100.

XGBoost: Extreme Gradient Boosting or popularly known as XGBoost is a tree-based ensemble learning technique that uses gradient boosting [23]. XGBoost has recently been widely used and has shown great performance in various machine learning competitions. Hyperparameters like *learning rate*, *maximum depth*, *minimum samples* that are required to be at a leaf node, and *number of estimators* are tuned using grid search. The values of parameters *learning rate*, *maximum depth*, *minimum samples,* and *number of estimators* for CN vs AD without MMSE are 0.01, 2, 3 and 1000, and with MMSE are 0.01, 2, 3, 1000. Similarly, the values of parameters *learning rate*, *maximum depth*, *minimum samples,* and *number of estimators* for CN vs MCI without MMSE are 0.01, 2, 3 and 1000, and with MMSE are 0.01, 2, 3, 1000. The value of parameters *learning rate*, *maximum depth*, *minimum samples,* and *number of estimators* for MCI vs AD without MMSE are 0.01, 2, 3 and 1000, and with MMSE are 0.01, 2, 3, 1000.

Random Forest: Random forest is a meta-estimator used for solving both regression and classification problems. The random forest as the name suggests is a forest of many decision trees that outputs the class that is calculated from aggregating the classification outputs of each decision tree of that forest [24]. There are certain parameters that affect the classification performance of the Random Forest classifier. Grid search is performed to find optimal values of the function to measure the quality of a split, and the number of estimators. The values of parameters *criterion* and *number of estimators* for CN vs AD, CN vs MCI and MCI vs AD are *gini* and 200, respectively for both without and with MMSE.

E. Performance Measures

The performance of the proposed approach is measured by different evaluation criteria such as accuracy, precision, recall, specificity, sensitivity, F1-score, the area under the curve (AUC) as shown in Eqs. (2)-(7).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1-Score = 2.\frac{Precision \times Recall}{Precision+Recall} \quad (5)$$

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (7)$$

where *TP*, *TN*, *FP,* and *FN* indicate True Positive, True Negative, False Positive and False Negative, respectively.

III. EXPERIMENTAL RESULTS

The feature selection techniques give significant attributes as shown in Table 2. This shows that atrophy in the hippocampal region is more significant in the distinction of different stages of AD when used with neuropsychological scores (MMSE). The classification accuracy with and without MMSE for different classifiers are shown in figs. 2-4. It is observed that SVM with RBF kernel outperforms the other classification techniques.

The classification performance in terms of accuracy after feature selection for SVM with RBF kernel is shown in Table 3. From Table 3, it can be observed that with MMSE, the performance is seen better in all three classification types viz. CN vs AD, CN vs MCI, and MCI vs AD. Table 4 compares AUC values with MMSE and without MMSE for various classifiers under the studied experiments. The best performance is shown by SVM with RBF kernel, Extreme Gradient Boosting, and Random Forest. While taking other measures of performance, it is observed that SVM with RBF kernel performs better than other classifiers.

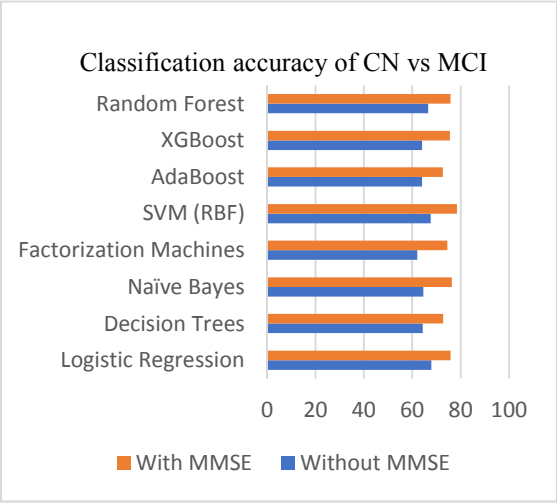

Fig. 2: Classification accuracy for CN vs MCI with and without MMSE

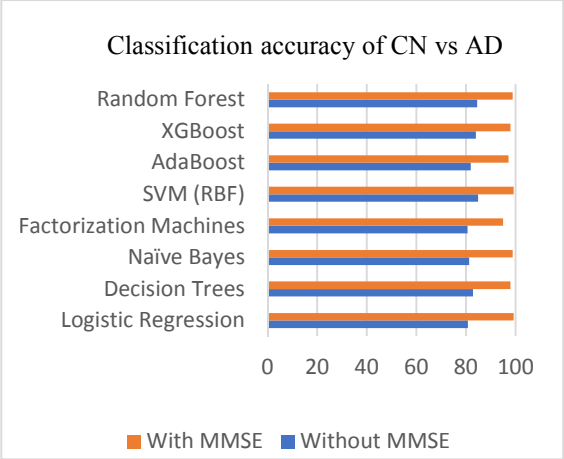

Fig. 3: Classification accuracy for CN vs AD with and without MMSE

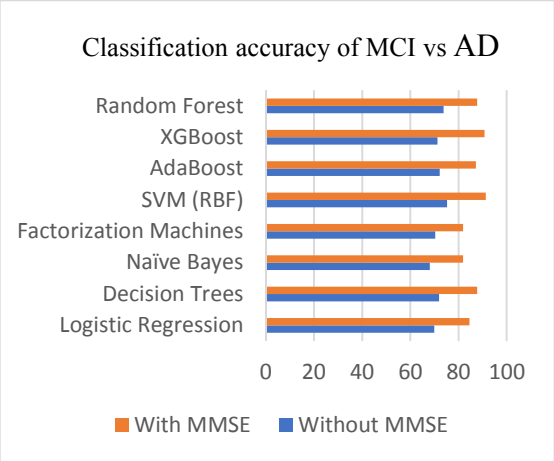

Fig. 4: Classification accuracy for MCI vs AD with and without MMSE

Table 5 shows that SVM with RBF performs better with MMSE in terms of all performance measures. It can be also observed that the performance of CN vs AD is better than MCI vs AD and CN vs MCI. Table 6 compares the performance of the proposed approach with other approaches for different stages of Alzheimer's disease. It can be observed that with the addition of neuropsychological tests to imaging modalities, the accuracy increases significantly. The performance of the proposed approach is better in terms of accuracy, sensitivity, and specificity with various MRI modalities and multimodal techniques as well. The addition of neuropsychological tests with imaging modalities provides better decisional space for the classification of different stages of Alzheimer's disease.

IV. CONCLUSION

The research for finding new biomarkers for Alzheimer's disease is undoubtedly of great importance but a combination of neuropsychological test scores like MMSE with MRI measures can also give more accurate results. This paper proposes a feature selection based model which enables to learn good patterns for the given data and enhances the performance of the model. This paper substantiates that the role of neuropsychological test scores can be vital in improving the performance of the classification models as it carries some important information about the brain. Also, the multimodal methods in early diagnosis of AD are proved to be more accurate such as image modality with neuropsychological scores. In the future, the proposed study can be used to find how atrophy in different regions of interest is correlated with test scores. Also, different neuropsychological tests other than MMSE can be used for proving more robust and accurate results.

Table 3: Performance in terms of accuracy by SVM based on feature selection

Classification Type	With MMSE	Without MMSE
CN vs AD	99.2	84.9
CN vs MCI	78.5	67.6
MCI vs AD	91.3	75.2

Table 4: Comparison of AUC values of different classifiers with and without MMSE

Model	CN vs AD		CN vs MCI		MCI vs AD	
	Without MMSE	With MMSE	Without MMSE	With MMSE	Without MMSE	With MMSE
Logistic Regression	0.807	0.992	0.677	0.752	0.70	0.845
Decision Trees	0.823	0.979	0.642	0.723	0.719	0.877
Naïve Bayes	0.811	0.987	0.645	0.67	0.681	0.819
Factorization Machines	0.807	0.949	0.621	0.745	0.704	0.821
Support Vector Machines	0.849	0.992	0.677	0.784	0.752	0.913
AdaBoost	0.819	0.971	0.639	0.726	0.723	0.819
XGBoost	0.840	0.979	0.639	0.755	0.713	0.906
Random Forest	0.845	0.987	0.665	0.758	0.739	0.877

Table 5: Performance of svm based on feature selection with and without MMSE as shown in table 2

Performance Measure	CN vs AD		CN vs MCI		MCI vs AD	
	Without MMSE	With MMSE	Without MMSE	With MMSE	Without MMSE	With MMSE
Accuracy	84.89	99.2	67.56	78.5	75.15	91.3
Precision	0.85	0.99	0.68	0.79	0.75	0.91
Recall	0.85	0.99	0.67	0.78	0.75	0.91
F1-Score	0.85	0.99	0.67	0.78	0.75	0.91
AUC	0.849	0.992	0.674	0.784	0.752	0.913
Sensitivity	0.84	1.00	0.59	0.72	0.81	0.92
Specificity	0.86	0.98	0.76	0.85	0.70	0.90

Table 6: Performance comparison of the proposed approach with other methods

Method	Target	Modality	Data	Performance		
				Acc (%)	Sens.	Spec.
Zhou et al., 2014 [25]	CN vs AD	MRI	(127 AD, 59 CN) Private	78.2	0.685	0.755
	aMCI vs CN			66.7	0.55	0.725
	naMCI vs CN			62.1	0.515	0.703
	CN vs AD	MRI + MMSE	(127 AD, 59 CN) Private	92.4	0.84	0.961
	aMCI vs CN			74.9	0.611	0.834
	naMCI vs CN			74.1	0.552	0.823
Hinrichs et al., 2011 [11]	CN vs AD	MRI + PET	(129 AD, 60 CN) ADNI	87.6	0.789	0.938
	CN vs AD	MRI+PET+ CSF+ APOE+ Cognitive Scores	(129 AD, 60 CN) ADNI	92.4	0.867	0.966
Zhou et al., 2014 [7]	CN vs AD	MRI	(129 AD, 60 CN) Private	83.1	0.779	0.856
	CN vs AD	MRI + MMSE	(129 AD, 60 CN) Private	92.3	0.882	0.942
Kim et al., 2017 [26]	CN vs AD	MRI	(160 AD, 208 CN) ADNI	92.84	0.885	0.961
	MCI vs CN			78.28	0.863	0.651
Cuingnet et al., 2011[#] [10]	CN vs AD	MRI	(162 CN, 76 MCIc, 137 AD) ADNI	-	0.81	0.95
	CN vs MCIc			-	0.73	0.85
Proposed approach	CN vs AD	MRI	(46 AD, 155 MCI, 119 CN) ADNI	84.89	0.84	0.86
	CN vs MCI			67.56	0.59	0.76
	MCI vs AD			75.15	0.81	0.70
	CN vs AD	MRI + MMSE	(46 AD, 155 MCI, 119 CN) ADNI	99.2	1.00	0.98
	CN vs MCI			78.5	0.72	0.85
	MCI vs AD			91.3	0.92	0.90

[#] Best performance out of ten different methods

REFERENCES

[1] R. A. Stelzmann, H. Norman Schnitzlein, and F. Reed Murtagh, "An English translation of Alzheimer's 1907 paper, "Über eine eigenartige Erkankung der Hirnrinde"," Clinical Anatomy: The Official Journal of the American Association of Clinical Anatomists and the British Association of Clinical Anatomists, vol. 8, no. 6, pp. 429-431, 1995.

[2] C. Patterson, "The state of the art of dementia research: New frontiers," World Alzheimer Report 2018, 2018.

[3] L. J. Herrera, I. Rojas, H. Pomares, A. Guillén, O. Valenzuela, and O. Baños, "Classification of MRI images for Alzheimer's disease detection," in 2013 International Conference on Social Computing, 2013, pp. 846-851: IEEE.

[4] D. H. Adler et al., "Characterizing the human hippocampus in aging and Alzheimer's disease using a computational atlas derived from ex vivo MRI and histology," Proceedings of the National Academy of Sciences, vol. 115, no. 16, pp. 4252-4257, 2018.

[5] K. Domoto-Reilly, D. Sapolsky, M. Brickhouse, B. C. Dickerson, and A. s. D. N. Initiative, "Naming impairment in Alzheimer's disease is associated with left anterior temporal lobe atrophy," Neuroimage, vol. 63, no. 1, pp. 348-355, 2012.

[6] S. P. Poulin, R. Dautoff, J. C. Morris, L. F. Barrett, B. C. Dickerson, and A. s. D. N. Initiative, "Amygdala

atrophy is prominent in early Alzheimer's disease and relates to symptom severity," Psychiatry Research: Neuroimaging, vol. 194, no. 1, pp. 7-13, 2011.
[7] Q. Zhou, M. Goryawala, M. Cabrerizo, W. Barker, R. Duara, and M. Adjouadi, "Significance of normalization on anatomical MRI measures in predicting Alzheimer's disease," The Scientific World Journal, vol. 2014, pp. 1-12, 2014.
[8] M. M. Williams, M. Storandt, C. M. Roe, and J. C. Morris, "Progression of Alzheimer's disease as measured by Clinical Dementia Rating Sum of Boxes scores," Alzheimer's & Dementia, vol. 9, no. 1, pp. S39-S44, 2013.
[9] J. Zhang, B. Yan, X. Huang, P. Yang, and C. Huang, "The diagnosis of Alzheimer's disease based on voxel-based morphometry and support vector machine," in 2008 Fourth International Conference on Natural Computation, 2008, vol. 2, pp. 197-201: IEEE.
[10] R. Cuingnet et al., "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.
[11] C. Hinrichs, V. Singh, G. Xu, S. C. Johnson, and A. D. N. Initiative, "Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population," Neuroimage, vol. 55, no. 2, pp. 574-589, 2011.
[12] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273-297, 1995.
[13] M. Ewers et al., "Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance," Neurobiology of aging, vol. 33, no. 7, pp. 1203-1214. e2, 2012.
[14] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician," Journal of psychiatric research, vol. 12, no. 3, pp. 189-198, 1975.
[15] (2019, Jul 15). Data Sharing and Publication Policy. Available: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_DSP_Policy.pdf
[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321-357, 2002.
[17] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.
[18] Z. Y. Algamal and M. H. Lee, "Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification," Expert Systems with Applications, vol. 42, no. 23, pp. 9326-9332, 2015.
[19] J. R. Quinlan, "Induction of decision trees," Machine learning, vol. 1, no. 1, pp. 81-106, 1986.
[20] I. Rish, "An empirical study of the naive Bayes classifier," in IJCAI 2001 workshop on empirical methods in artificial intelligence, 2001, vol. 3, no. 22, pp. 41-46.
[21] S. Rendle, "Factorization machines," in 2010 IEEE International Conference on Data Mining, 2010, pp. 995-1000: IEEE.
[22] J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson, "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, 2009.
[23] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 785-794: ACM.
[24] M. Pal, "Random forest classifier for remote sensing classification," International Journal of Remote Sensing, vol. 26, no. 1, pp. 217-222, 2005.
[25] Q. Zhou et al., "An optimal decisional space for the classification of Alzheimer's disease and mild cognitive impairment," IEEE Transactions on Biomedical Engineering, vol. 61, no. 8, pp. 2245-2253, 2014.
[26] J. Kim and B. Lee, "Automated discrimination of dementia spectrum disorders using extreme learning machine and structural T1 MRI features," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2017, pp. 1990-1993: IEEE.