Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2017 July ; 2017: 3914–3917. doi:10.1109/EMBC.2017.8037712.

^{11}C-PIB PET Image Analysis for Alzheimer's Diagnosis Using Weighted Voting Ensembles

Wenjun Wu [Student Member, IEEE], **Janani Venugopalan [Student Member, IEEE]**, **May D. Wang, Ph.D. [Senior Member, IEEE] Alzheimer's Disease Neuroimaging Initiative**

Wallace H. Countler Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, 30332, USA

Abstract

Alzheimer's Disease (AD) is one of the leading causes of death and dementia worldwide. Early diagnosis confers many benefits, including improved care and access to effective treatment. However, it is still a medical challenge due to the lack of an efficient and inexpensive way to assess cognitive function [1]. Although research on data from Neuroimaging and Brain Initiative and the advancement in data analytics has greatly enhanced our understanding of the underlying disease process, there is still a lack of complete knowledge regarding the indicative biomarkers of Alzheimer's Disease. Recently, computer aided diagnosis of mild cognitive impairment and AD with functional brain images using machine learning methods has become popular. However, the prediction accuracy remains unoptimistic, with prediction accuracy ranging from 60% to 88% [2,3,6]. Among them, support vector machine is the most popular classifier. However, because of the relatively small sample size and the amount of noise in functional brain imaging data, a single classifier cannot achieve high classification performance. Instead of using a global classifier, in this work, we aim to improve AD prediction accuracy by combining three different classifiers using weighted and unweighted schemes. We rank image-derived features according to their importance to the classification performance and show that the top ranked features are localized in the brain areas which have been found to associate with the progression of AD. We test the proposed approach on 11C- PIB PET scans from The Alzheimer's Disease Neuroimaging Initiative (ADNI) database and demonstrated that the weighted ensemble models outperformed individual models of K-Nearest Neighbors, Random Forests, Neural Nets with overall cross validation accuracy of 86.1% ± 8.34%, specificity of 90.6% ± 12.9% and test accuracy of 80.9% and specificity 85.76% in classification of AD, mild cognitive impairment and healthy elder adults.

I. Introduction

As one of the compelling unsolved medical problems, Alzheimer's Disease (AD) affects more than 5.3 million patients in the United States of America [1]. AD is an irreversible chronic neurodegenerative disease that is the most common form of dementia. The incidence of dementia caused by AD has become a significant social problem. There has been extensive ongoing research about early diagnosis and treatment of AD, but early diagnosis

wuwenjun@gatech.edu.

remains a medical challenge due to the absence of a definitive diagnosis test for AD. In fact, less than 50% of the people with AD are being diagnosed accurately with the disease based on clinical symptoms [1]. The positron emission tomography (PET) imaging is a non-invasive, three-dimensional imaging modality that uses radioactive substance to detect functional changes in the brain. PET imaging has been recently identified as a major advancement in the detection of AD [2]. The tracer, Carbon 11-labeled Pittsburgh Compound B (11C- PIB), has shown more uptake in the brains of patient with AD than in those of control group, especially in the area thalamus, putamen, caudate, hippocampus and subcortical white matter of the patients [2]. Thus, a region-based analysis of 11C- PIB PET scans that addressed those critical brain areas is expected to generate optimistic prediction performance.

Recent advances in computer aided diagnosis (CAD) systems have shown potentials in providing accurate diagnosis of the AD using brain function images [3] However, the prediction accuracy of AD, especially among patients with mild cognitive impairment (MCI) was approximately 70% [4]. Besides, most present CAD systems are based on support vector machine (SVM) [3,4,5]. Although SVM has been the most commonly used classifier, it has limited performance in the presence of noise and outliers, which is abundant in PET imaging data. Moreover, because of relatively small sample size and the amount of noise in functional PET imaging data, a single classifier cannot achieve good general performance. It is well-known in the artificial intelligence field that ensemble methods can be used for improving general classification performance and alleviate the potential data overfitting [6]. Previous study has demonstrated the potential of ensemble methods in improving prediction accuracy of AD in PET imaging data. C. Cabral et al [7] classified AD, MCI and Control (CN) in Fluorodeoxyglucose-(FDG) PET images using favorite class ensemble methods, which composed of three base classifiers, each trained with different feature subsets. However, their proposed ensemble method only utilized single type of classifier and average voting to generate final decision, which could be biased and prone to noise due to the limitation of unweighted voting and single type of classifier. In present study, we addressed this challenge by proposing an ensemble classification of 11C-PIB PET scans from Alzheimer’s disease neuroimaging initiative (ADNI) participants. In this approach, the classification produced in first iteration is used as “prior knowledge” to generate both weighted and unweighted ensemble of different classifiers.

II. Methodology

In this work, we performed a systematic analysis of 11C- PIB PET scans to refine the current knowledge regarding the indicative biomarkers of AD and to improve AD diagnosis precision (Figure 1). After obtaining data from ADNI database and perform image processing, we extracted volume, texture and voxel features across different brain areas segmented. We then performed classification using individual classifiers, such as Random forests, k-Nearest Neighbors and Neural Nets. Finally, we combined the decision of three individual classifiers using weighted and unweighted voting.

A. Data

Data used in the preparation of this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was initiated in 2003 and led by Principal Investigator Michael W. Weiner, MD. The study aims to test whether serial magnetic resonance imaging, PET, other biological markers and clinical and neuropsychological assessment can be used to measure the progression of MCI and early AD. In this study, the preprocessed 228 11C-PIB PET image volumes from 103 ADNI participants were acquired from the ADNI database. Preprocessing steps performed by ADNI are: co-registration, average, standard space transformation, voxel normalization and smoothing with 8mm FWHM filter.

B. Image Processing

PET scans were processed automatically (Figure 2), using FSL Toolbox (Oxford University, UK) [8]. First, the images were skull-stripped to remove non-brain tissue. To enable robust registration, all images were then aligned to standard space, ICBM152 space [9]. Next, tissue segmentation is performed to segment the images into Grey Matter (GM), White Matter (WM) and Cerebrospinal fluid (CSF). Finally, we performed volume segmentation to extract useful brain areas that have been found to show increased 11C- PIB uptake in AD patients by previous study [2, 10–13]. The extracted brain areas are Thalamus, Brainstem, Hippocampus, Amygdala, Putamen, Pallidum, Accumbens, and Caudate. In our study, a total of 208 out of 228 PET scans, which have sufficient quality to provide us with successful volume segmentation of the listed eight brain areas through the processing steps (Figure 2) were utilized for further analysis.

C. Feature Extraction

We extracted three types of features to be used as classification features: volume, voxel intensities and texture. Volumes of the eight extracted brain areas were calculated from the binary mask. We calculated volumes of WM, GW, CSF from the probability tissue maps [14]. Using segmented binary masks, the voxels within all tissue types and brain areas were extracted. We performed texture analysis to extract energy, entropy and 13 Haralick texture features [15]. Energy and entropy were calculated from multiwavelet transformation [15]. Haralick's texture features were calculated using 64 gray-level co-occurrence matrix (GLCM) in 8 directions [15]. Since there is a lack of established way to perform three-dimensional texture analysis, we extended two-dimensional texture analysis for three-dimensional texture analysis. For each volume, texture features of each slice were averaged to generate the final value.

C. Feature Selection

Minimum redundancy maximum relevance feature selection method (mRMR) was used to minimize redundancy and select features according to measures of relevance and dependence [16]. Up 300 features were selected and used by the classifiers. The number of features used by each classifier was optimized by 10-fold cross validation (CV). Considering the total sample size is 208, a maximum of 300 features would be appropriate.

D. Classification

Features selected by mRMR were used by baseline classifiers for prediction of CN, AD and MCI. The baseline classifiers adopted in this study were K-Nearest Neighbors (kNN), Random Forests (RF), and Neural Nets (NN). These classifiers are commonly used classifiers and are suitable for solving high dimensional, multi-class classification problems where there is relatively small amount of training samples. SVM is not selected since it is inherently binary classifier. Classification scores, which are the predicted class posterior probabilities, were generated from the classifiers.

1) Hyper parameter selection: hyper parameters of different classifiers, such as the hidden layers of neural nets, the number of nearest neighbors of kNN and the number of single decision trees in the RF, were optimized using grid search in 10-fold CV.

2) Ensemble decision: the classification score from baseline classifiers were assembled using weighted and unweighted voting schemes.

In unweighted voting scheme, the new classification scores were computed by averaging the classification scores from three baseline classifiers. The new decision label is the class that has largest classification score.

In weighted voting scheme, we first determined the weight for each classifier using the following equation [6]

$$\log(\frac{accuracy}{1-accuracy})$$

where accuracy refers to the classification accuracy of individual classifiers. The weight adjusted the relative importance of each classifiers so that the accuracy performance of each classifier is proportional to its weight. In weighted voting scheme, the weighted average of classification scores from baseline classifiers is the new classification scores, which were then used to compute new decision labels.

E. Evaluation

Dataset was separated into 20% testing data and 80% training data. Hyparameters and the number of features selected were optimized using 10-fold CV on training data. The final model was trained on entire set of training data with optimized hyperparameters and evaluated on 20% testing data. Performance metrics reported in this study are: Accuracy, Specificity and Pearson’s correlation coefficient (PCC). Mean values with standard deviations were reported (mean ± standard deviation).

III. Results

As described above, each classifier classified 208 PET image scans into categories of CN, MCI and AD. The performances of three individual classifiers and two ensemble classifiers are compared. In this dataset, there were 47 instances of CN, 99 instances of MCI, and 62 instances of AD.

A. Classification Results

The results obtained from classification experiments are shown in Table II. The best CV performance of 86.1% ± 8.34% accuracy, 90.6% ± 12.9% specificity and the best test performance of 80.9% accuracy, 85.76% specificity were achieved by weighted ensemble classifier.

In terms of individual classifiers, the results of RF, kNN and NN are drastically different. RF achieved 74.5% overall test accuracy and 82.0% specificity while the overall test accuracy of KNN and NN are 61.3%, 63.4% and specificity of 66.5% and 65.3% respectively. Besides, the CV performance of RF is much better than kNN and NN interms of accuracy, specificity and PCC across three classes. The superiority of RF over kNN and NN could attribute to the fact that when training set is small, high bias classifiers, such as single decision tree, which is the base unit of RF, have an advantage over low bias classifiers, such as kNN since the latter will overfit. Besides, RF by itself, is an ensemble method that uses a multitude of simple decision trees. Decision trees are weak learners and might have better prediction in regards to this classification problem.

In regards to ensemble methods, the unweighted ensemble classifier has overall test accuracy of 70.1% and 62.5% specificity. The weighted ensemble classifier achieved highest accuracy in CN, MCI and AD, and highest specificity in MCI and AD. The unweighted ensemble classifier outperforms NN and KNN in terms of overall accuracy, PCC and specificity.

However, RF and weighted ensemble classifier performs much better than unweighted ensemble classifier in almost all areas of measurements. The superiority of weighted ensemble classifier over unweighted ensemble classifier conforms with our expectation since unweighted average of decision values could lead to biased performance. The presented results indicate that the weighted ensemble method that combines multiple classifiers has great potential to enhance the overall diagnosis precision of AD.

B. Feature Analysis

To determine the most important features for the classification methods, top ranked features from mRMR were investigated, as shown in Table I. Although there is some variability in the feature ranked by different classification methods, there are several highly ranked common features. Haralick texture features are ranked the highest and among them, correlation feature [15], which measures the gray tone linear-dependencies and information measure of correlation [15], are the most important. Highly ranked brain regions are: Grey Matter [2], Caudate [10], Putamen [11] and Thalamus[12] in descending sequence. These highly ranked brain areas conform with the important brain areas found by previous study that exhibit marked 11C-PIB uptakes in patients with AD and MCI, comparing with that in normal elders [2, 10–13].

IV. Conclusion & Future work

Recently, computer aided diagnosis (CAD) systems using brain images has become popular in AD diagnosis [3]. However, the prediction accuracy of AD, especially among patients with mild cognitive impairment (MCI) was only approximately 70% [4].

In this study, we achieved high AD and MCI diagnosis accuracy with ensemble learning methods that combine different types of classifiers, such as NN, RF and kNN, as well as to refine current knowledge regarding brain areas associated with 11C- PIB. Both weighted ensemble methods and unweighted ensemble methods were tested on 11C- PIB PET image dataset from ADNI. The top features ranked by classifiers are in the brain areas that have been found to associate with the progression of AD [2]. We showed that the ensemble method, where the proportion of the decision was based on the performance of individual classifier outperformed individual classifiers, with best overall CV accuracy of 86.1% ± 8.34%, CV specificity of 90.6% ± 12.9%, best overall test accuracy of 80.9% and specificity of 85.76%. This result also outperforms most state-of-art computer-aided AD diagnosis systems with accuracy of 60% to 88%. Besides, we have also confirmed that the highly ranked common features are in brain areas that have been found to be related with the progression of AD. The results have demonstrated the potential value of 11C-PIB in improving AD diagnosis accuracy as an indicative biomarker of AD.

Our work, however, currently only addressed 11C- PIB PET image datasets while other tracers such as FDG and Florbetapir are also suggested as core biomarkers for AD. In the future, we would like to compare the performance of proposed methods on different PET imaging datasets. We would also like to develop ensemble methods that can integrate PET imaging datasets from different PET imaging tracers such as Florbetapir, FDG and 11C-PIB.

Acknowledgment

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

[1]. Alzheimer's Association. 2016 Alzheimer's Disease Facts and Figures. Alzheimer's & Dementia 2016;12(4).

[2]. Nordberg Agneta. "PET imaging of amyloid in Alzheimer's disease." The lancet neurology 3.9 (2004): 519–527. [PubMed: 15324720]

[3]. Ramírez J, Górriz JM, Salas-Gonzalez D, Romero A, López M, Álvarez I, and Gómez-Río M, "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features," Inf. Sci. (Ny), vol. 237, pp. 59–72, 2013.

[4]. Eskildsen SF, Coupé P, Fonov VS, Pruessner JC, and Collins DL, "Structural imaging biomarkers of Alzheimer's disease: predicting disease progression," Neurobiol. Aging, vol. 36, pp. S23–S31, 2015. [PubMed: 25260851]

[5]. Illán IA, Górriz JM, Ramírez J, Salas-Gonzalez D, López MM, Segovia F, Chaves R, Gómez-Rio M, and Puntonet CG, "18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis," Inf. Sci. (Ny), vol. 181, no. 4, pp. 903–916, 2011.

[6]. Rokach L, "Ensemble-based classifiers," Artif Intell Rev, vol. 33, pp. 1–39, 2010.

[7]. Cabral C, Silveira M, and Alzheimer's Disease Neuroimaging Initiative, "Classification of Alzheimer's disease from FDG-PET images using favourite class ensembles," in 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, vol. 2013, pp. 2477–2480.

[8]. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, and Smith SM, "FSL," Neuroimage, vol. 62, no. 2, pp. 782–790, Aug. 2012. [PubMed: 21979382]

[9]. Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL, and Brain Development Cooperative Group, "Unbiased average age-appropriate atlases for pediatric studies.," Neuroimage, vol. 54, no. 1, pp. 313–27, Jan. 2011. [PubMed: 20656036]

[10]. Benzinger TLS, Blazey T, Jack CR, Koeppe RA, Su Y, Xiong C, et al. "Regional variability of imaging biomarkers in autosomal dominant Alzheimer's disease.," Proc. Natl. Acad. Sci. U. S. A, vol. 110, no. 47, pp. E4502–9, Nov. 2013. [PubMed: 24194552]

[11]. Farid K, Almkvist O, Brueggen K, Carter SF, Wall A, Herholz K, and Nordberg A, "HIGH PUTAMEN 11C-PIB RETENTION IN MCI IS ASSOCIATED WITH AN INCREASED RISK OF CONVERSION TO AD," Alzheimer's Dement, vol. 10, no. 4, p. P16, 7 2014.

[12]. Koivunen J, Verkkoniemi A, Aalto S, Paetau A, Ahonen J-P, Viitanen M, et al. "PET amyloid ligand [11C]PIB uptake shows predominantly striatal increase in variant Alzheimer's disease," Brain, vol. 131, no. 7, pp. 1845–1853, Jul. 2008. [PubMed: 18583368]

[13]. Apostolova LG, Hwang KS, Andrawis JP, Green AE, Babakchanian S, Morra JH, et al. "3D PIB and CSF biomarker associations with hippocampal atrophy in ADNI subjects." Neurobiology of aging 318 (2010): 1284–1303. [PubMed: 20538372]

[14]. Zhang Y, Brady M, and Smith S, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," IEEE Trans. Med. Imaging, vol. 20, no. 1, pp. 45–57, Jan. 2001. [PubMed: 11293691]

[15]. Haralick RM, "Statistical and structural approaches to texture," Proc. IEEE, vol. 67, no. 5, pp. 786–804, 1979.

[16]. Peng Hanchuan, Long Fuhui, and Ding C, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," IEEE Trans. Pattern Anal. Mach. Intell, vol. 27, no. 8, pp. 1226–1238, Aug. 2005. [PubMed: 16119262]

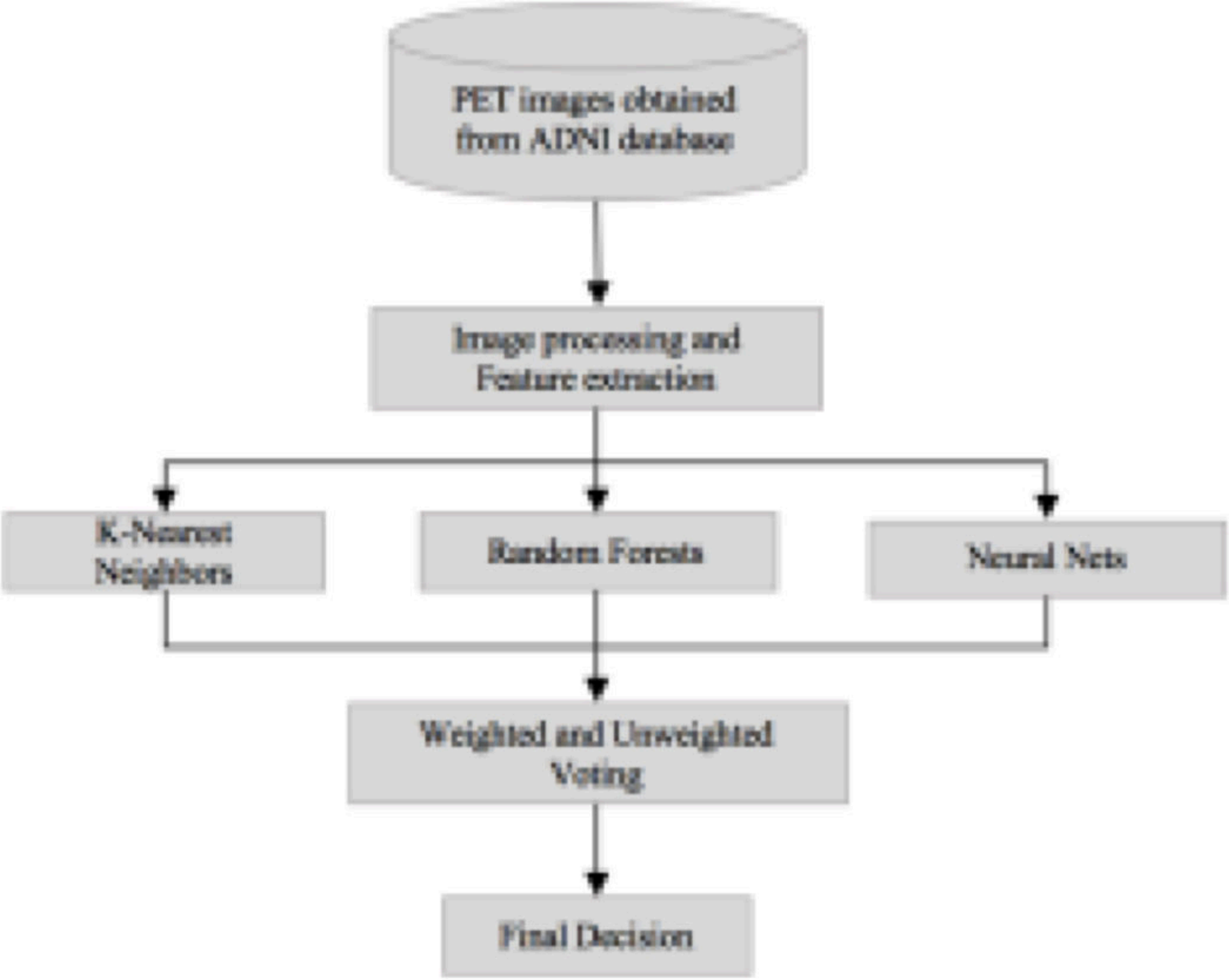
PET images obtained from ADNI database
Image processing and Feature extraction
K-Nearest Neighbors
Random Forests
Neural Nets
Weighted and Unweighted Voting
Final Decision

Figure 1.
Schematic diagram of Proposed Study

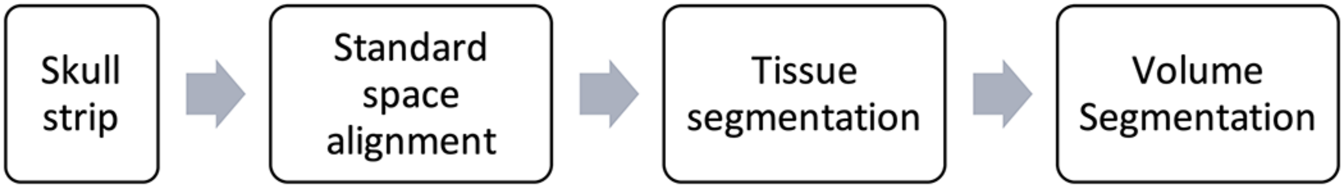

Figure 2.
Schematic diagram of iamge processing

TABLE I.

Top Features Ranked by mRMR Feature Reduction Method.

Rank	Region	Feature
1	Gray Matter	Haralick feature (correlation)
2	Gray Matter	Haralick feature (correlation)
3	Right Caudate	Haralick feature (correlation)
4	Right Caudate	Haralick feature (Information Measure of Correlation I)
5	Right Putamen	Voxel intensity
6	Right Thalamus	Haralick feature (Information Measure of Correlation I)
7	Right Thalamus	Voxel intensity
8	Left Thalamus	Voxel intensity
9	Right Caudate	Voxel intensity
10	Left Putamen	Voxel intensity

TABLE II.

CLASSIFICATION RESULTS OF ALL METHODS

Methods[a]	Classes[b]	CV[d] Accuracy	CV Specificity	CV PCC[c]	Test Accuracy	Test Specificity	Test PCC
kNN	CN	0.832 ± 0.046	0.916 ± 0.105	0.512 ± 0.127	0.714	0.800	0.546
	MCI	0.715 ± 0.135	0.542 ± 0.153	0.435 ± 0.277	0.548	0.524	0.155
	AD	0.787 ± 0.123	0.803 ± 0.129	0.520 ± 0.288	0.643	0.788	0.528
RF	CN	0.870 ± 0.042	0.893 ± 0.051	0.717 ± 0.123	0.819	**0.933**	0.443
	MCI	0.793 ± 0.065	0.786 ± 0.098	0.543 ± 0.128	0.676	0.781	0.202
	AD	0.822 ± 0.060	0.910 ± 0.074	0.695 ± 0.171	0.767	0.798	0.394
NN	CN	0.802 ± 0.048	0.716 ± 0.129	0.512 ± 0.083	0.714	0.685	0.393
	MCI	0.502 ± 0.134	0.542 ± 0.324	0.235 ± 0.129	0.500	0.582	0.178
	AD	0.638 ± 0.083	0.765 ± 0.149	0.320 ± 0.073	0.786	0.742	0.387
Unweighted	CN	0.861 ± 0.044	0.907 ± 0.033	0.702 ± 0.185	0.860	0.917	0.5949
	MCI	0.736 ± 0.103	0.737 ± 0.096	0.588 ± 0.168	0.553	0.333	0.236
	AD	0.861 ± 0.035	0.853 ± 0.112	0.714 ± 0.172	0.818	0.870	0.769
Weighted	CN	**0.901 ± 0.038**	**0.940 ± 0.083**	**0.718 ± 0.083**	**0.886**	0.918	**0.703**
	MCI	**0.827 ± 0.039**	**0.821 ± 0.073**	**0.694 ± 0.129**	**0.752**	**0.803**	**0.545**
	AD	**0.885 ± 0.060**	**0.888 ± 0.129**	**0.726 ± 0.073**	**0.843**	**0.899**	**0.778**

[a] k-Nearest Neighbors (kNN), Random Forests (RF), Neural Nets (NN), Ensemble Model with Unweighted Voting Scheme (unweighted), Ensemble Model with Weighted Voting Scheme (weighted)

[b] Control (CN), Mild Cognitive Impairment (MCI), Alzeimer’s disease,

[c] Pearson’s Correlation Coefficient(PCC)

[d] Cross Validation (CV)