^{11}C-PIB PET IMAGE ANALYSIS FOR AUTOMATED ALZHEIMER'S DIAGNOSIS USING WEIGHTED VOTING ENSEMBLES

WENJUN WU, JANANI VENUGOPALAN, MAY D WANG
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, USA

CREATING THE NEXT®

Overview

- Alzheimer's Disease – leading causes of death and dementia;
- Early Diagnosis is beneficial and important, but challenging;
- Computer-aided diagnosis prevails, but prediction accuracy varies: 60 – 88%;
- SVM has been the baseline classifier, but prone to noise in imaging data;
- Objective of this work
  - Combine different classifiers using weighted and unweighted schemes;
  - Improve AD prediction accuracy;
  - Identify important features
- Results
  - Weighted ensemble models outperforms individual models
  - Cross Validation Accuracy of 86.1% ± 8.34%, specificity of 90.6% ± 12.9%
  - Test accuracy of 80.9% and specificity 85.8%

I. INTRODUCTION

ALZHEIMER'S DISEASE (AD)

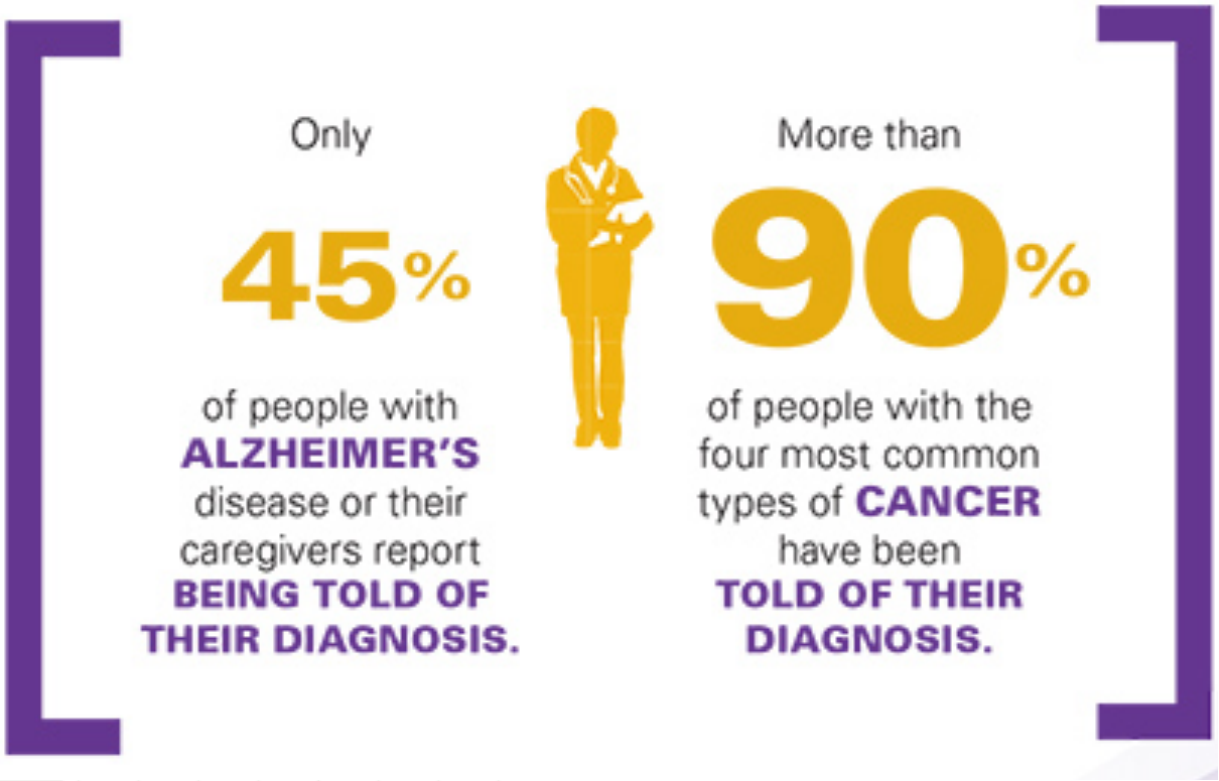


Figure 1. 2016 Alzheimer's Disease (AD) facts

Background

- PET imaging: major advancement in the assessment of AD[4].

- Carbon 11-labeled Pittsburgh Compound B (^{11}C- PIB), has shown more uptake in the AD patients' brain.

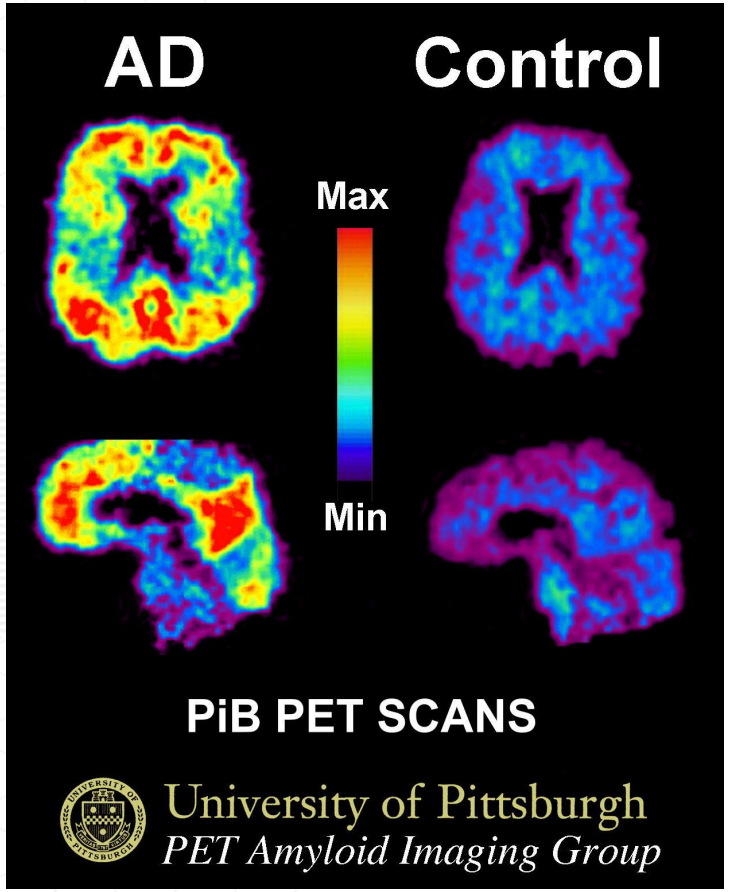


Figure 2. Difference in PIB uptake between AD patients and Control in PET images.

Background

- Prediction accuracy of AD was approximately 70% [4].
- Potential of Ensemble methods in improving classification performance.
  - noisy data and small sample size
- Previous study [7] used some ensemble method on FDG-PET scans, but accuracy remains unoptimistic.
  - Feature extraction not based on physiological brain areas
  - Single type of classifier
  - Average voting to generate final decision

Proposed Method

In present study, we proposed an ensemble classification of ^{11}C-PIB PET scans from Alzheimer's disease neuroimaging initiative (ADNI) participants. The classification produced in first iteration is used as "prior knowledge" to generate both weighted and unweighted ensemble of different classifiers.

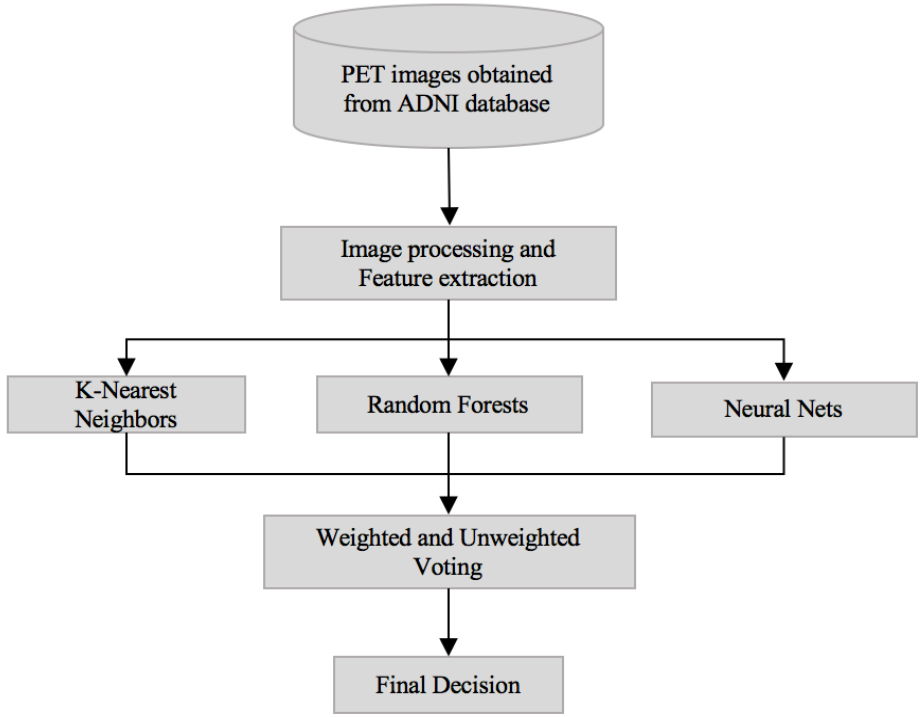


Figure 3. Proposed Method Overview

II. METHODOLOGY

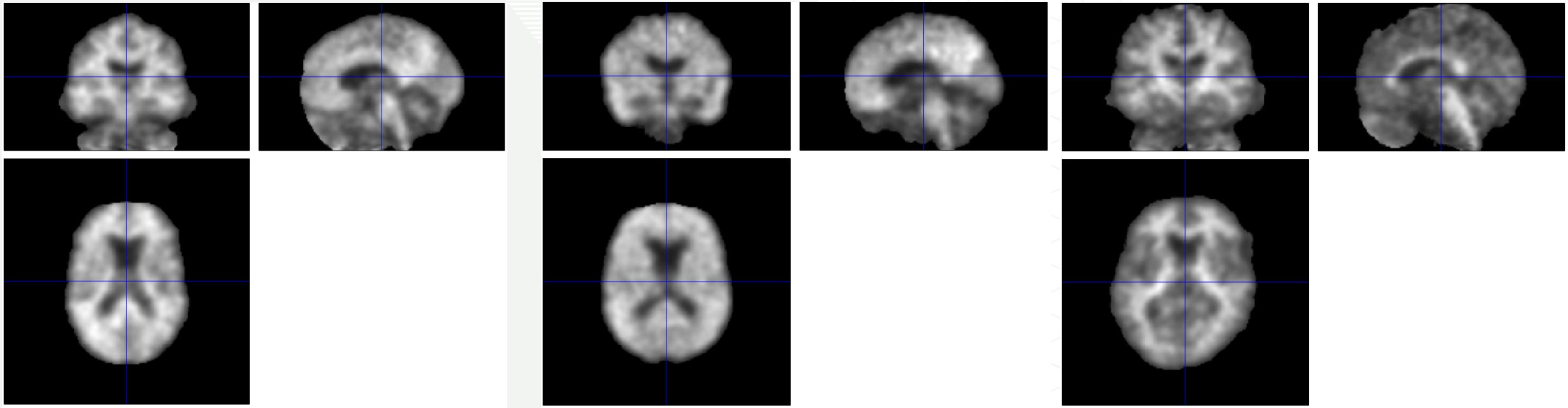
Data

- Data collection and sharing for this study was supported by the Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Inclusion/Exclusion Criteria
  - Most preprocessed 228 ^{11}C-PIB PET image volumes.
  - Only 208 out of 228 PET scans, which have sufficient quality to provide us with the listed eight brain areas through the processing pipelines, were utilized for further analysis.

Dataset

	Number of Subjects
Control	47
Mild Cognitive Impairment (MCI)	99
Alzheimer's Disease (AD)	62

TABLE I. CLASS DISTRIBUTION OF THE DATASET



Alzheimer's Disease (AD)

Mild Cognitive Impairment (MCI)

Control (CN)

Figure 4. Comparison of PET images of Alzheimer's Disease, Mild Cognitive Impairment and Control.

Image Processing

- Image Processing
  - All steps performed using FMRIB Software Library Toolbox (Oxford University, UK) [8]
  - Skull Stripping
  - Registering to standard space
  - Tissue segmentation
    - Gray Matter(GW), White Matter(WM), Cerebrospinal fluid(CF)
  - Volume segmentation

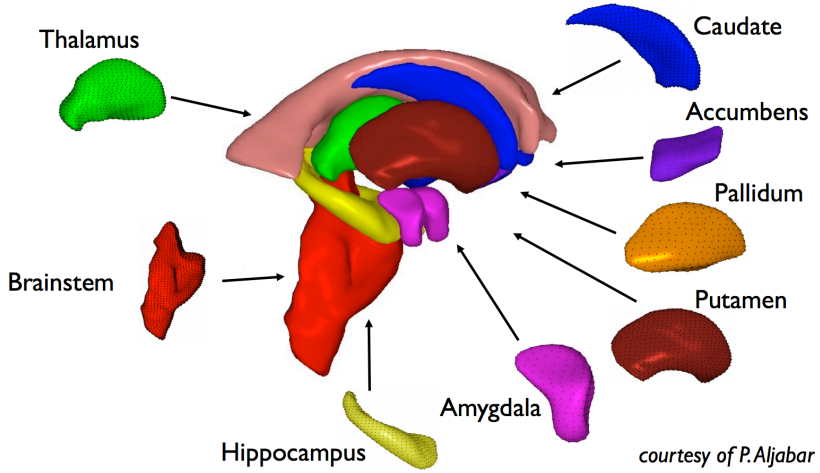


Figure 5. Volume Segmentation 3D Overview.

Figure 6. Schematic diagram of image processing

Feature Extraction

- Extracted Features
  - Volume
    - CSF, GW, MW, 8 brain areas
  - Voxel intensities
    - max, min, mean, median, mode, standard deviation, sum
    - within CSF, GW, MW and atlas volumes
  - Texture
    - 13 Haralick texture features [15]
    - Energy and Entropy [15]

Feature Selection for classification methods

- Minimum redundancy maximum relevance feature selection method (mRMR) was used to minimize redundancy and select features according to measures of relevance and dependence [16].
- The highest ranked 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).

Baseline Classification

- Baseline Method:
  - Neural nets
  - Random Forest
  - KNN
- Hyper parameter selection
  - hyper parameters of different classifiers were optimized using grid search in 10-fold CV
    - hidden layers of neural nets
    - the number of nearest neighbors of kNN
    - the number of single decision trees in the RF

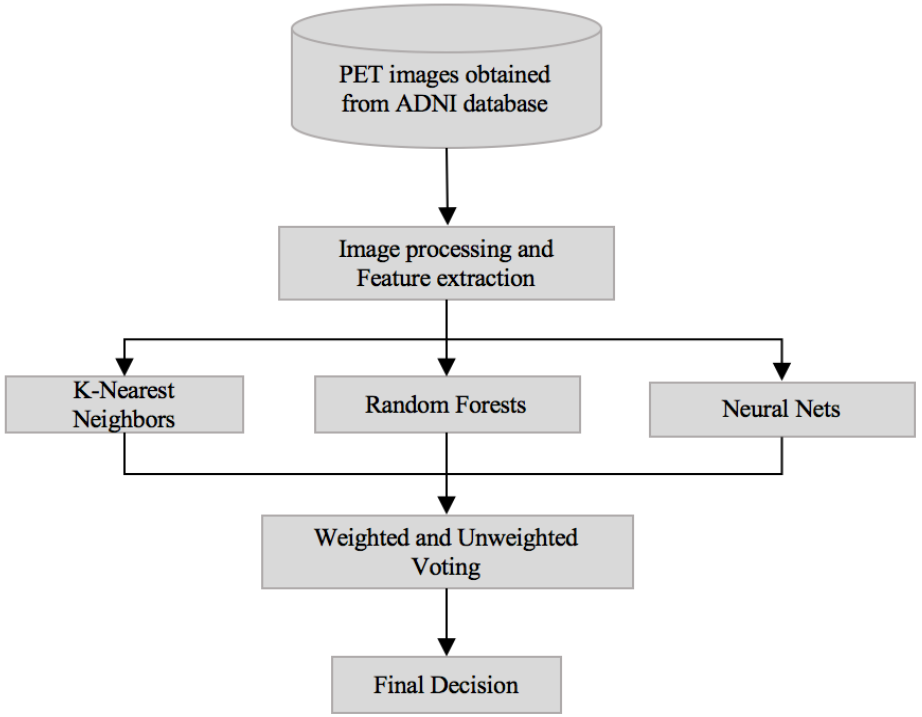


Figure 6. Proposed Method Overview

Ensemble Classification

- **<u>Unweighted voting schemes</u>**: the average of classification score (Predicted class posterior probabilities) is the new classification score.
- In **<u>Weighted voting schemes:</u>** the weight of each classifier is calculated using its resultant accuracy from cross-validation:

$$\text{weight} = \log\left(\frac{accuracy}{1 - accuracy}\right) \qquad (1)$$

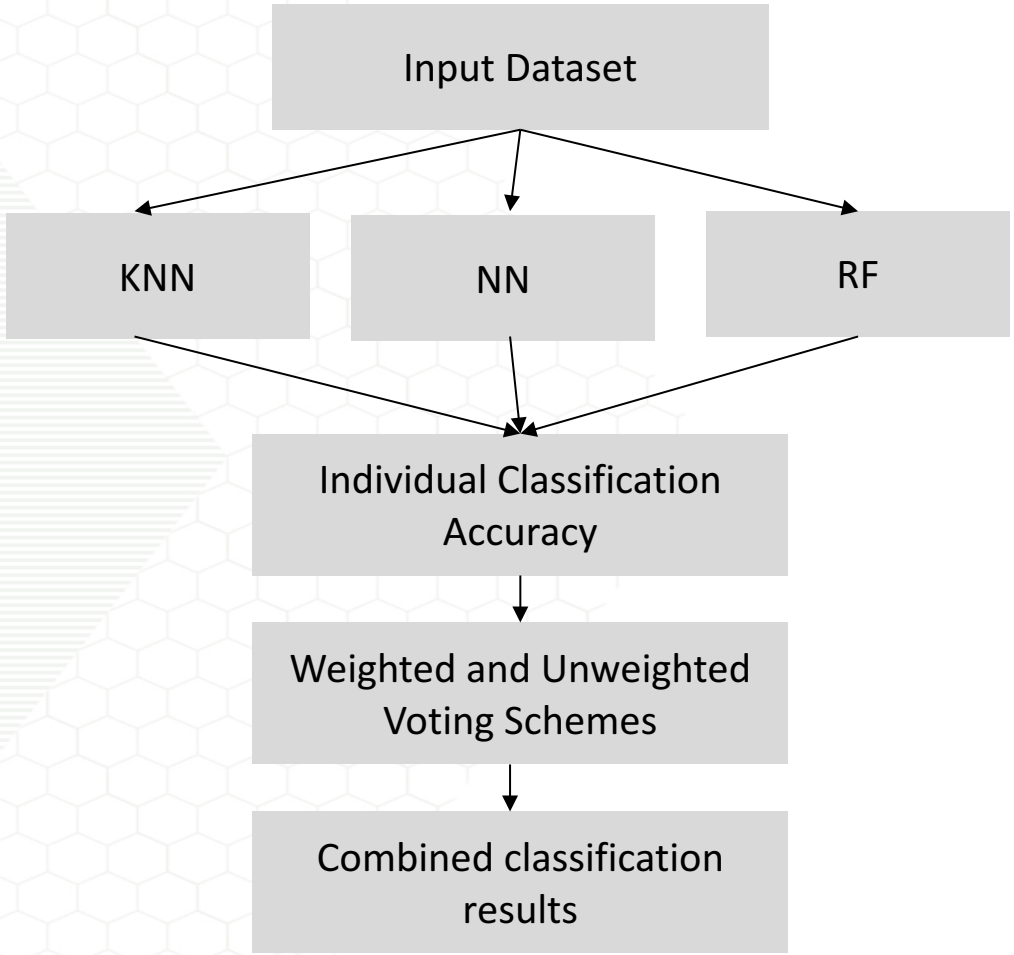


Figure 7. Ensemble Classification Scheme

Evaluation

- Dataset: 20% testing data, 80% training data.
- Final model
  - Hyper-parameters and the number of features selected were optimized using 10-fold CV on training data.
  - Trained on training data with optimized hyper-parameters
  - Evaluated on 20% testing data.
- Performance Metrics (mean ± standard deviation)
  - Pearson's Correlation Coefficients (PCC)
  - Accuracy
  - Specificity

III. RESULT

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**

TABLE II. CLASSIFICATION RESULTS.

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), Alzheimer's (AD)disease,
c. Pearson's Correlation Coefficient(PCC)
d. Cross Validation (CV)

Classification Results

- Best CV performance: Weighted Ensemble Classifier
  - Accuracy: 86.1% ± 8.34%;
  - Specificity: 90.6% ± 12.9%;
- Best test performance: Weighted Ensemble Classifier
  - Accuracy: 80.9%;
  - Specificity: 85.76%
- Individual Methods
  - RF > NN > KNN
  - Advantage of high bias classifiers
- Ensemble Methods
  - Weighted voting > Unweighted voting
  - Weighted voting could lead to biased performance

Feature Analysis

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 III. TOP FEATURES RANKED BY MRMR FEATURE REDUCTION METHOD.

- Haralick texture features are ranked the highest
  - correlation features [15], which measures the gray tone linear-dependencies and information measure of correlation [15], are the most important.

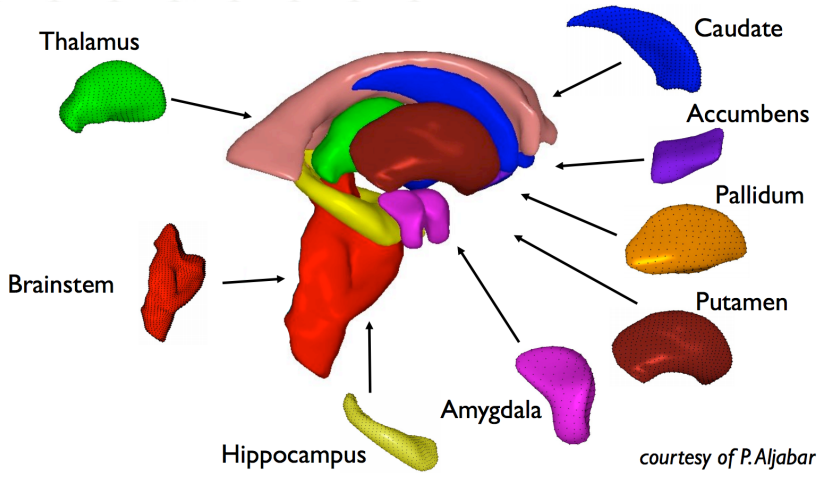


Figure 10. Volume Segmentation 3D Overview.

IV. Conclusion & Future work

- We showed that the ensemble method with weighted voting scheme 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%.
- Currently only addressed 11C- PIB PET image datasets.
- In the future:
  - Compare the performance of proposed methods on different PET imaging datasets.
  - Develop ensemble methods that can integrate PET imaging datasets from different PET imaging tracers such as Florbetapir, FDG and 11C- PIB.

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.
3) J. Ramírez, J. M. Górriz, D. Salas-Gonzalez, A. Romero, M. López, I. Álvarez, and M. Gómez-Río, "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) S. F. Eskildsen, P. Coupé, V. S. Fonov, J. C. Pruessner, and D. L. Collins, "Structural imaging biomarkers of Alzheimer's disease: predicting disease progression," *Neurobiol. Aging*, vol. 36, pp. S23–S31, 2015.
5) I. A. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. M. López, F. Segovia, R. Chaves, M. Gómez-Rio, and C. G. Puntonet, "18F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis," *Inf. Sci. (Ny).*, vol. 181, no. 4, pp. 903–916, 2011.
6) L. Rokach, "Ensemble-based classifiers," Artif Intell Rev, vol. 33, pp. 1–39, 2010.
7) C. Cabral, M. Silveira, 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) M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *Neuroimage*, vol. 62, no. 2, pp. 782–790, Aug. 2012.
9) V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, D. L. Collins, and Brain Development Cooperative Group, "Unbiased average age-appropriate atlases for pediatric studies.," *Neuroimage*, vol. 54, no. 1, pp. 313–27, Jan. 2011.
10) T. L. S. Benzinger, T. Blazey, C. R. Jack, R. A. Koeppe, Y. Su, C. Xiong, *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.
11) K. Farid, O. Almkvist, K. Brueggen, S. F. Carter, A. Wall, K. Herholz, and A. Nordberg, "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, Jul. 2014.
12) J. Koivunen, A. Verkkoniemi, S. Aalto, A. Paetau, J.-P. Ahonen, M. Viitanen, *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.
13) L. G. Apostolova, K. S. Hwang, J. P. Andrawis, A. E. Green, S. Babakchanian, J. H. Morra, *et al*. "3D PIB and CSF biomarker associations with hippocampal atrophy in ADNI subjects." Neurobiology of aging 31.8 (2010): 1284-1303.
14) Y. Zhang, M. Brady, and S. Smith, "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.
15) R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, 1979.
16) Hanchuan Peng, Fuhui Long, and C. Ding, "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.

CONTACT INFO

Wenjun Wu
+1 404 838-8084 (US)
wuwenjun@gatech.edu/wenjunw@uw.edu

Classification Results: Cross Validation (CV)

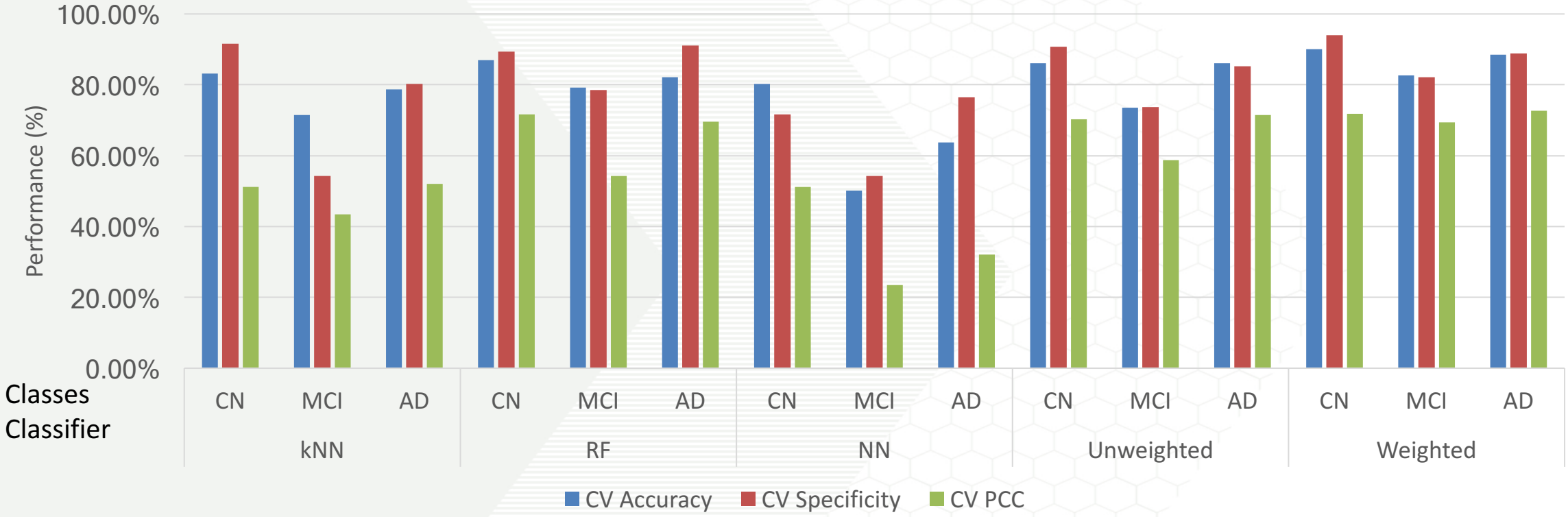


Figure 8. Cross Validation Result Overview. k-Nearest Neighbors (kNN), Random Forests (RF), Neural Nets (NN), Ensemble Model with Unweighted Voting Scheme (unweighted), Ensemble Model with Weighted Voting Scheme (weighted); Control(CN), Mild Cognitive Impairment (MCI), Alzheimer's (AD)disease; Pearson's Correlation Coefficient(PCC)

Classification Results: Test

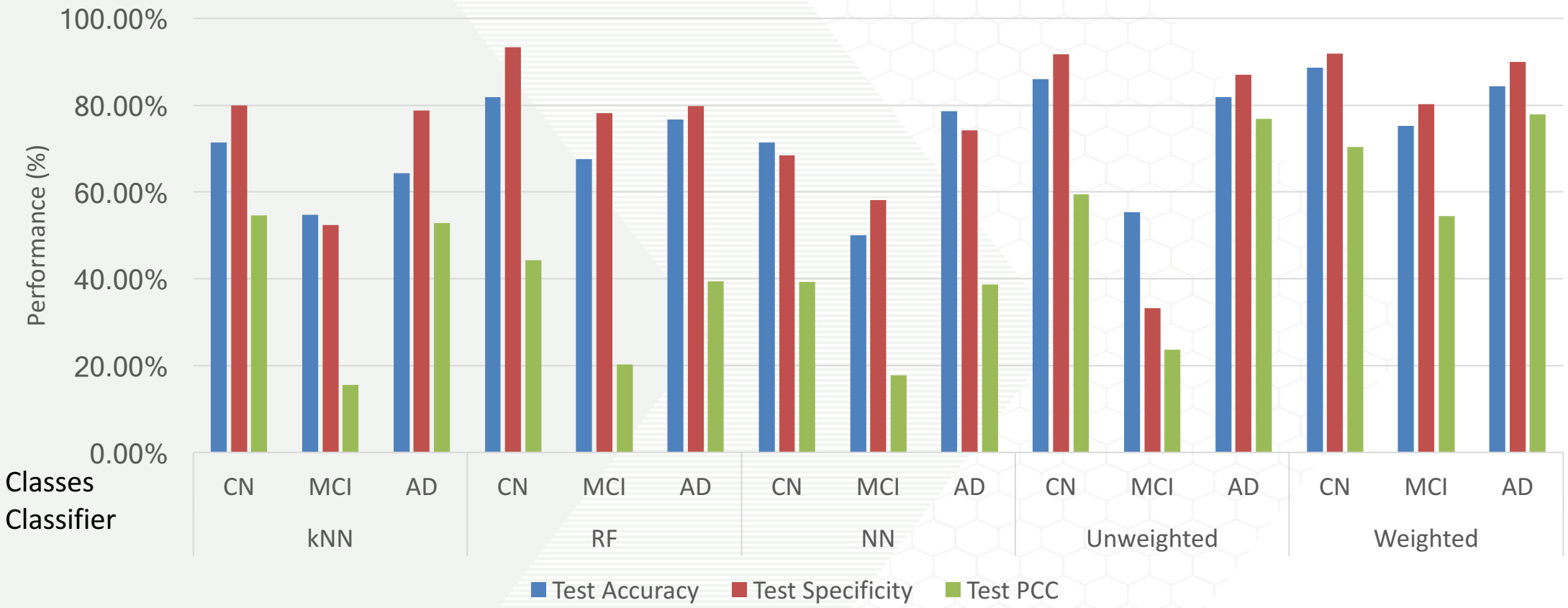


Figure 9. Test Result Overview. k-Nearest Neighbors (kNN), Random Forests (RF), Neural Nets (NN), Ensemble Model with Unweighted Voting Scheme (unweighted), Ensemble Model with Weighted Voting Scheme (weighted); Control (CN), Mild Cognitive Impairment (MCI), Alzheimer's (AD)disease; Pearson's Correlation Coefficient(PCC)