Journal of Neuroscience Methods 333 (2020) 108544

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

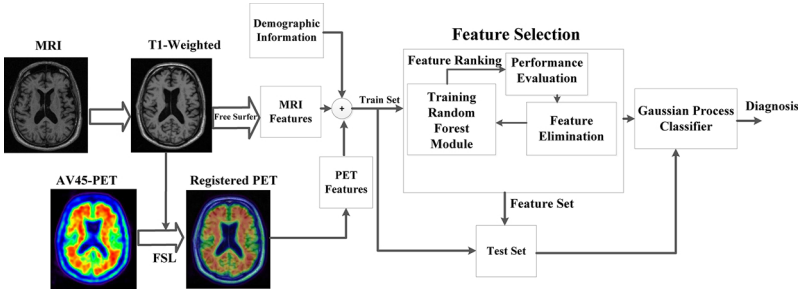
journal homepage: www.elsevier.com/locate/jneumeth

A Gaussian-based model for early detection of mild cognitive impairment using multimodal neuroimaging

Parisa Forouzannezhad[a], Alireza Abbaspour[a], Chunfei Li[a], Chen Fang[a], Ulyana Williams[a], Mercedes Cabrerizo[a], Armando Barreto[a], Jean Andrian[a], Naphtali Rishe[b], Rosie E. Curiel[c,d], David Loewenstein[c,d,e], Ranjan Duara[d,e], Malek Adjouadi[a,d]

[a] Department of Electrical and Computer Engineering, Florida International University, Miami, FL, USA
[b] School of Computing and Information Sciences, Florida International University, Miami, FL, USA
[c] Center for Cognitive Neuroscience and Aging, Department of Psychiatry and Behavioral Sciences, University of Miami School of Medicine, Miami, FL, USA
[d] 1Florida Alzheimer's Disease Research Center (ADRC), University of Florida, Gainesville, FL, USA
[e] Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, FL, USA

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Multimodal Neuroimaging
Alzheimer's disease
Early Mild Cognitive Impairment (EMCI)
Random Forest
Gaussian Process

ABSTRACT

Background: Diagnosis of early mild cognitive impairment (EMCI) as a prodromal stage of Alzheimer's disease (AD) with its delineation from the cognitively normal (CN) group remains a challenging but essential step for the planning of early treatment. Although several studies have focused on the MCI diagnosis, this study introduces the early stage of MCI to assess more thoroughly the earliest signs of disease manifestation and progression.
New method: We used random forest feature selection model with a Gaussian-based algorithm to perform method evaluation. This integrated method serves to define multivariate normal distributions in order to classify different stages of AD, with the focus placed on detecting EMCI subjects in the most challenging classification of CN vs. EMCI.
Results: Using 896 participants classified into the four categories of CN, EMCI, late mild cognitive impairment (LMCI) and AD, the results show that the EMCI group can be delineated from the CN group with a relatively high accuracy of 78.8% and sensitivity of 81.3%.
Comparison with existing method(s): The feature selection model and classifier are compared with some other prominent algorithms. Although higher accuracy has been achieved using the Gaussian process (GP) model (78.8%) over the SVM classifier (75.6%) for CN vs. EMCI classification, with 0.05 being the cutoff for significance, and based on student's t-test, it was determined that the differences for accuracy, sensitivity, specificity between the GP method and support vector machine (SVM) are not statistically significant.
Conclusion: Addressing the challenging classification of CN vs. EMCI provides useful information to help clinicians and researchers determine essential measures that can help in the early detection of AD.

https://doi.org/10.1016/j.jneumeth.2019.108544
Received 17 November 2018; Received in revised form 26 November 2019; Accepted 5 December 2019
Available online 12 December 2019
0165-0270/ © 2019 Published by Elsevier B.V.

1. Introduction

Alzheimer's disease (AD) is one of the most common neurodegenerative disorders among the elderly population. It is progressive, irreversible in nature and is considered the main cause of dementia. Therefore, detecting the earliest manifestations of AD is critical for the timely planning of treatment for this healthcare challenge. AD has become a world health problem affecting developed and developing nations alike, and the number of diagnosed AD patients is unfortunately increasing rather dramatically as both the life span of humans and the earth's population continue to increase. According to the 2018 report by the Alzheimer's Association, there are 5.7 million Americans living with Alzheimer's disease and by 2050 this number may reach 13.8 million (Association et al., 2018). Given the severity of the disease observed through the structural and functional measures as extracted through the magnetic resonance imaging (MRI) and positron emission tomography (PET) imaging modalities, most studies confirm classification accuracy of 90% and higher in delineating AD from the cognitively normal (CN) group; however, the chance for treatment planning and therapeutic intervention at a late stage of the disease may be too late given its irreversible nature. Hence to improve the chances for slowing down disease progression, any classification algorithm must consider classifying the disease in the earliest stage possible. Previous studies proved that between 10 and 15% of mild cognitive impairment (MCI) patients develop into AD annually (Petersen et al., 2009), while it is only 1 to 2% for the CN elderly people (Bischkopf et al., 2002). Although there is no certain treatment for AD, several types of medication are shown to postpone the onset of symptoms and slow the progression of AD from its early prodromal phase.

Diagnosis of AD is mostly based on the clinical history and on neuropsychological test scores such as Mini-mental state examination (MMSE) and clinical dementia rating (CDR). However, diagnosis based on neuroimaging have been widely increased due to the uncertainty and variability of such neuropsychological tests. These neuroimaging techniques are non-invasive and provide valuable information for both clinical and research purposes. MRI (Westman et al., 2011; Liu et al., 2013a), PET (Foster et al., 2007; Singh et al., 2017) and functional magnetic resonance imaging (fMRI) (Forouzannezhad et al., 2018b; Zhang et al., 2015; Wee et al., 2016) are the most common types of neuroimaging. Moreover, researchers use other biomarkers such as apolipoprotein (APOE) $\varepsilon 4$ allele (Li et al., 2017, 2017b), Amyloid β ($A\beta$) (Bateman et al., 2012; Clark et al., 2016), and combination of biomarkers (Liu et al., 2015; Westman et al., 2012; Yuan et al., 2012; Ithapu et al., 2015; Duara et al., 2013) to predict AD and its different conversion phases.

However, the high dimensionality of these neuroimaging datasets compounded with the low number of multimodal neuroimaging samples available make the analysis of these types of data quite challenging. Patterns of neuronal cell death, at least in the early stages of the disease, may not necessarily reflect the anatomical or functional abnormalities in the different regions of the brain. Therefore, the analysis should not only carefully scrutinize the different brain regions, with added focus on regions that are known to be disease prone but to also look at all potential biomarkers that are best suited to be combined in an optimal fashion to detect these subtle changes. In order to overcome such a problem, machine learning techniques were introduced to analyze medical imaging data (Donnelly-Kehoe et al., 2018; Liu et al., 2016b, 2015, 2014a; Ye et al., 2011). Indeed, machine learning algorithms try to find a low dimensional representation of the data which is embedded in the high dimensional space. Machine learning algorithms using discriminative features as MCI or AD biomarkers provide powerful models for computer-aided diagnosis (Davatzikos et al., 2011; Eskildsen et al., 2013).

Recently, several machine learning techniques have been proposed for the prediction and classification of AD and its prodromal stages; among them which are viewed as the most accurate and most applicable approaches are deep neural network (DNN), support vector machine (SVM), and Bayesian network (BN). SVM is one of the most popular supervised machine learning model with associated learning algorithms to analyze the data applied for classification (De Martino et al., 2008; Forouzannezhad et al., 2018a; Khedher et al., 2015; Dukart et al., 2013; Adjouadi et al., 2005; Nir et al., 2015). Through a set of training examples, the SVM algorithm makes a model that is capable of assigning new test data to one of the predefined classes. Bayesian prediction and classification models are another type of machine learning based on the Bayes theorem with the assumption of strong independence among the classification features (López et al., 2009; Liu et al., 2013b; Seixas et al., 2014; Plant et al., 2010). Deep neural networks and deep learning approach are able to analyze the high dimensional data such as MRI which has been widely used in recent decades (Yao et al., 2018; Brosch et al., 2013; Liu et al., 2014, 2015; Suk et al., 2015; Bengio et al., 2013; Amoroso et al., 2018; Forouzannezhad et al., 2018c).

Several investigations are reported on the specific diagnosis of MCI (Sørensen et al., 2018; Salvatore and Castiglioni, 2018; Yao et al., 2018). Sørensen et al. applied ensembles SVM with radial basis function (RBF) and linear kernels on MRI data (Sørensen et al., 2018). Yao et al. used relative importance as feature selection algorithm hierarchical grouping to do multi-way classification on T1-weighted MRI data (Yao et al., 2018). Salvatore et al. applied a combination of t-test and feature ranking as feature reduction technique while using SVM as the classifier (Salvatore and Castiglioni, 2018). Although such approaches are commendable and have their own merit in shedding light on the progression of the disease, the research community understands that for a complex disease like AD, the disease may have started for over a decade prior to any noticeable physical symptoms (Sperling et al., 2011; Buckner, 2004; Schmidt et al., 2002; Cummings et al., 1998). The complexity of this challenge in delineating the MCI group from the CN group is reflected in the type of classification results, often not exceeding the lower end of the 80% range, that several studies have endeavored to resolve. These attempts, regardless of the multimodal imaging approach and the integration of the different biomarkers along with genetic and other demographic factors could not bring forth any new measures that could potentially augment these classification results.

It is thus imperative to include the early mild cognitive impairment (EMCI) group in any classification study that is bound to assess the different progression phases of the disease with the intent to diagnose the disease in the earliest stage possible. Such a task demands examination all types of measures, structural, functional or metabolic, neuropsychological, demographic, and genetic to assess which measures characterize best progression from normal control to this early stage of mild cognitive impairment. There are current research efforts that attempt to establish new neuropsychological tests and new imaging techniques that could even precede the EMCI phase (Duara et al., 2011; Loewenstein et al., 2017, 2018; Curiel et al., 2018). Since EMCI can be considered as an earlier state of mild cognitive impairment, it is of great significance to detect this stage for potential early treatment planning and for designing subject-specific early curative/therapeutic intervention protocols. The EMCI stage has shown a milder degree of cognitive impairment as compared to the MCI group, making this phase of the disease more amenable to treatment or to a potential preventive process that will decelerate its progression and provide a longer and better quality of life for these patients; recall the aforementioned percentage of MCI patients that decline to AD.

In this study, a feature selection algorithm is intoduced to find the most relevant features and a probabilistic approach is developed for optimized classification results. Our focus is on the delineation of the EMCI group from the CN group due to the aforementioned importance of early detection, while most studies focused on the MCI diagnosis, combining both EMCI and late mild cognitive impairment (LMCI) groups. We propose a feature selection based on ranking the most

important features to help clinicians determine the essential features in classifying the EMCI group using a large number of subjects in the 4 groups considered.

The rest of this work is organized as follows. Section 2 provides the materials and method section, where we indicate the participants of the study and the image processing related to the MRI and PET modalities. Section 3 describes the feature selection process. Section 4 defines the Gaussian process and shows how the Gaussian-based model was derived. Section 5 shows the results obtained using the proposed integrated method. Section 6 provides a discussion and overall analysis of the results obtained. Finally, a retrospective on the results obtained using the proposed method and final remarks are provided in Section 7.

2. Materials and Methods

2.1. Participants and Data Acquisition

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu)[1] . The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Identification of biological markers at early stage of AD will help researchers and clinicians to plan for early treatment and therapeutic interventions. The EMCI subject inclusion criteria are as follows: MMSE scores between 24-30, CDR of 0.5, objective memory loss of 0.5-1.5 SD (standard deviation) below normal according to the education adjusted scores on delayed recall of one paragraph from Wechsler Memory Scale Logical Memory II, memory complaints, absence of significant level of impairment in other cognitive domain, absence of dementia, and essentially preserved activities of daily living. LMCI criteria are almost similar to the EMCI except for the memory loss scores by Wechsler Memory Scale Logical Memory II which is set at more than 1.5 SD below normal.

In this study, a total of 896 participants classified into the four categories of CN (248), AD (159), EMCI (296), and LMCI (193). All subjects had an MRI and a Florbetapir (18F-AV-45) amyloid PET scan within a 6-month time window. Table 1 illustrates the demographic information of the subjects considered.

2.2. MRI Processing

MRI images are 3T T1-weighted using a 3D sagittal volumetric magnetization prepared rapid gradient echo (MP-RAGE) sequence. Here, the following are assumed: repetition time is 2,300 *ms*, minimum full echo time, inversion time is 900 *ms* with a 256 × 256 × 170 acquisition matrix providing a voxel size of 1.0 × 1.0 × 1.2 mm^3. In this study the T1-weighted standard MNI 305 space MRI has been used to do feature extraction using FreeSurfer version 5.3. The T1-weighted image was used as the reference image in the registration process, which included skull-striping, segmentation, and demarcation of the different brain regions.

We extracted 3 measurements that included mean intensity, volume, and intensity standard deviation after dividing the image into 45 subcortical regions. In addition, 9 morphological variables including: gray matter volume, rectified mean curvature, folding index, surface area, intrinsic curvature index, average thickness, rectified Gaussian curvature index, white matter volume, and thickness standard deviation for 68 cortical regions were generated. Furthermore, the estimated total intracranial volume (eTIV) is calculated, which is used later for normalization of the volumetric measures (Sargolzaei et al., 2015).

Table 1
Demographic information of the participants.

	CN	EMCI	LMCI	AD
Number of subjects	248	296	193	159
Male/Female	123/125	164/132	109/84	92/67
Age-PET(year)	75.7(6.5)[a]	71.5(7.4)	73.8(8.1)	74.9(7.8)
Age-MRI (year)	75.2(6.5)	71.3(7.4)	73.6(8.1)	74.7(7.8)
Years of Education	16.4(2.5)	15.9(2.6)	16.2(2.7)	15.7(2.7)

[a] Values are represented as mean (standard deviation) for all continuous attributes.

2.3. AV-45 PET Processing

PET images used in this study were acquired of 370 MBq (10 mCi), dynamic 3D scan of four 5-minute frames from 50 to 70 minutes post-injection, co-registered, averaged, reoriented into a standard 160 × 160 × 96 voxel image grid with 1.5 *mm* cube voxels, and smoothed to a uniform isotropic resolution of 8 *mm* full width of maximum. In order to acquire the standard uptake value ratio (SUVR) for each participant's amyloid accumulation, first the AV-45 PET scan was linearly registered onto the T1-weighted image using FSL (FMRIB Software Library) (Jenkinson et al., 2012) with 12 degrees of freedom (DoF) as depicted in Fig. 1.

Registration phase plays an important role to get as much information as possible from PET scan due to the low resolution of this neuroimaging technique. This registration process provides the same MRI parcellation and segmentation for the AV-45 PET image. Then the mean intensity of each FreeSurfer region (ROI) for the 45 subcortical and the 68 cortical regions were assessed, which together identify the standardized uptake values (SUVs). These values extracted from each region were obtained by volume-weighted means as in equation (1).

$$S_n = \frac{\sum_{j=1}^{n} I_j}{M} \tag{1}$$

where S_n is the mean SUV of region n, with M defining the number of voxels found in region n, and I_j is the intensity of voxel j in the AV-45 PET. The whole cerebellum SUV including 4 subcortical regions (right/left cerebellum and right/left white matter) and global uptake value consisting of 68 cortical ROIs (34 ROIs for each hemisphere) were calculated as follows:

$$S_{\mathrm{CR}} = \frac{S_{R_1} \times V_{R_1} + S_{R_2} \times V_{R_2} + \ldots + S_{R_k} \times V_{R_k}}{V_{R_1} + V_{R_2} + \ldots + V_{R_k}} \tag{2}$$

where the S_{CR} represents the SUV of combined 4 cerebellum and 68 cortical regions, S_{R_i} defines the SUV corresponding to region ROI_i and V_{R_i} is a measure of the volume of ROI_i. In the end, all the computed global SUVs were normalized by the whole cerebellum SUV as the reference to compute standard uptake value ratio (SUVR) as expressed in equation (3) (Lizarraga et al., 2018).

$$\mathrm{SUVR} = \frac{\text{Global SUV}}{\text{Reference SUV}} \tag{3}$$

3. Feature Selection Model

The main challenge in analyzing high-dimensional data is the existence of a very large number of features or variables that may not all

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

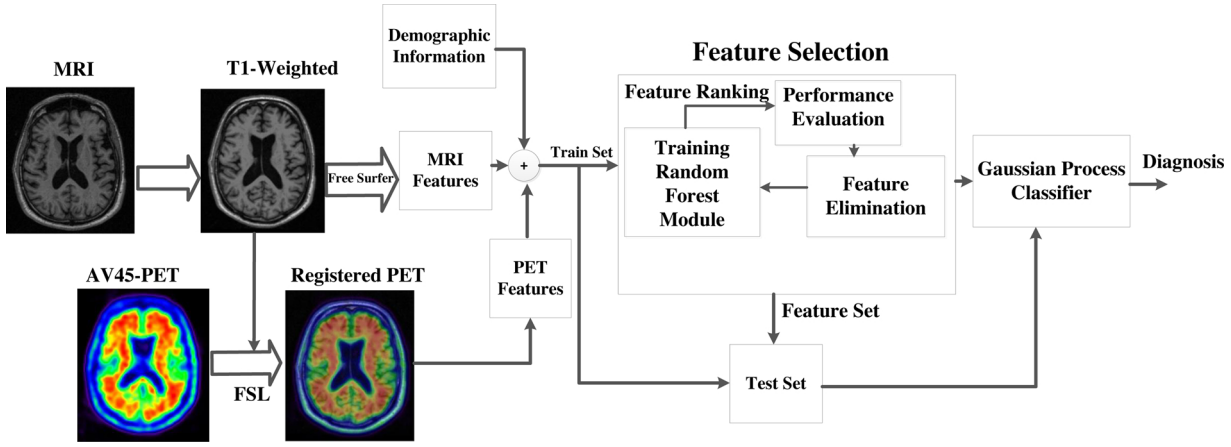


Fig. 1. Neuroimaging feature extraction and overall diagram of the proposed method.

be associated with the disease and could be contributing differently at any given phase of the disease. This problem is made even more difficult when the database suffers from a low sample size for multimodal neuroimaging. Hence, dimensionality reduction techniques or selection of prominent features can play an essential role in machine learning when seeking optimal classification results (Peng et al., 2005; Ota et al., 2015). In addition, ranking of these prominent or most relevant features can be appraised whenever the classification results in delineating challenging groups have been optimized, especially when the most subtle of changes differentiate them (like in CN vs. EMCI); in other words, these features are deemed relevant only because they do indeed reflect these subtle changes albeit at varying degrees. An optimal decision-making process need to be established when applying the feature (dimension) reduction techniques to guarantee that the relevant features are maintained. Although there is is a small probability in removing some relevant features during feature reduction (Mwangi et al., 2014); however, several studies used different feature selection techniques to find the most relevant features successfully, especially in kernel-based techniques such as SVM (Chu et al., 2012; Cui et al., 2011).

In this study, first we divided the data into 80% training and 20% testing data and then applied the feature selection process only on the training data and assessed the model on the 20% remaining test data. Due to the fact that random forest (RF) method is time-consuming, we applied ANOVA on the whole training data for each pairwise group separately to reduce the dimensionality considering a P-value of 0.05. Then, we used the RF model to obtain the most important features. Random forest is a tree-based approach, which facilitates multimodal imaging classification by deriving the similarity measures (Gray et al., 2013). The RF model combines re-sampling and random feature selection to construct the trees for both classification and regression purposes. On the other hand, RF methodology can provide the hierarchical importance of the different features using statistical permutation and Gini impurity index (Menze et al., 2009). The Gini importance score is a measure of variable relevance based on impurity reduction. The Gini impurity, $G(n)$, can be calculated as follows

$$G(j) = 1 - f_{+1}^2 - f_{-1}^2 \tag{4}$$

where $f_n = \frac{k_n}{k}$ defines the ratio of the k_n samples from the binary class out of the total samples k at the specific node of j. Then, the reduction of Gini impurity, $\Delta G(j)$ resulting from splitting the samples to j_l and j_r sub-nodes are then calculated as in (5)

$$\Delta G(j) = G(j) - f_l G(j_l) - f_r G(j_r) \tag{5}$$

where l and r subscripts specify the left and right sub-nodes at each sample splitting. The fraction of data points for the left and right subsets f_l and f_r are defined as $f_l = \frac{k_l}{k}$ and $f_r = \frac{k_r}{k}$. Finally, the Gini Index as an indicator for the selection of feature F is calculated based on aggregating the impurity reduction $\Delta G(j, T)$ for the nodes j and trees T as given in (6)

$$\text{Index}(F) = \sum_T \sum_j \Delta G_F(j, T) \tag{6}$$

The random forest feature elimination approach is implemented as in the given steps below. First, the algorithm is applied using all variables including age, education, structural and functional variables extracted from the MRI and PET imaging modalities. If we consider p as a sequence of probable number of variables to retain ($p_1 > p_2 > . ..$), at each iteration the variables are ranked based on the explained below process and the top ranked variables p_j are maintained. The performance of the model is evaluated and the number of variables is determined. Then using the 10-fold cross-validation resampling approach, the above process is encapsulated in one iteration of resampling. This process is repeated for every iteration of 10-fold cross-validation. Then the subsets with the highest accuracy for each iteration were selected and gathered in a pool. In the next step, the features were selected based on the most votes obtained. The RF recursive feature elimination (RF-RFE) model is demonstrated as below:

Random Forest Feature Selection

1 for each iteration of 10-fold cross validation
- Partition the train data into training and testing sets
- Train the random forest model using the training set
- Assess the performance of the model on the testing set
- Rank the variables according to their importance
- for each subset of variable numbers p_j for $j = 1, 2, .. ., p$
  - Retain the p_j highest ranked variables
  - Train the RF model on the training set using p_j variables
  - Assess the performance of the model
  - Recalculate the importance of each variable
  - Repeat the process since no variables can be removed
- end

2 end
3 Choose the subsets with the highest accuracy in each iteration
4 Gather all selected subsets in a pool and select the variables based on the most votes obtained
5 Specify the number of variables
6 Fit the model

In the end, we applied correlation versus class labels to the data in order to prevent the algorithm from overfitting by training with too many features. The features with highest absolute value of correlation are selected based on the Pearson's correlation coefficient.

4. Gaussian Process

Gaussian process models (GPs) are a class of supervised machine learning based on Bayes theory for updating probabilities on the assumed hypotheses. Like SVM approaches, GPs are kernel based, which makes them efficient for high-dimensional data analysis. The Gaussian process as implemented here is a probabilistic approach that uses the average predictive probability instead of a single model. This probabilistic approach can be adapted to the classification problem by transforming the output using the appropriate activation function (Christopher, 2016). The primary goal for training data points x_i for N samples with an associated binary class labeled as $y_i \in \{-1, +1\}$, is to predict the class for which the new testing data points belong. The basic aim of GP classifiers is to predict the probabilities of $p(y|x)$ for the test input. With the following four steps described in 4.1 through 4.4, we illustrate how to calculate the GP prediction (Challis et al., 2015).

4.1. GP Likelihood

The first step in GP prediction is to define a likelihood for the prediction output. Here, for a binary classification, $y \in \{-1, 1\}$, the probability can be described by a conditional distribution as follows:

$$p(y|z) = (1 + e^{-yz})^{-1} = \sigma(yz) \tag{7}$$

where $\sigma(\bullet)$ is the logistic sigmoid function presented in Fig. 2. As it can be seen in this distribution, in each label, the probability can be controlled by the magnitude of z. The GP considers z as an unobservable variable. The data likelihood can be defined by taking the product over all the classes in the training data as expressed in (8)

$$p(y|z) = \prod_{i=1}^{N} p(y^i|z^i) = \prod_{i=1}^{N} \sigma(y^i z^i) \tag{8}$$

where N is the number of training data.

4.2. GP Prior

The importance of the specification of GP prior is because of its ability to fix the properties of the functions for the inference. In order to use the specification of GP prior, it is assumed that the unobservable variables, $[z^1, ..., z^N]^T = z \in R^N$, are distributed based on a GP prior, $g(z|0, K)$, where

$$P(z|X) = g(z|0, K) = \frac{\exp(-\frac{1}{2} z^T K^{-1} z)}{\{\det(2\pi K)\}^{1/2}} \tag{9}$$

Here, X is the training input which is the output of feature selection that may contain any of the features of MRI, PET, age, education, with g being the probability density function with zero mean vector and its covariance matrix, $K \in R^{N \times N}$, is symmetric positive-semidefinite with K^{-1} defining the inverse of the covariance matrix. In order to obtain the covariance function of the GP prior, we used Gaussian kernel also known as radial basis function (RBF) or square exponential (ES) that can be computed by equation (10) as follows (Wilson, 2014):

$$[K]_{m,n} = k(x_m, x_n) = \exp(-\frac{\|(x_m - x_n)\|^2}{2\sigma^2}) \tag{10}$$

where $[K]_{mn}$ denotes the element situated in row m and column n of the covariance matrix K, k is the covariance kernel, and x_m and x_n are the input vectors that can be represented by RBF kernel as feature vectors.

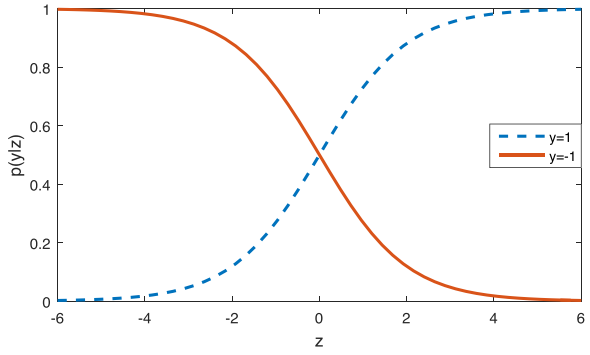


Fig. 2. The logistic sigmoid function used in the proposed GP learning is plotted based on the latent variable z and class labels of y.

4.3. Marginal likelihood

After obtaining the GP likelihood, $p(y|z)$, and the GP prior, $P(z|X) = g(z|0, K)$, the complete data likelihood can be defined as

$$p(y, z|X) = p(y|z)p(z|X) = g(z|0, K) \prod_{i=1}^{N} \sigma(y^i z^i) \tag{11}$$

Considering the fact that z is latent, and in order to obtain the marginal likelihood, equation (11) should be integrated with respect to z as expressed below:

$$p(y|X) = \int p(y, z|X)dz = \int g(z|0, K) \prod_{i=1}^{N} \sigma(y^i z^i)dz \tag{12}$$

The marginal likelihood value can thus be interpreted as the probability of observing the behavior of the training data based on the modeling assumptions.

4.4. Posterior distribution

Obtaining the marginal likelihood and the GP prior were the initial steps taken for obtaining the posterior distribution or the predictions for the test set, X'. This distribution is computed using the following equation:

$$p(\hat{y}'|X') = \int p(\hat{y}'|z')p(z'|z)p(z|X)dzdz' \tag{13}$$

where $p(\hat{y}'|z') = \sigma(\hat{y}'z')$ and the probabilities $p(z|X)$ and $p(z'|z)$ are as defined

$$p(z|X) = \frac{p(y, z|X)}{p(y|X)} \tag{14}$$

$$p(z'|z) = g(z'|\sigma' K^{-1} z, \sigma'' - \sigma'^T K^{-1} \sigma') \tag{15}$$

Here, $\sigma'' = K(X', X')$ and $\sigma' = K(X', X)$. In order to solve for equation (11) and equation (14), we used Laplace approximation for determining a Gaussian approximation (Williams and Barber, 1998). Having found the posterior distribution, $p(\hat{y}'|X')$, allows for making a prediction as well as for examining the accuracy of the model. In a binary classification, equation (16) is used to make a binary prediction $\hat{y}' \in \{-1, 1\}$ based on the posterior distribution as follows:

$$\hat{y}' = \begin{cases} +1 & \text{if } p(\hat{y}' = 1|X) > \eta, \\ -1 & \text{otherwise} \end{cases} \tag{16}$$

where $\eta \in (0, 1)$ is a coefficient used to compensate the class imbalances in the training data which is to equal to 0.5 here. For example, if we wish to make fewer false positive or false negative misclassifications, the threshold parameter, η, can be tuned in between 0 and 1.

5. Results

The feature selection process and the Gaussian model were implemented using R software (Team, 2017) to classify 6 binary groups of [CN vs. EMCI, CN vs. LMCI, CN vs. AD, EMCI vs. LMCI, EMCI vs. AD, and LMCI vs. AD]. Different metrics such as accuracy (ACC), sensitivity (SEN), and specificity (SPE) are determined to assess the performance

Table 2
Performance Comparison of the proposed method for individual modal and multi-feature classification. (ACC: accuracy, SEN: sensitivity, SPE: specificity, CN: cognitively normal, EMCI: early mild cognitive impairment, LMCI: late mild cognitive impairment, AD: Alzheimer's disease, MRI: magnetic resonance imaging, PET: positron emission tomography, DI: demographic information).

Modality	CN vs. EMCI			CN vs. LMCI			CN vs. AD			EMCI vs. LMCI			EMCI vs. AD			LMCI vs. AD		
	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE
MRI	75.9	77.9	75.5	62.1	48.1	77.5	83.6	80.6	85.7	72.1	80.3	61.9	85.6	88.8	87.4	62.3	55.3	73.0
PET	-[a]	-	-	76.1	66.4	85.8	90.0	90.3	89.8	62.8	61.5	64.3	69.1	80.3	45.7	69.7	76.0	61.3
MRI + PET	75.9	77.9	75.5	78.1	69.9	87.8	92.5	92.3	93.8	72.5	81.2	66.9	88.1	92.8	87.4	77.1	79.9	75.9
MRI + PET + DI[b]	78.8	81.3	76.8	79.8	70.2	89.9	94.7	92.3	95.9	73.2	81.2	69.9	88.1	92.8	87.4	77.1	79.9	75.9

[a] No PET features were selected by the RF-RFE process.
[b] DI: Demographic information consists of age and education.

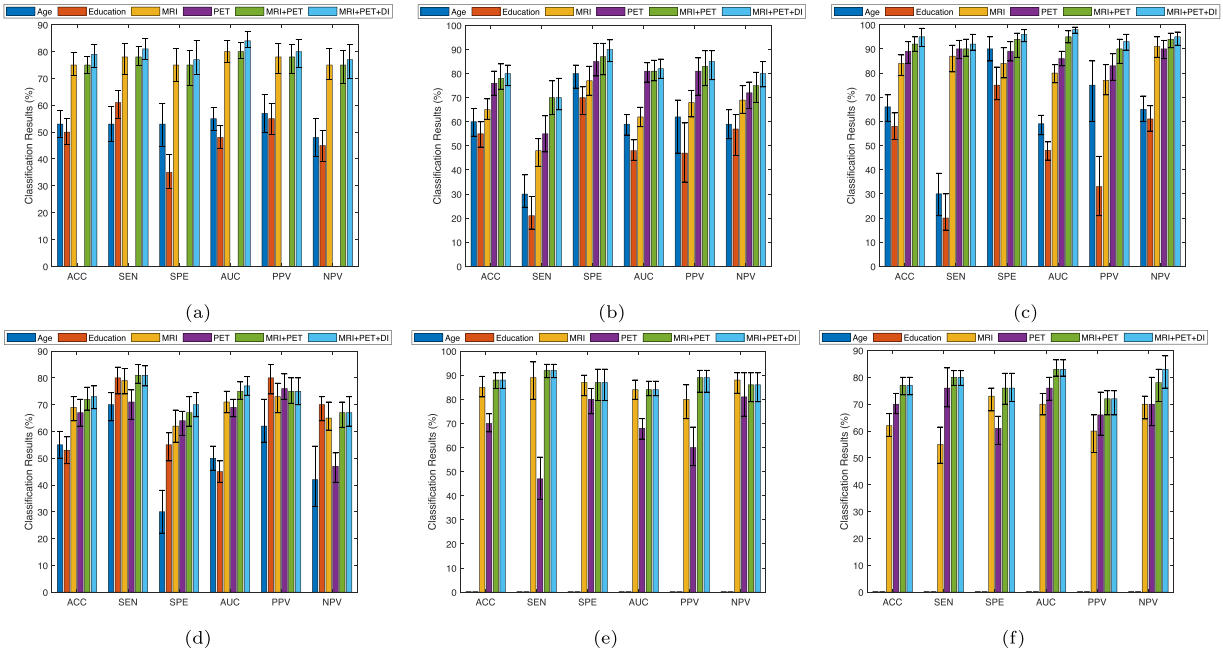


Fig. 3. Classification results with 95% confidence intervals for different modalities for the binary classifications of (a) CN vs. EMCI, (b) CN vs. LMCI, (c) CN vs. AD, (d) EMCI vs. LMCI, (e) EMCI vs. AD, and (f) LMCI vs. AD based on results of Table 2.

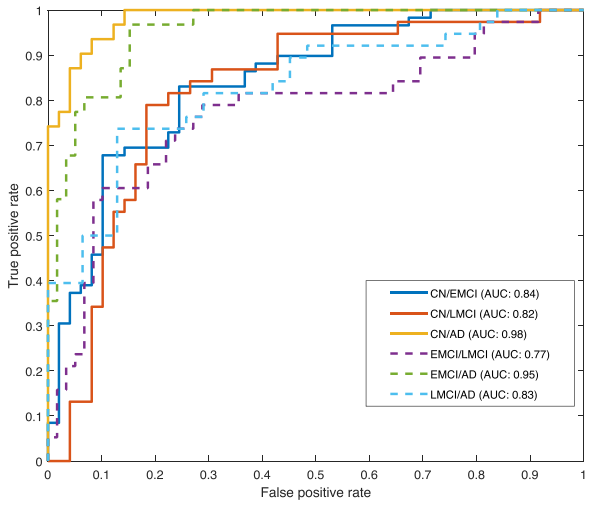


Fig. 4. Receiver Operating Characteristics (ROC) curve and Area Under the Curve (AUC) values for all binary classifications.

of the algorithm. Table 2 presents the classification results using the different imaging modalities when used separately and when combined. Here, MRI and PET features are selected based on the proposed feature selection algorithm. As can be seen from the results, for the most challenging CN vs. EMCI classification, an accuracy of 78.8%, a sensitivity of 81.4% and a specificity of 76.8% were obtained when combining MRI, PET and the demographic information. It should be noted that sensitivity is viewed as the most important metric among these parameters since it reflects the accuracy of diagnosing the true positive group in every binary classification. The EMCI (when considered) is assumed as true positive, $y = +1$, in every pairwise classification. In other pairs, except for CN vs. AD, LMCI is considered as the true positive and in the case of CN vs. AD classification, AD is assumed as the positive group.

In EMCI vs. CN classification, none of the PET features were selected by the proposed RF-RFE process, which may indicate that in this early stage SUVR measurements do not contribute to the classification results. This last assertion indicates that given the low resolution of PET, SUVR measurements are unable to extract the subtle changes that delineate the two groups of CN and EMCI. Except for the CN vs. EMCI classification, combining MRI and PET modalities enhanced all ACC, SEN and SPE results. Inclusion of *age* and *education level* to the multimodal (MRI and PET combination) imaging framework improved there results even further, albeit slightly in many cases. Interestingly, from these results we see that *age* and *education level* did not play any significant role in the classification of either EMCI and LMCI groups with AD. In addition, the sensitivity associated with EMCI detection is higher than specificity. However, the sensitivity in positive group of LMCI is

Table 3
The first eight most important features selected by the feature selection algorithm for all binary classification along with the P-value (lh: left hemisphere, rh: right hemisphere).

CN vs. EMCI	P-Value	CN vs. LMCI	P-Value	CN vs. AD	P-Value
age	$6.06e^{-5}$	lh superior frontal SUVR	$2.03e^{-14}$	lh entorhinal thickness	$1.3e^{-30}$
lh lateral ventricle volume	$3.66e^{-6}$	lh cortical SUVR	$1.47e^{-14}$	rh superior frontal SUVR	$9.49e^{-29}$
lh precuneus volume	$7.24e^{-5}$	lh frontal pole SUVR	$1.76e^{-12}$	lh cortical SUVR	$1.25e^{-29}$
lh superior parietal volume	$7.87e^{-6}$	rh cortical SUVR	$2.73e^{-12}$	rh medial orbitofrontal SUVR	$4.50e^{-27}$
lh superior frontal volume	$1.96e^{-8}$	rh superior frontal SUVR	$1.12e^{-14}$	lh rostral middle frontal SUVR	$6.27e^{-32}$
rh lingual volume	$1.46e^{-6}$	lh pars traingularis SUVR	$1.87e^{-14}$	rh frontal pole SUVR	$4.30e^{-30}$
rh lateral ventricle volume	$8.55e^{-5}$	lh middle temporal SUVR	$1.77e^{-12}$	lh caudal middle frontal SUVR	$1.69e^{-29}$
3rd ventricle volume	$2.31e^{-6}$	lh inferior parietal SUVR	$3.35e^{-11}$	lh accumbens SUVR	$1.44e^{-29}$
EMCI vs. LMCI	**P-Value**	**EMCI vs. AD**	**P-Value**	**LMCI vs. AD**	**P-Value**
lh precuneus volume	$1.45e^{-4}$	lh middle temporal volume	$2.50e^{-18}$	rh inferior temporal volume	$4.05e^{-6}$
lh superior frontal SUVR	$9.96e^{-6}$	lh middle temporal SUVR	$4.49e^{-17}$	rh lateral occipital SUVR	$2.70e^{-5}$
lh cortical SUVR	$4.63e^{-6}$	rh middle temporal SUVR	$1.10e^{-20}$	lh amygdala volume	$1.28e^{-7}$
lh hippocampus volume	$3.90e^{-8}$	lh entorhinal thickness	$3.02e^{-25}$	lh pericalcarine SUVR	$1.58e^{-6}$
lh amygdala volume	$1.26e^{-6}$	rh medial orbitofrontal SUVR	$7.92e^{-20}$	lh hippocampus volume	$3.22e^{-7}$
lh frontal pole SUVR	$1.69e^{-4}$	lh hippocampus volume	$1.07e^{-22}$	lh inferior temporal volume	$1.69e^{-5}$
lh parahippocampal volume	$2.82e^{-4}$	rh precauneus volume	$3.16e^{-14}$	rh pericalcarine SUVR	$9.80e^{-6}$

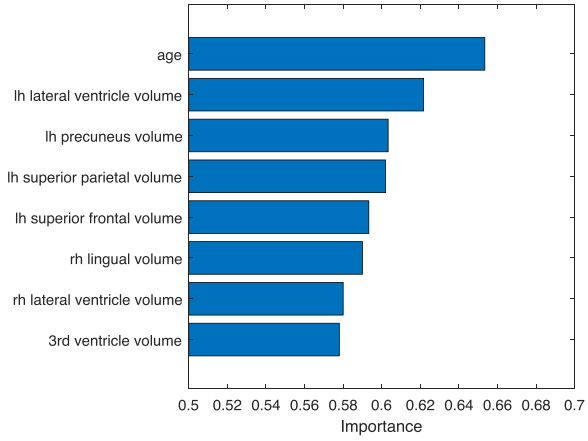


Fig. 5. Importance of the features listed in Table 3 for the most challenging CN vs. EMCI classification.

poor since the design of the algorithm was focused more on the early diagnosis of MCI. Based on our finding in this study, PET features begin to contribute to the overall classification accuracy at the LMCI stage.

Fig. 3 shows classification results including ACC, SEN, SPE, area under the curve (AUC), positive predictive value (PPV), and negative predictive value (NPV) for different modalities separately and combined (Liu et al., 2016a). Fig. 4 displays the receiver operating characteristic (ROC) for the pairwise classification of all the considered groups. The area under the plot of ROC (AUC), a plot of sensitivity versus specificity, can be a useful tool to evaluate the accuracy of the classifier. An AUC value is between 0 and 1, and an ideal classifier will associate a value of 1. Here, a high AUC of 0.84 was achieved for the challenging CN vs. EMCI classification as can be seen in Fig. 4. Evidently, and as expected, higher AUC values of 0.98 for the CN vs. AD classification, and of 0.95 for the EMCI vs. AD have been achieved. In addition, AUC of 0.77, 0.82, and 0.83 have been obtained for EMCI vs. LMCI, CN vs. LMCI, and LMCI vs. AD, respectively. An observation that can be made here is that the AUC value for the EMCI vs. LMCI is lower than its counterpart for the CN vs. EMCI classification, which could mean that the neuropsychological test scores used at baseline for this type of delineation (EMCI vs. LMCI) are more relevant than what neuroimaging measures could extract.

During the training and feature selection processes, we investigated what constituted the most important variables which are to be selected for training the classifier. Table 3 provides the eight most important features selected by the algorithm along with the P-value related to the analysis of variance. These P-values are produced in the first step before applying random forest. In addition, Fig. 5 shows the importance of the features based on the Gini importance measure and they are listed in Table 3 for the challenging classification of CN vs. EMCI.

It can be observed that different features from MRI and PET data have been selected for the different binary classifications. For example, all features selected for the CN vs. EMCI classification belonged to the MRI data, while most of the features for the binary classifications of CN vs. LMCI and CN vs. AD were selected from PET data. Indeed, using multimodal imaging enhances the accuracy differently for each binary classification as a function of the features that were deemed important and the modality they were extracted from. The results of CN vs. EMCI and CN vs. LMCI suggest that the amyloid deposition in its very early stage of Alzheimer's disease is not as significant as in the later stage of the disease. The amyloid burden will probably continue to increase during the transition between EMCI and LMCI. Fig. 6 presents box plot of different features from CN vs. EMCI, CN vs. LMCI and CN vs. AD classifications, which indicate the significance of the different neuroimaging modalities at the different stages of the disease. New insights into the merits of combining MRI and PET imaging modalities are postulated in (Duara et al., 2015).

To assess the performance of the proposed method, we compared our results with SVM classifier using the same kernel and same features. The results as provided in Table 4 indicate that GP with linear kernel do not provide higher accuracy in comparison to the GP with RBF in most groups except for CN vs. LMCI. GP with linear kernel is more successful to detect LMCI in group of CN vs. LMCI; however, the computation time for this method is significantly higher than the other methods as can be seen from Table 5. Generally, it was observed that both linear and RBF kernels for GP and SVM can be applied successfully based on the goal of the research study.

In addition, in order to assess the performance of the proposed method, the classification results of some well-established feature selection methods using SVM and GP classifiers for 3 binary classifications of CN vs. EMCI, CN vs. LMCI, and CN vs. AD are plotted in Fig. 7. Combination of t-test with GP classifier and random forest recursive feature elimination feature selections with GP and SVM classifiers have higher accuracy in comparison to the other combination algorithms.

6. Discussion

In this study, we evaluated a machine learning algorithm based on the Gaussian process for the delineation of the challenging EMCI from the CN group. The similarity of SVM and Gaussian process is in using the covariance kernel; however, the maximum margin approach is distinct in the SVM approach. The classification results of GP and SVM

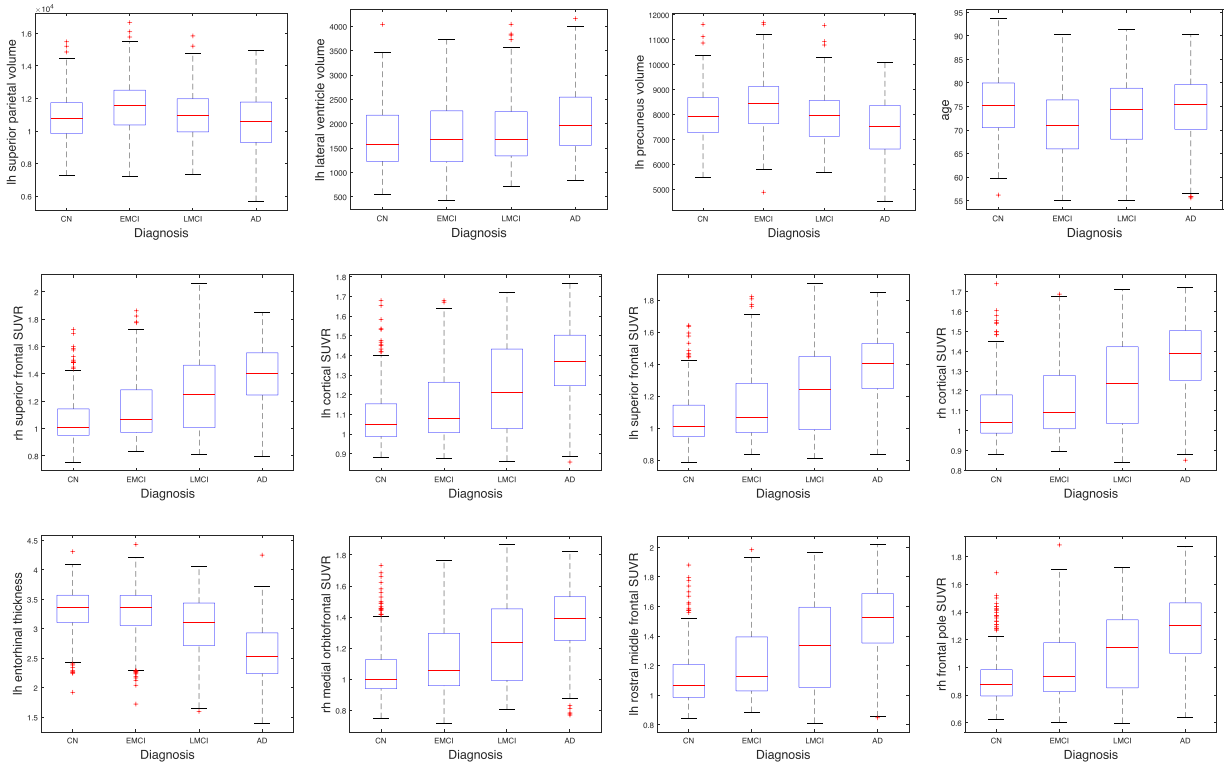


Fig. 6. Boxplot of different features for top: CN vs. EMCI, middle: CN vs. LMCI, and bottom: CN vs. AD.

Table 4
Performance Comparison of the Gaussian classifier with SVM using the same kernel and the same feature including the MRI, PET, and DI. (ACC: accuracy, SEN: sensitivity, SPE: pecificity, CN: cognitively normal, EMCI: early mild cognitive impairment, LMCI: late mild cognitive impairment, AD: Alzheimer's disease, SVM: support vector machine, GP: Gaussian process, RBF-K: radial basis function kernel, and L-K: linear kernel).

	CN vs. EMCI			CN vs. LMCI			CN vs. AD			EMCI vs. LMCI			EMCI vs. AD			LMCI vs. AD		
	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE
SVM (RBF-K)	75.6	78.9	70.6	76.9	69.3	85.7	91.2	90.1	93.9	70.1	80.0	60.0	85.5	90.5	80.8	75.3	73.3	75.6
GP (RBF-K)	**78.8**	**81.3**	**76.8**	79.8	70.2	**89.9**	**94.7**	**92.3**	**95.9**	**73.2**	**81.2**	**69.9**	**88.1**	**92.8**	**87.4**	77.1	79.9	75.9
SVM (L-K)	69.4	71.2	67.4	78.7	70.6	82.8	92.6	89.4	93.7	72.1	79.1	68.3	79.2	84.8	67.8	**80.2**	**80.1**	**80.7**
GP (L-K)	68.7	67.7	75.0	**81.5**	**76.3**	**86.6**	91.5	91.3	92.7	72.0	78.3	67.9	81.4	81.0	83.1	70.7	70.9	73.8

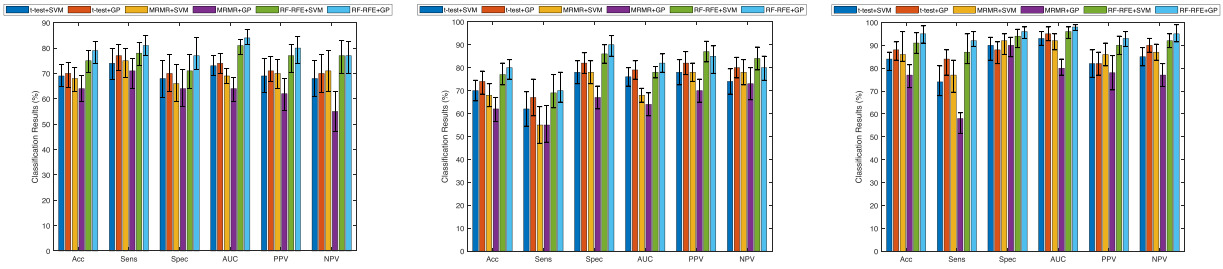
Table 5
Computation time for SVM and GP using linear and RBF kernels (numbers are in seconds).

	CN vs. EMCI	CN vs. LMCI	CN vs. AD	EMCI vs. LMCI	EMCI vs. AD	LMCI vs. AD
SVM (RBF-K)	96.58	74.52	61.22	90.12	67.69	62.88
GP (RBF-K)	44.48	28.8	18.3	38.22	27.6	25.18
SVM (L-K)	15.21	11.21	22.08	15.77	19.62	10.62
GP (L-K)	2453.97	4963.02	915.47	1756.32	1833.21	729.53

are not statistically significant; however, GP provides the predicted probability of the labels which could be beneficial in the clinical investigation while SVM provides the binary predicted labels. For example, the small predicted probability for a subject will suggest more tests are required in a clinical setting.

Moreover, temporal atrophy seems to be more relevant for the CN vs. LMCI binary classification rather than for the more challenging CN vs. EMCI. This indicates that memory deterioration of the medial temporal lobe mostly occurs at the transition from EMCI to the LMCI stage. In addition, we observed that the most important features are often selected from the left hemisphere of the brain which may suggest that more deterioration has taken place on the left hemisphere than the right hemisphere, specifically in the transition from EMCI to LMCI, although the right/left handed information of the participants is not available to make a stonger and more valid assertion.

So far, only a limited number of studies have considered EMCI and LMCI groups (Prasad et al., 2015; Goryawala et al., 2015; Shimada et al., 2013; Guerrero et al., 2014; Guo et al., 2017; Jie et al., 2018; Singh et al., 2017; Shakeri et al., 2016). Prasad et al. proposed an SVM model to rank brain connectivity features based on their importance in the classification process (Prasad et al., 2015). Using diffusion-weighted MRI together with connectivity metrics, an accuracy of 78.2% for CN vs. AD and of 59.2% for CN vs. EMCI classification were obtained by applying an SVM-based classification. They focused more on exploring features that are predictive of AD and used the classification process to better assess the information attained through the connectivity maps. Guerrero et al. reported a higher 65% accuracy for CN vs. EMCI classification using data from the ADNI-GO dataset and making use of the



(a) (b) (c)

Fig. 7. Classification results with 95% confidence intervals for combination of different feature selection using SVM and GP classifiers for the most important classifications of (a) CN vs. EMCI, (b) CN vs. LMCI, and (c) CN vs. AD. (MRMR: minimum redundancy maximum relevance, RF-RFE: random forest recursive feature elimination).

Table 6
Accuracy (ACC), sensitivity (SEN), and specificity (SPE) of the Gaussian classifier comparing to the previous works MRI: magnetic resonance imaging, dMRI: diffusion magnetic resonance imaging, fMRI: functional magnetic resonance imaging, PET: positron emission tomography, DI: demographic information.

		CN vs. EMCI			CN vs. LMCI			CN vs. AD		
	Modality	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE
(Shakeri et al., 2016)	MRI	56	52	60	59	52	64	84	73	89
(Prasad et al., 2015)	dMRI	59.2	-	-	62.8	-	-	78.2	-	-
(Guerrero et al., 2014)	MRI	65	61	69	-	-	-	86	86	85
(Jie et al., 2018)	fMRI	66.0	71.4	64.1	-	-	-	93.8	92.8	95.7
(Guo et al., 2017)	fMRI	72.8	78.3	67.1	78.6	**82.5**	72.2	88.9	91.7	85.7
proposed	MRI + PET + DI	**78.8**	**81.3**	**76.8**	**79.8**	**70.2**	**89.9**	**94.7**	**92.3**	**95.9**

		EMCI vs. LMCI			EMCI vs. AD			LMCI vs. AD		
	Modality	ACC	SEN	SPE	ACC	SEN	SPE	ACC	SEN	SPE
(Shakeri et al., 2016)	MRI	63	62	66	81	70	82	67	58	73
(Prasad et al., 2015)	dMRI	63.4	-	-	-	-	-	-	-	-
(Guerrero et al., 2014)	MRI	-	-	-	-	-	-	-	-	-
(Jie et al., 2018)	fMRI	-	-	-	-	-	-	-	-	-
(Guo et al., 2017)	fMRI	-	-	-	-	-	-	-	-	-
proposed	MRI + PET + DI	**73.2**	**81.2**	**69.9**	**88.1**	**92.8**	**87.4**	**77.1**	**79.9**	**75.9**

sparse regression for variable selection and manifold learning as a classifier (Guerrero et al., 2014). They used mini-mental state examination (MMSE) instead of disease labels to have a more continuous correlation of the disease stage and SVM with the linear kernel as the classification model.

Singh et al. proposed a feedforward deep neural network to perform classification on fluorodeoxyglucose positron emission tomography (FDG-PET) (Singh et al., 2017). They used probabilistic principal component analysis (PPCA) on max-pooled data from FDG-PET and some demographic information including age, gender, APOE $\varepsilon 1$ and $\varepsilon 2$ alleles, and functional activity questionnaire (FAQ). They achieved maximum F1-scores of 72% for the CN vs. EMCI classification and 98.14% for the CN vs. AD classification. Goryawala et al. introduced linear discriminative analysis (LDA) classifier with two-fold cross-validation using MRI data, demographic information, and neuropsychological test scores (Goryawala et al., 2015). Using MRI features they achieved an accuracy of 61.6% for CN vs. EMCI and 84.2% for CN vs. AD classification. Moreover (Zhou et al., 2014) used MRI features combined with MMSE to determine that the two most discriminative volumetric variables were the right hippocampus and the left inferior lateral ventricle and when combined with MMSE scores provided an average accuracy of 92.4% (sensitivity: 84.0%; specificity: 96.1%) for AD vs. CN classification. Their results also show for amnestic MCI (aMCI) and non-amnestic MCI (naMCI) that brain atrophy is almost evenly seen on both sides of the brain for AD subjects, which is different from right-side dominance for aMCI and left-side dominance for naMCI. However, since the ADNI subject's diagnosis is based on some neuropsychological tests such as MMSE, involving this parameter trains the algorithm based on clinical ground truth, which evidently increased the accuracy.

Shakeri et al. obtained an accuracy of 56% for the CN vs. EMCI classification on MRI data using a multilayer perceptron (MLP) on top of a so-called deep variational autoencoder (VAE) for feature selection and classification (Shakeri et al., 2016). Guo et al. proposed an approach using functional connectivity networks among different brain regions using fMRI data and a multi-kernel SVM classifier that combines multiple variations of functional MRI (fMRI) data (Guo et al., 2017). This approach resulted in an accuracy of 72.8% for the classification of the CN vs. EMCI groups and 88.9% for the classification of the CN vs. AD groups; however, one drawback of this study is that their results were based on a limited number of CN, EMCI, and AD subjects (28 CN, 29 AD, and 33 EMCI). Jie et al. used multi-kernel SVM with t-test feature selection algorithm for classification of fMRI data and obtained 66% accuracy with 71.4% sensitivity for delineating the CN vs. EMCI (Jie et al., 2018). Table 6 lists all the classification results from previous studies which are compared to the proposed method. One advantage of the proposed method over the previously reported methods is that the results offer both higher accuracy and higher sensitivity values for most classification groups although the confidence intervals have not been considered in those studies for full comparison purposes.

To address the chance level due to random variation, we compared the GP-based model to SVM. Although the GP model yielded a higher accuracy of 78.8% compared to the accuracy of 75.6% obtained with the SVM method, the differences in accuracy, sensitivity, specificity between them were found to be not statistically significant with P-values of 0.8, 0.6, 0.5, respectively. Nonetheless, although the results are encouraging showing higher accuracy and sensitivity than reported

from other related studies, there are some limitations to our study. First, there are obviously other biomarkers such as cerebrospinal fluid (CSF) and APOE that may potentially augment and improve these current classification results, especially in terms of the CN vs. EMCI classification. In addition, the presence of noise has not been tested in our study since the MRI and PET images considered underwent high level of quality control. It is essential for any study involving acquisition of any imaging modality to undergo quality control to ensure that no shading or aliasing is experienced, and that such images are free from the presence of impulse noise, Gaussian noise and any other source of noise that could affect the results of segmentation and the eventual extraction of key structural or functional features. Moreover, since finding the optimal threshold for the Pearson's correlation using optimization algorithms is time-consuming, we set the threshold manually in this study. We contemplate to investigate using the mean (μ), standard deviation (σ) and the upper percentile (z) to determine a practical threshold of what we assume will be normal distributions of the data under study as $T = \mu + z\sigma$, a formula we discovered to work well with electroencephalography (EEG) data in epilepsy (Rajaei et al., 2019). Therefore, finding new approaches to do classification using a more comprehensive multimodal neuroimaging platform will be beneficial for these type of studies. These limitations will guide the development of new machine learning algorithms that would improve these early but good results presented through the Gaussian process-based method.

7. Conclusion

In this paper, a probabilistic approach for finding the most important features augmented with a Gaussian-based model is designed to address the challenging classification of the EMCI group from the CN group. The intent was to detect and extract these subtle changes in the transition phase from CN to EMCI in order to plan for early treatment and for the design of patient-specific curative/therapeutic protocols. This approach evaluated the merits of using the Gaussian process and integrating the Bayesian prediction and classification model as another research direction for the application of machine learning in AD. To this end, we have used a feature selection method based on the random forest algorithm and applied a recursive feature elimination (RF-RFE) approach. Many of the related studies involve the MCI group as a whole in their classification without distinction of the early and late stages of MCI (EMCI and LMCI), which takes away the ability to detect the earliest signs of AD, a challenge which this study has aimed to resolve. In addition, a large number of subjects was considered, proving the ability of the proposed method to be generalized for other classification purposes.

Acknowledgments

We are grateful for the continued support from the National Science Foundation (NSF) under NSF grants CNS-1920182, CNS-1532061, CNS-133892, and CNS-1551221. We also greatly appreciate the support of NIH-NIA grants (1R01AG055638 - 01A1, 1R01AG061106- 01), State of Florida grant (8AZ23), the 1Florida Alzheimer's Disease Research Center (ADRC) (NIA 1P50AG047266-01A1) and the Ware Foundation.

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).

References

Adjouadi, M., Zong, N., Ayala, M., 2005. Multidimensional pattern recognition and classification of white blood cells using support vector machines. Particle & Particle Systems Characterization 22, 107–118.

Amoroso, N., Diacono, D., Fanizzi, A., La Rocca, M., Monaco, A., Lombardi, A., Guaragnella, C., Bellotti, R., Tangaro, S., Initiative, A.D.N., et al., 2018. Deep learning reveals Alzheimer's disease onset in MCI subjects: Results from an international challenge. Journal of neuroscience methods 302, 3–9.

Association, A., et al., 2018. 2018 Alzheimer's disease facts and figures. Alzheimer's & Dementia 14, 367–429.

Bateman, R.J., Xiong, C., Benzinger, T.L., Fagan, A.M., Goate, A., Fox, N.C., Marcus, D.S., Cairns, N.J., Xie, X., Blazey, T.M., et al., 2012. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. New England Journal of Medicine 367, 795–804.

Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence 35, 1798–1828.

Bischkopf, J., Busse, A., Angermeyer, M., 2002. Mild cognitive impairment-a review of prevalence, incidence and outcome according to current approaches. Acta Psychiatrica Scandinavica 106, 403–414.

Brosch, T., Tam, R., Initiative, A.D.N., et al., 2013. Manifold learning of brain MRIs by deep learning. International Conference on Medical Image Computing and Computer-Assisted Intervention 633–640.

Buckner, R.L., 2004. Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. Neuron 44, 195–208.

Challis, E., Hurley, P., Serra, L., Bozzali, M., Oliver, S., Cercignani, M., 2015. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. Neuroimage 112, 232–243.

Christopher, M.B., 2016. PATTERN RECOGNITION AND MACHINE LEARNING. Springer-Verlag, New York.

Chu, C., Hsu, A.L., Chou, K.H., Bandettini, P., Lin, C., Initiative, A.D.N., et al., 2012. Does feature selection improve classification accuracy? impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage 60, 59–70.

Clark, L.R., Racine, A.M., Koscik, R.L., Okonkwo, O.C., Engelman, C.D., Carlsson, C.M., Asthana, S., Bendlin, B.B., Chappell, R., Nicholas, C.R., et al., 2016. Beta-amyloid and cognitive decline in late middle age: findings from the wisconsin registry for Alzheimer's prevention study. Alzheimer's & dementia: the journal of the Alzheimer's Association 12, 805–814.

Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., Zhu, W., Park, M., Jiang, T., Jin, J.S., et al., 2011. Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. PloS one 6, e21896.

Cummings, J.L., Vinters, H.V., Cole, G.M., Khachaturian, Z.S., 1998. Alzheimer's disease etiologies, pathophysiology, cognitive reserve, and treatment opportunities. Neurology 51, S2–S17.

Curiel, R.E., Loewenstein, D.A., Rosselli, M., Penate, A., Greig-Custo, M.T., Bauer, R.M., Guinjoan, S.M., Hanson, K.S., Li, C., Lizarraga, G., et al., 2018. Semantic intrusions and failure to recover from semantic interference in mild cognitive impairment: relationship to amyloid and cortical thickness. Current Alzheimer Research 15, 848–855.

Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q., 2011. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiology of aging 32, 2322–e19.

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., Formisano, E., 2008. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. Neuroimage 43, 44–58.

Donnelly-Kehoe, P.A., Pascariello, G.O., Gómez, J.C., Initiative, A.D.N., et al., 2018. Looking for Alzheimer's disease morphometric signatures using machine learning techniques. Journal of neuroscience methods 302, 24–34.

Duara, R., Barker, W., Loewenstein, D., Greig, M.T., Rodriguez, R., Goryawala, M., Zhou, Q., Adjouadi, M., 2015. Insights into cognitive aging and Alzheimer's disease using amyloid PET and structural MRI scans. Clinical and Translational Imaging 3, 65–74.

Duara, R., Loewenstein, D.A., Potter, E., Barker, W., Raj, A., Schoenberg, M., Wu, Y., Banko, J., Potter, H., Greig, M.T., et al., 2011. Pre-MCI and MCI: neuropsychological, clinical, and imaging features and progression rates. The American Journal of Geriatric Psychiatry 19, 951–960.

Duara, R., Loewenstein, D.A., Shen, Q., Barker, W., Potter, E., Varon, D., Heurlin, K., Vandenberghe, R., Buckley, C., 2013. Amyloid positron emission tomography with 18F-flutemetamol and structural magnetic resonance imaging in the classification of mild cognitive impairment and Alzheimer's disease. Alzheimer's & dementia: the journal of the Alzheimer's Association 9, 295–301.

Dukart, J., Mueller, K., Barthel, H., Villringer, A., Sabri, O., Schroeter, M.L., 2013. Meta-analysis based SVM classification enables accurate detection of Alzheimer's disease across different clinical centers using FDG-PET and MRI. Psychiatry Research: Neuroimaging 212, 230–236.

Eskildsen, S.F., Coupé, P., García-Lorenzo, D., Fonov, V., Pruessner, J.C., Collins, D.L., Initiative, A.D.N., et al., 2013. Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. Neuroimage 65, 511–521.

Forouzannezhad, P., Abbaspour, A., Cabrerizo, M., Adiouadi, M., 2018a. Early diagnosis of mild cognitive impairment using random forest feature selection. In: 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS). IEEE. pp. 1–4.

Forouzannezhad, P., Abbaspour, A., Fang, C., Cabrerizo, M., Loewenstein, D., Duara, R., Adjouadi, M., 2018. A survey on applications and analysis methods of functional magnetic resonance imaging for Alzheimer's disease. Journal of neuroscience methods.

Forouzannezhad, P., Abbaspour, A., Li, C., Cabrerizo, M., Adjouadi, M., 2018c. A deep neural network approach for early diagnosis of mild cognitive impairment using multiple features. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE 1341–1346.

Foster, N.L., Heidebrink, J.L., Clark, C.M., Jagust, W.J., Arnold, S.E., Barbas, N.R., DeCarli, C.S., Scott Turner, R., Koeppe, R.A., Higdon, R., et al., 2007. FDG-PET improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease. Brain 130, 2616–2635.

Goryawala, M., Zhou, Q., Barker, W., Loewenstein, D.A., Duara, R., Adjouadi, M., 2015. Inclusion of neuropsychological scores in atrophy models improves diagnostic classification of Alzheimer's disease and mild cognitive impairment. Computational intelligence and neuroscience 2015 56.
Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D., Initiative, A.D.N., et al., 2013. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. NeuroImage 65, 167–175.
Guerrero, R., Wolz, R., Rao, A., Rueckert, D., ADNI, A.D.N.I., et al., 2014. Manifold population modeling as a neuro-imaging biomarker: application to ADNI and ADNI-GO. NeuroImage 94, 275–286.
Guo, H., Zhang, F., Chen, J., Xu, Y., Xiang, J., 2017. Machine learning classification combining multiple features of a hyper-network of fMRI data in Alzheimer's disease. Frontiers in neuroscience 11, 615.
Ithapu, V.K., Singh, V., Okonkwo, O.C., Chappell, R.J., Dowling, N.M., Johnson, S.C., 2015. Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment. Alzheimer's & dementia: the journal of the Alzheimer's Association 11, 1489–1499.
Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. Neuroimage 62, 782–790.
Jie, B., Liu, M., Shen, D., 2018. Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease. Medical Image Analysis 47, 81–94.
Khedher, L., Ramírez, J., Górriz, J.M., Brahim, A., Segovia, F., s Disease Neuroimaging Initiative, A., et al., 2015. Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. Neurocomputing 151, 139–150.
Li, C., Fang, C., Adjouadi, M., Cabrerizo, M., Barreto, A., Andrian, J., Duara, R., Loewenstein, D., 2017. A neuroimaging feature extraction model for imaging genetics with application to Alzheimer's disease. In: Bioinformatics and Bioengineering (BIBE). 2017 IEEE 17th International Conference on, IEEE. pp. 15–20.
Li, C., Loewenstein, D.A., Duara, R., Cabrerizo, M., Barker, W., Adjouadi, M., 2017b. The relationship of brain amyloid load and apoe status to regional cortical thinning and cognition in the adni cohort. Journal of Alzheimer's Disease 59, 1269–1282.
Liu, F., Zhou, L., Shen, C., Yin, J., 2014a. Multiple kernel learning in the primal for multimodal Alzheimer's disease classification. IEEE journal of biomedical and health informatics 18, 984–990.
Liu, M., Zhang, D., Adeli, E., Shen, D., 2016a. Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis. IEEE Transactions on Biomedical Engineering 63, 1473–1482.
Liu, M., Zhang, D., Shen, D., 2016b. Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment. IEEE transactions on medical imaging 35, 1463–1474.
Liu, S., Cai, W., Wen, L., Feng, D., 2013a. Multi-channel brain atrophy pattern analysis in neuroimaging retrieval. In: Biomedical Imaging (ISBI). 2013 IEEE 10th International Symposium on, IEEE. pp. 202–205.
Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M.J., et al., 2015. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE Transactions on Biomedical Engineering 62, 1132–1140.
Liu, S., Liu, S., Cai, W., Pujol, S., Kikinis, R., Feng, D., 2014. Early diagnosis of Alzheimer's disease with deep learning. In: Biomedical Imaging (ISBI). 2014 IEEE 11th International Symposium on, IEEE. pp. 1015–1018.
Liu, S., Song, Y., Cai, W., Pujol, S., Kikinis, R., Wang, X., Feng, D., 2013b. Multifold bayesian kernelization in Alzheimer's diagnosis. International Conference on Medical Image Computing and Computer-Assisted Intervention 303–310.
Lizarraga, G., Li, C., Cabrerizo, M., Barker, W., Loewenstein, D.A., Duara, R., Adjouadi, M., 2018. A neuroimaging web services interface as a cyber physical system for medical imaging and data management in brain research: Design study, JMIR Medical Informatics 6, e26.
Loewenstein, D.A., Curiel, R.E., DeKosky, S., Bauer, R.M., Rosselli, M., Guinjoan, S.M., Adjouadi, M., Pe nate, A., Barker, W.W., Goenaga, S., et al., 2018. Utilizing semantic intrusions to identify amyloid positivity in mild cognitive impairment. Neurology 91, e976–e984.
Loewenstein, D.A., Curiel, R.E., DeKosky, S., Rosselli, M., Bauer, R., Grieg-Custo, M., Penate, A., Li, C., Lizagarra, G., Golde, T., et al., 2017. Recovery from proactive semantic interference and MRI volume: A replication and extension study. Journal of Alzheimer's Disease 59, 131–139.
López, M., Ramírez, J., Górriz, J., Salas-Gonzalez, D., Alvarez, I., Segovia, F., Puntonet, C., 2009. Automatic tool for Alzheimer's disease diagnosis using PCA and bayesian classification rules. Electronics Letters 45, 389–391.
Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., Hamprecht, F.A., 2009. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC bioinformatics 10, 213.
Mwangi, B., Tian, T.S., Soares, J.C., 2014. A review of feature reduction techniques in neuroimaging. Neuroinformatics 12, 229–244.
Nir, T.M., Villalon-Reina, J.E., Prasad, G., Jahanshad, N., Joshi, S.H., Toga, A.W., Bernstein, M.A., Jack, C.R., Weiner, M.W., Thompson, P.M., 2015. Diffusion weighted imaging-based maximum density path analysis and classification of Alzheimer's disease. Neurobiology of aging 36, S132–S140.
Ota, K., Oishi, N., Ito, K., Fukuyama, H., Group, S.J.S., Initiative, A.D.N., et al., 2015. Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer's disease. Journal of neuroscience methods 256, 168–183.
Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on pattern analysis and machine intelligence 27, 1226–1238.
Petersen, R.C., Roberts, R.O., Knopman, D.S., Boeve, B.F., Geda, Y.E., Ivnik, R.J., Smith, G.E., Jack, C.R., 2009. Mild cognitive impairment: ten years later. Archives of neurology 66, 1447–1455.
Plant, C., Teipel, S.J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., Ewers, M., 2010. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. Neuroimage 50, 162–174.
Prasad, G., Joshi, S.H., Nir, T.M., Toga, A.W., Thompson, P.M., 2015. Brain connectivity and novel network measures for Alzheimer's disease classification. Neurobiology of aging 36, S121–S131.
Rajaei, H., Adjouadi, M., Janwattanapong, P., Pinzon, A., Gonzales-Arias, S., Barreto, A., Andrian, J., Rishe, N., Yaylali, I., 2019. Dynamics and Distant Effects of Frontal/Temporal Epileptogenic Focus using Functional Connectivity Maps. IEEE Transactions on Biomedical Engineering.
Salvatore, C., Castiglioni, I., 2018. A wrapped multi-label classifier for the automatic diagnosis and prognosis of Alzheimer's disease. Journal of neuroscience methods 302, 58–65.
Sargolzaei, S., Sargolzaei, A., Cabrerizo, M., Chen, G., Goryawala, M., Noei, S., Zhou, Q., Duara, R., Barker, W., Adjouadi, M., 2015. A practical guideline for intracranial volume estimation in patients with Alzheimer's disease, BMC bioinformatics 16, S8.
Schmidt, R., Schmidt, H., Curb, J.D., Masaki, K., White, L.R., Launer, L.J., 2002. Early inflammation and dementia: A 25-year follow-up of the honolulu-asia aging study. Annals of neurology 52, 168–174.
Seixas, F.L., Zadrozny, B., Laks, J., Conci, A., Saade, D.C.M., 2014. A bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment. Computers in biology and medicine 51, 140–158.
Shakeri, M., Lombaert, H., Tripathi, S., Kadoury, S., Initiative, A.D.N., et al., 2016. Deep spectral-based shape features for Alzheimer's disease classification. In: International Workshop on Spectral and Shape Analysis in Medical Imaging. Springer. pp. 15–24.
Shimada, H., Makizako, H., Yoshida, D., Shimokata, H., Ito, K., Washimi, Y., Endo, H., Suzuki, T., et al., 2013. Characteristics of cognitive function in early and late stages of amnestic mild cognitive impairment. Geriatrics & gerontology international 13, 83–89.
Singh, S., Srivastava, A., Mi, L., Caselli, R.J., Chen, K., Goradia, D., Reiman, E.M., Wang, Y., 2017. Deep-learning-based classification of FDG-PET data for Alzheimer's disease categories. In: 13th International Conference on Medical Information Processing and Analysis. International Society for Optics and Photonics. pp. 105720J.
Sørensen, L., Nielsen, M., Initiative, A.D.N., et al., 2018. Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination. Journal of neuroscience methods 302, 66–74.
Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., Kaye, J., Montine, T.J., et al., 2011. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the national institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia: the journal of the Alzheimer's Association 7, 280–292.
Suk, H.I., Lee, S.W., Shen, D., Initiative, A.D.N., et al., 2015. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Structure and Function 220, 841–859.
Team, R.C., 2017. R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. 2016.
Wee, C.Y., Yang, S., Yap, P.T., Shen, D., Initiative, A.D.N., et al., 2016. Sparse temporally dynamic resting-state functional connectivity networks for early mci identification. Brain imaging and behavior 10, 342–356.
Westman, E., Muehlboeck, J.S., Simmons, A., 2012. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. Neuroimage 62, 229–238.
Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J.S., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B., et al., 2011. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. Neuroimage 54, 1178–1187.
Williams, C.K., Barber, D., 1998. Bayesian classification with gaussian processes. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 1342–1351.
Wilson, A.G., 2014. Covariance kernels for fast automatic pattern discovery and extrapolation with gaussian processes. University of Cambridge.
Yao, D., Calhoun, V.D., Fu, Z., Du, Y., Sui, J., 2018. An ensemble learning system for a 4-way classification of Alzheimer's disease and mild cognitive impairment. Journal of neuroscience methods 302, 75–81.
Ye, J., Wu, T., Li, J., Chen, K., 2011. Machine learning approaches for the neuroimaging study of Alzheimer's disease. Computer 44, 99–101.
Yuan, L., Wang, Y., Thompson, P.M., Narayan, V.A., Ye, J., Initiative, A.D.N., et al., 2012. Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. NeuroImage 61, 622–632.
Zhang, X., Hu, B., Ma, X., Xu, L., 2015. Resting-state whole-brain functional connectivity networks for MCI classification using l2-regularized logistic regression. IEEE transactions on nanobioscience 14, 237–247.
Zhou, Q., Goryawala, M., Cabrerizo, M., Wang, J., Barker, W., Loewenstein, D.A., Duara, R., Adjouadi, M., 2014. An optimal decisional space for the classification of Alzheimer's disease and mild cognitive impairment. IEEE Transactions on Biomedical Engineering 61, 2245–2253.