Diagnosis of Alzheimer’s Disease Based on Structural MRI Images Using a Regularized Extreme Learning Machine and PCA Features

Ramesh Kumar Lama,1,2 Jeonghwan Gwak,1,3 Jeong-Seon Park,4 and Sang-Woong Lee1,2

1National Research Center for Dementia, Gwangju, Republic of Korea
2Department of Software, Gachon University, 1342 Seongnamdaero, Sujeonggu, Seongnam, Gyeonggido 13120, Republic of Korea
3School of Electrical Engineering and Computer Science (EECS), Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Republic of Korea
4Department of Multimedia, Chonnam National University, 50 Daehakro, Yeosu, Jeollanamdo 59626, Republic of Korea

Correspondence should be addressed to Sang-Woong Lee; slee@gachon.ac.kr

Received 4 December 2016; Revised 30 March 2017; Accepted 13 April 2017; Published 18 June 2017

1. Introduction

Alzheimer’s disease (AD) is a progressive, neurodegenerative brain disorder that attacks neurotransmitters, brain cells, and nerves, affecting brain functions, memory, and behaviors and then finally causing dementia on elderly people. Despite its significance, there is currently no cure for it. However, there are medicines available on prescription that can help delay the progress of the condition. Thus, early diagnosis of AD is essential for patient care and relevant researches. Major challenges in proper diagnosis of AD using existing classification schemes are the availability of a smaller number of training samples and the larger number of possible feature representations. In this paper, we present and compare AD diagnosis approaches using structural magnetic resonance (sMR) images to discriminate AD, mild cognitive impairment (MCI), and healthy control (HC) subjects using a support vector machine (SVM), an importance vector machine (IVM), and a regularized extreme learning machine (RELM). The greedy score-based feature selection technique is employed to select important feature vectors. In addition, a kernel-based discriminative approach is adopted to deal with complex data distributions. We compare the performance of these classifiers for volumetric sMR image data from Alzheimer’s disease neuroimaging initiative (ADNI) datasets. Experiments on the ADNI datasets showed that RELM with the feature selection approach can significantly improve classification accuracy of AD from MCI and HC subjects.
limbic system, and (3) neocortical areas are affected at the final stage. Therefore, the study of medial temporal lobe atrophy (MTA), particularly in the hippocampus, the entorhinal cortex, and the amygdala provides the evidence of the progression of AD. Generally, MTA is measured in terms of voxel-based [5], vertex-based [6], and ROI-based [7] approaches. However, as the disease progresses, other regions in the brain are also affected. In such cases, whole-brain methods are preferred rather than a specific region-based method; then, the characterization of brain atrophy for differentiating AD and MCI patients can be performed more efficiently.

In recent years, major advances in neuroimaging have provided opportunities to study neurological-related diseases, resulting improvements in early and accurate detection of AD [5, 6, 8]. Magnetic resonance imaging (MRI) is more widely used in AD-related studies because of its noninvasive nature and lack of pain to patients. In addition, MRI provides an excellent spatial resolution and good contrast [5–7, 9]. Thus, several studies have used structural MRI- (sMRI-) based biomarkers to classify AD [10–19], which describes brain atrophy and change in the size of brain tissues. Similarly, functional MRI (fMRI) [20] can be utilized to characterize the hemodynamic response relevant to neural activity and functional/structural connectivity [21–23], which can be used to describe neurological disorders in the whole brain at the connectivity level. In this paper, we focused only on AD classification using sMRI. The intensity and stage of the neurodegeneration can be identified by the help of atrophy measured by sMRI [24]. Thus, sMRI-based feature extraction has attracted the attention for researchers of AD classification. These studies include morphometric methods such as region of interest (ROI)/volume of interest (VOI) grey matter voxels in the automatic segmentation of images [25] and the sMRI measurement of the hippocampus and the medial temporal lobe [26].

Several machine learning techniques have been used to distinguish AD subjects from elderly control subjects using different biomarkers. The commonly used classifiers include support vector machine (SVM), artificial neural network (ANN), and other ensemble classifiers. Among them, SVM and the variants have been widely studied due to its relatively good accuracy and ability to deal with high-dimensional data. A SVM-type classifier (e.g., Magnin et al. [27]) begins with a learning stage from the training dataset consisting of well-characterized subjects with known states (i.e., labels for the subjects are given). Then, the classifier aims to maximize the margin of the training data by constructing the optimal separating hyperplane or a set of hyperplanes in a single- or higher-dimensional space. At a testing stage, classification is performed for test dataset based on the learned hyperplane(s). In general, three-dimensional (3D) T1-weighted MR images of each subject were automatically parcellated into ROIs. Grey matter from each ROI is extracted, as shown in Figure 1, as a feature for classification.

Zhang et al. [10] proposed a multimodal classification approach by utilizing multi-kernel SVM based on the biomarkers including sMRI [18, 19], positron emission tomography (PET) [6], and cerebrospinal fluid (CSF) [28] to discriminate AD (or MCI) and normal control (NC) subjects. From the binary classification (i.e., AD versus NC and MCI versus NC) results, their proposed model could obtain a good accuracy for AD classification and an encouraging accuracy for MCI classification. Liu et al. [29, 30] proposed deep learning-based multiclass classification among normal controls (NC), MCI nonconverters (ncMCI), MCI converters (cMCI), and AD subjects based on 83 ROIs of sMRI images and the corresponding registered PET images. Stacked autoencoders (SAE) were used as unsupervised learning to obtain high-level features, and then softmax logistic regression was adopted as the classifier. While the experimental results showed reasonably good performance, it is still arguing that the denoising nature of SAE can increase the difficulty of suitable feature learning and thus it may be difficult for practical use. Li et al. [31] proposed fine-grained new features based on principle component analysis (PCA), stability selection, dropout, and multitask learning, where restricted Boltzmann machine (RBM) model was used as the deep learning architecture. 93 ROIs of MRI and PET images, together with CSF biomarkers, are used. Ye et al. [32] introduced conceptual machine learning-based multimodal data fusion approach using MRI, PET, genetic, CSF, demographic for AD-related research, and functional connectivity analysis. Recently, Rama et al. [33] proposed IVM-based classification approach for multiclass classification. In this method, only the subset of features from structural MRI was used as input to kernel logistic regression thus reducing the computational cost. This method used total 65 ROIs as features for training and testing and achieved the accuracy of up to 70% while classifying AD, MCI, and HC and 76.9% for binary classification of HC and AD [33].
While several approaches have been proposed for classification of different AD stages, with relatively small dataset, it is very difficult to extract effective information. This work focuses on comparing and presenting efficient classification approaches working robustly for a relatively small dataset. To this end, we present and compare three representative classifiers, with an efficient feature selection approach, including SVM, an import vector machine (IVM) and a regularized extreme learning machine (RELML) for the multiclass classification of different stages of AD progression.

2. Materials for Study

2.1. sMRI Dataset. Data used in preparation of this paper were obtained from the Alzheimer’s disease neuroimaging initiative database (ADNI) (http://adni.loni.usc.edu/). The ADNI database was launched in 2003 as a public-private partnership. The primary goal of ADNI has been to test whether the serial MRI, PET, other biological markers, and clinical and neurophysiological assessment can be combined to measure the progression of the cognitive impairment and the early AD.

2.2. Subjects. The ADNI dataset consists of more than 6000 subjects aged from 18 to 96. From it, we selected 214 subjects aged between 65 and 96. The selected participants met the criteria defined in the ADNI protocol. We constructed balanced dataset consisting of 214 subjects as follows:

- (1) 70 NC subjects: 33 males, 37 females; age ± SD = 76.3 ± 5.4 years, range = 60–90 years; mini-mental state estimation (MMSE) score = 29.2 ± 1.0, range = 25–30.
- (2) 74 MCI subjects who had not converted to AD within 18 months: 38 males, 36 females; age ± SD = 74.5 ± 7.2 years, range = 58–88 years; MMSE score = 27.2 ± 1.7, range = 24–30.
- (3) 70 AD subjects: 39 males, 31 females; age ± SD = 76.0 ± 7.3 years, range = 55–91 years; MMSE = 23.2 ± 2.0, range = 18–27.

Table 1 shows a summary of demographic status of the selected subjects.

<table>
<thead>
<tr>
<th></th>
<th>NC</th>
<th>MCI</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>70</td>
<td>74</td>
<td>70</td>
</tr>
<tr>
<td>Average age</td>
<td>76.3</td>
<td>74.5</td>
<td>76.0</td>
</tr>
<tr>
<td>Average education points</td>
<td>16.19</td>
<td>15.96</td>
<td>15.53</td>
</tr>
<tr>
<td>MMSE</td>
<td>29.2 ± 1.0</td>
<td>27.2 ± 1.7</td>
<td>23.2 ± 2.0</td>
</tr>
</tbody>
</table>

Table 1: Summary of subject’s demographic status.

2.3. Preprocessing of sMRI Data. We used a fully automated pipeline of the FreeSurfer 5.3.0 software package for reconstruction and volumetric segmentation from all the sMRI images and extracted the pattern of useful data. The software performs a series of preprocessing operations with the FreeSurfer’s recon-all processing pipeline on the original sMRI data as shown in Figure 2. The preprocessing steps include motion correction, T1-weighted image averaging, registration of volume to the Talairach space, skull striping with a deformable template model. The white surface and the pial surface are generated for each hemisphere using encoding the shape of the corpus callosum and pons in the Talairach space and following the intensity gradients from the white matter. The accurate matching of the morphologically homologous cortical locations across subjects was estimated using the mapping of the atlas based on a cortical surface to a sphere aligning the cortical patterns. Cortical thickness at each vertex of the cortex is denoted by the average shortest distance between white and pial surfaces. The area of every triangle in a standardized spherical surface tessellation provides the surface area. Similarly, the registration surface based on the folding pattern was used to compute the local curvature. The method developed by Schaefer [34] was used to measure the folding index over the whole cortical surface. All the extracted features are explained in terms of feature measures as in Table 2.

2.4. Details of the sMRI Data. We perform binary classification for NC versus AD and multiclass classification using the one-versus-all (OVA) class setting for NC, MCI, and AD. For the subjects and groups chosen as in Table 2, volumetric features, fM5, in Table 2, were used for the study, and for each feature, we computed the grey matter tissue volume from the individual subject’s sMRI image. Block brain regions selected for the classification are shown in Figure 3. Each tissue is discriminated from other tissues by using color code defined by FreeSurfer software package. The left column presents the coronal view followed by the sagittal view in the middle column and the axial view at the rightmost column. We followed neurological convention for the view. All sMRI scans used in this paper were acquired from 3T scanners.

3. Proposed Methods: Classification of Stages of AD Progression

We used the three representative machine-learning classification algorithms, SVM, IVM, and RELM. The stepwise block diagram of the classification of stages of AD progression is shown in Figure 4.

3.1. Efficient Feature Selection. In neuroimaging analysis, the number of features per subject can be very high compared to the number of subjects, which is commonly referred to as the curse of dimensionality. We perform an efficient feature selection method based on PCA which is a method widely used to reduce the dimensionality of a high-
dimensional (imaging) data [25]. As the result, the most information representative dimensions are kept while the least important ones are excluded. PCA generates new features which are a linear combination of the initial features and maps each instance of the given dataset present in a $d$-dimensional space to a $k$-dimensional subspace such that $k < d$. The set of $k$ new dimensions generated are called the principal components (PCs), and each PC is directed towards maximum variance excluding the variance already accounted for in all its preceding components. Subsequently, the first component covers the maximum variance, and each component that follows it covers a lesser value of variance.

PCs can be represented as

$$PC_i = a_1X_1 + a_2X_2 + \cdots + a_dX_d,$$

where $PC_i$ is the $i$th PC, $X_j$ is the original feature $j$, and $a_j$ represents the numerical coefficient for $X_j$.

3.2. SVM Classifier. SVM [35] is basically a binary classifier which is useful for the classification of both separable and nonseparable data. It has been used in the neuroimaging field and considered as one of the most popular machine learning tools in the neuroscience domain in the last decade. It is a supervised classification algorithm and finds the optimal hyperplane that separates both classes with maximum margin from support vectors during the training phase. For the testing of new data points, the classifier’s decision is based on the estimated hyperplane. For the linearly separable patterns, linear SVM is used. However, linear SVM cannot guarantee better performance in complex cases with nonseparable patterns. In such scenario, linear SVM is extended using kernel trick. The input patterns are mapped into a higher dimensional space using linear and nonlinear functions known as kernels. Linear and nonlinear radial basis function (RBF) kernels are widely used SVM kernels.

3.3. IVM Classifier. The fundamental principle of IVM proposed by Zhu and Hastie [36] is built on kernel logistic regression (KLR). It has not merely performed well in the binary classification as SVM, and it can be naturally generalized to the multiclass classification. Thus, we begin with the explanation of logistic regression. Let $x_j = (x_{j1}, \ldots, x_{jn})^T$ represent observed samples with class labels $y_j \in C \{j = 1, \ldots, K\}$ pattern classes. The training set is represented as $(x_i, y_j), i = 1, \ldots, n$. For the binary class problem, where input samples $x_i$ are independent and identically distributed, the conditional class posterior probability $P(Y_i | x_i; w)$ is estimated using the following logistic regression model:

$$P(Y_i | x_i; w) = \frac{1}{1 + \exp(w^T x_i)}.$$  \hspace{1cm} (2)

The logistic regression predicts the class based on probabilities which are either $p$ for $y_i = 1$ or $1 - p$ for $y_i = 0$. Thus, we can express the cost function of logistic regression as

$$Q_0(w) = \prod_{i=1}^{n} p(y_i | x_i)^{y_i} (1 - p(y_i | x_i))^{1-y_i}. \hspace{1cm} (3)$$

In order to fit the parameters for the given model by training the given data points, we try to find the parameter $w$ that minimizes $Q_0$. As a result, $w$ is selected, which is most likely to generate the labels as the same as in the training set. The minimization can be obtained by using the gradient and the Hessian. In order to prevent overfitting, one may introduce a prior over the parameters and optimize

$$Q(w) = Q_0(w) + \frac{\lambda}{2} w^T L w. \hspace{1cm} (4)$$

Therefore, the iteration scheme could simply be formulated with the Newton-Raphson iteration method. To extend the linear model to a nonlinear one, the original features $x_n$ are transformed into the higher dimensional space $k_n$ using a kernel function

$$k_n = k(x_n, x_n). \hspace{1cm} (5)$$

The model of kernel logistic regression now presumes the a posteriori probabilities are given by

$$P_{nc}(w) = \frac{\exp(w^T k_n)}{\sum_{c} \exp(w^T k_n)}, \hspace{1cm} (6)$$

with $k_n$ as the $n$th column of the kernel matrix $K$, and the unknown parameter $w = [\ldots, w_c, \ldots]$ refers to $c$ classes. The parameters are determined in an iterative way by optimizing the regularized objective function. One of the limitations of the standard KLR is that all possible training samples are used to evaluate the kernel function, thus increasing the computational complexity and the memory.
Figure 3: Block brain regions selected for AD classification using sMRI images.

MRI images

White/pial surface

Feature measures
- Cortical thickness
- Surface area
- Folding index
- Intrinsic curvature
- Volume

Diagnosis output
- NC
- MCI
- AD

Classifier
- SVM
- IVM
- RELM

Cross validation (CV)
- 70/30 CV
- 10-fold CV
- Leave-one-out CV

Figure 4: Block diagram of automatic diagnosis system.
requirement for large datasets. Meanwhile, the complexity of the classifier can be controlled by enforcing the sparseness in the learning model. The sparse kernel machine uses only the kernel function evaluated at a subset of the training data points for prediction of new inputs. The most common methods to implement sparseness are by introducing a suitable prior or by a subset selection. One of the popular examples for sparse kernel machine is SVM, which only supports that vectors are used to predict new inputs. The main idea of incorporating sparseness into KLR is to select a subset of the hidden nodes and output nodes. The subset is determined by a greedy manner. This method begins with empty set and then constructs the set of import vectors by successively adding data samples. The construction process of sets stops once the convergence criterion is reached. The convergence criterion is used by the ratio of feature vectors out of the training set. Thus, the kernel matrix only consists of the selected a subset of the training data points for prediction of new inputs. The sparse kernel machine uses only the kernel function evaluated at a subset of the training data points for prediction of new inputs. The main idea of incorporating sparseness into KLR is to select a subset of the hidden nodes and output nodes. The subset is determined by a greedy manner. This method begins with empty set and then constructs the set of import vectors by successively adding data samples. The construction process of sets stops once the convergence criterion is reached. The convergence criterion is used by the ratio of feature vectors out of the training set. Thus, the kernel matrix only consists of the selected a subset of the training data points for prediction of new inputs.

\[
H(w_1, \ldots, w_N, b_1, \ldots, b_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_L \cdot x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_L \cdot x_N + b_L) \end{bmatrix},
\]

\[
\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \\ T = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix} \end{bmatrix},
\]

Here, \( H \) is called the hidden layer output matrix. The output weights \( \beta \) can be calculated as

\[
\beta = H^T T,
\]

where \( H^+ \) is the Moore-Penrose generalized inverse of the matrix \( H \) with the advantage of speed. ELM is well-suited for the tasks related to neuroimaging and big data for the classification of binary and multiclass settings. However, the decrease in computation time increases the error in the output, thus decreasing the accuracy. To increase the accuracy, ELM is combined with sparse representation. This hybrid algorithm performs classification in two fundamental steps [38–40]. In the first stage, the ELM network is trained with the conventional training approach. However, in the testing stage, reliability-based classification is used. In reliability-based classification, the ELM classifier is employed if the test data is correctly classified; otherwise, the sparse representation-based classification is used [41]. Additionally, a regularization term is added to improve generalization performance and make the solution more robust. Finally, the output weight of the RELM can be expressed as

\[
\beta = \left( \frac{I}{C} + H^T H \right)^{-1} H^T T.
\]

4. Experimental Results and Analysis

4.1. Permutation Testing. Permutation testing can be applied to assess the statistical significance of the classifier [42]. The assessment proceeds with the selection of the test statistic of the classifier and assigns random labels to the classifier by permuting the class labels for the training dataset. Permutation testing involves performing cross validation (CV) on data for which the diagnostic label has been randomly permuted. This leads to a distribution of classification results under the null hypothesis that the classifier cannot accurately predict the clinical labels from the data. The \( p \) value of the permuted prediction rate against the prediction rate with the original data labels indicates the significance of the classifier. In this work, we used 70/30 CV, 10-fold CV, and leave-one-out (LOO) CV methods. Experiments for both binary and multiclass classification were carried out with the same setup.
Similarly, the number of wrongly classified negatives, which represent correctly identified controls, can be further divided into true positive (TP) and true negative (TN), which represent correctly identified controls. Similarly, the number of wrongly classified subjects may be represented by false positive (FP) and false negative (FN).

The accuracy measures the proportion of examples that are correctly labeled by the classifier.

\[
\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{12}
\]

However, for dataset with very unbalanced class distribution, accuracy in (12) may be a misleading performance metric. Thus, two performance metrics known as sensitivity and specificity are also used.

\[
\text{SEN} = \frac{TP}{TP + FN}, \tag{13}
\]

\[
\text{SPE} = \frac{TN}{TN + FP}. \tag{14}
\]

The sensitivity in (13) measures the rate of true positives while the specificity in (14) measures the rate of true negatives. The performance metrics for the multiclass classification are easily extended as the averaged ones on the OVA setting.

4.3. Binary Classification: Results and Analysis. The experimental results of binary classification (NC versus AD) are shown in Table 4, and those with feature selection are in Table 5. 141 subjects were randomly selected for the binary classification. Initially, we randomly segregated the training and testing dataset and used the first 111 randomly chosen subjects from each group for training and the remaining 30 subjects for testing the classifier. Similarly, in 10-fold cross-validation, all 141 subjects were randomly divided into equally sized subsets, that is, 10% testing subjects and 90% training subjects, for each of the 10-fold sets of the CV. In addition, for the nested validation, we repeated the classification experiment 10 times in the case of the 10-fold CV and leave-one-out CV, and 100 times in the case of the conventional 70/30 CV, to ensure the robustness of the classification results. The mean accuracy of all the repetitions was calculated by the final results.

4.4. Multiclass Classification: Results and Analysis. For multiclass classification, we adopted all labeled 214 subjects in Table 1. The same subjects were used in binary classification, and we adopted three CV methods. From Tables 6 and 7, it is easily observed that RELM outperforms SVM and IVM in terms of accuracy. From Figure 6, we also could see the effectiveness of the feature selection approach in 10-fold
Similar to binary classification cases, on the 70/30 CV case, we obtained the experimental results with large variance, and thus, they were also excluded from the analysis. From the results, it is naturally driven that multiclass classification (which is the general form in clinical diagnosis of AD) assisted by RELM is effective compared to the other considered representative classifiers.

4.5. Discussion on the Results. It has been known that in many problem tasks, IVM generally performs similar with SVM in terms of accuracy and provides probabilistic output. From our experiments, we could confirm that SVM generates better accuracy compared to IVM, which is mainly attributed to the robustness of SVM to outliers. The main impetus of this study was to compare representative classifiers, SVM, IVM, and RELM for binary and multiclass classification tasks. Trivially, the accuracy of the binary classification cases was higher than the corresponding multiclass classification cases. Also, the experimental results on large dataset of 214 subjects verified that RELM-based AD diagnosis framework (significantly) outperform the others with higher accuracy. To the best of our knowledge, this is the first study in which the RELM framework was used for multiclass classification on sMRI data obtained from the ADNI dataset. To classify the effectiveness of feature selection in combination with the classifiers, we utilized the PCA-based feature selection method as an efficient approach to validate its efficiency. It selects features that represent higher degrees of significance based on the internal linear SVM-based classification scores, and thus, it has the possibility of making classification significantly more accurate. The experimental results also support that such adoption of feature selection can be beneficial to improve the accuracy of the classifiers, SVM, IVM, and RELM. From the noteworthy results, we could conclude that the approaches for the stage classification can be used as an effective assistive tool for the establishment of a clinical diagnosis.

5. Conclusions and Future Work

The early diagnosis of AD and MCI is essential for patient care and research, and it is widely accepted that preventive measures play an important role to delay or alleviate the progression of AD. For the classification task of different stages of AD progression, the smaller number of training samples and the larger number of feature representations are the major challenges. In this study, we investigated

**Table 6: Performance of multiclass classification.**

<table>
<thead>
<tr>
<th>CV method</th>
<th>Classifier</th>
<th>Performance metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SVM</td>
<td>ACC (%)</td>
</tr>
<tr>
<td>10-fold CV</td>
<td>SVM</td>
<td>52.63</td>
</tr>
<tr>
<td></td>
<td>IVM</td>
<td>54.90</td>
</tr>
<tr>
<td></td>
<td>RELM</td>
<td>57.56</td>
</tr>
<tr>
<td>LOO CV</td>
<td>SVM</td>
<td>57.40</td>
</tr>
<tr>
<td></td>
<td>IVM</td>
<td>55.50</td>
</tr>
<tr>
<td></td>
<td>RELM</td>
<td>61.20</td>
</tr>
</tbody>
</table>

**Table 7: Performance of multiclass classification with feature selection.**

<table>
<thead>
<tr>
<th>CV method</th>
<th>Classifier</th>
<th>Performance metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SVM</td>
<td>ACC (%)</td>
</tr>
<tr>
<td>10-fold CV</td>
<td>SVM</td>
<td>56.60</td>
</tr>
<tr>
<td></td>
<td>IVM</td>
<td>56.14</td>
</tr>
<tr>
<td></td>
<td>RELM</td>
<td>59.81</td>
</tr>
<tr>
<td>LOO CV</td>
<td>SVM</td>
<td>58.30</td>
</tr>
<tr>
<td></td>
<td>IVM</td>
<td>56.80</td>
</tr>
<tr>
<td></td>
<td>RELM</td>
<td>61.58</td>
</tr>
</tbody>
</table>

Figure 5: Performance comparison of binary classification in terms of accuracy: (a) binary classification and (b) binary classification with feature selection.
SVM, IVM, and RELM for the classification problem. In IVM, only the subsets of the input vectors of KLR are selected by minimizing the regularized cost function to reduce computation time. RELM is an effective solution for SLFNs implemented without iteratively tuning the artificial hidden nodes and adopts reliability-based classification where ELM is adopted if the test data is correctly classified, and sparse representation is selected for the other cases. Experiments on the ADNI dataset showed that RELM-based classifier could significantly improve accuracy in both binary and multiclass classification tasks. In addition, we could observe that adoption of the PCA-based feature selection could improve the accuracy slightly. While this study is focusing on the stage diagnosis of AD progression using sMRI alone, further study is still being carried out to improve the accuracy by elaborating the classifiers, possibly using a model ensemble approach, and feature selection. Also, the studies of adding more modalities such as fMRI and PET in combination with sMRI are also one of our future researches.

Conflicts of Interest

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Research Foundation (NRF) grant funded by the Korean government (NRF-2015R1D1A1A01060422, NRF-2014M3C7A1046050, and NRF-2016M3C7A1905477) and partially supported by the MISP (Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW supervised by the ITIP (Institute for Information & Communications Technology Promotion). Data collection and sharing were funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense, Award no. W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging and the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; EuroImmmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co. Inc.; Mesoc Scale Diagnostics LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provide funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org/). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuroimaging at the University of Southern California. The Alzheimer’s Disease Neuroimaging Initiative data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (http://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 the 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.
References


