Predicting the prognosis of MCI patients using longitudinal MRI data

Fusun Er, Dionysis Goularas and the Alzheimer's Disease Neuroimaging Initiative

Abstract—The aim of this study is to develop a computer-aided diagnosis system with a deep-learning approach for distinguishing "Mild Cognitive Impairment (MCI) due to Alzheimer's Disease (AD)" patients among a list of MCI patients. In this system we are using the power of longitudinal data extracted from magnetic resonance (MR). For this work, a total of 294 MCI patients were selected from the ADNI database. Among them, 125 patients developed AD during their follow-up and the rest remained stable. The proposed computer-aided diagnosis system (CAD) attempts to identify brain regions that are significant for the prediction of developing AD. The longitudinal data were constructed using a 3D Jacobian-based method aiming to track the brain differences between two consecutive follow-ups. The proposed CAD system distinguishes MCI patients who developed AD from those who remained stable with an accuracy of 87.2%. Moreover, it does not depend on data acquired by invasive methods or cognitive tests. This work demonstrates that the use of data in different time periods contains information that is beneficial for prognosis prediction purposes that outperform similar methods and are slightly inferior only to those systems that use invasive methods or neuropsychological tests.

Index Terms— Convolutional Neural Network, Alzheimer's disease, Voxel-based Morphometry, Pooling

1 Introduction

ALZHEIMER'S disease (AD) is a progressive neurodegenerative disease that most frequently occurs in the elderly population. Neurodegeneration is initially characterized by synaptic damage and followed by neuronal loss [1]. In the course of time, as a result of neurodegeneration, brain tissue shrinks in volume, thus causing cognitive impairment. Mild Cognitive Impairment (MCI) is a clinical diagnosis indicating abnormal decline in cognitive abilities more than expected and is been associated with a significant risk of developing AD. MCI is in many cases characterized by a decline in memory, and unlike AD, MCI patients may remain stable for years in terms of memory condition. However, neurological studies showed that some individuals with MCI develop AD [2]. Although MCI and AD are distinct disorders, because of the relation between them, it has been of great interest to predict the risk of developing AD for MCI patients.

Since AD shares many clinical symptoms with MCI [3], the existence of reliable biomarkers plays an important role in the diagnosis of "MCI due to AD" [4,5]. Previous researches have defined several biomarkers based on neuroimaging [5] (i.e., structural magnetic resonance imaging (MRI), functional MRI, fluorodeoxyglucose - positron emission tomography (FDG-PET) imaging), cerebrospinal fluid (CSF) protein test [6] and blood test [7]. Although several modalities are being used to develop biomarkers for dementias, it should be considered preferable to use non-invasive methods [8] because ethical issues may arise when considering the use of them in the diagnosis and treatment of a patient with impaired mental function [9]. This fact oriented some researchers towards studies aiming to discover and use less invasive biomarkers for patients having dementias [10].

Blood-based biomarkers are considered as a minimally invasive option compared to the CSF-based biomarkers acquired by invasive methods, thus presenting a higher risk for patients [11]. However, MRI is a non-invasive method for acquiring images representing the morphology of brain structures in high resolution [12]. By consequence, structural MRI became the most studied modality for neurodegenerative diseases.

In neurodegenerative diseases, structural MRI-based biomarkers reflect the structural changes in grey and white matter tissue. It has been proven that MCI patients have significantly more atrophy than normal elderly people but less than AD patients [5]. More specifically, hippocampal and entorhinal atrophy plays an important role in designing biomarkers for "MCI due to AD" patients [13]. However, detecting such slight changes is most of the times hard to be done by physicians only by visual assessment [14]. By consequence, the use of computational models for the definition of biomarkers can contribute significantly to the prognosis prediction for MCI patients [15].

In computer-aided diagnosis studies, various deep learning based solutions are proposed showing promising results in the diagnosis of several diseases (i.e., breast cancer [16], liver cancer [17] and gliomas [18]). *Cheng et al.* utilized a stacked denoising autoencoder based deep learning schema for breast lesion classification with 89.6% accuracy in ultrasound images [16]. A previous study showed that deep convolutional neural network models give promising results in the segmentation of liver lesions on CT images [17]. A convolutional neural

- *F.Er is with the department of biotechnology, Yeditepe University, Istanbul, TR 34755. E-mail: fusun.er@std.yeditepe.edu.tr*
- *D.Goularas is with the department of computer engineering, Yeditepe University, Istanbul, TR 34755. E-mail: goularas@cse.yeditepe.edu.tr—*

xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

network achieved an accuracy of 78.7% for predicting the prognosis of MCI patients [19]. Finally, in another study, the use of a deep convolutional neural model reached an accuracy of 96% for differentiating glioblastoma multiforme (Grade IV) from low grade gliomas (Grade II and III) [18].

The present study proposes a CAD model that utilizes a deep learning approach with an autoencoder and a convolutional neural network for an early prognosis prediction of MCI patients running a risk of developing AD. The system takes advantage of a longitudinal analysis with a 3D Jacobian-based method by using only non-invasive structural MRI of two different time periods, the baseline and a 12 months later follow-up. In summary, this method takes the differences between the two MRI volumes and generates a new one, called jacobian volume. Based on these new volumes, a voxel-based morphometric analysis is implemented for detecting the regions of interest that are statistically important between the MCI-patients who remained stable and those who developed AD. Then, these regions are separated and mapped according to individual brain atlas regions in order to be used for the creation of filters during the training of the autoencoder and the pooling process for the CNN.

Finally, the resulted features extracted from the CNN are processed in an SVM classifier for the prognosis prediction yielding an accuracy of 87.2% (sensitivity of 92.4% and specificity of 80.4%). Fig. 1 illustrates the steps of the proposed system. In the following section, the above steps and methods will be described analytically.

2 Related Work

Several longitudinal neuroimaging studies explored the structural changes of the brain associated with a disease or a condition through a particular period of time. First-episode schizophrenia [20], prion disease [21], Friedreich ataxia [22], peripheral vestibular dysfunction following vestibular neuritis [23] and chronic fatigue syndrome [24] are some of the diseases that have been investigated by longitudinal analysis.

A number of pre-processing procedures and analysis of longitudinal data were proposed and validated. Tensor-based morphometry (TBM) is one of the widely used longitudinal approaches that are based on the structural deformation patterns of MRI data. TBM is originally proposed by Kipps et al. [25] and applied in many areas like first-episode schizophrenia [20,26], primary progressive multiple sclerosis [27] and logopenic variant of primary progressive aphasia [28]. Basically, TBM utilizes the deformation fields obtained during the registration procedure of images acquired in two different time periods and the implementation in the literature may slightly differ: in some studies, instead of registering two different time periods images, an average image of these two images is calculated to be used as an registration template [21,27]. In order to increase the accuracy of inter-subject alignment for the normalization procedure, DARTEL normalization is applied [23,26,27].

Several recent CAD studies attempted to identify MCI patients who will develop AD. In these works, different kinds of data were utilized (e.g. magnetic resonance imaging (MRI), clinical, neuropsychological and PET scan data) that were either cross-sectional or longitudinal. Longitudinal data can vary from study to study: In the work of Ardekani et al. [29], a measure of the hippocampal volumetric integrity (HVI) from a baseline and one-year follow-up structural MRI was utilised together with cognitive tests, genetic and demographic information. Minhas et al. [30] used longitudinal MRI derived measures together with neuropsychological tests for the purpose of differentiating converter MCI to non-converter MCI patients. In another work, Nazeri et al. [31] utilized plasma proteins measurements (a minimal invasive method) together with tensor based morphometry (TBM) for MRI data to detect changes over time.

The scope of this study aims to contribute to this field of research by proposing a CAD system based on longitudinal data and a CAD system for the prediction prognosis of MCI patients that can eventually develop AD. In the next section the subjects and methods of the CAD system will be detailed.

3 Subjects and Methods

3.1. Subjects

The data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). 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).

Ethical Statement: All procedures performed in studies involving the ADNI participants were carried out in accordance with the Helsinki declaration. The ADNI study was approved by the Institutional Review Boards of all of the participating institutions. Informed written consent was obtained from all participants and their legal representatives at each site prior to the collection of clinical, genetic, and imaging data.

In order to prevent double-dipping [33], for this study, three different datasets are chosen, composed of structural MR images and obtained from different patients. The first one (Dataset-I) is used for a voxel-based morphometric (VBM) analysis in order to create a SPM-F contrast map

(SPM: Statistical Parametric Mapping) of the important longitudinal changes between two groups of patients, those that will develop AD and those that will remain stable. This dataset contains 126 MCI patients with a single baseline and a single follow-up MRI exam for each of them. It is retrospectively known that 74 patients among them remained stable in MCI state and the rest (52 patients) converted diagnostically to AD at some point after their follow-up examination. The second one (Dataset-II) is used to generate a set of filters that will be used in a Computer-Aided Diagnosis model. This dataset has a total of 51 ADNI patients where each patient was diagnosed with MCI and has a baseline and a 12-months follow-up MRI exam. Among them, 22 patients who developed AD were grouped as Converted-MCI (C-MCI) and the 29 patients who remained stable in MCI were grouped as Non-Converted-MCI (NC-MCI). Similarly, the third dataset (Dataset-III) contains a total of 117 MCI patients, where 51 of them were in the C-MCI group and the remaining patients were in the NC-MCI group. Table I presents the use of each dataset. The average(±standard deviation) of conversion times after baseline (in months) of all total of 125 patients of the C-MCI groups of the Dateset-I, Dataset-II and Dataset-III were 9.57±2.97, 28.36±9.29 and 30.94±16.55, respectively, with an overall value of 21.6±15.27. The data having higher conversion times were reserved for training (Dataset-II) and test (Dataset-III) datasets.

It is worth to mention that the patients from the ADNI dataset are diagnosed with MCI after an evaluation based on multiple criteria. More particularly, according to information given by the ADNI group, patients with mini-mental state examination (MMSE) scores between 24-30 (inclusive), having a memory complaint, an objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, a Clinical Dementia Rating (CDR) scale of 0.5, absence of significant levels of impairment in other cognitive domains, essentially preserved activities of daily living, and an absence of dementia, are grouped as MCI.

The pre-processed images of the selected baseline and one-year follow-up T1-weigthed 3-Tesla MRI scans were downloaded from the ADNI data archive, whih are acquired with a magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence with data access permission from the ADNI. The ADNI uses several image pre-processing correction steps to produce the pre-processed images of the MRIs (i.e., gradwarp, B1 correction, and N3 specification) in order to reduce the risk of scanner bias and the effect of heterogeneity of protocols [32].

3.2. Statistical Description of Datasets

Descriptive statistical methods were used to analyse the demographic and clinical data of the groups taken from the baseline examination, in terms of age in years, score of mini-mental state examination (MMSE) test, years of education and gender. The age, years of education and MMSE score were provided as mean (standard deviation) and minimum-maximum value range for each group of patients. Further, the gender information was provided as a categorical variable where “M” stands for the number of male and “F” stands for the number of female patients. A Wilcoxon rank sum test was implemented to compare the medians of age, years of education and MMSE scores of the two groups. The existence of significant difference between the groups was assessed by a Pearson’s chi-square test. A further longitudinal statistical analysis was performed to compare the MMSE scores at the baseline and the 12-months follow-up exams for each group using one-tailed paired student t-test. All statistical tests of descriptive statistics were assessed with a significance level of 0.01. The statistical analysis was performed using Statistical Toolbox 9.1 of Matlab R2014b (Matlab, The MathWorks, Inc., Natick, Massachusetts, USA).

The descriptive statistics and statistical comparison of the NC-MCI and C-MCI groups in gender, age, years of education and MMSE score at baseline are shown in Table II for the Dataset-I, Dataset-II and Dataset-III. On all datasets, the statistical analysis showed that the three groups did not differ significantly in gender, age and years of education at the significance level of 0.01. On the other hand, the median of MMSE scores of the NC-MCI patients were significantly higher than those of the C-MCI patients ($p<0.01$) on all three datasets.

No significant difference in MMSE score between the baseline and the 12-months follow-up examination was found for the NC-MCI group of the Dataset-I (t-value: -0.94; df: 73; p-value = 0.34), the Dataset-II (t-value:1.27; df: 28; p-value = 0.21) and the Dataset-III (t-value:2.29; df: 65; p-value =0.02). On the other hand, the mean MMSE score at the 12-months follow-up examination of the converted MCI group was significantly reduced at the Dataset-I (t-value: 05.66; df: 51; p-value < 0.01), the Dataset-II (t-value: 3.47; df: 21; p-value <0.01) and the Dataset-III (t-value: 3.6; df: 50; p-value <0.01).

The average number of months from baseline to the month that conversion occurred are 10 months, 28 months, 31 months among C-MCI group of the Dataset-I, the Dataset-II, the Dataset-III, respectively. As we can observe, for the two groups (Dataset-II, the Dataset-III), the patients developed AD more than one year after the baseline. In order to increase the data size, for the Dataset-I, we included some patients who developed AD in almost less than a year after the baseline.

3.3. Methods

In this study, we used the Statistical Parametric Mapping (SPM) software version 12 (SPM12) built in Matlab (Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London) [34] and the

Computational Anatomy Toolbox (CAT) software version 12 (CAT12) to pre-process the acquired scans. SPM is a statistical method invented by Karl Friston to test hypotheses about brain imaging data [35] and CAT is a freely available toolbox developed as an extension to SPM for incorporating morphometric neuroimaging methods like voxel-based morphometry (VBM), surface-based morphometry (SBM), deformation-based morphometry (DBM), and region-of-interest (ROI)-based or label-based morphometry.

The proposed system aims to predict whether a particular MCI patient will remain stable in the diagnosis of MCI or will develop AD. It comprises of an autoencoder, a convolutional neural network and a support vector machines classifier.

1) Longitudinal pre-processing pipeline

The longitudinal pre-processing phase comprises of the following steps: the creation of two normalized volumes from the scans belonging to the baseline and the follow-up examination, the creation of a 3D jacobian determinant volume and a volume normalization and dilation procedure.

a) Generation of normalized volumes

The baseline and the 12-months follow-up normalized volumes were created from the patients' MR images using SPM12 toolbox of MATLAB. In total, we created 294 (all three datasets) modulated normalized volumes.

b) Generation of a 3D jacobian determinant volume

In this step, we compute a 3D jacobian determinant volume of longitudinal changes for each patient using their baseline and 12-months follow-up exam. For the rest of this article, this volume will be named as "jacobian volume". In summary, we attempt to detect the differences between the baseline and the 12-months follow-up.

The jacobian volumes are generated as defined in [53]. Firstly, the displacement fields from co-registered baseline to follow-up volumes were calculated followed by the estimation of directional gradients $u = (u_x, u_y, u_z)$. Then, the determinant, J_d, of the Jacobian matrix at each voxel position (x, y, z) is defined as in Eq.1.

$$J_d(x,y,z) = \begin{vmatrix} \frac{du_x(x,y,z)}{dx} & \frac{du_x(x,y,z)}{dy} & \frac{du_x(x,y,z)}{dz} \\ \frac{du_y(x,y,z)}{dx} & \frac{du_y(x,y,z)}{dy} & \frac{du_y(x,y,z)}{dz} \\ \frac{du_z(x,y,z)}{dx} & \frac{du_z(x,y,z)}{dy} & \frac{du_z(x,y,z)}{dz} \end{vmatrix} \quad (1)$$

Fig. 2 presents one particular slice of the generated displacement fields from the baseline to the corresponding 12-months follow-up, belonging to a convert patient (a) and a stable patient (b).

c) Volume normalization and dilation process

Each jacobian volume has values in the range [-4000, +4000]. A normalization procedure using a sigmoid function based on the previous range was carried out such that the jacobian volume values were mapped in the range [0, 1]. Then, a dilation operation with a spherical structural element of radius equal to 3 voxels was applied to the jacobian volumes. The dilation operation allowed to remove small dark spots because they were filled with the surrounding intensity values. It helped to enlarge clusters within the SPM-F contrast map that includes statistically significantly different voxels between two patient groups.

2) Voxel-based morphometric data analysis

In order to extract the differences between the groups of NC-MCI and C-MCI patients, a voxel-based morphometric analysis was performed on their normalized and dilated jacobian volumes of the Dataset-I. For this purpose, a second level analysis between subjects including also gender, age, and total intracranial volume (TIV) as additional covariates was designed using SPM12 to specify statistically significantly different voxel positions between the two groups. TIV was calculated as the sum of the normalized modulated tissue segments, the grey matter (GM), the white matter (WM) and the cerebrospinal fluid (CSF) using "Estimate TIV" implementation of the CAT12 toolbox. Gender was described as a categorical binary variable where zero indicates a male and one indicates a female patient. Finally, the age parameter was rounded to the nearest year. It is generally recommended to use all these three covariates together in order to acquire the optimal detection of volume loss [36].

After the VBM analysis, a SPM-F contrast map of the changes was obtained with an uncorrected p-value threshold of 0.05, an extent threshold of 400 and an F-contrast of parameters. During the experimental tests, we observed that when we were associating the volumes of interest with anatomical regions of the brain, the system produced better results in terms of accuracy. By consequence, an atlas-based region-of-interest (ROI) analysis was generated on the SPM-F contrast map using the Automated Anatomical Labeling (AAL) atlas. In fig. 3, we can see an illustration of associating the regions of interest according to particular regions of the brain based on the AAL atlas (45th slices are demonstrated). As it will be described later, for every volume region only its mean value will be taken in consideration. These values are representatives of the differences occurring in individual anatomical regions only, and not to a

combination of one or more regions. In total, 75 regions were generated based on AAL atlas. As the brain has 116 anatomical regions, 41 of them were not represented, as they did not present a statistically significantly difference for the progression of AD.

3) Autoencoder

An autoencoder is an unsupervised neural network to learn efficient data encoding that are trained to reconstruct themselves. The hidden layer outputs of a trained autoencoder provide the encoded data of the input and the weights connected to each hidden layer compose a convolutional filter. Finally, a set of convolutional filters are produced with a number equal to the number of hidden layer nodes.

The purpose of the autoencoder is to generate filters from patches of size 7×7×7 that are extracted from the regions of interest. These filters can be considered as an abstraction or a general representation of the features included in the regions of interest. During the training process, we used 51.000 patches from the Dataset-II to train an auto encoder with three hidden layers (with 64, 27 and 8 hidden units, respectively) as described in fig. 4.

4) Convolutional neural network

A convolutional neural network (CNN) is a particular kind of a feed-forward neural network that is used to learn latent representation of grid-structured data, like images or time-series data. CNN was originally proposed by LeCun et al. [37] for the recognition of handwritten zip code. Nowadays, it is widely used in neuroimaging studies due to the fact that neuroimaging data are generally acquired and analysed in image series. A CNN architecture can be formed in a variety of ways, but the simplest one is composed of a set of convolutional and subsampling/pooling layers and optionally a supervised classifier as a final layer. In this study, the proposed CNN comprises of an input, three convolutional layers, and a pooling layer for acquiring the prognostic features that will be used later as an input for the SVM classifier.

Fig. 5 presents the CNN architecture that generates a vector of prognostic features for a patient. After the acquisition of patients' jacobian volumes, we apply the following convolution procedure: for every patient, its jacobian volume is convolved with three convolutional layers using the three level filters (64, 27 and 8) calculated before with the auto encoder architecture. As mentioned above, these filters may be considered as an abstraction of the characteristics indicating if an AD conversion will happen or not. For a particular brain volume, the convolution is performed with each slice of the volume. In total, eight new brain volumes are generated (named as "feature maps"). Then, for each of them, using the mapped SPM-F contrast map, 75 volumes of interest are determined based on Anatomical atlas. These regions, as explained before, indicate the parts of individual brain anatomical regions that are considered statistically important after performing the VBM analysis. The next step is to calculate the mean value of each volume of interest. In total, 75 values are generated for each feature map named as "feature vector". As we have eight convolved brain volumes (or eight feature maps), in the end, a total of 600 values are produced named as "prognostic features".

5) Support vector machine classifier

As described above, for every patient a vector of prognostic features is calculated. Then, for the prognosis prediction a support vector machine (SVM) with a linear kernel classifier is chosen because it produces the best results compared to other classifiers as it will be detailed later.

6) Recursive feature elimination

A further evaluation was performed. We applied a support vector machine-based recursive feature elimination method (SVM-RFE) [38] in order to select the most significant features among the prognostic features. This method is based on a feature scoring procedure indicating the contribution of each feature to the construction of the SVM hyperplane during the training of the SVM classifier. For this purpose, "SVMAttributeEval" implementation of the Weka software [39] is utilized.

7) Classification performance metrics

Classification performances were measured in terms of accuracy (ACC), sensitivity (TPR), specificity (SPC), positive predictive value (PPV) and negative predictive value (NPV). The accuracy is defined as the percentage of all correct identifications among the whole population (2). The sensitivity measures the percentage of correctly identified C-MCI patients among the C-MCI groups (3), and similarly, the specificity measures the percentage of correctly identified NC-MCI patients among the NC-MCI groups (4). Furthermore, the PPV shows the percentage of correctly identified C-MCI patients to the number of patients predicted as in the C-MCI group (5). Finally, the NPV is the percentage of correctly identified NC-MCI patients to the total number of NC-MCI patients (6). Confidence intervals of repeated measurements are calculated as in (7).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2020.3017872, IEEE/ACM Transactions on Computational Biology and Bioinformatics

$Accuracy = (TP + TN)/(TP + FN + FP + TN)$	(2)
$Sensitivity = TP\ /\ (TP + FN)$	(3)
$Specificity = TN\ /\ (FP + TN)$	(4)
$Positive\ Predictive\ Value\ = TP\ /\ (TP + FP)$	(5)
$Negative\ Predictive\ Value = TN/(TN + FN)$	(6)
$Confidence\ Interval\ = \bar{x} \pm ((z * \sigma)/(\sqrt{n}))$	(7)

In the above equations, TP (true positives) and TN (true negatives) are the number of correctly classified C-MCI and NC-MCI patients respectively. Finally, FP (false positives) and FN (false negatives) are the numbers of incorrectly classified patients.

4 Results

4.1. VBM Analysis

Fig. 6 shows the SPM-F contrast map of statistically significantly different voxels between the two groups with an uncorrected p-value threshold of 0.05 and an extent threshold of 400 (F=3.92).

The SPM-F contrast map defines a total of 15410 significant voxels that have higher F-values than the selected threshold value. A total of 938 voxels of the SPM-F contrast map are located in the left middle temporal gyrus. Similarly, some other prevailing anatomical regions are left superior temporal gyrus (n=864), right superior temporal gyrus (n=775), left postcentral gyrus (n=606) and right middle temporal gyrus (n=604).

4.2. Evaluation of the CAD System

In this study two experiments were designed to assess the predictive power of the CAD system based on two different patch sizes: 5×5×5 and 7×7×7. Smaller or bigger patch sizes were also tested but the results were not satisfactory. In these experiments, an autoencoder with three layers was trained on randomly selected patches of SPM-F Contrast Map voxels with three hidden layers of 64, 27 and 8 hidden units, respectively.

The predictive power of the prognostic features was tested with three different classifications methods using the Weka software [39]: a support vector machine classifier (SVM) using a linear kernel, a linear discriminant analysis (LDA) and a multilayer perceptron (MLP) using a 10-fold cross-validation (see Table III). In this method the dataset is divided in ten parts and the classifier is trained ten times where for each time the current part among the nine is used as a test set and the rest as a training set. Next, we applied a SVM-RFE procedure, and similar to the previous procedure, the most significant 120 features were selected.

The results are given in Table III in terms of sensitivity, specificity, accuracy, precision and negative predictive value for different size of patches and classifiers for all prognostic features based on the AAL atlas. Furthermore, Table IV shows SVM-RFE procedure based on AAL atlas.

The best performance was achieved in the configuration using patches of size 7 ×7×7 and a subset of selected features, where an accuracy of 87.2%, a sensitivity of 92.4% and a specificity of 80.4% was achieved..

4.3. Evaluation of the jacobian volume

In order to measure the performance of the jacobian volume, we compared the results of the CAD system for the best configuration in terms of patch size for the baseline only and the follow up exam only.

As shown in Table V, the jacobian volume performed better than the baseline or the follow up thus proving that the differences between the two exams contain important information for the prediction prognosis on before and after SVM-RFE feature selection procedure.

4.4. Assessing the SPM-F Contrast Map

In order to show the efficiency of the SPM-F contrast map (VBM mask), we performed several tests using patches of size 7 ×7×7, with or without the VBM mask. Table VI presents the test results that the performance of the best case decreased from 65% to 62.4% in accuracy (with a dramatic decrease in SPC from 59.9 to 43.1%) without using VBM mask.

4.5. Comparison with Other CAD Systems

The proposed CAD system was compared with similar systems that present one or more common characteristics such as the use of longitudinal data or not, data from ADNI or another database, and the use of invasive minimal invasive or non-invasive methods. Table VII presents the most successful ones in terms of accuracy, sensitivity and specificity. Devanand *et al.* [40] proposed a non-ADNI study with high accuracy results (94.9%) but with relatively low specificity (85.2%) combining MRI, clinical and cognition data obtained in the baseline. *Green et al.* [41] proposed a CAD system with a very high accuracy of 94.4% (94.7% sensitivity, 94.1% specificity) based on EEG and CSF data (an invasive method) of a non-ADNI dataset. In both studies, the number of C-MCI patients was relatively small.

Among the studies that utilized the ADNI dataset, the study of Nazeri *et al.* achieved an accuracy of 93.5% using longitudinal data of the MRI and plasma protein data [31] which is a minimal invasive method. The study of Douaut *et al.* was interesting due to the employment of Difussion Tensor Imaging (DTI) data [42]. When comparing our study in terms of ADNI dataset and the use of invasive methods or additional tests, the closest among the studies of Table VII is the one of Minhas *et al.* [30] which has a

similar success rates (accuracy 89.7%) and still uses cognitive tests. Our approach achieved high success rates with an accuracy of 87.2%, a sensitivity of 92.4% and a specificity of 80.4%.

As a general observation, the current study presents lower results when compared with similar systems that use invasive methods but similar results when compared with studies using clinical and cognition tests. Moreover, it performed much better when compared with systems where no additional tests or invasive methods where utilized.

5 Discussion

In neuroimaging, the performance of longitudinal data over cross sectional data has been validated by numerous studies: Xu *et al.* [48] demonstrated that for the detection of brain imaging phenotypes, longitudinal data outperform cross-sectional data. For neurodegenerative diseases, longitudinal ADNI studies [29–31] performed better than the cross-sectional studies [7,46,47]. According to the above observations, the exploration of longitudinal data and the research for new methods in order to extract information from them could be beneficial for the prognosis and treatment of dementias. Within this scope, this work attempted to contribute to the information extraction of this type of data by proposing a method aiming to estimate the differences between two MRI volume exams based on the generation of a 3D jacobian volume. This method was applied in a CAD system and the results showed that the information extracted by this process contributed to the success of this approach when used in combination with VBM, anatomical brain atlas and deep learning techniques.

In comparison with other systems, the acquired results were higher from all systems that use only MRI data, they were lower only with those that utilized invasive methods like CSF tests [41,42], minimal invasive methods like plasma [31] and they were similar with studies that were using additional cognitive tests [30]. In particular, the CSF test is one of the diagnostic methods for brain or spine related diseases. During this test, a sample of cerebrospinal fluid is removed in a procedure called "Lumbar puncture" often performed in anaesthesia. The CSF sampling is performed by inserting a needle into the spinal canal where several trials may be needed with different sizes of needles [49] thus elevating the risk of back pain and trauma. In addition, lumbar puncture may have some post-operation complications like cerebral and spinal herniation post lumbar puncture headache [50]. From that point of view, the proposed method presents a significant advantage as it does not need invasive procedures which may be uncomfortable, inappropriate or tests that sometimes are difficult to be executed by the elderly people [51].

Using a system for the prognostic diagnosis of Alzheimer's disease in MCI patients is vital to start the treatment immediately using preventive or therapeutic medicines for slowing down the progression to AD or treating the symptoms of dementia [52]. However, the benefit of medication treatment for preventive purposes should be considered carefully in a geriatric population due to the comorbidities and side-effects of drugs. The use of preventive medication in MCI patients who actually will not develop AD may expose those patients to the unnecessary burden of medication. For this reason, identifying MCI patients who are most likely to convert to AD with a high accuracy is essential in terms of preventing unnecessary medication-related problems.

6 Conclusion

To summarize, we proposed a CAD system for MCI patients to predict their prognosis based on longitudinal data from T1-weighted structural brain MR images data only, without using invasive methods or cognitive tests. The longitudinal data were processed with the generation of a 3D jacobian volume and the resulted MRI volumes were used as an input for a system comprising a VBM analysis, an autoencoder architecture, a convolutional neural network and an atlas-based pooling method. Finally, the resulted features were directed to an SVM classifier for the prognosis prediction process. The CAD system yielded a 87.2% accuracy, thus contributing to the efforts towards the creation of a prognosis prediction system which will decide about the type of treatment of MCI patients without necessarily involving the use of invasive methods or cognitive tests. At the same time, the limitation of this method is that it gives results that are slightly inferior of studies that are using additional tests. As a future direction, we would like to further improve this method in order to include different prognostic groups in an effort to investigate if it is possible with this methodology to predict other types of dementias in earlier stages.

Acknowledgment

We would like to thank Dr. Burcu Örmeci for her significant help and guidance through this work.
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.

This work was supported by TUBITAK BIDEB 2211-C Priority Areas National Scholarship Program for PhD Students (No: 1649B031402382).

REFERENCES

[1] L. Crews and E. Masliah, "Molecular mechanisms of neurodegeneration in Alzheimer's disease," *Hum. Mol. Genet.*, 2010.

[2] A. J. Mitchell and M. Shiri-Feshki, "Rate of progression of mild cognitive impairment to dementia - Meta-analysis of 41 robust inception cohort studies," *Acta Psychiatr. Scand.*, 2009.

[3] R. Tarawneh and D. M. Holtzman, "The clinical problem of symptomatic Alzheimer disease and mild cognitive impairment," *Cold Spring Harb. Perspect. Med.*, 2012.

[4] J. W. Ashford, a Rosen, and M. Adamson, *Handbook of Imaging the Alzheimer Brain*. 2011.

[5] S. L. Risacher and A. J. Saykin, "Neuroimaging biomarkers of neurodegenerative diseases and dementia," *Semin. Neurol.*, 2013.

[6] J. Hugon, J. Dumurgier, E. Cognat, and C. Paquet, "Cerebrospinal fluid biomarkers in Alzheimer's disease," *Bull. Acad. Natl. Med.*, 2018.

[7] S. J. Kiddle *et al.*, "Plasma protein biomarkers of Alzheimer's disease endophenotypes in asymptomatic older twins: early cognitive decline and regional brain volumes," *Transl. Psychiatry*, 2015.

[8] A. D. Watt, N. L. Jenkins, G. McColl, S. Collins, and P. M. Desmond, "Ethical issues in the treatment of late-stage Alzheimer's disease," *Journal of Alzheimer's Disease*. 2019.

[9] S. Gauthier, A. Leuzy, E. Racine, and P. Rosa-Neto, "Diagnosis and management of Alzheimer's disease: Past, present and future ethical issues," *Progress in Neurobiology*. 2013.

[10] B. Klimova and K. Kuca, "Alzheimer's disease: Potential preventive, non-invasive, intervention strategies in lowering the risk of cognitive decline - A review study," *Journal of Applied Biomedicine*. 2015.

[11] A. L. Baird, S. Westwood, and S. Lovestone, "Blood-based proteomic biomarkers of Alzheimer's disease pathology," *Frontiers in Neurology*. 2015.

[12] A. Del Sole, S. Malaspina, and A. Magenta Biasina, "Magnetic resonance imaging and positron emission tomography in the diagnosis of neurodegenerative dementias," *Funct. Neurol.*, 2016.

[13] T. R. Stoub, E. J. Rogalski, S. Leurgans, D. A. Bennett, and L. deToledo-Morrell, "Rate of entorhinal and hippocampal atrophy in incipient and mild AD: Relation to memory function," *Neurobiol. Aging*, 2010.

[14] J. M. Ringman, W. Pope, and N. Salamon, "Insensitivity of visual assessment of hippocampal atrophy in familial Alzheimer's disease," *J. Neurol.*, 2010.

[15] L. Harper *et al.*, "MRI visual rating scales in the diagnosis of dementia: Evaluation in 184 post-mortem confirmed cases," *Brain*, 2016.

[16] J. Z. Cheng *et al.*, "Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans," *Sci. Rep.*, 2016.

[17] W. Li, F. Jia, and Q. Hu, "Automatic Segmentation of Liver Tumor in CT Images with Deep Convolutional Neural Networks," *J. Comput. Commun.*, 2015.

[18] M. G. Ertosun and D. L. Rubin, "Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks.," in *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2015.

[19] F. Er, D. Goularas, and B. Ormeci, "A novel convolutional neural network model based on voxel-based morphometry of imaging data in predicting the prognosis of patients with mild cognitive impairment," *J Neurol Sci*, vol. 34, pp. 52–69, 2017.

[20] L. Kong, S. Bachmann, P. A. Thomann, M. Essig, and J. Schröder, "Neurological soft signs and gray matter changes: a longitudinal analysis in first-episode schizophrenia.," *Schizophr. Res.*, 2012.

[21] E. De Vita *et al.*, "Neuroanatomical correlates of prion disease progression - a 3T longitudinal voxel-based morphometry study," *NeuroImage Clin.*, 2017.

[22] W. Santner, M. Schocke, S. Boesch, W. Nachbauer, and K. Egger, "A longitudinal VBM study monitoring treatment with erythropoietin in patients with Friedreich ataxia," *Acta Radiol. Short Reports*, 2014.

[23] S. K. Hong, H. J. Kim, and H. J. Lee, "Changes in the gray matter volume during compensation after vestibular neuritis: A longitudinal VBM study," *Restor. Neurol. Neurosci.*, 2014.

[24] Z. Y. Shan *et al.*, "Progressive brain changes in patients with chronic fatigue syndrome: A longitudinal MRI study," *J. Magn. Reson. Imaging*, 2016.

[25] C. M. Kipps, A. J. Duggins, N. Mahant, L. Gomes, J. Ashburner, and E. A. McCusker, "Progression of structural neuropathology in preclinical Huntington's disease: A tensor based morphometry study," *J. Neurol. Neurosurg. Psychiatry*, 2005.

[26] T. Asami, S. Bouix, T. J. Whitford, M. E. Shenton, D. F. Salisbury, and R. W. McCarley, "Longitudinal loss of gray matter volume in patients with first-episode schizophrenia: DARTEL automated analysis and ROI validation," *Neuroimage*, 2012.

[27] A. Eshaghi *et al.*, "Temporal and spatial evolution of grey matter atrophy in primary progressive multiple sclerosis," *Neuroimage*, 2014.

[28] J. Hugon, J. Dumurgier, E. Cognat, and C. Paquet, "Cerebrospinal fluid biomarkers in Alzheimer's disease," *Bull. Acad. Natl. Med.*, 2018.

[29] B. A. Ardekani, E. Bermudez, A. M. Mubeen, and A. H. Bachman, "Prediction of incipient Alzheimer's disease dementia in patients with mild cognitive impairment," *J. Alzheimer's Dis.*, 2016.

[30] S. Minhas, A. Khanum, F. Riaz, A. Alvi, and S. A. Khan, "A Nonparametric Approach for Mild Cognitive Impairment to AD Conversion Prediction: Results on Longitudinal Data," *IEEE J. Biomed. Heal. Informatics*, 2017.

[31] A. Nazeri, H. Ganjgahi, T. Roostaei, T. Nichols, and M. Zarei, "Imaging proteomics for diagnosis, monitoring and prediction of Alzheimer's disease," *Neuroimage*, 2014.

[32] C. R. Jack *et al.*, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging*. 2008.

[33] N. Kriegeskorte, W. K. Simmons, P. S. Bellgowan, and C. I. Baker, "Circular analysis in systems neuroscience: The dangers of double dipping," *Nat. Neurosci.*, 2009.

[34] J. Ashburner and K. Friston, "Multimodal image coregistration and partitioning - A unified framework," *Neuroimage*, 1997.

[35] W. Penny, K. Friston, J. Ashburner, S. Kiebel, and T. Nichols, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. 2007.

[36] G. S. Pell, R. S. Briellmann, C. H. (Patrick) Chan, H. Pardoe, D. F. Abbott, and G. D. Jackson, "Selection of the control group for VBM analysis: Influence of covariates, matching and sample size," *Neuroimage*, 2008.

[37] Y. Lecun and Y. Bengio, "Convolutional Networks for Images , Speech , and Time Series Variable-Size Convolutional Networks : SDNNs," *Processing*, 2010.

[38] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, 2002.

[39] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. 2016.

[40] D. P. Devanand *et al.*, "Combining Early Markers Strongly Predicts Conversion from Mild Cognitive Impairment to Alzheimer's Disease," *Biol. Psychiatry*, 2008.

[41] D. L. Green, L. Payne, R. Polikar, P. J. Moberg, D. A. Wolk, and J. Kounios, "P50: A candidate ERP biomarker of prodromal Alzheimer's disease," *Brain Res.*, 2015.

[42] D. G. *et al.*, "Brain microstructure reveals early abnormalities more than two years prior to clinical progression from mild cognitive impairment to Alzheimer's disease," *J. Neurosci.*, 2013.

[43] F. Peters, S. Villeneuve, and S. Belleville, "Predicting progression to dementia in elderly subjects with mild cognitive impairment using both cognitive and neuroimaging predictors," *J. Alzheimer's Dis.*, 2014.

[44] J. Dukart, F. Sambataro, and A. Bertolino, "Accurate prediction of conversion to Alzheimer's disease using imaging, genetic, and neuropsychological biomarkers," *J. Alzheimer's Dis.*, 2015.

[45] A. Ortiz, J. Munilla, J. M. Górriz, and J. Ramírez, "Ensembles of

Deep Learning Architectures for the Early Diagnosis of the Alzheimer's Disease," *Int. J. Neural Syst.*, 2016.

[46] I. O. Korolev, L. L. Symonds, and A. C. Bozoki, "Predicting progression from mild cognitive impairment to Alzheimer's dementia using clinical, MRI, and plasma biomarkers via probabilistic pattern classification," *PLoS One*, 2016.

[47] A. A. Willette, V. D. Calhoun, J. M. Egan, and D. Kapogiannis, "Prognostic classification of mild cognitive impairment and Alzheimer's disease: MRI independent component analysis," *Psychiatry Res. - Neuroimaging*, 2014.

[48] Z. Xu, X. Shen, and W. Pan, "Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes," *PLoS One*, 2014.

[49] C. Morgan, J. Pearson, and G. Fuller, "Lumbar punctures and cerebrospinal fluid analysis," *Medicine (United Kingdom)*. 2016.

[50] F. H. Duits *et al.*, "Performance and complications of lumbar puncture in memory clinics: Results of the multicenter lumbar puncture feasibility study," *Alzheimer's Dement.*, 2016.

[51] N. Sharma and A. N. Singh, "Exploring biomarkers for Alzheimer's disease," *Journal of Clinical and Diagnostic Research*. 2016.

[52] G. B. Frisoni *et al.*, "Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers," *The Lancet Neurology*. 2017.

[53] R.G. Boyes et al. "Cerebral atrophy measurements using Jacobian integration: Comparison with the boundary shift integral", NeuroImage, 2006.

Fusun Er received her PhD degree in biotechnology from Yeditepe University, Istanbul, Turkey, in 2017. She holds bachelor and master degree in computer science. She is currently working as an assistant professor in the software engineering department of Istanbul Okan University, Istanbul, Turkey. Her research interests include developing efficient algorithms and software for assisting neuroradiologists in detection of brain diseases, with a particular focus on magnetic resonance imaging data

Dionysis Goularas received his PhD degree in computer science from the University of Paris-SUD, Paris, France, in 2005. He holds bachelor and master degree in computer science. He is currently working as an assistant professor in the computer engineering department of Istanbul Yeditepe University, Istanbul, Turkey. His research interests include artificial intelligence, deep learning, image processing, computer graphics, computer vision, virtual reality, medical imaging and 3D reconstruction.