Journal of Neuroscience Methods xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination

Lauge Sørensen[a,b,*], Mads Nielsen[a,b], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Department of Computer Science, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark
[b] Biomediq A/S, DK-2100 Copenhagen Ø, Denmark

HIGHLIGHTS

- 3rd place in International Challenge for Automated Prediction of MCI from MRI data.
- Multi-class classification of control, MCI, MCI-converters, and Alzheimer's disease.
- Ensemble SVM with bagging and sequential feature selection outperforms single SVMs.
- Left presubiculum and right subiculum volume shown to be important MRI features.
- Benefit from selecting more features and increasing the number of base classifiers.

ARTICLE INFO

Article history:
Received 21 July 2017
Received in revised form 2 January 2018
Accepted 19 January 2018
Available online xxx

Keywords:
Alzheimer's disease
Computer-aided diagnosis
Ensemble support vector machine
Mild cognitive impairment
Mini-mental state examination
Structural MRI

ABSTRACT

Background: The International Challenge for Automated Prediction of MCI from MRI data offered independent, standardized comparison of machine learning algorithms for multi-class classification of normal control (NC), mild cognitive impairment (MCI), converting MCI (cMCI), and Alzheimer's disease (AD) using brain imaging and general cognition.
New method: We proposed to use an ensemble of support vector machines (SVMs) that combined bagging without replacement and feature selection. SVM is the most commonly used algorithm in multivariate classification of dementia, and it was therefore valuable to evaluate the potential benefit of ensembling this type of classifier.
Results: The ensemble SVM, using either a linear or a radial basis function (RBF) kernel, achieved multi-class classification accuracies of 55.6% and 55.0% in the challenge test set (60 NC, 60 MCI, 60 cMCI, 60 AD), resulting in a third place in the challenge. Similar feature subset sizes were obtained for both kernels, and the most frequently selected MRI features were the volumes of the two hippocampal subregions left presubiculum and right subiculum. Post-challenge analysis revealed that enforcing a minimum number of selected features and increasing the number of ensemble classifiers improved classification accuracy up to 59.1%.
Comparison with existing method(s): The ensemble SVM outperformed single SVM classifications consistently in the challenge test set.
Conclusions: Ensemble methods using bagging and feature selection can improve the performance of the commonly applied SVM classifier in dementia classification. This resulted in competitive classification accuracies in the International Challenge for Automated Prediction of MCI from MRI data.

1. Introduction

The combination of image analysis and machine learning to construct structural magnetic resonance imaging (MRI) biomarkers of dementia is an active research area (Falahati et al., 2014; Rathore et al., 2017; Arbabshirani et al., 2017). Many different methods have been proposed and evaluated with promising results, however, there is a need for standardized comparisons. Several studies have

* Corresponding author at: University of Copenhagen, Department of Computer Science, Universitetsparken 1, DK-2100 Copenhagen Ø, Denmark.
E-mail address: lauges@diku.dk (L. Sørensen).

[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.

https://doi.org/10.1016/j.jneumeth.2018.01.003
0165-0270/

empirically compared different methods (Cuingnet et al., 2011; Aguilar et al., 2013; Sabuncu et al., 2015) providing some insight as to which MRI features and/or which multivariate methods are beneficial. More recently, challenges in dementia classification have been organized (Simmons et al., 2014; Bron et al., 2015) providing diverse, independent, standardized comparisons. The International Challenge for Automated Prediction of MCI from MRI Data (Sarica et al., 2016), henceforth referred to as "the challenge", offered an opportunity to compare different machine learning methods using precomputed MRI features and mini-mental state examination (MMSE) scores supplied by the challenge organizers. The challenge relied on data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Petersen et al., 2010), and a notable characteristic, in comparison with previous challenges, were the multi-class classification of normal control (NC), Alzheimer's disease (AD), mild cognitive impairment that did not convert to AD at follow-up (MCI), and MCI that converted to AD at follow-up (cMCI) as evaluation metric.

This paper presents our algorithm submitted for the challenge. The algorithm used an ensemble of support vector machines (SVMs), i.e., a combination of several differently trained SVMs. An SVM is the most commonly used multivariate method in MRI-based dementia classification (Falahati et al., 2014; Rathore et al., 2017; Arbabshirani et al., 2017), and the classifier has also been widely and successfully applied in studies using data from the ADNI cohort (Weiner et al., 2015). Ensemble classification methods such as the ones that use different subsets of the data, e.g., bagging (Breiman, 1996), or different feature subsets, e.g., the random subspace method (Ho, 1998), may in many cases improve classification performance over a single classifier (Kuncheva, 2014), and ensemble SVMs have previously been successfully applied for dementia classification using different types of MRI measurements and ensemble methods (Shen et al., 2012; Chincarini et al., 2011; Varol et al., 2012; Simpson et al., 2013).

The proposed ensemble method was inspired by the random forest algorithm that uses a combination of bagging and random feature subsets (Breiman, 2001). In particular, we combined bagging without replacement with sequential forward feature selection (SFS) to obtain feature subsets optimal for the SVM classifier. To the best of our knowledge, this is a novel way of constructing the SVM ensemble. Previous feature subset ensemble SVM studies, both within MRI-based dementia classification and within other application areas, were either purely feature subset-based using some form of feature selection or ranking (Chincarini et al., 2011; Varol et al., 2012), random subspace (Waske et al., 2010; Xia et al., 2016), or a combination of selection/ranking and random subspace (Nanni, 2006; Lienemann et al., 2007; Kuncheva et al., 2010; Chen et al., 2014), or combined bagging and feature subsets either using ranking (Shen et al., 2012), random subspace (Tao et al., 2006) or recursive feature elimination based on linear SVM weights (Anaissi et al., 2016). The last is non-trivial to extend to non-linear SVM kernels.

We experimented with using both a linear kernel and a radial basis function (RBF) kernel in the SVMs, and these two configurations were submitted for the challenge. A detailed analyses of the classification results and of the selected feature subsets is presented for the ensembles submitted to the challenge, in addition to a post-challenge analysis of the performance of different feature subset methods and ensemble sizes.

2. Materials and methods

2.1. Data

The challenge used data 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 MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org.

Table 1
Characteristics of the challenge datasets.

	n	Age Mean (SD)	Sex % Male	MMSE score Mean (SD)
Challenge learning set				
NC	60	72.3 (5.7)	50.0	29.1 (1.1)
MCI	60	72.2 (7.5)	46.7	28.3 (1.6)
cMCI	60	73.0 (7.3)	58.3	27.2 (1.9)
AD	60	74.8 (7.4)	48.3	23.4 (2.1)
Challenge test set				
NC	40	74.9 (5.6)	45.0	29.0 (1.1)
MCI	40	72.4 (8.1)	57.5	27.6 (1.9)
cMCI	40	71.7 (6.3)	62.5	27.6 (1.8)
AD	40	73.1 (8.2)	57.5	22.7 (2.0)

The challenge organizers selected a total of 400 subjects from ADNI; 100 NC, 100 MCI, 100 cMCI, and 100 with AD. The subjects were split in a learning set with 240 observations and a test set with 160 observations (Table 1). The subject selection and data set definition procedures are described on the challenge website (Sarica et al., 2016). Information about time to follow-up diagnosis, used to determine MCI or cMCI, was not provided for the challenge data.

2.2. Features

The available features in the challenge consisted of 426 T1-weighted structural MRI measures computed using the cross-sectional pipeline of the FreeSurfer software package (version 5.3) (Fischl and Dale, 2000; Fischl et al., 2002), the age and sex of the subjects, and their baseline MMSE score. The challenge organizers performed all MRI processing and made the resulting MRI measures available to the challenge participants. Among the available MRI measures, we selected 33 brain volumetric measures, 14 hippocampal subregional volumetric measures, 66 regional cortical thickness measures, and the volume of white matter hypointensities. In addition, we computed 10 lobar cortical thickness measures as the mean of the individual regional cortical measures representing each lobe according to the grouping defined by Schmansky et al. (2017). See Table 2 for a detailed specification of the 124 MRI features considered in this study.

The supplied hippocampal subregional volumetric measures and regional cortical thickness measures contained unrealistically large values in some cases. An automatic MRI feature pre-processing step was therefore implemented to bring the order of magnitude to a realistic range (e.g., such that a mean cortical thickness of 2000.0 mm became 2.0 mm). This step was performed prior to the computation of the 10 lobar cortical thickness measures.

FreeSurfer's estimate of the intra-cranial volume (ICV) was also provided among the MRI measurements, and it was included in the feature vector to allow the algorithm to automatically select it if beneficial.

The MMSE score was part of the information used to obtain the clinical diagnosis in ADNI (Petersen et al., 2010) which in turn served as the label in the challenge. We therefore, in addition to the raw MMSE score, made an encoded version using the ADNI thresholds as follows: MMSE < 24:0 (we know this an AD subject); MMSE ≥ 24 and MMSE ≤ 26:1 (this is a gray zone); MMSE > 26:2 (we know this is not an AD subject).

Please cite this article in press as: Sørensen, L., Nielsen, M., Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination. J Neurosci Methods (2018), https://doi.org/10.1016/j.jneumeth.2018.01.003