New approach for automatic classification of Alzheimer's disease, mild cognitive impairment and healthy brain magnetic resonance images

Salim Lahmiri, Mounir Boukadoum

Department of Computer Science, University of Quebec at Montreal, Montreal, Quebec, H3C 3P8, Canada
E-mail: lahmiri.salim@courrier.uqam.ca

Published in Healthcare Technology Letters; Received on 30th October 2013; Revised on 19th March 2014; Accepted on 20th March 2014

Explored is the utility of modelling brain magnetic resonance images as a fractal object for the classification of healthy brain images against those with Alzheimer's disease (AD) or mild cognitive impairment (MCI). More precisely, fractal multi-scale analysis is used to build feature vectors from the derived Hurst's exponents. These are then classified by support vector machines (SVMs). Three experiments were conducted: in the first the SVM was trained to classify AD against healthy images. In the second experiment, the SVM was trained to classify AD against MCI and, in the third experiment, a multiclass SVM was trained to classify all three types of images. The experimental results, using the 10-fold cross-validation technique, indicate that the SVM achieved $97.08\% \pm 0.05$ correct classification rate, $98.09\% \pm 0.04$ sensitivity and $96.07\% \pm 0.07$ specificity for the classification of healthy against MCI images, thus outperforming recent works found in the literature. For the classification of MCI against AD, the SVM achieved $97.5\% \pm 0.04$ correct classification rate, 100% sensitivity and $94.93\% \pm 0.08$ specificity. The third experiment also showed that the multiclass SVM provided highly accurate classification results. The processing time for a given image was 25 s. These findings suggest that this approach is efficient and may be promising for clinical applications.

1. Introduction: The problem of distinguishing Alzheimer disease (AD) from mild cognitive impairment (MCI), and MCI from healthy conditions in brain magnetic images (MRI), has received increasing attention in recent years. For instance, Chincarini *et al.* [1] sampled the brain with seven relatively small volumes that were filtered to give intensity and textural MRI-based features. Each filtered region was analysed with a random forest classifier to extract relevant features that were fed to a support vector machine (SVM) for classification. The system's performance was evaluated on the classification of 144 patients with AD, 322 with MCI and 189 controls (healthy). Using receiver operating curve analysis and 20-fold cross-validation, the result was 0.97 area under curve (AUC) for discriminating the AD images from the normal ones, with 89% sensitivity and 94% specificity. In addition, the classification of controls against MCI converters (patients with MCI who progressed to clinically probable AD) achieved an AUC of 0.92, with 89% sensitivity and 80% specificity. As for the classification of MCI converters against MCI non-converters, the results indicated an AUC of 0.74, with 72% sensitivity and 65% specificity. Zhang *et al.* [2] proposed a multimodal data fusion and classification method based on features extracted from structural MRI, functional imaging, and cerebrospinal fluid (CSF). The three modalities were used to measure brain atrophy, quantify hypo-metabolism and specific proteins linked to AD, respectively. Ninety-three volumetric features were extracted from 93 regions of interest (ROI) and automatically labelled by an atlas warping algorithm. The linear SVM was used to evaluate the classification accuracy using 10-fold cross-validation. The result of classifying 51 AD against 52 normal controls yielded a classification accuracy of 93.2%, with 93% sensitivity and 93.3% specificity. In the case of 99 MCI against 52 healthy controls, the SVM achieved a classification accuracy of 76.4% (with 81.8% sensitivity and 66% specificity) with the multimodal approach, and only 72% using the best individual modality (biomarkers). Moreover, the sensitivity of the multimodal approach indicated that 91.5% of 43 MCI converters and 73.4% of 56 MCI non-converters were classified correctly. Liu *et al.* [3] presented a random patch-based subspace ensemble classification system where local patches are

extracted from relevant regions to capture spatial consistency by uniformly dividing the tissue density maps into patches of a fixed size without overlapping. Then, these local patches were randomly sampled to construct a feature subspace for designing each classifier in the ensemble system. Finally, the classifiers were combined for AD classification. The presented sparse representation-based classifier (SRC) was evaluated on 652 T1-weighted MR brain images that included 198 images with AD, 225 with MCI and 229 normal images. The SRC classifier achieved an accuracy of 90.8% and an AUC of 94.86% for AD classification, and an accuracy of 87.85% and AUC of 92.90% for MCI classification. Zhang and Shen [4] proposed a multimodal multi-task learning system to jointly predict multiple variables from multimodal data composed of clinical and categorical variables. The approach was based on a multi-task feature selection that selects the common subset of relevant features for multiple variables from each modality, and a multimodal support vector regression used to fuse features from all modalities to classify multiple variables. The proposed system was tested on a set of 1.5T MRI, PET and CSF data obtained from 45 AD patients, 91 MCI patients and 50 healthy controls. Using 10-fold cross-validation, the SVM regression achieved an accuracy of 93.3% in the classification of AD against healthy images, and 83.2% in the classification of MCI against healthy images. In [5], Cuingnet *et al.* evaluated the performance of ten different approaches (five voxel-based methods, three methods based on cortical thickness and two methods based on the hippocampus) for the classification of 509 1.5T MRIs that include 162 healthy brains, 137 with AD, 134 with MCI and 76 MCI-converters within 18 months. The whole-brain methods (voxel-based or cortical thickness based) achieved the highest accuracies in the classification of AD against healthy images (with up to 81% sensitivity and 95% specificity). For the detection of prodromal AD (healthy against MCI), the obtained sensitivity was substantially lower; all methods reached a specificity over 85%, but the sensitivity ranged between 51 and 73%. Finally, in the problem of classifying MCI non-converters against MCI converters, the thickness-based approach reached 32 to 57% sensitivity and 62 to 91% specificity, and the Hippocampus-based volume

approach distinguished MCI converters from MCI non-converters with 62% sensitivity and 69% specificity.

Although the previous methods were proved capable of distinguishing AD, MCI and healthy images, this was achieved at the cost of complex architectures and less than perfect performance, leaving the door open for alternative approaches and improvements. For instance, the system presented in [1] was based on identifying multiple ROI, using numerous filtering protocols and extracting a large feature set to prune for the most suitable features. The works presented in [2–4] used sophisticated multimodal or ensemble systems that require several preprocessing steps, with only average classification accuracy at the end, as is also the case in [5].

In this exploratory study, we investigate a simpler and potentially more effective algorithm to distinguish: (i) AD from MCI images; and (ii) MCI from healthy brain MRIs. The system is based on fractal multi-scale analysis (MSA) [6] to estimate the Hurst's exponent (HE) of a MRI at different scales of analysis. HE evaluates the self-affinity of a signal, i.e. how it can be made self-similar by an anisotropic, affine transformation (a self-similar image is one whose whole is similar to its parts). We make the hypothesis that a healthy brain biological tissue exhibits more self-affinity than a diseased brain tissue, and that this remains true for one or more levels of analysis. Intuitively, this is similar to saying that the pixel distribution of a healthy brain MRI shows more regularity (i.e. is less rough or rugged) than in that of a brain with AD or MCI. Based on this assumption, the HE extracted at different scales of analysis could help distinguish the three types of MRI. Computing it via MSA offers the advantage of combining sensitivity to any type of signal dependence with a high computational efficiency because of the relatively simple algorithm [6].

A SVM [7] is used to classify a given image based on the estimated HEs. This approach has already been found to be effective to classify healthy brain MRIs against those with AD [8]. In the study by Lahmiri and Boukadoum [8], perfect classification accuracy was obtained for a database of 51 healthy brain MRIs and 42 AD MRIs, using MSA, a SVM classifier and the leave-one-out method of validation. This Letter extends the work in [8] with three types of experiments. The first one addresses the classification of AD against MCI. The second experiment tests the algorithm on the classification of healthy images against MCI. Finally, the third experiment investigates the ability of the algorithm to classify all of AD, MCI and healthy MRIs using a multiclass SVM.

This Letter is organised as follows: Section 2 presents the fractal MSA, the SVM, and the metrics used for performance evaluation. The experimental results are presented in Section 3 and Section 4 concludes the work.

2. Methods: The overall approach is described in the block diagram of Fig. 1. It consists of the following four-step sequence:

1. The brain MRI is converted to a one-dimensional (1D) signal by row concatenation. This operation carries the risk of introducing artefacts in the obtained 1D signal, as the black pixels located outside the brain region of the MRI may create a pseudo regular component in the 1D signal. However, the impact on the estimated

HEs may not be significant given the large size of the obtained 1D signal (256^2 pixels), which preserves the efficiency of the regression estimates.

2. The MSA is employed to estimate the HE of the 1D signal at six different analysis scales to analyse the signal at different levels of granularity.

3. The obtained HE values form the components of a feature vector that is fed to the SVM for classification.

4. Finally, the classification performance is evaluated.

The MSA, SVM and performance measures are presented in the next subsections.

2.1. Fractal MSA: Consider a signal $s(t)$ defined at discrete time intervals $t = v, 2v, \dots, T$ over a period T that is an integer multiple of v . The q th-order moments of the distribution that characterise the statistical evolution of $s(t)$ is defined as follows [9]

$$K_q(d) = \frac{\langle \|s(t+d) - s(t)\|^q \rangle}{\langle |s(t)|^q \rangle} \quad (1)$$

where $d \in [v, d_{\max}]$ is a time interval and $d_{\max}$ is its predetermined upper limit. The generalised HE $H(q)$ is defined from the scaling behaviour of $K_q(d)$ according to the following empirical relation [9]

$$K_q(d) \propto \left(\frac{d}{v}\right)^{qH(q)} \quad (2)$$

If $K_q(d)$ and d satisfy a linear relationship in log-log scale for a given order q , the HE $H(q)$ can be estimated by running a linear regression of $\log(K_q(d))$ against $\log(d)$. The generalised HE $H(q)$ describes the long-memory dependence or persistence in the signal $s(t)$. The multi-scaling structure of signal $s(t)$ is related to different the orders q of $H(q)$. In general, when $H(q) > 0.5$, the signal fluctuations related to order q are persistent. When $H(q) < 0.5$, the signal fluctuations related to order q are anti-persistent. Finally, the signal fluctuations are those of a random walk if $H(q) = 0.5$ [9]. Notice that $H(q=2)$ corresponds to the classic HE [6]. In this Letter, the range of q is arbitrarily fixed to the interval from 1 to 6. These moments are considered because they were found to be effective in [8]. To determine $H(q)$, the original MRI is transformed first into a 1D signal by row concatenation. Then, the HEs $H(q)$ for $q = 1, \dots, 6$ are estimated by applying the MSA algorithm. As mentioned above, the resulting six-component feature vector forms the input of the SVM classifier to classify the MR images.

2.2. Support vector machines: Introduced by Vapnik [7], the SVM classifier is based on the statistical learning theory. It implements the principle of structural risk minimisation and has excellent generalisation ability as a result, even when the data sample is small. The SVM performs a classification task by constructing an optimal separating hyper-plane that maximises the margin between the two or more nearest data points belonging to two separate classes. Given a training set $\{(x_i, y_i), i = 1, 2, \dots, m\}$, where the input $x_i \in R^d$, and class labels $y_i \in \{+1, -1\}$, the separation hyper-plane for a linearly separable binary classification problem is given by

$$f(x) = \langle w \cdot x \rangle + b \quad (3)$$

where w is a weight vector and b is a bias, both determined from the training set. For a nonlinearly separable binary classification problem, the optimal discriminant function between the two

Figure 1 Flowchart of proposed approach

classes is given by

$$g(x) = \text{sign} \left(\sum_{i=1}^m y_i \alpha_i^* K(x, x_i) + b^* \right) \quad (4)$$

where α^* is the optimal Lagrange multiplier, Φ is a kernel function and $K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ is used to map the training points to a high-dimensional feature space where linear separation can be possible.

In this Letter, a polynomial kernel was used for the SVM. As a global kernel, it allows data points that are far away from each other to also have an influence on the kernel values contrary to a radial basis kernel alternative. The general polynomial kernel is given by

$$K(x, x_i) = (x_i \cdot x + 1)^d \quad (5)$$

where d is the order of the polynomial. It was varied from two to four in this study, with higher orders ignored because of the higher computational burden with no substantial gain in classification accuracy from our experience.

Extending the binary classifier defined in (4) to a multiclass SVM with k classes can be done by either building k one-against-all SVM or $k(k-1)$ one-class-against-another SVMs [10].

3. Experimental results: We used a collection of 33 axial, T2-weighted, MR brain images of size 256×256 pixels, downloaded from the Harvard Medical School webpage [11]. The set consisted of equal numbers of healthy, MCI and AD magnetic resonance images. We had no indication regarding image depth, the AD stage and MCI type: conversion or no conversion to AD as would be provided in more structured databases such as ADNI [12]. However, this was not an issue in this position work (a more comprehensive study using ADNI [12] and OASIS [13] MRI databases is under planning). Figs. 2–4 show examples of the magnetic resonance images of healthy brains, brains with MCI and brains with AD used in our experiments.

The validation of all experiments was conducted with the 10-fold cross-validation method, after which the average and standard deviation of the correct classification accuracy rate, sensitivity and specificity were computed to evaluate the SVM performance.

All the experiments were implemented on a 1.6 GHz CPU with 1 GB of RAM, running under the Windows XP operating system and using Matlab® R2009a [14]. The average processing time of an image was 25 s, regardless of type.

The boxplots of the estimated HEs of MCI, AD and healthy control images at scales 1–6 are presented in Figs. 5–10. The Figures show clearly distinguishable median values, with different spreads, at all scales of analysis and all categories of brain MRIs. This difference in distributions was confirmed by two statistical

Figure 2 Healthy brain MRI

Figure 3 MCI brain MRI

Figure 4 Brain MRI affected with AD

tests. The first one tested the null hypothesis of equal means for the HE values of healthy, MCI and AD data based on the F -statistic of an analysis of variance (ANOVA) test [15]. The second test checked the null hypothesis of equal distribution variances based on the Brown-Forsythe statistic [16, 17]. The results of the two tests are provided in Tables 1 and 2, and they reject the two hypotheses at the 5% level of significance for all scales of analysis, since the obtained p -values were all inferior to 0.05. We conclude from these results that the distributions of HEs among healthy, MCI and AD are statistically different and, therefore, these fractal measures are potentially discriminative to distinguish healthy, MCI and AD images.

The experimental results based on 10-fold cross-validation indicate that for the classification of healthy against MCI, the SVM

Figure 5 Boxplot of first scale fractals (1: MCI, 2: AD and 3: healthy)

Figure 6 Boxplots of second scale fractals (1: MCI, 2: AD and 3: healthy)

Figure 9 Boxplots of fifth scale fractals (1: MCI, 2: AD and 3: healthy)

Figure 7 Boxplots of third scale fractals (1: MCI, 2: AD and 3: healthy)

Figure 10 Boxplots of sixth scale fractals (1: MCI, 2: AD and 3: healthy)

achieved $97.08\% \pm 0.05$ correct classification rate with $98.09\% \pm 0.04$ sensitivity and $96.07\% \pm 0.067$ specificity. On the other hand, for the classification of MCI against AD, the results were $97.5\% \pm 0.04$ correct classification rate, 100% sensitivity and $94.93\% \pm 0.08$ specificity.

Figure 8 Boxplots of fourth scale fractals (1: MCI, 2: AD and 3: healthy)

Table 1 Results from test of equality of means

Scales	F-statistic	p-value
1	17.7225	0
2	19.1857	0
3	21.8999	0
4	24.8498	0
5	26.3930	0
6	25.9038	0

Table 2 Results from test of equality of variances

Scales	Brown-Forsythe statistic	p-value
1	7.9525	0
2	14.1925	0
3	20.5371	0
4	22.7932	0
5	20.8122	0
6	16.7263	0

Table 3 MCI against healthy: comparison with previous works

Study	Approach	Performance
[1]	intensity and textural features + SVM	accuracy not mentioned 89% sensitivity 80% specificity
[2]	multimodal data fusion + SVM	76.4% accuracy 81.8% sensitivity 66% specificity
[3]	local patches + ensemble classifier	87.8% accuracy
[4]	clinical and categorical variables + SVM regression	83.2% $\pm$ 0.1 accuracy
[5]	comparative study of ten different methods	accuracy not mentioned $\geq$ 85% specificity 51–73% sensitivity
our approach	multi scale HE + SVM	97.1% $\pm$ 0.1 accuracy 98.1% $\pm$ 0.0 sensitivity 96.1% $\pm$ 0.1 specificity

Table 4 Confusion matrix of the multiclass SVM

	MCI	AD	Healthy
MCI	10	0	1
AD	0	11	0
healthy	0	1	10

Table 3 compares the performance of our approach to the results obtained by other works in the literature in the case of classifying healthy controls against MCI. As the table shows, our presented system outperforms recent studies [1–5] in terms of accuracy, sensitivity and specificity. However, it should be noted that the comparison is based on databases of different sizes. Moreover, we did not consider the problem of classification of MCI non-converters against MCI converters as in [1, 5]. In this respect, our result should only be considered promising before further investigation with a larger database with more modalities.

The results of the third experiment using the multi-class SVM approach are provided in the confusion matrix shown in Table 4. They were obtained with a polynomial kernel of order three for the SVM. As indicated, all AD images were correctly classified. On the other hand, One MCI image was misclassified as healthy and one healthy image was misclassified as AD. This translates to a 100, 91 and 91% correct classification rate, respectively. Notwithstanding the relatively small database, these finding indicate that MSA-based fractals may also be suitable features for a multiclass automated diagnosis system.

4. Summary and conclusion: An algorithm for the classification of AD, MCI and healthy brain magnetic resonance images was presented. It uses MSA to obtain HE at various scales to characterise the brain MRI, and a SVM is used to classify the resulting exponents. The obtained preliminary results indicate that the proposed algorithm is highly accurate with a relatively low processing time that can be shortened further with a faster processor. More notably, our approach appears to outperform the works reported in the literature regarding the classification of healthy against MCI images. This preliminary evidence of the effectiveness of our approach suggests that it might be interesting for clinical applications, should a larger database of MR image such as ADNI [12] confirm the present findings. In addition, the method should be extended to 3D brain magnetic resonance images to follow the current trend for even more comprehensive analysis and diagnosis tools.

5 References

- [1] Chincarini A., Bosco P., Calvini P., *ET AL.*: ‘Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer’s disease’, *NeuroImage*, 2011, **58**, pp. 469–480
- [2] Zhang D., Wang Y., Zhou L., Shen D.: ‘The Alzheimer’s disease neuroimaging initiative: ‘multimodal classification of Alzheimer’s disease and mild cognitive impairment’, *NeuroImage*, 2011, **55**, pp. 856–867
- [3] Liu M., Zhang D., Shen D.: ‘Ensemble sparse classification of Alzheimer’s disease’, *NeuroImage*, 2012, **60**, pp. 1106–1116
- [4] Zhang D., Shen D.: ‘Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer’s disease’, *NeuroImage*, 2012, **59**, pp. 895–907
- [5] Cuingnet R., Gerardin E., Tessieras J., *ET AL.*: ‘Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database’, *NeuroImage*, 2011, **56**, pp. 766–781
- [6] Di Matteo T.: ‘Multi-scaling in finance’, *Quant. Financ.*, 2007, **7**, (1), pp. 21–36
- [7] Vapnik V.: ‘The nature of statistical learning theory’ (Springer, Berlin, 1995)
- [8] Lahmiri S., Boukadoum M.: ‘Alzheimer disease detection in brain magnetic resonance images using multi-scale fractal analysis’, *ISRN Radiol.*, 2013, Article ID 627303, doi: 10.5402/2013/627303
- [9] Peng C.K., Buldyrev S.V., Havlin S., Simons M., Stanley H.E., Goldberger A.L.: ‘Mosaic organization of DNA nucleotides’, *Phys. Rev. E*, 1994, **42**, (2), pp. 1685–1689
- [10] Hsu C.-W., Lin C.-J.: ‘A comparison of methods for multiclass support vector machines’, *IEEE Trans. Neural Netw.*, 2002, **13**, (2), pp. 415–425
- [11] <http://www.med.harvard.edu/aanlib/> (consulted in 18 March 2014)
- [12] <http://adni.loni.ucla.edu> (consulted in 18 March 2014)
- [13] <http://www.oasis-brains.org/> (consulted in 18 March 2014)
- [14] www.mathworks.com (consulted in 18 March 2014)
- [15] Rees D.G.: ‘Essential statistics’ (Chapman & Hall/CRC, 2001, 4th edn.)
- [16] Brown M.B., Forsythe A.B.: ‘Robust tests for the equality of variances’, *J. Am. Stat. Assoc.*, 1974, **69**, pp. 364–367
- [17] Brown M.B., Forsythe A.B.: ‘The small sample behavior of some test statistics which test the equality of several means’, *Technometrics*, 1974, **16**, pp. 129–132