Natural Image Bases to Represent Neuroimaging Data

Ashish Gupta ASHISH@LOUISIANA.EDU
Murat Seçkin Ayhan MSAYHAN@LOUISIANA.EDU
Anthony S. Maida MAIDA@CACS.LOUISIANA.EDU
University of Lousiana at Lafayette,
104 University Circle,
Lafayette, LA 70504 USA

Abstract

Visual inspection of neuroimagery is susceptible to human eye limitations. Computerized methods have been shown to be equally or more effective than human clinicians in diagnosing dementia from neuroimages. Nevertheless, much of the work involves the use of domain expertise to extract hand–crafted features. The key technique in this paper is the use of cross–domain features to represent MRI data. We used a sparse autoencoder to learn a set of bases from *natural images* and then applied convolution to extract features from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Using this new representation, we classify MRI instances into three categories: Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI) and Healthy Control (HC). Our approach, in spite of being very simple, achieved high classification performance, which is competitive with or better than other approaches.

1. Introduction

Alzheimer's disease (AD) is a major cause of dementia. It is progressive and fatal. A clinical diagnosis largely depends on the experience of clinicians and it can be inaccurate even among experienced clinicians (Matsuda, 2007; Kloppel et al., 2008). Conclusive diagnosis requires an autopsy.

Neuroimaging provides a variety of *biomarkers* that respond to biological and structural changes due to onset and progression of the disease. However, a visual inspection is susceptible to human eye limitations and other factors including subjectivity and experience of the clinician (Imabayashi et al., 2004; Matsuda, 2007; Kloppel et al., 2008).

Three-Dimensional Stereotactic Surface Projection (3D-SSP) can be used to perform both statistical analysis and standardization of brain imagery (Matsuda, 2007; Minoshima et al., 1995). 3D-SSP analysis is superior to visual inspection in discriminating AD from HC (Imabayashi et al., 2004). However, in addition to a unique pipeline for data processing, other decisions have to be made. For instance, z-score analysis after pixel normalization requires a choice of reference region, which has an immediate impact on results. While Minoshima et al. used the thalamus as a reference region, Imabayashi et al. referred to global activity for routine clinical diagnosis of early AD.

Kloppel et al. provided a direct comparison of diagnostic accuracy between radiologists and a computerized method that utilizes a support vector machine (SVM). They studied diagnostic classification in two different settings: i) AD versus HC and ii) AD versus fronto-temporal lobar degeneration. In each task, the computerized method either made a tie with or outperformed radiologists. Kloppel et al. and Imabayashi et al. suggest a general adoption of computerized methods for visual image interpretation for dementia diagnosis.

Using spatial independent component analysis (sICA), Yang et al. decomposed MRI data into a set of bases and corresponding coefficients. Each base captures local information and a linear combination of the bases represents a brain. The coefficients are used as features for discriminating i) AD from HC and ii) MCI from HC using an SVM. In essence, their results[1] are based on the separation of dementia and

Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR: W&CP volume 28.

[1] For the ADNI dataset.

healthy brain scans. They did not demonstrate the efficacy of their approach for the discrimination of AD from MCI.

Representation learning is concerned with obtaining meaningful and potentially high-level abstractions from sensory input. It is useful for higher-level cognitive tasks, such as computer vision and object recognition. Recently, breakthrough results have been reported in these domains (Hinton et al., 2006; Bengio & Lecun, 2007; Krizhevsky et al., 2012). In our work, we learn such meaningful and low dimensional representations of high dimensional MRI imagery. Reduction in regional cerebral blood flow (rCBF) and loss of gray matter are strong indicators of AD (Matsuda, 2007; Imabayashi et al., 2004). Our approach exploits the intuition that the discriminating power of neuroimaging biomarker should improve as the disease progresses. We model the putative biomarkers by capturing structural brain deformations that progress along with the severity of AD. Thus, our approach strives to maximize information about the putative biomarker which are correlated with dementia, instead of capturing the details about the brain that may not be as pertinent in diagnosing AD.

Our approach for modeling lesions is motivated by self-taught learning (STL) (Raina et al., 2007). They made a cogent argument that basic structural patterns are shared among data in similar spaces. Once we learn to recognize such patterns, they could be used for the task of interest on any data in some other but similar space. In this respect, we hypothesize that natural images can be leveraged in order to learn such basic patterns and to recognize them in MRI data, even though the image domains are different. In this setting, natural images are *unlabeled*, whereas MRI data are *labeled* according to their dementia levels. As in STL, the unlabeled data do not share the class labels or follow from the generative distribution of the labeled data. Moreover, the datasets are not completely irrelevant to each other as the unlabeled data will help the classification of MRI data.

2. Preliminaries

2.1. Sparse Autoencoders (SAE)

An autoencoder (Bourlard & Kamp, 1988) uses an unsupervised learning algorithm that exploits inputs as if they were targets. The algorithm consists of two modules: i) *an encoder* and ii) *a decoder*. Encoding maps the data from the input space to a representation space and decoding maps it back into the input space, thus *reconstructing* the data. The algorithm minimizes reconstruction error by back-propagation. Values of parameters W minimizing the error give rise to feature detectors (*bases*).

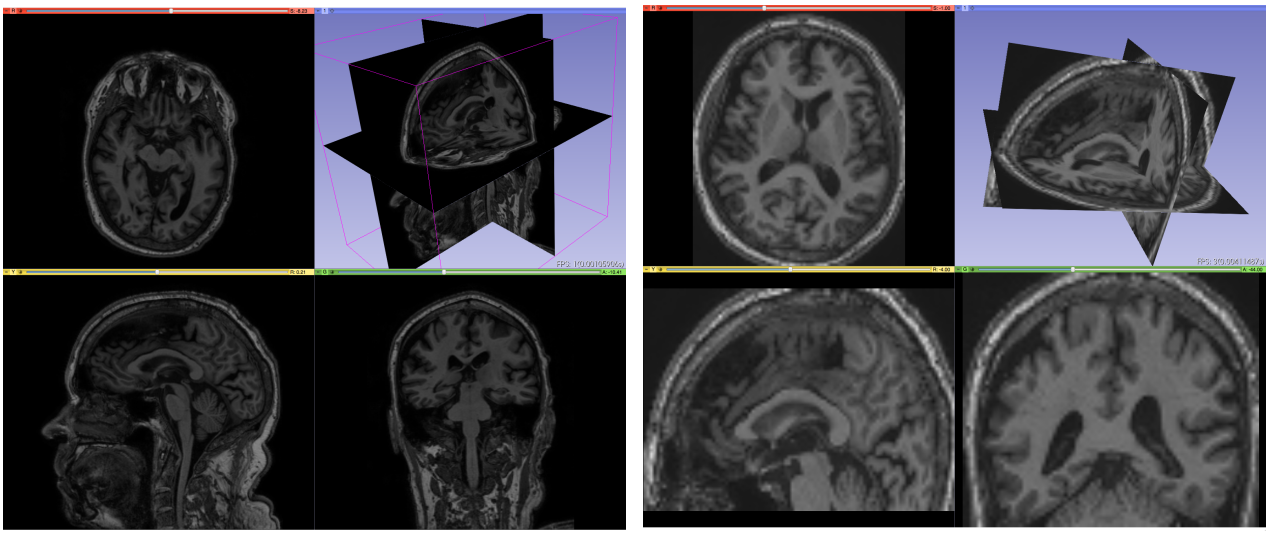
Representations learned using autoencoders are very similar those learned by PCA and ICA. However, the advantages of sparse-overcomplete representation, which can be achieved using autoencoders, have been advocated in recent work (Olshausen, 2001; Teh et al., 2003). The sparsity in the representation is enforced by sparsifying the logistic (Ranzato et al., 2006) using an additional penalty term. The penalty term encourages units to maintain low average activations. Formally, the cost function is defined as:

$$J(W,b) = \frac{1}{2m}\sum_{\mathbf{x}\in\mathcal{D}} L(\mathbf{x},\hat{\mathbf{x}}) + \beta\sum_{j=1}^{k} KL(\rho||\hat{\rho}_j) + \lambda||W||_2$$
$$\mathbf{h} = f_{enc}(\mathbf{x}) = \sigma(\mathbf{W_1}\mathbf{x} + \mathbf{b_1})$$
$$\hat{\mathbf{x}} = f_{dec}(\mathbf{h}) = \sigma(\mathbf{W_2}\mathbf{h} + \mathbf{b_2})$$

where, $\mathcal{D}$ is data matrix, $\mathbf{h}$ is the hidden representation of the data, $\hat{\mathbf{x}}$ is reconstructed data, $L(\cdot,\cdot)$ is squared loss error, $\sigma(\cdot)$ is sigmoid function, ρ is sparsity parameter, $\hat{\rho}$ is average activation, and $KL(\cdot||\cdot)$ is Kullback-Leibler divergence, $\{W,b\} = \{W_1, W_2; b_1, b_2\}$ are the parameters to learn and k is the number of bases. The hyper–parameters $(\beta, \lambda, k, \rho)$ are determined by grid search on validation data. The autoencoder is applied to a set of natural images and MRI patches to create a set of bases

2.2. Convolutional Network

The idea behind convolutional networks is neurobiologically motivated and dates back to the pioneering work of (Hubel & Wiesel, 1962) on locally sensitive and orientation-selective neural feature detectors in the visual cortex of the cat. It is specifically designed to recognize patterns and is invariant to translation, scaling, skewing, and other forms of distortions. The network includes the following forms of structural constraints (Lecun & Bengio, 1995): i) Feature extraction: Neurons take input from a local receptive field in the previous layer. The position of the feature relative to other features is preserved. ii) Feature mapping: helps in reducing the number of free parameters and obtaining shift invariance. iii) Subsampling layer: performs local averaging and subsampling thus reducing the resolution of the feature maps. It reduces the sensitivity of the feature map to various distortions.



(a) Raw MRI scan. (b) Processed MRI scan.

Figure 1. Visualization of MRI scan of an AD patient using 3D Slicer.

3. Dataset and Preprocessing

3.1. Natural Images

We used a set of *ZCA whitened* natural images[2]. The collection contains 10 different images of size 512×512, from each of which we sampled 1000 patches of size 8×8. Pixel values of the patches were scaled to $[0.1 - 0.9]$.

3.2. MRI Data

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

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 MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

Table 1 describes the demographics of the patients in our collection, which is broken into three groups: AD, MCI and HC. For a balanced dataset, we sampled 755 scans from each group. Figure 1(a) shows the MRI scan of an AD patient. The dimensions of the raw MRI data are $170 \times 256 \times 256$.

[2] http://redwood.berkeley.edu/bruno/sparsenet/

Table 1. Demographics of the ADNI dataset.

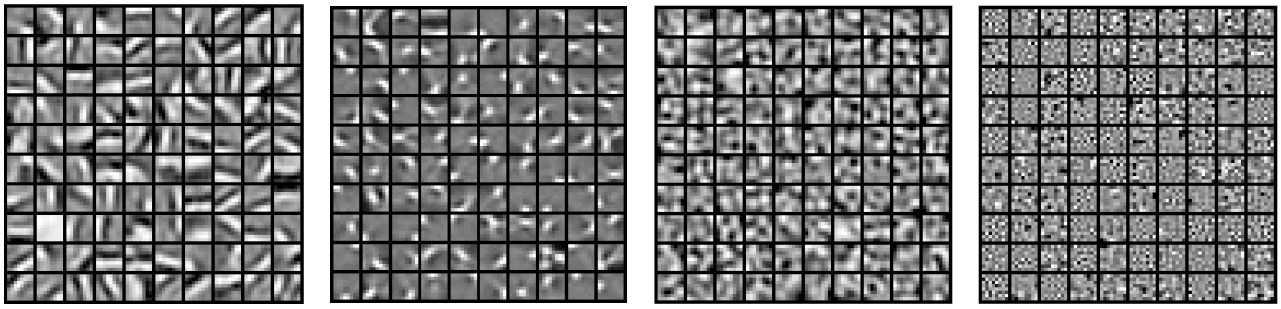
Class	# of	Sex		Age		# of
	Subjects	M	F	mean	sd	MRI scans
AD	200	103	97	76.05	7.66	755
MCI	411	267	114	75.50	7.38	2282
HC	232	113	119	76.18	4.97	1278

3.3. Stereotactic Normalization

Statistical Parametric Mapping[3] was used to normalize the image data into an International Consortium for Brain Mapping template. Our configuration also includes a positron density template with no weighting image, and a 7^{th}-order B-spline for interpolation. The remaining parameters were set to their default. After normalization, dimensionality reduces to $79 \times 95 \times 68$. Figure 1(b) shows a scan after normalization. In addition, we perform min-max normalization to scale the data to the range of $[0.1 - 0.9]$.

Note that we do not extract anatomical structures such as *gray matter* or *white matter*. As a result, no structural knowledge is injected into learning. Instead, we expect the learner to figure out whatever is informative.

[3] The SPM toolbox can be downloaded from http://www.fil.ion.ucl.ac.uk/spm/



(a) (b) (c) (d)

Figure 2. Bases learned by Autoencoder. (a) The bases learned from Natural Images. Our result is consistent with (Ranzato et al., 2006). (b) Bases learned from MRI scans. (c)-(d) Bases from natural image and MRI scans respectively that were chosen based on the performance on validation set.

4. Learning of Bases

Following the stereotactic normalization, MRI analysis enables an unbiased assessment of gray matter loss (Matsuda, 2007). The goal of visual inspection is to assess the severity of lesions; however, it is virtually impossible to detect a slight decrease in regional cerebral blood flow or glucose metabolism in the case of early AD (Imabayashi et al., 2004).

Lesions are the objects of interest. These are the *causes* of structural deformation. Exploiting statistical regularities in images, a sparse autoencoder yields a set of feature detectors. Using these detectors, we aim to identify markers, so to speak, present in the scans. These markers are used to represent lesions along the surface and ventricles of the brain.

All complex visual features share some primary features. These features are viewed as parts of objects and constellation models where model for parts can be reused in different instances. We learn bases from two distinct sets of data i) MRI imagery and ii) natural images. Since we rescale MRI and natural image pixel values to the same range, we ensure proper activation when using natural image bases.

5. Experiments

Our experiments involve three steps: i) learning a basis set, ii) extraction of features from MRI data and iii) classification of MRI data. The steps of the algorithm are given in Algorithm 1. The diagnostic classifier is a neural network. In order to learn various hyper-parameters for the autoencoder and the neural network, we split the MRI data into three subsets: i) 15% for validation ii) 10% for testing and iii) the rest for training.

Algorithm 1

Input 1: patches P
Input 2:MRI_Scans {Train TrD, Val VD, Test TD}
$(\beta, \lambda, k, \rho) \leftarrow$ SAE–validate (P, TrD, VD)
{basis set:B}$\leftarrow$ SAE $(\text{P}, \beta, \lambda, k, \rho)$
$gridSize\ gs = [3, 3]$
$features\ f = \emptyset$
$D = \text{TrD} \cup \text{VD} \cup \text{TD}$
for $s \leftarrow 1$ **to** # of scans **do**
  for $b \leftarrow 1$ **to** # of basis **do**
    $convScan \leftarrow$ CONVOLVE(D(s), B(b))
    $f(s, b) \leftarrow$ MAX–POOL(convScan, gs)
  end for
end for
$model \leftarrow$ TRAIN–NETWORK(f_{TrD}, f_{VD})

We divided the parameter search into two phases. In the first phase, we search for the best parameters $(\rho, \beta, \lambda, k$, see section 2.1) for the autoencoder. Two autoencoders were trained, one on MRI data and the other on natural images. The performances of two sets of bases learned were evaluated on the *validation set* using a classifier. To make the search faster we used soft-max as the classifier in this phase. Figure 3 shows the performance of the soft–max on various hyper–paramenter settings. For number of bases, k, we have used these values: [24, 48, 72, 100, 150, 200, 300, 400]. In our experiments, weight decay constant λ, was set to 0.001. Any larger value led to extremely poor performace. Once the parameters for the autoencoder were determined $(\rho = 0.01, \beta = 3, \lambda = 0.001, k = 100)$ and features (described below) were extracted, parameters for the neural network were learned.

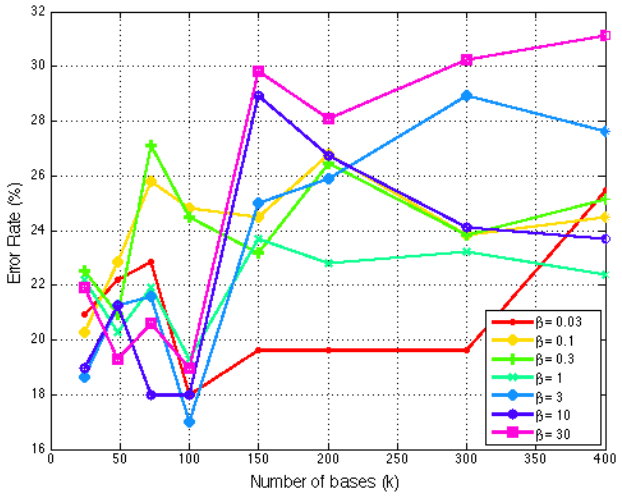


Figure 3. Performance of softmax on validation set. β signifies importance of sparsity penalty term. We have used λ (weight decay constant) = 0.001 for all the experiments.

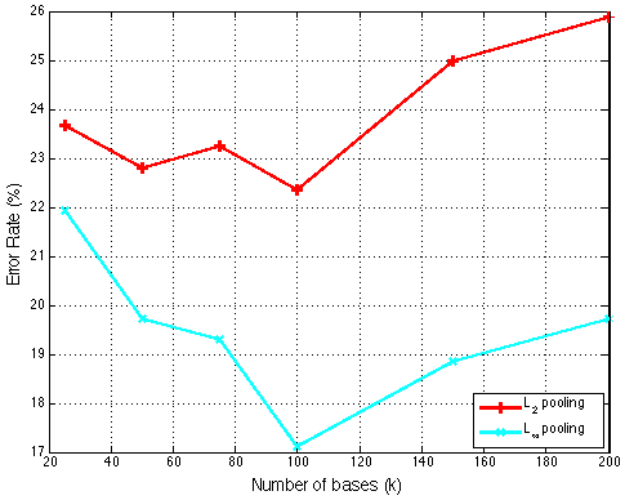


Figure 4. Comparision of L_2 pooling and L_∞ pooling over larger number of variables.

Bases

We sampled 8×8 patches from natural images (Section 3.1). More specifically, 1000 patches from each natural image give rise to 10000 patches in total. We sampled ten times as many from MRI scans. Using sparse autoencoder (Section 2.1), we learn the sparse-overcomplete representation with 100 bases in both cases (based on cross–validation). Figure 2 shows all the bases. These bases are eventually used to represent lesions in MRI data. Our results (figure 2a) are consistent with the (Ranzato et al., 2006; Olshausen, 2001; Teh et al., 2003) which have shown remarkable performance in other object recognition tasks. However, in our case, the basis sets that perform well are different. They are shown in figure 2c-d. They were generated using the parameters learned in our first phase of validation.

Convolutional Feature Extraction

For each basis, we apply 2D valid convolution[4] on each slice of each scan. Convolution with a 8×8 base yields $72 \times 88 \times 68$ dimensional feature maps (Figure 5(b)). Then, feature activations are obtained via *sigmoid* function (Figure 5(c)). For a complete list of feature activations for the median slice, see figure 9 in supplement. By a collection of feature activations for 100 bases per slice, we could obtain a new representation of the actual MRI data. However, this representation would be further blown up since we have 100 bases. To overcome this issue, we applied pooling by dividing each slice into 3×3 grid resulting in $100(bases) \times 3 \times 3 \times 68(slices) = 61,200$ feature activations for each scan. We tried different pooling strategies: L_2 pooling and L_∞ pooling. Mean pooling was not used because it smooths everything out and does not capture lesion feature. Theoretical analysis (Boureau et al., 2010) suggests that L_∞ pooling (max–pooling) works best when the features are sparse which agrees with our experiments (figure 4).

Figure 5(c) shows higher activations (shown in red) of the basis around the area of the brain filled with cerberospinal fluid (CSF). Due to shrinkage in an AD brain, CSF accounts for higher volume. Thus, the feature extraction process captures this information, which relates to the severity of lesions. Different bases strive to capture this information from different parts of the brain (for details see figure 9 in supplement).

Classification

We deal with i) binary classification and ii) three-way classification problems. Binary classification tasks allow for a comparison of our results with others. In three-way classification, we consider all three groups, AD, MCI and HC, at once.

The 61,200 max–pooled feature activations are fed to a neural network that uses back-propagation with scaled conjugated gradient. We used a single hidden layer with logistic units. The size of the hidden layer was determined based on validation accuracy and it ranges from 800 to 1600 units depending on the classification task.

[4]The valid 2D convolution of p × q on m × n matrix results with an m-p+1 × n-q+1 matrix.

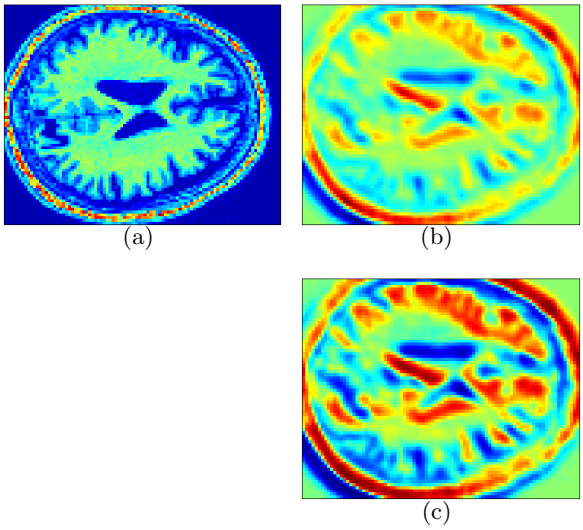


Figure 5. The effect of a basis on the median (34^{th}) slice. (a) Original median slice. (b) Result of convolution with the 55^{th} base. (c) Feature activations using sigmoid function.

Our performance metrics are sensitivity, specificity and classification accuracy. Sensitivity indicates our ability to recognize positive results, whereas specificity does it for negative results. For instance, given an AD patient, sensitivity is the probability that our classifier will assign the correct diagnosis. Specificity is the probability that our classifier will tell that the patient is healthy, given that he/she is, in fact, healthy.

Table 2 shows the comparison of our method for the binary classification task with one of the recent studies. We performed our experiment with basis sets learned from two different sources viz. natural images and MRI data. Our method shows significant improvements across all the measures compared to Yang et al. Figure 6 presents the ROC curves for the binary classifiers using natural image bases.

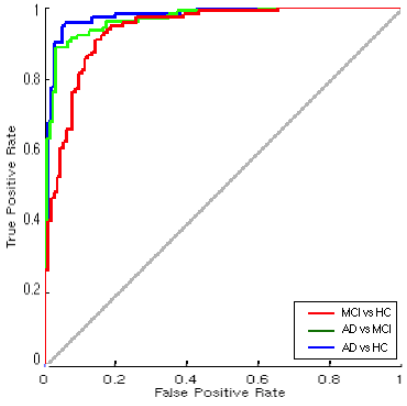


Figure 6. ROC plots of binary classifiers.

Table 2. Binary classification results on test set(%).

		Sensitivity	Specificity	Accuracy
Nat.Img. bases	AD/HC	**95.24**	94.26	**94.74**
	MCI/HC	**92.23**	**81.45**	**86.35**
	AD/MCI	84.07	**92.11**	**88.10**
MRI bases	AD/HC	92.67	**94.92**	93.80
	MCI/HC	85.85	80.99	83.30
	AD/MCI	**84.55**	88.46	86.30
Yang et al.	AD/HC	81.90	79.50	80.70
	MCI/HC	73.20	68.60	71.10
	AD/MCI	-	-	-

Table 3. Three-way classification results on test set(%).

		Sensitivity	Specificity	Accuracy
Nat.Img. bases	AD/others	**95.90**	**91.80**	
	MCI/others	**74.20**	**92.70**	**85.00**
	HC/others	**87.70**	**91.30**	
MRI bases	AD/others	81.36	89.95	
	MCI/others	67.89	87.67	78.20
	HC/others	84.96	85.86	

Table 4. Test confusion matrix for three-way classification.

Output Class \ Target Class	**HC**	**MCI**	**AD**	
HC	57 **25.1%**	13 5.7%	0 0.0%	81.4% 18.6%
MCI	7 3.1%	66 **29.1%**	3 1.3%	86.8% 13.2%
AD	1 0.4%	10 4.4%	70 **30.8%**	86.4% 13.6%
	87.7% 12.3%	74.2% 25.8%	95.9% 4.1%	**85.0%** 15.0%

Someone likely to develop AD progresses from being healthy to the point where they have mild dementia to their final stage of severe dementia. Moreover, within the MCI class, depending upon how severe the dementia is, subjects can be in *early* or *late* phases of MCI. Thus, discriminating AD from MCI seems imperative for early detection of AD. To demonstrate the effi-

cacy of our method for early detection, we performed a three-way classification task. Table 3 shows the performance of our method on this task. It also shows that natural image bases outperformed the bases learned from MRI scans. Table 4 shows the confusion matrix for 3-way classification with natural image bases.

6. Discussion

Utilizing either MRI or natural images, our approach improves over the sICA-based method (Yang et al., 2011). To compare MRI bases with natural image bases, over half of the MRI image surface area of a coronal section is white matter, which, because of its uniformity, conveys no meaningful features, except at the boundaries. However, the informative indicators for the progression of AD are unspecified putative biomarker features. Using a small database of natural images gives a much more comprehensive basis set that has the ability to capture lesion features.

To further analyse the reason for better performance, we replaced SAE in our pipeline (Algorithm 1) with ICA[5]. Using ICA bases learned from *natural images* and *fMRI* scan patches we obtained the accuracy of 67.98% and 69.29% respectively. Compared to ICA bases, the SAE bases performed (table 3) significantly better. Figure 7 shows the convolved and sigmoid activation of the median slice (figure 5(a)) using ICA bases. Due to the small weights of ICA bases, the activation falls in the linear range so there is no effect of applying the sigmoid nonlinearity. This is in contrast to figure 5(c), where more details about the lesions are revealed after applying sigmoid activation and thus, allowing the pooling layer to capture much richer features. The pooling strategy seems to have impact on the performance (figure 4) as well. In its entirety, the superior performance can be attributed to representational power of SAE, the invariance property of convolution and the pooling strategy.

However, there is more room for improvement. The various design choices were made for simplicity and computational efficiency. They lead to the following unexplored avenues:

- features are expected to be highly correlated.
- we have not exploited the 3D spatial information present in MRI imagery.
- we used only one convolutional layer with a coarse 3×3 grid for max pooling.

[5] http://research.ics.aalto.fi/ica/fastica/

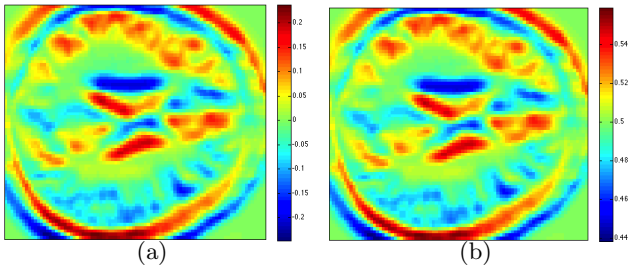


Figure 7. The effect of a ICA basis on the median (34^{th}) slice (figure 5(a)). (a) Result of convolution. (b) Feature activations using sigmoid function.

Feature correlation is the price we pay for the simplicity we adopted by considering each MRI slice individually (see figure 8 in the supplement for details). It is worth exploring the potential of various dimensionality reduction methods. Alternatively, methods to exploit the 3D spatial information that innately manifests in MRI data can be used. Adding a few more convolutional layers with finer subsampling might improve the classification performance but at the cost of increasing the computational requirements.

7. Conclusion

Visual inspection can be ineffective in identifying slight structural and metabolic changes in human brain. By focusing on an expressive bassis set for putative biomarkers, we achieve remarkably high diagnostic accuracy. Our data processing steps do not incorporate prior domain-knowledge, such as extraction of gray matter. Furthermore, we do not have to make prior choices, as opposed to 3D-SSP studies, in order to analyze statistical characteristics of data.

We have experimentally demonstrated that our approach for computerized-diagnosis of AD is i) significantly better than the method we compared with, and ii) shows promises for building a model for early detection of AD.

The HC and MCI classes can be very subtle in their differences. In spite of that, our approach, as simple as it is, was able to uncover such subtlety with very high sensitivity. This is encouraging for a future investigation of *early MCI* and *late MCI* cases.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01

AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

We would also like to thank anonymous reviewers for their helpful feedback. We are also grateful to LINC lab for allowing us to use their computing resources.

References

Bengio, Y. and Lecun, Y. *Scaling learning algorithms towards AI*. MIT Press, 2007.

Boureau, Y., Ponce, J., and LeCun, Y. A theoretical analysis of feature pooling in vision algorithms. In *ICML*, 2010.

Bourlard, H. and Kamp, Y. Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics*, 59(4):291–294, 1988.

Hinton, G. E., Osindero, S., and Teh, Y. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, July 2006. ISSN 0899-7667.

Hubel, D. H. and Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160: 106–154, 1962.

Imabayashi, E., Matsuda, H., Asada, T., Ohnishi, T., Sakamoto, S., Nakano, S., and Inoue, T. Superiority of 3-dimensional stereotactic surface projection analysis over visual inspection in discrimination of patients with very early Alzheimer's Disease from controls using brain perfusion SPECT. *Journal of Nuclear Medicine*, 45(9):1450–1457, 2004.

Kloppel, S., Stonnington, C. M., and Barnes, J. et. al. Accuracy of dementia diagnosis - a direct comparison between radiologists and a computerized method. *Brain: A Journal of Neurology*, 131(11): 2969–2974, 2008.

Krizhevsky, A., Sutskever, I., and Hinton, G. Imagenet classification with deep convolutional neural networks. In *NIPS' 25*, pp. 1106–1114. 2012.

Lecun, Y. and Bengio, Y. *Convolutional Networks for Images, Speech and Time Series*, pp. 255–258. The MIT Press, 1995.

Matsuda, H. Role of neuroimaging in Alzheimer's Disease, with emphasis on brain perfusion SPECT. *Journal of Nuclear Medicine*, 48(8):1289–1300, 2007.

Minoshima, S., Frey, K. A., Koeppe, R. A., Foster, N. L., and Kuhl, D. E. A diagnostic approach in Alzheimer's Disease using three-dimensional stereotactic surface projections of Fluorine-18-FDG PET. *Journal of Nuclear Medicine*, 36(7):1238–1248, 1995.

Olshausen, B. A. Sparse codes and spikes. In *Probabilistic Models Of The Brain: Perceptron Aand Neural Function*, pp. 257–272. MIT Press, 2001.

Raina, R., Battle, A., Lee, H., Packer, B., and Ng, A. Y. Self-taught learning: Transfer learning from unlabeled data. In *ICML*, pp. 759–766, 2007.

Ranzato, M., Poultney, C, Chopra, S., and Lecun, Y. Efficient learning of sparse representations with an energy-based model. In *NIPS*, 2006.

Teh, Y. W., M., Welling, S., Osindero, Hinton, G. E., Lee, T., F., Cardoso. J., Oja, E., and Amari, S. Energy-based models for sparse overcomplete representations. *JMLR*, 4:2003, 2003.

Yang, W., Lui, R. L. M., Gao, J. H., Chan, T. F., Yau, S., Sperling, R. A., and Huang, X. Independent component analysis-based classification of Alzheimer's Disease MRI Data. *Journal of Alzheimer's Disease*, 24(4):775–783, 2011.