Chapter 14
Deep Learning in Diagnosis of Brain Disorders

Heung-Il Suk, Dinggang Shen, and the Alzheimer's Disease Neuroimaging Initiative*

Abstract In this chapter, we introduce our recent work on neuroimaging-based AD diagnosis with machine learning techniques, especially deep learning. Specifically, we focus on the problems of feature representation and complementary information fusion from different modalities, e.g., MRI and PET. In our experimental results on the publicly available ADNI dataset, we could validate the effectiveness of the deep learning-based feature representation and its superiority to the competing methods. We also present the importance of collaborating communities of machine learning and clinical neuroscience for clinical interpretation of the learned feature representations.

Keywords Alzheimer's disease (AD) • Mild cognitive impairment (MCI) • Deep learning • Stacked auto-encoder • Deep Boltzmann machine

14.1 Introduction

As the population becomes older, the world is now facing an epidemic of dementia. Among various causes of dementia, Alzheimer's Disease (AD) is the most prevalent in elderly people, which rises significantly every year in terms of the proportion of

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI). 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 list of ADNI investigators is available at http://adni.loni.ucla.edu/wpcontent/uploads/how_to_apply/ADNI_Authorship_List.pdf.

H.-I. Suk (✉)
Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea
e-mail: hisuk@korea.ac.kr

D. Shen
Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea

Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
e-mail: dgshen@med.unc.edu

S.-W. Lee et al. (eds.), *Recent Progress in Brain and Cognitive Engineering*,
Trends in Augmentation of Human Performance, DOI 10.1007/978-94-017-7239-6_14

cause of death. A recent study by Alzheimer's Association reported that 10∼20 % of people aged 65 or older have Mild Cognitive Impairment (MCI), a prodromal stage of AD [1]. But there is no treatment to halt its progression to AD yet. In this regard, it has been one of the major issues to understand the underlying mechanisms that develop such devastative neurodegenerative disease in the fields of neuroscience, neuropsychiatry, etc.

The current scientific technologies of medical imaging, such as Magnetic Resonance Imaging (MRI) and Positron Emission Topography (PET), provide paths to investigate the structure and function of the brain in vivo. With the help of such tools, researchers have made a great leap in understanding the disease. However, the group-level analysis prevalently used for investigation and understanding of the disease is not clinically applicable for individual diagnosis. In the meantime, machine learning techniques, which can efficiently analyze the complex patterns in observations, help pave the way for a computer-aided AD diagnosis system by building computational models that can discriminate patients with the disease from healthy normal subjects.

The conventional computer-aided diagnostic systems mostly considered neuroimaging features such as voxel intensities of predefined regions, gray matter volumes, cortical thickness, to name a few, all of which can be considered as simple low-level features. From a machine learning point of view, it is beneficiary to exploit the latent high-level features inherent in data to enhance the diagnostic performance. Deep learning [3], which has already proved its effectiveness by showing promising results in various fields including speech/object recognition [9, 23] and medical imaging analysis [14, 26], can discover latent or abstract high-level information in neuroimaging data, and thus be useful for the disease diagnosis. In this chapter, we introduce our recent work on neuroimaging-based AD diagnosis with deep learning.

14.2 Background

Figure 14.1 illustrates the general framework of machine learning-based AD diagnosis, composed of four main steps, namely, (1) neuroimaging data acquisition, (2) image preprocessing (including registration, tissue segmentation, Regions-

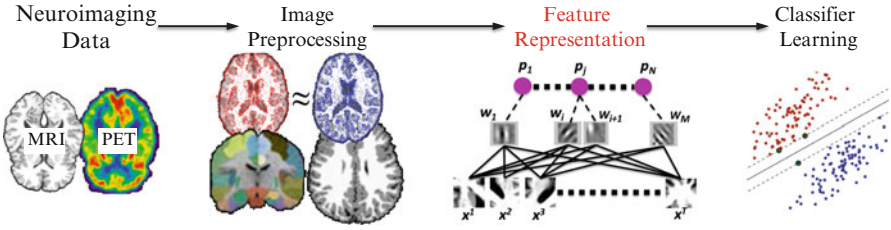

Fig. 14.1 A general framework for machine learning-based AD diagnosis using neuroimaging data

Of-Interest (ROIs) parcellation, etc.), (3) feature extraction/representation, and (4) classifier learning. Although machine learning techniques can be involved in all of these steps [11, 24], in this chapter, we focus on the step of feature representation.

As for feature extraction or representation, the existing methods can be categorized into voxel-based approach, ROI-based approach, and patch-based approach. A voxel-based approach directly uses the voxel intensities of MRI or PET as features in classification [2, 10]. Although the voxel-based approach can reflect small changes in structure or function and it is easy to interpret the results, its main limitation comes from the high-dimensionality of feature vectors and also no consideration of the inter-region relation information. On the other hand, the ROI-based approach can handle the issue of high-dimensionality by extracting representative features from the structurally or functionally parcellated brain regions. Thanks to the relatively low feature-dimension and the whole brain coverage, this approach has been most widely used in the literature [6, 12, 20, 29, 32, 34]. However, the features extracted from ROIs are very coarse in the sense that they cannot reflect small or subtle changes involved in the brain diseases. In the meantime, a patch-based approach dissects a brain into small 3D patches from which it extracts features and trains a classifier for each patch location, and then combines classifiers' outputs in a hierarchical manner [15, 31]. The patch-based approach has the advantages of (1) reflecting subtle changes by using voxel-wise features as the voxel-based approach does and (2) also considering a whole brain information as the ROI-based approach does by hierarchically integrating regional information.

14.3 Deep Learning for AD Diagnosis

Although the existing methods described in Sect. 14.2 have shown their effectiveness for AD diagnosis in the literature, they mostly used the simple low-level features without considering the high-level information latent in those features. Inspired from the biological model of the human visual cortex [7, 25], recent studies in machine learning have shown that a deep architecture composed of multiple non-linear transformations is useful to find highly non-linear and complex patterns in the data [3, 18]. This motivated us to apply deep learning techniques to neuroimaging-based AD diagnosis in [30, 31], where we used Stacked Auto-Encoder (SAE) and Deep Boltzmann Machine (DBM), respectively. In the following, we introduce these studies and further discuss the future research issues that should be tackled for clinical interpretation.

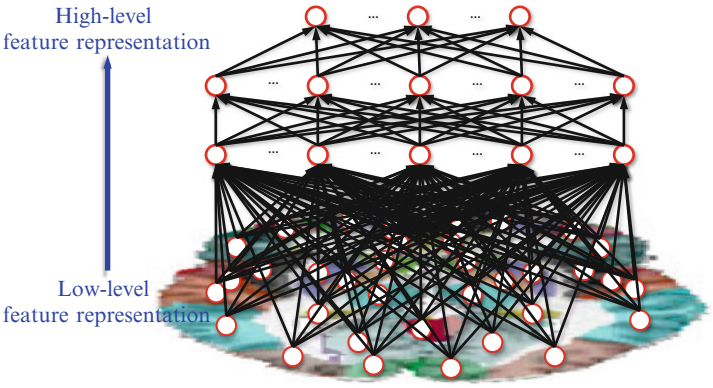

Fig. 14.2 Stacked auto-encoder that discovers the latent high-level information inherent in ROI-based features

14.3.1 Stacked Auto-Encoder (SAE)

As a pioneering study of the application of deep learning for AD diagnosis, we used an SAE [4] to discover a latent feature representation in neuroimaging or biological data. Specifically, as the name says, we stacked auto-encoders, one after another by taking the outputs from the hidden units of the lower layer as the input to the upper layer's input units, and so on. Figure 14.2 shows the network structure of our SAE model with three auto-encoders stacked. As illustrated in Fig. 14.2, in this study, we took an ROI-based approach by extracting representative features from ROIs, which set the values of the input units in the bottom layer of our SAE.

Thanks to the hierarchical nature in structure, one of the most important characteristics of the SAE is to learn or discover highly non-linear and complicated patterns such as the relations among input features. Another important characteristic of the SAE is that the latent representation can be learned directly from the data. Utilizing its representational and self-taught learning properties, we could find a latent representation of the original low-level features, directly extracted from neuroimaging data. When an input sample is presented to an SAE model, the different layers of the network represent different levels of information. That is, the lower the layer in the network, the simpler patterns (e.g., linear relations of features); the higher the layer, the more complicated or abstract patterns inherent in the input feature vector (e.g., non-linear relations among features) [17].

To find the optimal parameters, we performed unsupervised layer-wise pre-training [8] and supervised fine-tuning during the auto-encoding task via back-propagation [5, 13] sequentially. It is noteworthy that, in order to obtain the complicated non-linear relations among neuroimaging features, we considered a number of hidden units larger than the number of input features, from which we can

still find an interesting structure by imposing a sparsity constraint via a Kullback-Leibler (KL) divergence. Specifically, in our pre-training step, we optimized the following objective function:

$$E(\mathbf{Y}_{h-1}, \hat{\mathbf{Y}}_{h-1}) + \gamma \sum_{j=1}^{D_h} KL(\rho||\widehat{\rho_j}) \tag{14.1}$$

where $E(\mathbf{Y}_{h-1}, \hat{\mathbf{Y}}_{h-1})$ denotes an error between the input $\mathbf{Y}_{h-1}$ (i.e., the output from $(h-1)$-th layer) and its reconstruction $\hat{\mathbf{Y}}_{h-1}$, D_h is the number of units in the h-th hidden layer, and γ is a control parameter. In Eq. (14.1), KL divergence controls the sparseness of the hidden units based on the average activation $\hat{\rho}_j$ of the j-th hidden unit over the training samples and the target average activation ρ.

14.3.1.1 Experiments and Performance Comparison

To validate the effectiveness of the SAE-based feature representation, we conducted experiments with ADNI dataset (available at 'http://www.loni.ucla.edu/ADNI'). Specifically, we considered the baseline MRI, 18-fluoro-deoxyglucose PET, and CerebroSpinalFluid (CSF) data acquired from 51 subjects with AD, 99 subjects with MCI (including 43 progressive MCI (pMCI) and 56 stable MCI (sMCI))[1], and 52 Healthy normal Controls (HC). Along with the neuroimaging and biological data, two types of clinical scores, Mini-Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog), were also provided for each subject.

We built one SAE model[2] per modality and concatenated the original Low-Level Features (LLF) and the SAE-learned Features (SAEF) to construct an augmented feature vector (LLF+SAEF), which thus includes both low-level and high-level information. To fuse the complementary information from multiple modalities, we used a multi-kernel Support Vector Machine (SVM) [27], preceded by feature selection with a sparse regression method [33]. We considered three binary classification problems: AD vs. HC, MCI vs. HC, and pMCI vs. sMCI. In the classification of MCI vs. HC, both pMCI and sMCI data were used as the MCI class. Due to a limited small number of training samples, we applied a 10-fold cross validation technique.

We summarized the classification accuracies in Table 14.1. In AD vs. HC, compared to the accuracy of 0.970 with an LLF-based method, the proposed method improved the accuracy by 0.009. In the classification of MCI and HC, the

[1] In our work, 'progressive' and 'stable' denote whether the subjects with MCI progressed to AD in 18 months.

[2] The number of hidden units were manually determined proportional to the input dimension. As for the sparsity target and the weighting parameter of the sparsity penalty in Eq. (14.1), we set to $\rho = 0.05$ and $\gamma = 0.01$.

Table 14.1 Performance comparison in an ROI-based feature representation. As for statistical significance, a paired *t*-test was performed (LLF: Low-Level Features; SAEF: SAE-learned Feature representations; pMCI: progressive MCI; sMCI: stable MCI)

Method	AD/HC	MCI/HC	pMCI/sMCI
LLF	0.970±0.010	0.848±0.014	0.760±0.020
LLF+SAEF	**0.979±0.007**	**0.888±0.012**	**0.779±0.027**
p-value	0.0432	2.2693e-06	0.0904

proposed method showed the best classification accuracy of 0.888. The performance improvement compared to the classification accuracy of 0.848 with the LLF-based method was 0.04. In discriminating pMCI from sMCI, our method also outperformed the LLF-based method. While the LLF-based method showed the classification accuracy of 0.760, our method achieved the classification accuracy of 0.779. Based on these results, we argue that the SAE-based feature representation helped enhance the diagnostic accuracies, justifying the importance of using high-level information latent in the observation.

14.3.2 Deep Boltzmann Machine (DBM)

While an ROI-based approach helps alleviate the high-dimensionality problem in neuroimaging pattern analysis, it fails to handle subtle changes within an ROI or across ROIs. In this regard, Liu et al. proposed a patch-based approach that can efficiently handle both the high-dimensionality problem and subtle changes in an image and gradually integrated a number of local patches of a Gray Matter (GM) density map hierarchically [15]. Although they showed the efficacy of their method for AD/MCI diagnosis, it is well known that the structural or functional images are susceptible to acquisition noise, intensity inhomogeneity, artifacts, etc. Furthermore, the raw voxel density or intensity values in a patch can be considered as low-level features that do not efficiently capture more informative high-level features. To this end, we proposed a deep learning-based high-level structural and functional feature representation from MRI and PET, respectively, for AD/MCI classification. Furthermore, for multiple modalities fusion, unlike the existing methods that first extracted features from each modality independently and then mostly combined heterogeneous features via either simple feature concatenation or kernel machines, we designed a multi-modal deep learning architecture using DBM.

A DBM is structured by stacking multiple Restricted Boltzmann Machines (RBMs) in a hierarchical manner. The rationale of using DBM for feature representation is as follows: It can learn the internal latent representations that capture non-linear complicated patterns and/or statistical relations in a hierarchical manner [3, 16]. However, unlike many other deep network models such as deep belief network [8] and SAE [26], the approximate inference procedure after the initial bottom-up

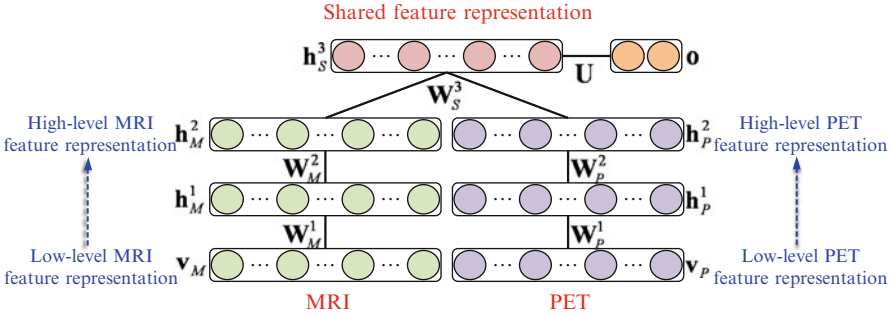

Fig. 14.3 Multimodal deep Boltzmann machine that integrates the structural and functional information and finds the shared feature representations

pass incorporates top-down feedback, which allows DBM to use higher-level knowledge to resolve uncertainty about intermediate-level features, thus creating better data-dependent representations and statistics [22]. Thanks to this two-way dependencies, i.e., bottom-up and top-down, it was shown that DBMs achieved the state-of-the-art performance in computer vision [21, 28]. To this end, we used a DBM to discover hierarchical feature representations from neuroimaging in our work.

Regarding multiple modalities fusion, different modalities will have different statistical properties. Thus, simple concatenation of the features of multiple modalities in a shallow architecture can cause strong connections among the variables of an individual modality, but failed to find inter-modality relations [19]. In order to tackle this problem, we devised a discriminative Multi-Modal DBM (MM-DBM), in which the top hidden layer had multiple entries of the lower hidden layers and the label layer, to extract a shared feature representation by fusing neuroimaging information of MRI and PET. Figure 14.3 presents a network of our MM-DBM, where one pathway represents the statistical properties of MRI and the other pathway represents those of PET, and the top shared hidden layer finally discovers the shared properties of the modalities in a supervised manner. The joint distribution over the multimodal inputs of MRI ($\mathbf{v}_M$) and PET ($\mathbf{v}_P$) and the output label ($\mathbf{o}$) can be estimated as follows:

$$P(\mathbf{v}_M, \mathbf{v}_P, \mathbf{o}; \Theta) = \sum_{\mathbf{h}_M^2, \mathbf{h}_P^2, \mathbf{h}_S} P(\mathbf{h}_M^2, \mathbf{h}_P^2, \mathbf{h}_S^3, \mathbf{o}) \left(\sum_{\mathbf{h}_M^1} P(\mathbf{v}_M, \mathbf{h}_M^1, \mathbf{h}_M^2) \right) \left(\sum_{\mathbf{h}_P^1} P(\mathbf{v}_P, \mathbf{h}_P^1, \mathbf{h}_P^2) \right) \tag{14.2}$$

where $\Theta = \{\mathbf{W}_M^1, \mathbf{W}_M^2, \mathbf{W}_P^1, \mathbf{W}_P^2, \mathbf{W}_S^3, \mathbf{U}\}$, $\mathbf{h}$ denotes a hidden layer, the subscripts M, P, and S denote, respectively, units of the MRI path, the PET path, and the shared hidden layer. For the parameters learning, we performed two consecutive steps: (1)

Table 14.2 Performance comparison in a patch-based feature representation (pMCI: progressive MCI, sMCI: stable MCI)

Method	AD/HC	MCI/HC	pMCI/sMCI
Intensity [15]	0.903±0.070	0.839±0.006	0.733±0.125
MM-DBM	**0.954±0.052**	**0.857±0.052**	**0.759±0.154**

a greedy layer-wise pre-training for a good initial setup of the model parameters and (2) iterative alternation of variational mean-field approximation to estimate the posterior probabilities of hidden units and stochastic approximation to update model parameters [22].

14.3.2.1 Experiments and Performance Comparison

We used the baseline MRI and PET data of the ADNI dataset: 93 subjects with AD, 204 subjects with MCI including 76 pMCI and 128 sMCI, and 101 HC. After conducting the preprocessing of anterior commissure-posterior commissure correction, skull stripping, cerebellum removal, registration to a common space, and tissue segmentation, we obtained spatially normalized GM volumes (i.e., GM tissue densities) and the PET images rigidly aligned to the corresponding MR images. For computational efficiency, we further down-sampled images to $64 \times 64 \times 64$ voxels. As for a patch size, we set it to $11 \times 11 \times 11$ by following Liu et al.'s work [15] for fair comparison, and thus the input dimension of each modality patch in our MM-DBM was 1,331.

We applied a 10-fold cross validation technique. The outputs of the top shared layer were used as features, which represent the fused information of structural and functional images. In order to combine the distributed patch information over an image and to build an image-level classifier, we used a hierarchical classifier learning scheme, described in [15]. In our work, we used a linear SVM for classification.

As presented in Table 14.2, in the classification of AD and HC, our method showed the mean accuracy of 0.954. Compared to the intensity-based method [15] that showed the accuracy of 0.903, we improved by 0.051. In the discrimination of MCI from HC, the proposed method achieved the accuracy of 0.857. Meanwhile, the intensity-based method [15] achieved the accuracies of 0.839. Again, the proposed method outperformed the competing method by making performance improvements of 0.018. In the classification between pMCI and sMCI, which is the most important for early diagnosis and treatment, the intensity-based method achieved the accuracy of 0.733. Compared to this result, our method improved the accuracy by 0.026. Concisely, in our three binary classifications, based on the classification accuracy, our deep learning-based method clearly outperformed the competing method.

14.4 Discussions and Conclusions

We applied deep learning methods for high-level feature representations and validated their efficacy by showing their superiority to the competing methods in terms of the diagnostic performance. Specifically, in Sect. 14.3.1, we applied an SAE to discover latent relations among the ROI-based features and then combined multiple modalities via kernel machine. Although the SAE model can be considered as the conventional multi-layer neural network, by initializing our model parameters with a greedy layer-wise pre-training and then fine-tuning the whole model, we could learn parameters to represent the inherent information better. Meanwhile, in Sect. 14.3.2, we devised a systematic method for a joint feature representation with an MM-DBM. Unlike the SAE model that learns parameters in a top-down manner, the DBM finds the optimal parameters in a bi-directional (i.e., bottom-up and top-down) manner. We utilized this favorable characteristic and successfully applied to find the shared representation from MRI and PET.

However, from a neurophysiological perspective, it is still very hard or impossible to interpret the learned representations and to understand the trained model parameters. In other words, there is no general or intuitive way to interpret the latent feature representations or the trained models. The problem of effective interpretation of the latent feature representations is a big challenge that should be tackled by the communities of machine learning and clinical neuroscience collaboratively. Furthermore, to our best knowledge, the existing disease diagnosis systems including ours output simply the clinical status of a testing subject, e.g., AD, MCI, or HC, with no presentation of the basis that supports their decision. In other words, when a subject is identified by a diagnosis system as a patient with either AD or MCI, it is clinically important to present its basis for the decision, e.g., structurally abnormal brain regions or abnormal functional connectivities observed in the neuroimaging data. Thus, all these would be our forthcoming research issues.

Acknowledgements This chapter follows closely the prior published papers [30, 31] by the authors. This work was supported in part by NIH grants EB006733, EB008374, EB009634, AG041721, MH100217, and AG042599, and partial supported by ICT R&D program of MSIP/IITP [B0101-15-0307, Basic Software Research in Human-level Lifelong Machine Learning (Machine Learning Center)].

References

1. Alzheimer's Association (2012) 2012 Alzheimer's disease facts and figures. Alzheimer's & Dement 8(2):131–168
2. Baron J, Chételat G, Desgranges B, Perchey G, Landeau B, de la Sayette V, Eustache F (2001) In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. NeuroImage 14(2):298–309
3. Bengio Y (2009) Learning deep architectures for AI. Found Trends Mach Learn 2(1):1–127

4. Bengio Y, Lamblin P, Popovici D, Larochelle H (2007) Greedy layer-wise training of deep networks. In: Schölkopf B, Platt J, Hoffman T (eds) Advances in neural information processing systems, vol 19. MIT Press, Cambridge, pp 153–160
5. Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press, New York
6. Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ (2011) Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol Aging 32(12):2322.e19–2322.e27
7. Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biolog Cybern 36(4):93–202
8. Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554
9. Hinton G, Deng L, Yu D, rahman Mohamed A, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Dahl TSG, Kingsbury B (2012) Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Process Mag 29(6):82–97
10. Ishii K, Kawachi T, Sasaki H, Kono AK, Fukuda T, Kojima Y, Mori E (2005) Voxel-based morphometric comparison between early- and late-onset mild Alzheimer's disease and assessment of diagnostic performance of z score images. Am J Neuroradiol 26:333–340
11. Kim M, Wu G, Wang Q, Lee SW, Shen D (2015) Improved image registration by sparse patch-based deformation estimation. NeuroImage 105:257–268
12. Kohannim O, Hua X, Hibar DP, Lee S, Chou YY, Toga AW Jr, Jack CR, Weiner MW, Thompson PM (2010) Boosting power for clinical trials using classifiers based on multiple biomarkers. Neurobiol Aging 31(8):1429–1442
13. LeCun Y, Bottou L, Orr G, Müller KR (1998) Efficient backprop. In: Orr G, Müller KR (eds) Neural networks: tricks of the trade. Lecture notes in computer science, vol 1524. Springer, Berlin/Heidelberg, pp 9–50
14. Liao S, Gao Y, Oto A, Shen D (2013) Representation learning: a unified deep learning framework for automatic prostate MR segmentation. In: Medical image computing and computer-assisted intervention (MICCAI 2013), Nagoya. Lecture notes in computer science, vol 8150, pp 254–261
15. Liu M, Zhang D, Shen D, the Alzheimer's Disease Neuroimaging Initiative (2013) Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis. Hum Brain Mapp 35(4):1305–1319
16. Mohamed A, Dahl GE, Hinton GE (2012) Acoustic modeling using deep belief networks. IEEE Trans Audio Speech Lang Process 20(1):14–22
17. Montavon G, Braun ML, Müller KR (2011) Kernel analysis of deep networks. J Mach Learn Res 12:2563–2581
18. Montavon G, Orr GB, Müller KR (eds) (2012) Neural networks: tricks of the trade. Lecture notes in computer science, vol 7700, 2nd edn. Springer, Berlin/Heidelberg
19. Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AY (2011) Multimodal deep learning. In: Proceedings of the 28th international conference on machine learning, Bellevue, pp 689–696
20. Nordberg A, Rinne JO, Kadir A, Langstrom B (2010) The use of PET in Alzheimer disease. Nat Rev Neurol 6(2):78–87
21. Salakhutdinov R, Hinton GE (2009) Deep Boltzmann machines. In: Proceedings of the international conference on artificial intelligence and statistics, Clearwater Beach, pp 448–455
22. Salakhutdinov R, Hinton G (2012) An efficient learning procedure for deep Boltzmann machines. Neural Comput 24(8):1967–2006
23. Salakhutdinov R, Tenenbaum J, Torralba A (2013) Learning with hierarchical-deep models. IEEE Trans Pattern Anal Mach Intell 35(8):1958–1971
24. Sanroma G, Wu G, Gao Y, Shen D (2014) Learning to rank atlases for multiple-atlas segmentation. IEEE Trans Med Imaging 33(10):1939–1953
25. Serre T, Wolf L, Poggio T (2005) Object recognition with features inspired by visual cortex. In: Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition, San Diego, vol 2, pp 994–1000

26. Shin HC, Orton MR, Collins DJ, Doran SJ, Leach MO (2013) Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. IEEE Trans Pattern Anal Mach Intell 35(8):1930–1943
27. Sonnenburg S, Rätsch G, Schäfer C, Schölkopf B (2006) Large scale multiple kernel learning. J Mach Learn Res 7:1531–1565
28. Srivastava N, Salakhutdinov R (2012) Multimodal learning with deep Boltzmann machines. In: Pereira F, Burges C, Bottou L, Weinberger K (eds) Advances in neural information processing systems, vol 25, Curran Associates, Inc., pp 2231–2239
29. Suk HI, Lee SW (2013) A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces. IEEE Trans Pattern Anal Mach Intell 35(2):286–299
30. Suk HI, Lee SW, Shen D (2015) Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct Funct 220(2):841–859
31. Suk HI, Lee SW, Shen D (2014) Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. NeuroImage 101:569–582
32. Suk HI, Lee SW, Shen D (2014) Subclass-based multi-task learning for Alzheimer's disease diagnosis. Front Aging Neurosci 6(168):1–12
33. Yuan M, Lin Y (2006) Model selection and estimation in regression with grouped variables. J R Stat Soc Ser B 68(1):49–67
34. Zhang D, Shen D (2012) Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 59(2):895–907