



Use of a Sparse-Response Deep Belief Network and Extreme Learning Machine to Discriminate Alzheimer's Disease, Mild Cognitive Impairment, and Normal Controls Based on Amyloid PET/MRI Images

Ping Zhou, Shuqing Jiang, Lun Yu, Yabo Feng, Chuxin Chen, Fang Li, Yang Liu and Zhongxiong Huang* on behalf of the Alzheimer's Disease Neuroimaging Initiative

Department of PET-CT Center, Chenzhou No.1 People's Hospital, Chenzhou, China

OPEN ACCESS

Edited by:

Chuantao Zuo,
Fudan University, China

Reviewed by:

Woon-Man Kung,
Chinese Culture University, Taiwan
Juanjuan Jiang,
Shanghai University, China

*Correspondence:

Zhongxiong Huang
hzx1968@yeah.net

Specialty section:

This article was submitted to
Nuclear Medicine,
a section of the journal
Frontiers in Medicine

Received: 25 October 2020

Accepted: 26 November 2020

Published: 18 January 2021

Citation:

Zhou P, Jiang S, Yu L, Feng Y,
Chen C, Li F, Liu Y and Huang Z
(2021) Use of a Sparse-Response
Deep Belief Network and Extreme
Learning Machine to Discriminate
Alzheimer's Disease, Mild Cognitive
Impairment, and Normal Controls
Based on Amyloid PET/MRI Images.
Front. Med. 7:621204.
doi: 10.3389/fmed.2020.621204

In recent years, interest has grown in using computer-aided diagnosis (CAD) for Alzheimer's disease (AD) and its prodromal stage, mild cognitive impairment (MCI). However, existing CAD technologies often overfit data and have poor generalizability. In this study, we proposed a sparse-response deep belief network (SR-DBN) model based on rate distortion (RD) theory and an extreme learning machine (ELM) model to distinguish AD, MCI, and normal controls (NC). We used [¹⁸F]-AV45 positron emission computed tomography (PET) and magnetic resonance imaging (MRI) images from 340 subjects enrolled in the ADNI database, including 116 AD, 82 MCI, and 142 NC subjects. The model was evaluated using five-fold cross-validation. In the whole model, fast principal component analysis (PCA) served as a dimension reduction algorithm. An SR-DBN extracted features from the images, and an ELM obtained the classification. Furthermore, to evaluate the effectiveness of our method, we performed comparative trials. In contrast experiment 1, the ELM was replaced by a support vector machine (SVM). Contrast experiment 2 adopted DBN without sparsity. Contrast experiment 3 consisted of fast PCA and an ELM. Contrast experiment 4 used a classic convolutional neural network (CNN) to classify AD. Accuracy, sensitivity, specificity, and area under the curve (AUC) were examined to validate the results. Our model achieved 91.68% accuracy, 95.47% sensitivity, 86.68% specificity, and an AUC of 0.87 separating between AD and NC groups; 87.25% accuracy, 79.74% sensitivity, 91.58% specificity, and an AUC of 0.79 separating MCI and NC groups; and 80.35% accuracy, 85.65% sensitivity, 72.98% specificity, and an AUC of 0.71 separating AD and MCI groups, which gave better classification than other models assessed.

Keywords: computer-aided diagnosis, Alzheimer's disease, mild cognitive impairment, sparse-response deep belief network, extreme learning machine

INTRODUCTION

Alzheimer’s disease (AD) is a neurodegenerative disease characterized by cognitive dysfunction and associated with advanced age. Because there are currently no therapies that can reverse the course of AD, it is important to diagnose AD and its prodromal stage, mild cognitive impairment (MCI) as early as possible (1).

In recent years, neuroimaging techniques have been shown to be effective tools for the diagnosis of AD. Magnetic resonance imaging (MRI) and positron emission tomography (PET) are two common neuroimaging methods. For example, Hua et al. proposed a powerful tool to monitor structural atrophy in incipient stages of AD using MR images (2). Mosconi et al. demonstrated that PET scans may provide objective and sensitive support to clinical diagnosis in early dementia (3). In addition, deep learning methods have shown great promise for image analysis and disease prediction. For instance, Hu et al. utilized a targeted autoencoder network to classify functional connectivity matrices across brain regions, which was able to distinguish MCI from NC with 87.5% accuracy (4). Liu et al. designed a deep learning architecture to more accurately differentiate AD, MCI, and normal controls (NC). The architecture, including stacked autoencoders and a softmax output layer, achieved 87.76% accuracy, 88.57% sensitivity, and 87.22% specificity distinguishing AD from NC and exhibited 76.92% accuracy, 74.29% sensitivity, and 78.13% specificity distinguishing MCI from NC (5). In addition, a few of deep learning studies based on PET/MRI could also be observed (6, 7).

However, the methods mentioned above had some disadvantages. For instance, gradient diffusion and gradient explosion may emerge with deepening of the autoencoder stack depth, resulting in decreased classification accuracy. To mitigate this limitation, we proposed a sparse-response deep belief network (SR-DBN) based on the rate distortion (RD) theory model. Our SR-DBN used the contrastive divergence algorithm to maximize the retention of data distribution, in case gradient diffusion and gradient explosion became factors. In addition, the SR-DBN model included sparsity. Compared to DBN models without sparsity, sparse representations allow changing the significant bits for each example in a fixed-size representation, which are more efficient from the point of view of information theory (8). Subsequently, we used an extreme learning machine (ELM) as a classifier to get the performance of the classification. Meanwhile, to evaluate the effectiveness of our method, we compared our model with other models.

Abbreviations: AD, Alzheimer’s disease; ADNI, Alzheimer’s disease neuroimaging initiative; AUC, area under curve; CAD, computer-aided diagnosis; CNN, convolutional neural network; ELM, extreme learning machine; GM, gray matter; MCI, mild cognitive impairment; MRI, magnetic resonance imaging; NC, normal controls; PCA, principal components analysis; PET, positron emission computed tomography; RBM, restricted Boltzmann machine; RD, rate distortion; SR-DBN, sparse-response deep belief network; SVM, support vector machine; WM, white matter.

MODEL DESIGN

Model Framework

As shown in **Figure 1**, the framework of the model consists of four parts: (1) original image data underwent standard preprocessing; (2) data dimensionality was reduced using fast principal component analysis (PCA); (3) features were extracted by three SR-DBNs based on rate distortion theory; and (4) processed data were classified by the ELM.

Mathematical Fundamentals of the Proposed Model

SR-DBN Model Based on RD Theory

Restricted Boltzmann machines (RBM) are neural perceptrons composed of a visible layer and a hidden layer. Several RBMs can form a DBN. Similar to the structure of a DBN, the SR-DBN also consists of several sparse-response restricted Boltzmann machines (SR-RBMs). In the model, the Kullback–Leibler divergence $KL(p^0 \| p_\theta^\infty)$ (9) between the original data’s distribution p^0 and the equilibrium distribution p_θ^∞ defined by RBM served as a distortion function. Considering the RD theory, we can deduce the following formulation:

$$\min_{w_{ij}, c_i, b_j} KL p^0 \| p_\theta^\infty + \lambda \sum_{l=1}^m \left\| p \left(\mathbf{h}^{(l)} | \mathbf{v}^{(l)} \right) \right\|_1 \quad (1)$$

where $\sum_{l=1}^m \left\| p \left(\mathbf{h}^{(l)} | \mathbf{v}^{(l)} \right) \right\|_1$ denotes the sparseness of representation and λ is a regularization parameter. Then we replaced Kullback–Leibler divergence with $p^0 \| p_\theta^\infty - p_\theta^1 \| p_\theta^\infty$ to simplify calculations (10). Suppose w is the weight matrix of RBM, b is the bias vector of the input layer, and c is the bias vector of the output layer, giving the updated rules below:

$$\begin{cases} w_{ij} := w_{ij} + \epsilon \left(\langle v_i h_j \rangle_{p^0} - \langle v_i h_j \rangle_{p_\theta^1} \right) \\ c_i := c_i + \epsilon \left(\langle v_i \rangle_{p^0} - \langle v_i \rangle_{p_\theta^1} \right) \\ b_j = b_j + \epsilon \left(\langle h_j \rangle_{p^0} - \langle h_j \rangle_{p_\theta^1} \right) \end{cases} \quad (2)$$

where ϵ denotes a learning rate. Additionally, we added another update with the gradient of the regularization term in each iteration. The term is as follows:

$$\begin{aligned} - \frac{\partial}{\partial w_{ij}} \sum_{l=1}^m \left\| p \left(\mathbf{h}^{(l)} | \mathbf{v}^{(l)} \right) \right\|_1 &= - \sum_{l=1}^m \frac{\partial}{\partial w_{ij}} \text{sigmoid} \left(b_j + \sum_i w_{ij} v_i^{(l)} \right) \\ &= - \sum_{l=1}^m p_j^{(l)} \left(1 - p_j^{(l)} \right) v_i^{(l)} \end{aligned} \quad (3)$$

$$\begin{aligned} - \frac{\partial}{\partial b_j} \sum_{l=1}^m \left\| p \left(\mathbf{h}^{(l)} | \mathbf{v}^{(l)} \right) \right\|_1 &= - \sum_{l=1}^m \frac{\partial}{\partial w_{ij}} \text{sigmoid} \left(b_j + \sum_i b_j v_i^{(l)} \right) \\ &= - \sum_{l=1}^m p_j^{(l)} \left(1 - p_j^{(l)} \right) \end{aligned} \quad (4)$$

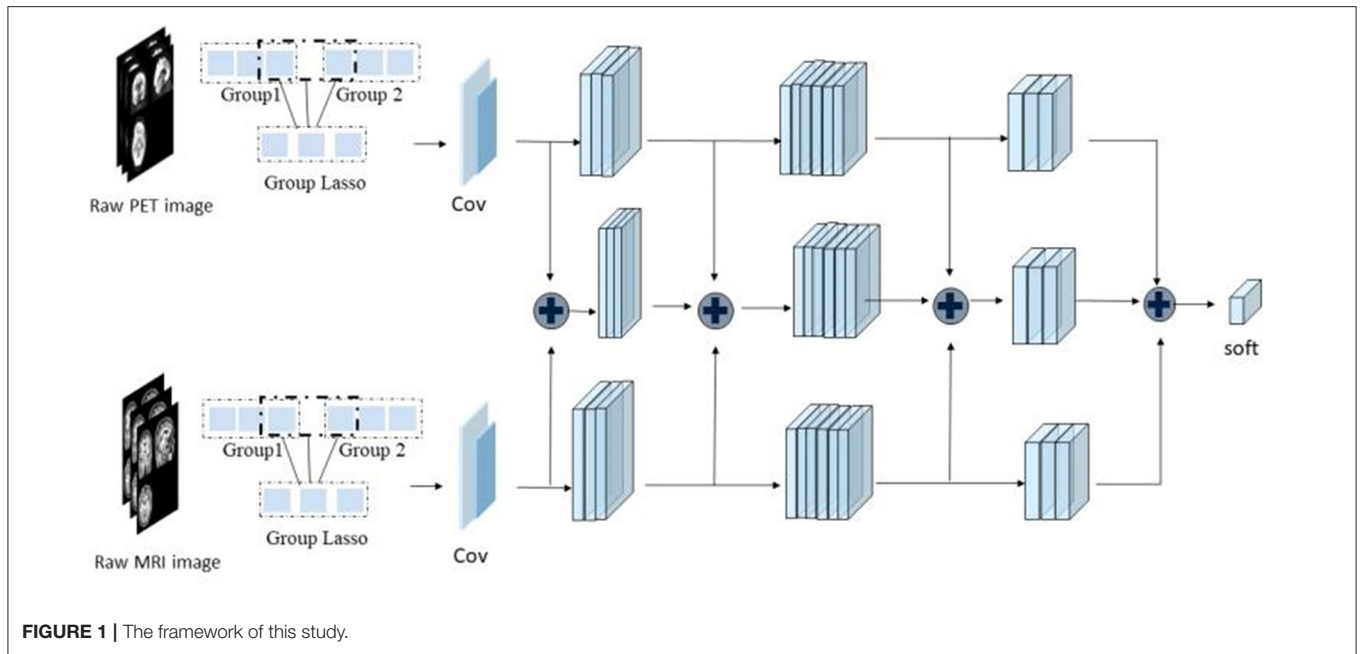


FIGURE 1 | The framework of this study.

where $p_j^{(l)} = \text{sigmoid}\left(\sum_i v_i^{(l)} w_{ij} + b_j\right)$, and $\text{sigmoid}(\cdot)$ represents the sigmoid function.

In this study, we employed one input layer, three hidden layers and one output layer.

ELM Model for Classification

An ELM is a neural network algorithm for a single hidden layer feedforward neural network. Its input weights and hidden node bias are generated randomly within a given range. The only optimal solution can be obtained by setting the number of hidden layer neurons (11). When the input weights and hidden layer bias are determined randomly, the output matrix of the hidden layer, H is also determined (12):

$$\hat{\beta} = H^+ T \tag{5}$$

where H^+ is the Moore–Penrose pseudoinverse matrix of H and the T notes the expected output.

MATERIALS AND METHODS

Materials

The data used in this study were access through the Alzheimer’s Disease Neuroimaging Initiative (ADNI) public database. ADNI is a consortium study initiated in 2004 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations (13). For additional information about ADNI, please see www.adni-info.org.

TABLE 1 | The clinical data of three cohorts.

	AD (n = 116)	MCI (n = 82)	NC (n = 142)
Gender (M/F)	68/48	50/32	52/90
Ages (years)	75.37 ± 5.61	75.6 ± 5.87	73.86 ± 7.06
MMSE	21.63 ± 3.97	23.8 ± 5.98	28.92 ± 1.27
MOCA	15.89 ± 5.85	18.8 ± 6.49	26.01 ± 2.76

In this study, we selected AV45 PET and structural MRI images of 340 subjects enrolled in ADNI, including 116 AD, 82 MCI, and 142 NC subjects. The clinical data for each of these diagnostic groups is shown in **Table 1**.

Image Preprocessing

MRI data were acquired on multiple 3T MRI scanners using scanner-specific T1-weighted sagittal 3D MPRAGE sequences. In order to increase signal uniformity across the multicenter scanner platforms, original MPRAGE acquisitions underwent standardized image preprocessing steps. The current study implemented the following steps: (1) segmentation of the images into gray matter (GM), white matter (WM) and cerebrospinal fluid (14), of which gray matter and white matter were used for further analysis; (2) normalization of all GM and WM images into Montreal Neurological Institute space; and (3) spatial smoothing using a Gaussian kernel of 4 mm³.

[18F]-AV45 PET data were acquired on multiple instruments of varying resolutions and following different platform-specific acquisition protocols. Similar to the MRI data, ADNI PET data underwent standardized image preprocessing steps aimed at increasing data uniformity across the multicenter acquisitions (15). The preprocessing steps included realignment, spatial

normalization to MNI space, and smoothing using a 7-mm³ Gaussian kernel. We performed a voxel-based partial volume effects correction of the normalized functional image using the Müller-Gärtner method. Lastly, the partial volume effect-corrected functional image was smoothed to reduce noise and improve image quality using an isotropic Gaussian smoothing kernel with a full width at half maximum setting of 7 mm³. The image was scaled up to obtain a standard uptake value rate map of the entire cerebellum.

Both MRI and ¹⁸F-AV45 PET images were preprocessed using statistical parametric mapping software (SPM12, <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>) on Matlab 2016b.

Dimension Reduction and Feature Extraction

Fast PCA was used to describe the data with a small number of linearly independent features under the principle of ensuring the minimum loss of data information.

In the study, the SR-DBN model undertook feature extraction. Compared with DBN, the SR-DBN is more efficient from the perspective of information theory, which allows changing the effective number of bits per example in a fixed size representation (8).

The SR-DBN model used in the study was made up of multiple basic SR-RBMs with the same numbers of nodes. The output of each SR-RBM was the input of the next basic SR-RBM at successive levels. In the last layer of the SR-DBN model, a back propagation network was set, receiving the output feature vector of SR-RBM as learned features, and adopting a gradient descent algorithm to fine-tune the weight of the whole network, thereby coordinating and optimizing the parameters of the whole SR-DBN.

Classification & Comparative Experiments

Three kinds of images were used as input: [¹⁸F]-AV45 PET, and GM and WM segmentations from the MRI. Correspondingly, we used three SR-DBNs to extract features. Following feature extraction, the ELM classified the three diagnostic groups. After obtaining the predicted labels, the accuracy, sensitivity, specificity, and area under curve (AUC) were calculated to evaluate the practicability of the model.

The model was evaluated using five-fold cross-validation, repeated 200 times. In the case of “lucky trails,” we randomly sampled the training and testing instances from each class to ensure they had similar distributions as the original dataset. The entire network was trained and fine-tuned with 80% of the data and then tested with the remaining 20% of the samples in each validation trial.

To evaluate the effectiveness of our method, we performed several comparative trials. In contrast experiment 1, ELM was replaced by a support vector machine (SVM). Contrast experiment 2 utilized DBN without sparsity. Contrast experiment 3 consisted of fast PCA and ELM. Contrast experiment 4 used a classic convolutional neural network (CNN) to classify AD, MCI, and NC. The experimental platform is based on Matlab 2016b.

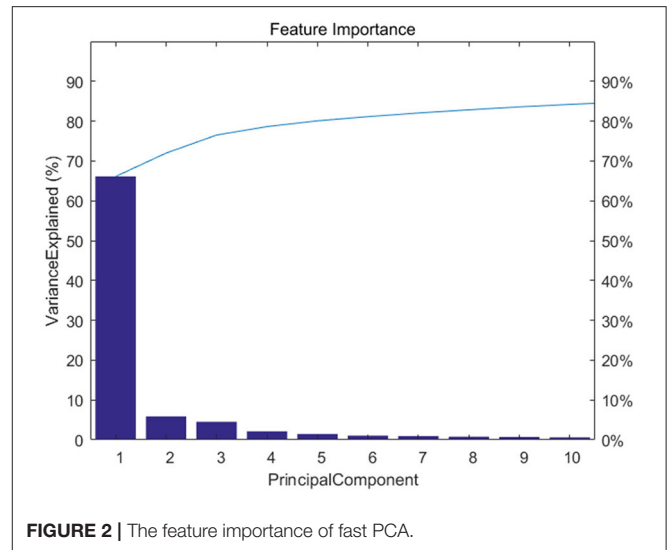


FIGURE 2 | The feature importance of fast PCA.

RESULTS

Results of Dimension Reduction

Figure 2 shows feature importance. We extracted the top 20 features which represent 90% information of the original data.

Results of Classification

The classification and comparative results are shown in Table 2 and Figure 3. In the classification of AD and NC, our model achieved 91.68% accuracy, 95.47% sensitivity, 86.68% specificity, and an AUC of 0.87. In the classification between MCI and NC, the model achieved 87.25% accuracy, 79.74% sensitivity, 91.58% specificity, and an AUC of 0.79. When separating between AD and MCI, the model achieved 80.35% accuracy, 85.65% sensitivity, 72.98% specificity, and an AUC of 0.71. Moreover, the time cost for image processing and classification in our proposed method and four compared methods were 36.2 s, 37.4 s, 491.7 s, 25 s, and 1,386.5 s. This result means that our method is faster than the classical CNN model, and similar to machine learning models.

Table 3 shows the comparative results of our model and results from the literature, including Hu’s model, Liu’s model, and Suk’s model (4, 5, 16). Specifically, Hu’s model used a single image modality (MRI) and Liu’s model used both MRI and PET. Our proposed model achieved the best classification result of all models compared.

DISCUSSION

In this paper, we used a SR-DBN and ELM for the classification of AD, MCI, and CN. In Table 2 and Figure 3, the superiority of our model compared to other models can be seen, as evidenced by the highest values for accuracy, sensitivity, specificity, and AUC.

Table 3 presents a comparison of our model with previous deep learning models from the literature. Hu’s model, Liu’s model, and Suk’s model adopted the stacked autoencoders and softmax classifier to classify AD. The thickness of the method

TABLE 2 | Results of the experiment.

		Proposed	Experiment 1	Experiment 2	Experiment 3	Experiment 4
AD vs. NC	ACC (%)	91.68 ± 1.09	85.68 ± 1.6	86.28 ± 1.02	76.43 ± 0.5	77.12 ± 0.04
	SEN (%)	95.47 ± 1.73	90.58 ± 2.43	90.92 ± 1.19	80.01 ± 0.62	74.02 ± 0.08
	SPE (%)	86.68 ± 2.61	79.6 ± 3.81	80.31 ± 2.24	71.99 ± 0.76	79.53 ± 0.13
	AUC	0.87 ± 0.01	0.82 ± 0.01	0.83 ± 0.02	0.75 ± 0.02	0.77 ± 0.03
MCI vs NC	ACC (%)	88.25 ± 1.38	80.34 ± 1.79	80.18 ± 1.03	67.64 ± 0.54	63.15 ± 0.02
	SEN (%)	79.74 ± 3.44	68.85 ± 4.77	68.33 ± 3.16	45.72 ± 1.1	78 ± 0.09
	SPE (%)	91.58 ± 2.14	87.19 ± 2.73	87.22 ± 1.55	80.05 ± 0.53	41.96 ± 0.11
	AUC	0.79 ± 0.01	0.72 ± 0.03	0.73 ± 0.03	0.60 ± 0.04	0.60 ± 0.02
AD vs MCI	ACC (%)	80.35 ± 1.8	73.07 ± 2.52	71.95 ± 1.61	65.77 ± 0.6	63.71 ± 0.03
	SEN (%)	85.65 ± 3.63	76.3 ± 3.93	78.51 ± 2.42	74.49 ± 0.72	79.44 ± 0.08
	SPE (%)	72.98 ± 4.71	69.41 ± 5.79	63.33 ± 4.06	54.15 ± 1.07	41.27 ± 0.09
	AUC	0.71 ± 0.08	0.69 ± 0.01	0.68 ± 0.02	0.59 ± 0.04	0.6 ± 0.02

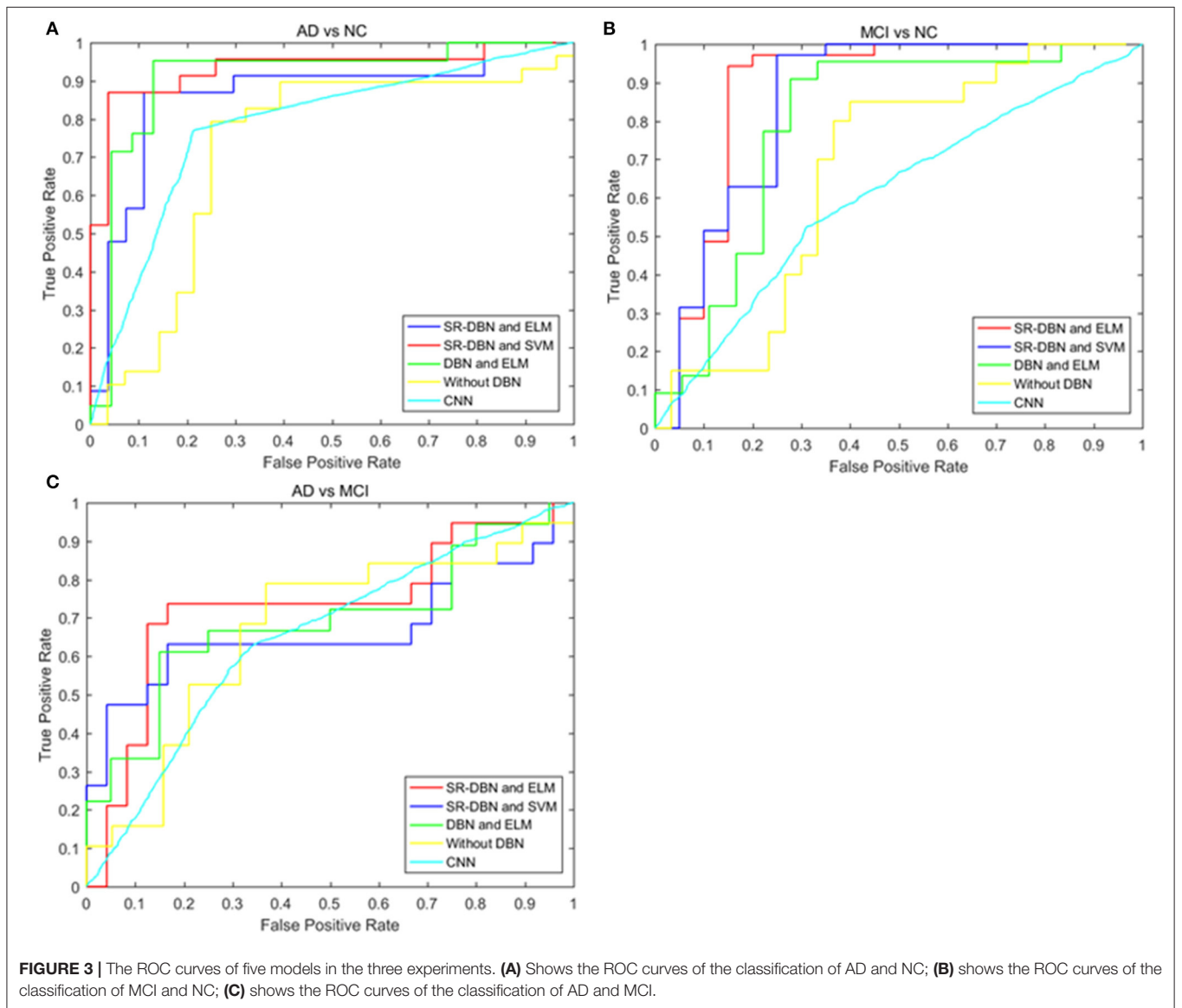


TABLE 3 | Comparison of classification performances (%).

	AD vs. NC			MCI vs. NC		
	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)
Our model	91.68	95.47	86.68	88.25	79.74	91.58
Hu's model	-	-	-	87.5	-	-
Liu's model	87.76	88.57	87.22	76.92	74.29	78.13
Suk's model	83.2	-	-	70.1	-	-

likely contributed to gradient diffusion and gradient explosion, which was successfully avoided by using CD in our model. In addition, Hu's model used a single imaging modality (MRI) and Liu's model was a multimodal example. As shown in **Table 3**, the performance of our model was superior to the two models, reflecting the potential utility of our model to aid in early AD diagnosis.

However, the study has several limitations. Firstly, the parameters of the model ought to be modified to obtain better performance. Secondly, the method is based on multimodal data, but subjects with missing image data points are excluded, limiting the sample size. Thirdly, we only compared classification results among our proposed SR-DBN model, machine learning models, and classical CNN models in our dataset. Other deep learning models, such as recurrent neural networks, and deep neural network models were not compared in the same dataset. They will be implemented and compared in the future. Finally, the data used were from Western patients, which could potentially affect the results. Data from Eastern patients should be included in future studies to optimize the model and make it more generalizable to Eastern populations.

CONCLUSION

In the study, we proposed a SR-DBN combined with ELM to classify AD, MCI, and NC. Our model achieved 91.68% accuracy, 95.47% sensitivity, 86.68% specificity, and an AUC of 0.87 on the classification between AD and NC participants; 87.25% accuracy, 79.74% sensitivity, 91.58% specificity, and an AUC of 0.79 on classification between MCI and NC participants; and 80.35% accuracy, 85.65% sensitivity, 72.98% specificity, and an AUC of 0.71 on the classification between AD and MCI patients. Our model obtained better classification compared other models examined, indicating its effectiveness.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database.

REFERENCES

- Zhu X, Suk HI, Wang L, Lee SW, Initiative AsDN. A novel relational regularization feature selection method for joint regression and

AUTHOR CONTRIBUTIONS

PZ, SJ, and LY are responsible for writing experimental procedures, organizing experimental results and writing the paper. YF, CC, FL, and YL are responsible for experimental data collection and preprocessing. ZH was responsible for proposing experimental plans and guiding the writing of the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by grants; this program is sponsored by scientific development projects from ChenZhou Municipal Science and Technology Bureau (No. yfzx201906).

Data collection and dissemination for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI): the National Institutes of Health (grant number U01 AG024904), and the Department of Defense (award number W81XWH-12-2-0012). ADNI was funded by the National Institute of Aging and the National Institute of Biomedical Imaging and Bioengineering as well as through generous contributions from the following organizations: AbbVie, Alzheimer's Association, Alzheimer's Drug Discovery Foundation, Araclon Biotech, BioClinica Inc., Biogen, Bristol-Myers Squibb Company, CereSpir Inc., Eisai Inc., Elan Pharmaceuticals Inc., Eli Lilly and Company, EuroImmun, F. Hoffmann-La Roche Ltd. and its affiliated company Genentech Inc., Fujirebio, GE Healthcare, IXICO Ltd., Janssen Alzheimer Immunotherapy Research & Development LLC., Johnson & Johnson Pharmaceutical Research & Development LLC., Lumosity, Lundbeck, Merck & Co. Inc., Meso Scale Diagnostics LLC., NeuroRx Research, Neurotrack Technologies, Novartis Pharmaceuticals Corporation, Pfizer Inc., Piramal Imaging, Servier, Takeda Pharmaceutical Company, and Transition Therapeutics. The Canadian Institutes of Health Research are 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 was coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego, CA, USA. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California, CA, USA.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to EditSprings (<https://www.editsprings.com/>) for the expert linguistic services provided.

classification in AD diagnosis. *Med Image Anal.* (2015) 75:570–7. doi: 10.1016/j.media.2015.10.008

- Hua X, Leow AD, Parikshak N, Lee S, Chiang M-C, Toga AW, et al. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease:

- an MRI study of 676 AD, MCI, and normal subjects. *Neuroimage*. (2008) 43:458–69. doi: 10.1016/j.neuroimage.2008.07.013
3. Mosconi L, Tsui WH, Herholz K, Pupi A, Drzezga A, Lucignani G, et al. Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *J Nucl Med*. (2008) 49:390–8. doi: 10.2967/jnumed.107.045385
 4. Hu C, Ju R, Shen Y, Zhou P, Li Q. Clinical decision support for Alzheimer's disease based on deep learning and brain network. In: *2016 IEEE International Conference on Communications (ICC)*. (2016).
 5. Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D. Early diagnosis of Alzheimer's disease with deep learning. In: *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. (2014).
 6. Shen T, Jiang JH, Lin W, Ge JJ, Wu P, Zhou YJ, et al. Use of overlapping group LASSO sparse deep belief network to discriminate Parkinson's disease and normal control. *Front Neurosci*. (2019) 13:396. doi: 10.3389/fnins.2019.00396
 7. Wang M, Jiang JH, Yan ZZ, Alberts I, Ge JJ, Zhang HW, et al. Individual brain metabolic connectome indicator based on Kullback-Leibler divergence similarity estimation predicts progression from mild cognitive impairment to Alzheimer's dementia. *Eur J Nucl Med Mol Imaging*. (2020) 47:2753–64. doi: 10.1007/s00259-020-04814-x
 8. Ranzato MA, Boureau Y-L, Cun YL. Sparse feature learning for deep belief networks. *Adv Neural Inf Process Syst*. (2007) 20:1185–92.
 9. Hinton GE. Training products of experts by minimizing contrastive divergence. *Neural Comput*. (2002) 14:1771–800. doi: 10.1162/089976602760128018
 10. Ji N-N, Zhang J-S, Zhang C-X. A sparse-response deep belief network based on rate distortion theory. *Pattern Recognit*. (2014) 47:3179–91. doi: 10.1016/j.patcog.2014.03.025
 11. Liang D, Pan P. Research on intrusion detection based on improved DBN-ELM. In: *2019 International Conference on Communications, Information System and Computer Engineering (CISCE)*, Haikou (2019).
 12. Ribeiro B, Lopes N. Extreme learning classifier with deep concepts. In: *Iberoamerican Congress on Pattern Recognition*, Havana (2014).
 13. Zhang D, Shen D, Initiative AsDN. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage*. (2012) 59:895–907. doi: 10.1016/j.neuroimage.2011.09.069
 14. Gonzalez-Escamilla G, Lange C, Teipel S, Buchert R, Grothe MJ, Initiative A, et al. PETPVE12: an SPM toolbox for partial volume effects correction in brain PET—application to amyloid imaging with AV45-PET. *Neuroimage*. (2017) 147:669–77. doi: 10.1016/j.neuroimage.2016.12.077
 15. Teipel SJ, Kurth J, Krause B, Grothe MJ, Initiative AsDN. The relative importance of imaging markers for the prediction of Alzheimer's disease dementia in mild cognitive impairment—beyond classical regression. *Neuroimage Clin*. (2015) 8:583–93. doi: 10.1016/j.nicl.2015.05.006
 16. Suk HI, Shen D. Deep learning-based feature representation for ad/mci classification. *Med Image Comput Comput Assist Interv*. (2013) 16(Pt 2) 583–90. doi: 10.1007/978-3-642-40763-5_72

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Zhou, Jiang, Yu, Feng, Chen, Li, Liu and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.