Journal of Biomedical Informatics 121 (2021) 103863

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Original Research

Alzheimer's disease diagnosis framework from incomplete multimodal data using convolutional neural networks

Mohammed Abdelaziz [a,b], Tianfu Wang [a,*], Ahmed Elazab [a,c]

[a] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China
[b] Department of Communications and Electronics, Delta Higher Institute for Engineering and Technology (DHIET), Mansoura 35516, Egypt
[c] Computer Science Department, Misr Higher Institute of Commerce and Computers, Mansoura 35516, Egypt

ARTICLE INFO

Keywords:
Alzheimer's disease
Mild cognitive impairment
Multimodal data classification and regression
Convolutional neural networks

ABSTRACT

Alzheimer's disease (AD) is a severe irreversible neurodegenerative disease that has great sufferings on patients and eventually leads to death. Early detection of AD and its prodromal stage, mild cognitive impairment (MCI) which can be either stable (sMCI) or progressive (pMCI), is highly desirable for effective treatment planning and tailoring therapy. Recent studies recommended using multimodal data fusion of genetic (single nucleotide polymorphisms, SNPs) and neuroimaging data (magnetic resonance imaging (MRI) and positron emission tomography (PET)) to discriminate AD/MCI from normal control (NC) subjects. However, missing multimodal data in the cohort under study is inevitable. In addition, data heterogeneity between phenotypes and genotypes biomarkers makes learning capability of the models more challenging. Also, the current studies mainly focus on identifying brain disease classification and ignoring the regression task. Furthermore, they utilize multistage for predicting the brain disease progression. To address these issues, we propose a novel multimodal neuroimaging and genetic data fusion for joint classification and clinical score regression tasks using the maximum number of available samples in one unified framework using convolutional neural network (CNN). Specifically, we initially perform a technique based on linear interpolation to fill the missing features for each incomplete sample. Then, we learn the neuroimaging features from MRI, PET, and SNPs using CNN to alleviate the heterogeneity among genotype and phenotype data. Meanwhile, the high learned features from each modality are combined for jointly identifying brain diseases and predicting clinical scores. To validate the performance of the proposed method, we test our method on 805 subjects from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Also, we verify the similarity between the synthetic and real data using statistical analysis. Moreover, the experimental results demonstrate that the proposed method can yield better performance in both classification and regression tasks. Specifically, our proposed method achieves accuracy of 98.22%, 93.11%, and 97.35% for NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. On the other hand, our method attains the lowest root mean square error and the highest correlation coefficient for different clinical scores regression tasks compared with the state-of-the-art methods.

1. Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disease which frequently affects the people over 65 years old and it is considered as the one of the most frequently common cause of dementia [1–3]. In addition, the Alzheimer's Association (2016) reported that, the AD is considered as the sixth-leading cause of death in the United States [4]. Furthermore, there are over 26 million people living with AD all over the world, and it is expected to be 114 million by 2050 [4].

The patients who suffer from subjective memory decline are likely progressed to a transitional prodromal stage which is commonly known as mild cognitive impairment (MCI) stage that eventually becomes AD [5]. Furthermore, the MCI stage can be further categorized into stable MCI (sMCI) and progressive MCI (pMCI). The former likely does not progress and patients keep good life unlike the latter that eventually leads to AD [6]. To date, there is no cure for AD in clinics. Thus, all efforts are devoted towards the early diagnosis of AD especially the transitional stages that are vital for both patients and their caregivers to

* Corresponding author.
E-mail address: tfwang@szu.edu.cn (T. Wang).

https://doi.org/10.1016/j.jbi.2021.103863
Received 26 February 2021; Received in revised form 9 June 2021; Accepted 1 July 2021
Available online 3 July 2021

Table 1
Characteristics of the subjects in the ADNI dataset (mean ± standard deviation (SD)).

	NC	sMCI	pMCI	AD	Total
M/F	118/108	150/76	106/61	103/83	477/328
Age (years)	76.15 ± 5.34	75.18 ± 7.34	74.83 ± 6.64	75.19 ± 7.47	75.38 ± 6.73
MMSE	29.11 ± 1.00	27.28 ± 1.77	26.59 ± 1.71	23.28 ± 2.02	26.72 ± 2.67
CDR-SOB	0.03 ± 0.12	0.88 ± 0.72	1.24 ± 0.85	3.34 ± 1.40	1.29 ± 1.49
CDR-GLOB	0	0.50 ± 0.03	0.50 ± 0.00	0.75 ± 0.25	0.42 ± 0.30
ADAS	6.21 ± 2.92	10.33 ± 4.30	13.30 ± 4.04	18.43 ± 6.70	11.66 ± 6.42

limit the progression of AD [7].

Recently, there are several biomarkers that can improve the understanding of the disease progression. Neuroimaging (e.g., magnetic resonance imaging (MRI) and positron emission tomography (PET)) and genetic (single nucleotide polymorphisms, SNPs) biomarkers are the most common ones that can be utilized to predict the progression of AD.

The structured MRI is one of the common imaging techniques that is used to measure the brain morphometry besides its sensitivity to neurons degeneration. Hence, the structured MRI can be utilized to predict the disease progression [8,9]. On the other hand, PET is a functional molecular imaging modality that is used to assess the functional changes in the brain and can be employed for understanding the neurophysiological mechanisms of the disease [7,10]. In recent years, the imaging genetic has been widely used to discover the association between the genetic basis (e.g., SNPs) and the quantitative traits of imaging data [11].

Since the breakthrough of deep learning [12], plenty of models have been successfully applied in different computer vision and image processing tasks. The most widely used model is the convolutional neural networks (CNNs) that are used to extract high level features of the entire object [13].

The recent work of multimodality data has attracted considerable attention due to its capability and efficiency in alleviating the limitations of a single modality and hence improves the accuracy of prediction models in disease diagnosis. However, there are still several limitations in the current studies. First, the conventional methods that use multimodal data often utilize several stages for classification and regression tasks. For instance, Zhou et al. [14] proposed stage-wise deep neural network (SWDNN) to discriminate AD and MCI stages. Second, the existing methods do not often use the maximum samples due to incomplete modality. Hence, the performance of the classifier is degraded due to a few training samples [15,16]. Third, the existing methods focus on identifying brain disease classification and ignore the integration of identifying the brain diseases and estimating clinical scores [14,17,18].

To address the aforementioned problems, we propose a novel multimodal neuroimaging and genetic data fusion for joint classification and clinical score regression tasks using the maximum number of available samples in one unified framework using CNN. Specifically, we fill the missing features in PET and SNPs based on linear interpolation to further utilize the maximum number of available samples in our dataset. Then, we apply the CNN for the neuroimaging and genetic data to learn the high-level features from each modality to alleviate the heterogeneity among features then concatenate these features to ultimately classify the disease and estimate the clinical scores. The clinical scores are usually used by physicians to determine the level of dementia of the patients. In our study, the clinical scores are usually used by physicians to determine the level of dementia of the patients. In our study, the clinical score regression includes the following scores: 1) Mini-mental state exam (MMSE) which is used to assess the cognitive functions in areas of memory, attention, language, and visual construction of patient [19]. The MMSE score ranges between 0 and 30, the lower score indicates the sever level of dementia. 2) Clinical dementia rating-sum of boxes (CDR-SOB) is a clinician-rated staging scheme for cognitive and functional abilities and it ranges between 0 and 18, the lower score indicates greater cognitive dysfunction [20]. 3) Clinical dementia rating-global (CDR-GLOB) is the global version of CDR and ranges from 0 to 3 [20]. 4) The last score is AD assessment scale-cognitive subscale (ADAS-Cog) which also assess the severity of AD and it ranges from 0 to 70, higher score suggests greater impairment [20].

The major contributions of this paper are as follows. First, we propose a novel framework to utilize the multimodal data including neuroimaging and genetic data for joint disease classification and clinical scores estimation in one unified framework. It is significantly different from previous works which learn each modality independently and then concatenate the learned features from each modality in independent stages. Second, we propose to utilize the maximum number of available samples in our dataset by filling the missing features for each incomplete sample using linear interpolation. Third, we propose to alleviate data heterogeneity by applying CNN to each modality and concatenate the final features from each modality in series for brain disease classification and the clinical scores regression.

The rest of the paper is organized as follows. In Section 2, we briefly recall the relevant researches in literature. In Section 3, we describe the proposed framework and the filling mechanism, in details. We further describe experimental results and discussion in Section 4. Finally, we conclude this paper in Section 5.

2. Related work

2.1. Machine learning-based methods

In recent years, various machine learning techniques exploiting different aspects of the disease. Zhang and Shen [21] developed a multimodal multi-task sparse learning method to jointly predict multiple clinical variables and discover brain disease status. Liu et al. [22] extracted multiple sets of features from each registered MRI to multiple templates and then modeled the relationships among templates and among individual subjects using the feature selection technique followed by a support vector machine (SVM). Lei et al. [23] introduced a discriminative sparse learning method with relational regularization for joint classification and regression using multimodal features. They improved the diagnosis of AD by developing a loss function that expands the distance to get the geometrical information and also makes use of the inherent information in the observations.

2.2. Deep learning-based methods

In recent years, deep learning methods have been widely investigated for AD classification. Pan et al. [24] developed a method such that they combined CNN and ensemble learning to discriminate AD/MCI patients from normal controls (NC). Lei et al. [25] designed two scenarios to predict clinical scores based on longitudinal multiple time points data. The first scenario utilizes the baseline data to obtain the longitudinal scores prediction while the second scenario uses all the previous time points data to get the predicted scores at the next time point. Liu et al. [26] extracted multiple MRI patches around the discriminative anatomical landmarks to jointly predict the clinical scores and classify AD disease stages using a deep multi-task multi-channel CNN. Hong et al. [27] developed a deep model for predicting

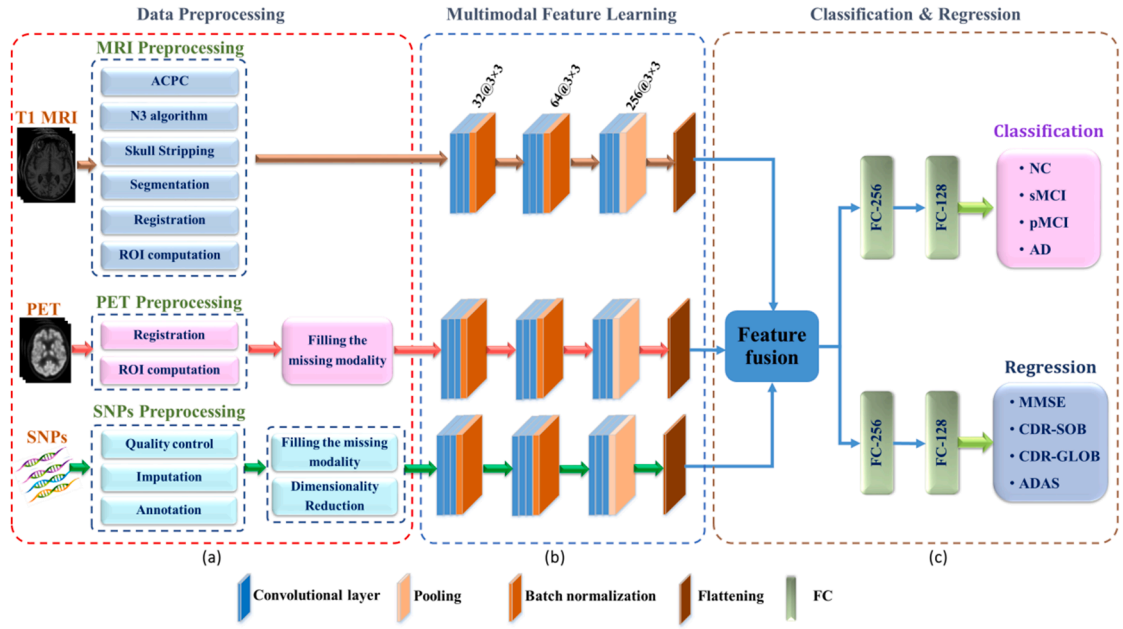

Fig. 1. Overview of the proposed method for identifying brain disease classification and predicting clinical scores using CNN.

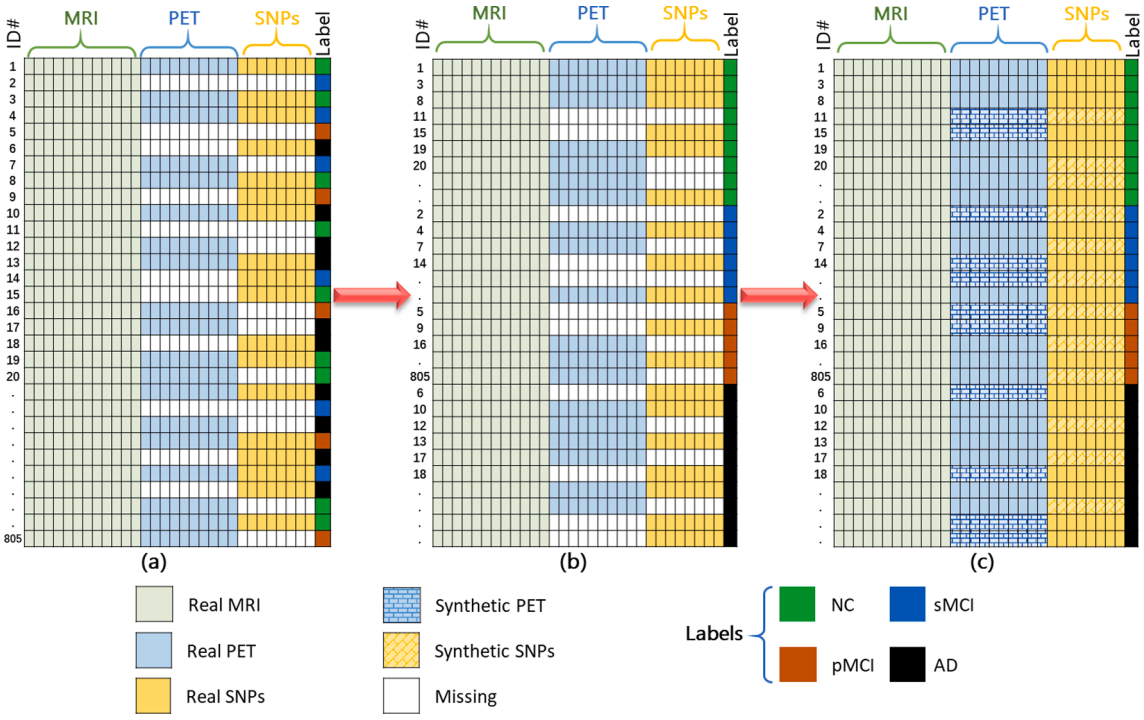

Fig. 2. The proposed mechanism for filling the missing neuroimaging and genetic data.

the development of AD disease using LSTM by utilizing the longitudinal MRI and PET images to predict the progression of the AD. However, they used the average of each modality for the missing samples. Zhou et al. [14] developed an SWDNN that utilizes neuroimaging and genetic data for classification. They learned neuroimaging and genetic data in three stages to utilize the maximum number of available samples. However, they used three stages for AD diagnosis only without considering clinical scores regression.

3. Methodology

3.1. Subjects

The Alzheimer's disease Neuroimaging Initiative (ADNI) dataset is used in this paper to evaluate the performance of our proposed method (http://adni.loni.usc.edu) [28]. The ADNI dataset contains different multimodal data (MRI, PET, SNPs, and clinical scores). Table 1 summarizes the characteristics of 805 subjects in the ADNI dataset. In this dataset, the number of available samples is 805, 396, and 687 subjects for MRI, PET, and SNPs, respectively, while the number of subjects with

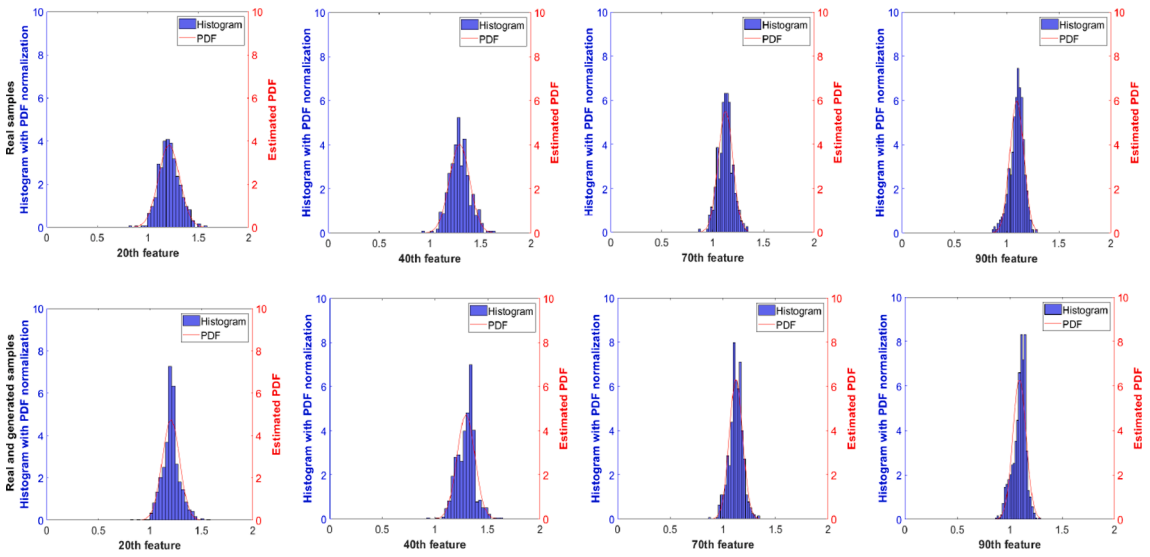

Fig. 3. Normalized histograms and estimated PDFs of the real and complete PET features.

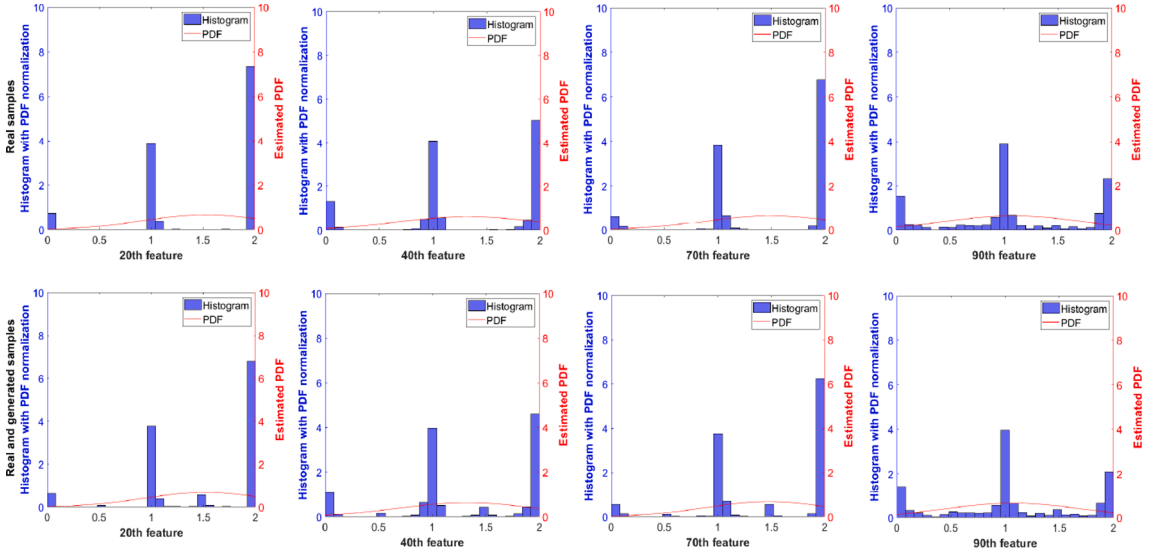

Fig. 4. Normalized histograms and estimated PDFs of the real and complete SNPs features.

Table 2
Independent samples *t*-test results for the real and complete dataset (α = 0.05).

	PET		SNPs	
	Real dataset	Complete dataset	Real dataset	Complete dataset
Mean	1.043	1.043	1.817	1.818
Variance	0.00052	0.00053	0.00025	0.00022
Hypothesized mean difference	0		0	
Degree of freedom	795		1430	
t statistic	−0.132		−0.217	
t Critical two-tail	1.963		1.962	

complete modalities is 396. To overcome the problem of incomplete modalities, we propose a filling mechanism for the missing PET and SNPs using linear interpolation to be identical with the number of MRI samples.

3.2. Data preprocessing

The goal of the neuroimaging data preprocessing is to extract the regions of interest (ROIs). For ROIs extraction, we followed the same pipeline in [14]. "The preprocessing pipeline of the neuroimaging data is as shown in Fig. 1 (a). For the MR images, we first corrected the anterior commissure-posterior commissure using the MIPAV software [29] then corrected the intensity inhomogeneity using the N3 algorithm [30]. Afterwards, the skull-stripping method in [31] was employed to extract the brain and then we removed the cerebellum. To segment the brain tissues (white matter (WM), gray matter (GM), and cerebrospinal fluid) we used the FAST algorithm in the FSL package[32] . Then, we adopted the HAMMER algorithm [33] to register the segmented brain of each subject to the atlas [34] to get the corresponding ROI labels. Eventually, for each ROI, we computed its GM volume and normalized it with the intracranial volume to obtain the final feature vector of every subject. On the other hand, for the PET images we aligned them to their corresponding T1 MR images using the affine registration. Then, the average PET intensity value of each ROI was calculated as a feature. Using the same atlas, we got the 93 ROIs of each subject. Ultimately, we totally

Table 3
Mean and SD of the real and complete PET and SNPs features.

Features	PET				SNPs			
	real		Real and synthetic		real		Real and synthetic	
	mean	SD	mean	SD	mean	SD	mean	SD
1st	1.154	0.072	1.175	0.071	1.892	0.307	1.882	0.309
10th	1.114	0.065	1.129	0.060	1.997	0.037	1.997	0.035
15th	1.416	0.100	1.416	0.087	1.992	0.065	1.990	0.066
20th	1.210	0.104	1.205	0.084	1.530	0.601	1.520	0.579
30th	0.878	0.070	0.874	0.057	1.998	0.013	1.998	0.012
35th	0.944	0.071	0.943	0.061	1.908	0.274	1.909	0.262
40th	1.293	0.099	1.295	0.084	1.331	0.657	1.324	0.630
50th	1.085	0.068	1.106	0.070	1.991	0.037	1.990	0.036
55th	1.132	0.057	1.125	0.051	1.397	0.627	1.417	0.606
60th	1.064	0.069	1.078	0.071	1.615	0.541	1.605	0.524
65th	1.112	0.122	1.130	0.101	1.736	0.456	1.740	0.438
70th	1.127	0.072	1.124	0.063	1.490	0.602	1.475	0.582
75th	1.153	0.064	1.173	0.060	1.615	0.540	1.605	0.522
80th	1.058	0.076	1.059	0.058	1.809	0.387	1.813	0.371
90th	1.099	0.067	1.092	0.063	1.086	0.626	1.073	0.610

Table 4
Brain disease classification comparisons between the proposed method and the four state-of-art methods for different classification tasks (%).

Tasks	Method	ACC	SEN	SPEC	PRE	F1
NC vs. AD	MTFL	83.38 ± 10.51	87.95 ± 10.51	79.25 ± 14.31	82.19 ± 15.46	84.47 ± 11.72
	LPP	81.54 ± 10.63	82.03 ± 12.09	79.03 ± 16.53	82.58 ± 16.48	81.25 ± 11.65
	PCA	81.58 ± 9.58	84.21 ± 11.12	78.96 ± 15.02	81.84 ± 14.89	82.39 ± 11.0
	SWDNN	91.35 ± 0.98	91.75 ± 1.20	90.90 ± 1.47	91.86 ± 1.23	91.80 ± 0.93
	Ours	**98.22 ± 1.26**	**97.78 ± 2.50**	**98.76 ± 1.26**	**98.99 ± 1.01**	**98.35 ± 1.22**
NC vs. sMCI	MTFL	68.54 ± 10.42	62.08 ± 15.48	74.85 ± 15.33	66.66 ± 21.21	63.15 ± 16.17
	LPP	66.87 ± 7.80	62.35 ± 8.98	71.29 ± 14.66	65.0 ± 18.73	62.64 ± 11.02
	PCA	67.98 ± 9.30	60.83 ± 13.59	74.85 ± 15.33	66.48 ± 21.02	62.40 ± 15.04
	SWDNN	73.62 ± 1.51	72.0 ± 1.51	75.0 ± 2.12	71.08 ± 1.79	71.51 ± 1.78
	Ours	**93.11 ± 1.64**	**92.65 ± 3.24**	**93.57 ± 2.65**	**93.60 ± 2.38**	**93.07 ± 1.42**
NC vs. pMCI	MTFL	75.21 ± 16.03	79.36 ± 17.51	69.67 ± 17.30	76.97 ± 14.45	77.72 ± 15.08
	LPP	70.12 ± 18.92	76.16 ± 22.72	64.55 ± 25.29	71.19 ± 18.57	72.31 ± 19.57
	PCA	74.87 ± 17.80	81.69 ± 19.23	67.40 ± 20.06	73.82 ± 17.39	76.94 ± 17.93
	SWDNN	84.97 ± 1.32	87.09 ± 1.56	82.28 ± 2.34	86.18 ± 1.59	86.62 ± 1.55
	Ours	**97.35 ± 0.99**	**97.82 ± 1.39**	**96.71 ± 1.46**	**97.58 ± 1.04**	**97.69 ± 0.87**

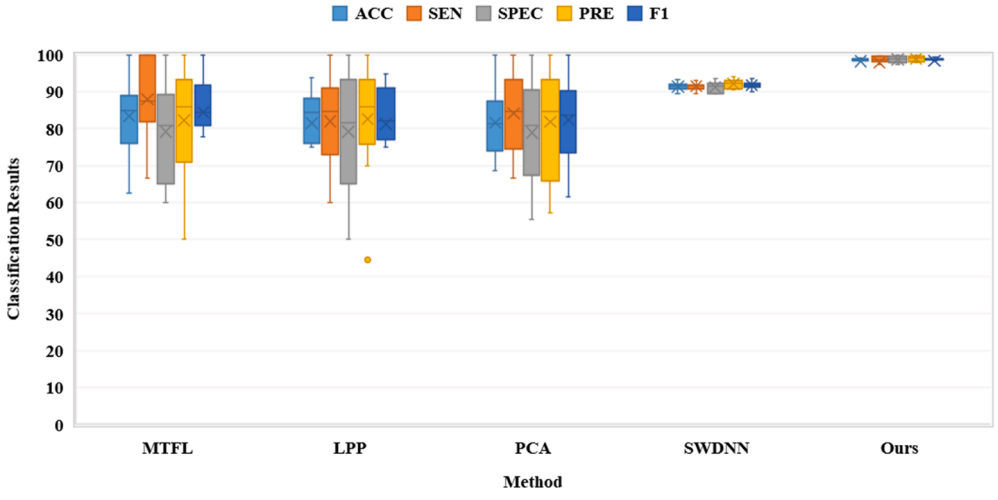

Fig. 5. Evaluation of the proposed method against competing methods for NC vs. AD task.

have 93 features for each neuroimaging data (MRI or PET) for every subject.

For genetic data, the preprocessing of the SNPs went through the following steps. Firstly, the standard quality control criteria (minor allele frequency < 0.01, poor call rate $< 95\%$, and Hardy-Weinberg equilibrium $< 10–6$) were initially applied to filter the genotype data using the PLINK software (http://zzz.bwh.harvard.edu/plink/) [35]. Secondly, the missing individuals' genotypes were imputed using the MACH program [36]. Thirdly, we used the SnpEff software to annotate all the SNPs for top risk 27 genes within ± 5 k base pairs as in [11]. Finally, we obtained 23,523 SNPs. Compared to dimension of MRI and PET images (i.e., 93 ROIs), the dimension of SNPs is too high and likely

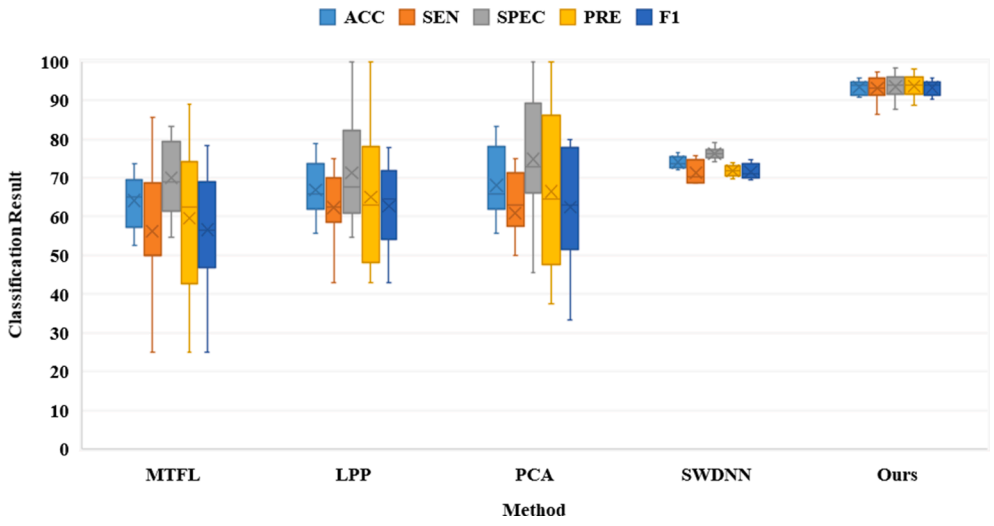

Fig. 6. Evaluation of the proposed method against competing methods for NC vs. sMCI task.

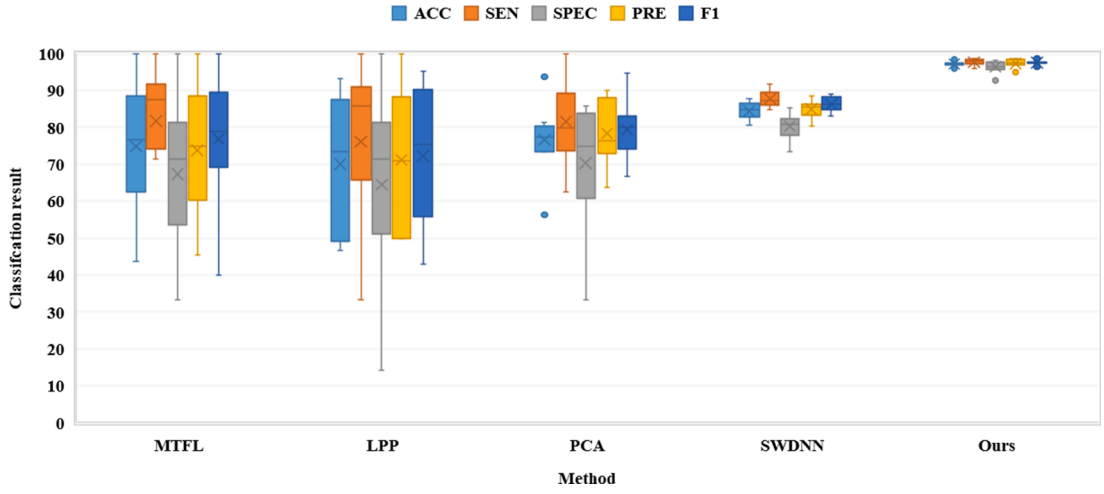

Fig. 7. Evaluation of the proposed method against competing methods for NC vs. pMCI task.

Table 5
Clinical scores regression comparisons between the proposed method and the four state-of-art methods for different tasks (%).

Task	Method	MMSE		CDR-SOB		CDR-GLOB		ADAS	
		RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC
NC vs. AD	MTFL	21.98 ± 4.52	63.20 ± 20.49	19.0 ± 3.10	75.78 ± 7.48	29.41 ± 4.82	75.67 ± 6.99	14.76 ± 3.50	65.64 ± 11.46
	LPP	24.10 ± 4.28	50.05 ± 18.57	23.80 ± 5.33	55.38 ± 17.17	36.10 ± 6.14	57.19 ± 14.0	13.37 ± 3.31	57.24 ± 10.83
	PCA	21.89 ± 4.39	62.92 ± 18.0	18.69 ± 3.58	75.75 ± 9.77	29.25 ± 4.80	76.08 ± 6.67	14.82 ± 3.59	65.39 ± 11.59
	SWDNN	16.64 ± 0.05	52.38 ± 9.11	18.57 ± 0.10	79.53 ± 1.96	18.27 ± 0.10	82.46 ± 1.11	17.81 ± 0.05	71.03 ± 3.70
	Ours	**10.23 ± 0.15**	**84.49 ± 1.89**	**11.03 ± 0.19**	**89.02 ± 1.31**	**11.15 ± 0.16**	**89.87 ± 2.44**	**10.64 ± 0.22**	**81.48 ± 2.10**
NC vs. sMCI	MTFL	26.14 ± 4.35	32.73 ± 21.16	20.65 ± 5.20	22.22 ± 10.19	48.36 ± 6.18	42.58 ± 13.52	16.94 ± 4.02	39.50 ± 17.84
	LPP	27.42 ± 4.18	26.89 ± 22.60	21.14 ± 5.41	16.34 ± 9.74	47.13 ± 7.36	43.73 ± 17.53	17.72 ± 5.20	40.11 ± 18.84
	PCA	27.98 ± 3.82	23.90 ± 20.69	20.77 ± 5.49	21.18 ± 12.41	49.08 ± 5.77	42.04 ± 13.34	17.24 ± 4.64	39.0 ± 23.33
	SWDNN	16.47 ± 0.04	28.75 ± 4.57	19.87 ± 18.04	44.41 ± 5.62	18.05 ± 0.20	55.90 ± 2.88	16.98 ± 0.06	36.85 ± 8.12
	Ours	**9.85 ± 0.06**	**37.88 ± 4.35**	**11.62 ± 0.16**	**71.60 ± 3.71**	**10.67 ± 0.17**	**85.96 ± 3.03**	**10.10 ± 0.18**	**60.18 ± 4.89**
NC vs. pMCI	MTFL	28.80 ± 6.58	35.83 ± 19.83	22.49 ± 3.70	38.69 ± 20.98	43.53 ± 5.91	53.95 ± 13.80	22.88 ± 3.45	25.98 ± 23.93
	LPP	31.51 ± 7.36	28.97 ± 17.22	21.55 ± 4.56	42.36 ± 23.51	42.42 ± 7.0	55.91 ± 16.94	22.77 ± 3.27	23.16 ± 19.97
	PCA	31.40 ± 6.19	29.77 ± 17.95	24.35 ± 3.66	30.71 ± 16.09	43.78 ± 5.82	53.24 ± 13.90	23.82 ± 3.77	25.07 ± 23.27
	SWDNN	18.23 ± 0.18	45.04 ± 4.28	20.85 ± 0.17	62.34 ± 2.34	19.21 ± 0.14	77.01 ± 1.36	18.66 ± 0.10	52.54 ± 3.62
	Ours	**10.64 ± 0.06**	**46.30 ± 5.63**	**10.98 ± 0.22**	**77.65 ± 5.36**	**10.54 ± 0.10**	**94.45 ± 4.50**	**11.06 ± 0.16**	**75.19 ± 2.70**

leads to overfitting. Therefore, we reduced the dimensionality of the SNPs to be compatible with the dimension of MRI features using the *t*-test as a pre-feature selection via the following Matlab code "https://www.mathworks.com/help/stats/selecting-features-for-classifying-high-dimensional-data.html".

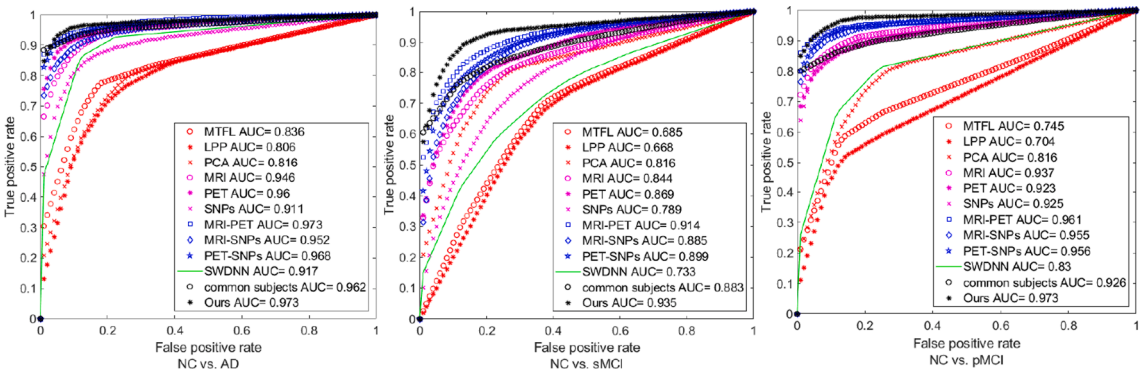

Fig. 8. ROC curves of the proposed method and SWDNN using different modality combinations for the three classification tasks.

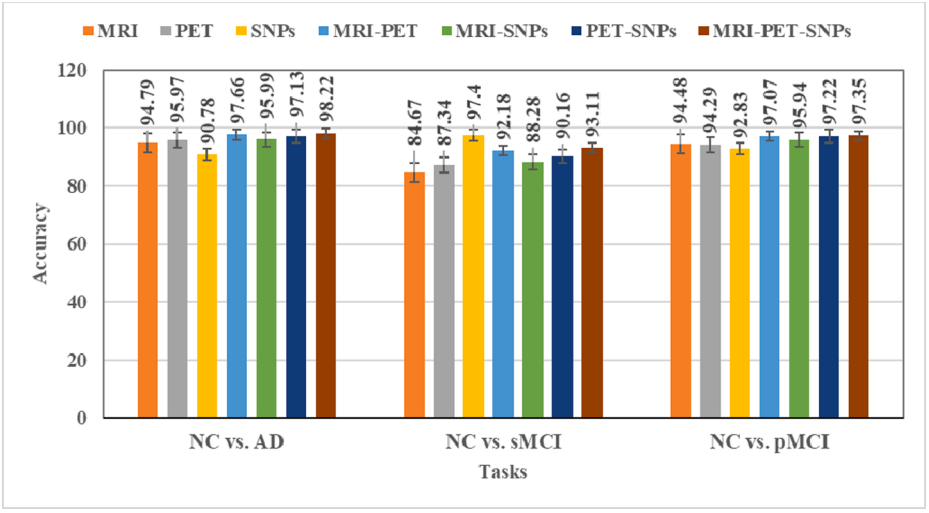

Fig. 9. Classification accuracies of the proposed method using different modality combinations for the three classification tasks.

Table 6
Results of the three classification tasks using different configurations (%).

Tasks	Method	ACC	SEN	SPE	PRE	F1
NC vs. AD	MRI	94.79 ± 1.99	94.55 ± 3.46	95.09 ± 3.30	96.00 ± 2.51	95.20 ± 1.86
	PET	95.97 ± 2.45	94.68 ± 4.31	97.54 ± 1.88	97.92 ± 1.51	96.21 ± 2.41
	SNPs	90.78 ± 1.77	91.47 ± 2.98	89.93 ± 2.73	91.75 ± 2.00	91.57 ± 1.67
	MRI-PET	97.66 ± 1.51	97.13 ± 1.89	98.31 ± 2.11	98.62 ± 1.66	97.85 ± 1.38
	MRI-SNPs	95.99 ± 2.15	95.88 ± 3.11	96.11 ± 2.58	96.82 ± 2.02	96.32 ± 2.02
	PET-SNPs	97.13 ± 1.92	96.54 ± 2.86	97.85 ± 1.88	98.23 ± 1.48	97.35 ± 1.80
	Common	96.92 ± 1.42	**99.16 ± 0.93**	94.41 ± 3.46	95.29 ± 2.84	97.16 ± 1.28
	Ours	**98.22 ± 1.26**	97.78 ± 2.50	**98.76 ± 1.26**	**98.99 ± 1.01**	**98.35 ± 1.22**
NC vs. sMCI	MRI	84.67 ± 3.48	83.34 ± 5.62	86.01 ± 4.40	85.78 ± 3.84	84.40 ± 3.74
	PET	87.34 ± 3.82	85.75 ± 6.41	88.93 ± 4.72	88.76 ± 4.35	87.07 ± 4.11
	SNPs	**97.40 ± 2.83**	74.26 ± 6.28	84.53 ± 3.67	82.92 ± 3.02	78.16 ± 3.64
	MRI-PET	92.18 ± 2.00	91.18 ± 3.63	93.17 ± 3.03	93.14 ± 2.75	92.08 ± 2.10
	MRI-SNPs	88.28 ± 2.68	84.81 ± 5.00	91.76 ± 3.16	91.25 ± 3.02	87.81 ± 3.03
	PET-SNPs	90.16 ± 2.22	88.34 ± 4.78	91.98 ± 4.28	91.92 ± 3.82	89.95 ± 2.40
	Common	89.79 ± 2.85	89.62 ± 5.44	89.92 ± 4.53	88.62 ± 4.79	88.95 ± 3.14
	Ours	93.11 ± 1.64	**92.65 ± 3.24**	**93.57 ± 2.65**	**93.60 ± 2.38**	**93.07 ± 1.42**
NC vs. pMCI	MRI	94.48 ± 2.02	95.26 ± 2.58	93.41 ± 3.09	95.19 ± 2.13	95.20 ± 1.75
	PET	94.29 ± 2.54	96.14 ± 2.95	91.80 ± 4.44	94.15 ± 3.00	95.09 ± 2.19
	SNPs	92.83 ± 1.41	94.95 ± 1.94	89.95 ± 2.51	92.78 ± 1.65	93.84 ± 1.22
	MRI-PET	97.07 ± 1.15	97.83 ± 1.71	96.03 ± 1.88	97.13 ± 1.27	97.46 ± 1.02
	MRI-SNPs	95.94 ± 1.42	96.85 ± 1.86	94.71 ± 2.29	96.15 ± 1.60	96.48 ± 1.23
	PET-SNPs	97.22 ± 1.39	**98.53 ± 0.85**	95.45 ± 3.05	96.76 ± 2.00	97.62 ± 1.14
	Common	93.98 ± 2.18	96.05 ± 4.11	91.36 ± 4.81	93.54 ± 3.31	94.67 ± 1.97
	Ours	**97.35 ± 0.99**	97.82 ± 1.39	**96.71 ± 1.46**	**97.58 ± 1.04**	**97.69 ± 0.87**

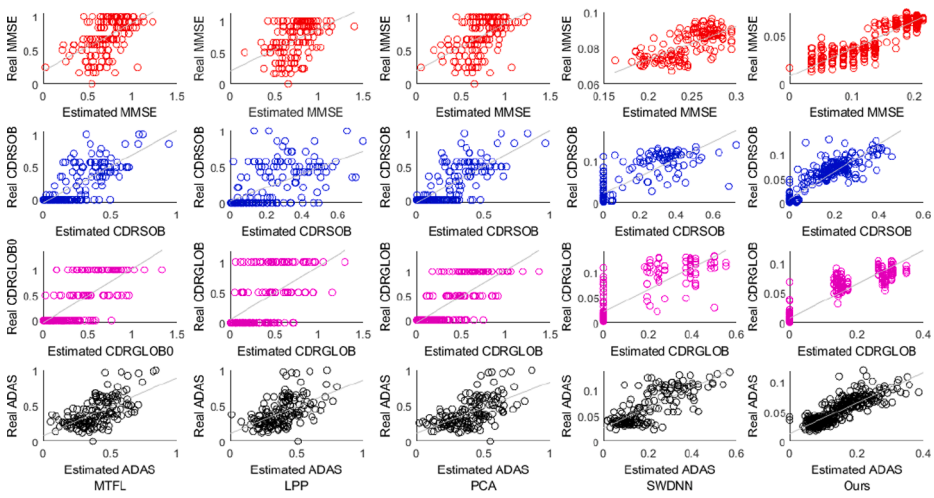

Fig. 10. Scatter plots of the estimated and the real clinical scores for NC vs. AD using five different methods.

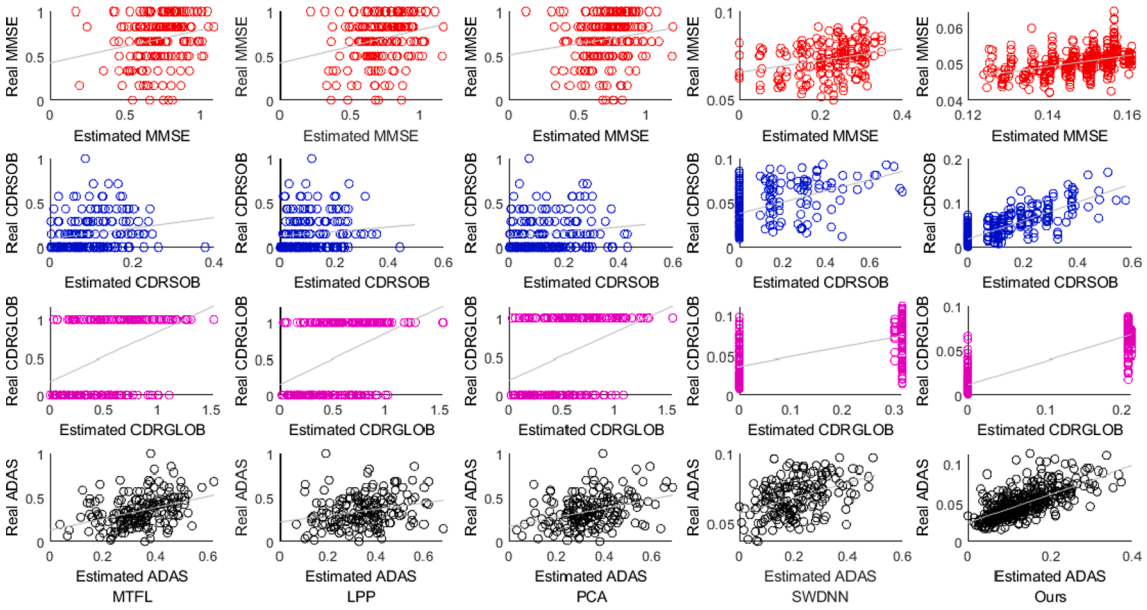

Fig. 11. Scatter plots of the estimated and the real clinical scores for NC vs. sMCI using five different methods.

3.3. Proposed framework

The general framework of our proposed method for joint classification and regression task using multimodal neuroimaging and genetic data is shown in Fig. 1. Our proposed framework maximally utilizes all the available samples from neuroimaging and genetic data by learning all the process in one unified framework. To alleviate the problem of the missing data, we initially filled the incomplete modality (PET and/or SNPs) using linear interpolation as shown in Fig. 2. It is significantly different from previous work which use the average of each reclassified modality to fill the missing of every sample in each modality [37]. Specifically, firstly, we reorganized the whole dataset (Fig. 2(a)) into four independent groups (i.e., NC, sMCI, pMCI, and AD) according to label for each sample (Fig. 2(b)). In this figure, the missing modality is represented as a blank (white) square. Secondly, since the missing modality of the samples was randomly distributed, some consecutive empty rows may appear in the reorganized feature matrix. Thus, we redistributed the missing rows to the closest real samples (rows) in the same category in order to avoid any potential bias in the synthesized data. Finally, we performed the linear interpolation to fill the missing samples.

After filling the incomplete samples, we propose a supervised CNN for identifying brain disease classification and predicting the clinical scores. Our framework utilizes three different types of modalities in one unified framework. We trained each modality using CNN such that the initially learned space representations from each modality are concatenated and then applied to a series of fully connected (FC) layers to predict both classification and the clinical score regression tasks. In this way, the data heterogeneity is alleviated by sequentially concatenating the high-level feature representation from MRI, PET, and SNPs, respectively.

3.4. CNN-Based classification

The CNN is a category of deep neural network which is used to extract the high-level feature representation of the input data [38]. In our study, we used three stacked 2D CNN for extracting the high-level features from neuroimaging and genetic data. Specifically, we used five layers in our network (input, convolution, activation, batch normalization, pooling, and fully connected (FC)) such that volume of

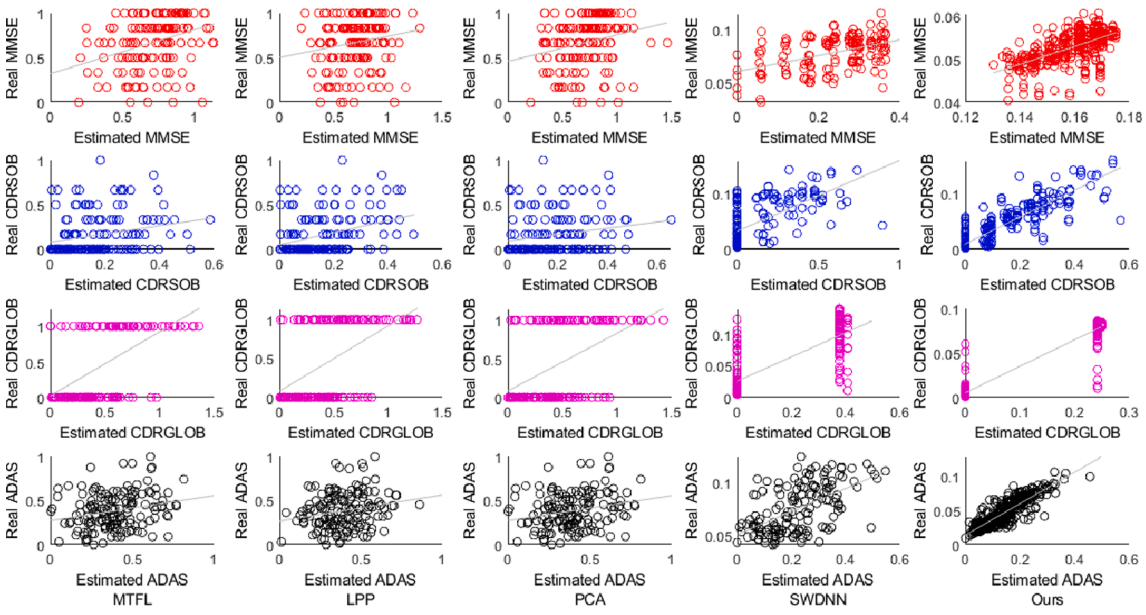

Fig. 12. Scatter plots of the estimated and the real clinical scores for NC vs. pMCI using five different methods.

Table 7
Clinical score regression results of the three tasks using different configurations (%).

Tasks	Method	MMSE		CDR-SOB		CDR-GLOB		ADAS	
		RMSE	CC	RMSE	CC	RMSE	CC	RMSE	CC
NC vs. AD	MRI	10.33 ± 0.12	81.26 ± 3.24	11.23 ± 0.16	84.91 ± 3.26	11.39 ± 0.23	85.55 ± 2.35	10.74 ± 0.18	78.46 ± 3.31
	PET	10.31 ± 0.10	82.11 ± 2.15	11.21 ± 0.13	84.65 ± 2.76	11.24 ± 0.13	86.61 ± 2.56	10.79 ± 0.09	73.45 ± 3.37
	SNPs	10.24 ± 0.10	73.66 ± 3.03	11.95 ± 0.20	71.64 ± 4.89	11.75 ± 0.13	76.85 ± 3.96	10.98 ± 0.12	70.04 ± 3.23
	MRI-PET	10.30 ± 0.08	**84.57 ± 2.68**	**11.03 ± 0.19**	88.48 ± 1.14	**11.12 ± 0.15**	90.16 ± 1.32	10.65 ± 0.15	80.19 ± 3.05
	MRI-SNPs	10.24 ± 0.13	80.36 ± 2.06	11.14 ± 0.17	86.03 ± 3.01	11.25 ± 0.25	87.55 ± 3.30	10.71 ± 0.11	82.12 ± 2.31
	PET-SNPs	10.27 ± 0.10	81.81 ± 3.85	11.25 ± 0.35	83.29 ± 3.08	11.28 ± 0.12	86.73 ± 2.42	10.74 ± 0.16	75.15 ± 3.03
	Common	17.07 ± 0.09	74.97 ± 5.04	17.76 ± 0.15	**94.55 ± 0.81**	17.47 ± 0.23	**94.72 ± 1.31**	17.84 ± 0.19	**89.09 ± 2.48**
	Ours	**10.23 ± 0.15**	84.49 ± 1.89	**11.03 ± 0.19**	89.02 ± 1.31	11.15 ± 0.16	89.87 ± 2.44	**10.64 ± 0.22**	81.48 ± 2.10
NC vs. sMCI	MRI	9.95 ± 0.06	40.51 ± 4.73	12.37 ± 0.17	62.19 ± 3.88	11.17 ± 0.15	70.94 ± 4.81	10.25 ± 0.13	51.12 ± 6.37
	PET	10.02 ± 0.10	33.23 ± 4.37	12.49 ± 0.26	59.82 ± 3.55	10.84 ± 0.26	80.19 ± 5.86	10.52 ± 0.15	41.06 ± 3.98
	SNPs	10.02 ± 0.07	31.59 ± 6.49	12.90 ± 0.14	47.83 ± 4.17	11.38 ± 0.20	67.56 ± 4.49	10.39 ± 0.10	40.33 ± 4.18
	MRI-PET	10.00 ± 0.07	39.02 ± 4.46	11.83 ± 0.19	69.04 ± 3.61	10.73 ± 0.19	**86.87 ± 2.54**	10.21 ± 0.18	60.88 ± 2.47
	MRI-SNPs	9.90 ± 0.05	42.77 ± 5.42	11.99 ± 0.25	67.31 ± 3.70	10.88 ± 0.22	78.11 ± 3.5	**10.06 ± 0.12**	59.66 ± 5.42
	PET-SNPs	9.94 ± 0.06	33.61 ± 6.48	12.21 ± 0.31	62.27 ± 4.69	10.70 ± 0.18	82.02 ± 5.93	10.27 ± 0.17	51.86 ± 4.06
	Common	16.22 ± 0.19	**67.19 ± 6.13**	17.45 ± 0.32	**77.72 ± 4.08**	16.41 ± 0.17	86.44 ± 3.67	16.44 ± 0.17	**74.54 ± 3.28**
	Ours	**9.85 ± 0.06**	37.88 ± 4.35	**11.62 ± 0.16**	71.60 ± 3.71	**10.67 ± 0.17**	85.96 ± 3.03	10.10 ± 0.18	60.18 ± 4.89
NC vs. pMCI	MRI	10.64 ± 0.04	**79.79 ± 6.88**	11.39 ± 0.25	71.28 ± 5.95	10.78 ± 0.19	91.41 ± 4.21	11.19 ± 0.14	67.83 ± 4.95
	PET	**10.57 ± 0.05**	46.10 ± 6.66	11.89 ± 0.16	59.62 ± 6.01	10.81 ± 0.21	90.21 ± 7.88	11.41 ± 0.11	63.76 ± 3.54
	SNPs	10.67 ± 0.06	50.03 ± 4.72	11.53 ± 0.21	64.13 ± 3.52	11.02 ± 0.16	85.75 ± 3.51	11.25 ± 0.09	68.10 ± 3.19
	MRI-PET	10.63 ± 0.07	45.57 ± 5.04	10.95 ± 0.20	76.51 ± 3.14	10.56 ± 0.08	**94.53 ± 2.32**	11.13 ± 0.13	70.06 ± 5.03
	MRI-SNPs	10.65 ± 0.05	47.53 ± 4.70	11.12 ± 0.24	74.25 ± 4.12	10.60 ± 0.09	92.68 ± 3.08	11.12 ± 0.14	74.13 ± 4.50
	PET-SNPs	10.65 ± 0.05	42.89 ± 8.32	11.15 ± 0.21	74.14 ± 3.12	10.57 ± 0.12	96.36 ± 0.88	11.17 ± 0.9	72.83 ± 3.79
	Common	18.03 ± 0.17	78.87 ± 3.15	19.28 ± 0.30	**87.58 ± 1.72**	18.03 ± 0.16	94.12 ± 1.66	18.08 ± 0.22	**82.71 ± 3.55**
	Ours	10.64 ± 0.06	46.30 ± 5.63	**10.98 ± 0.22**	77.65 ± 5.36	**10.54 ± 0.10**	94.45 ± 4.50	**11.06 ± 0.16**	75.19 ± 2.70

the input data is transformed from one to another according to the current layer. First, the input data is convolved to a set of kernel filters for extracting high-level feature followed by rectified linear unit as an activation function. In addition, the pooling layer is applied to reduce the space size and hence, reduce the computation complexity and alleviate overfitting. On the other hand, the convolved output is followed by batch normalization to normalize the features and hence speeding up the learning process. Furthermore, the FC layer is used to flatten our matrix to a vector and then the flattened matrix goes through a FC layer for either computing the class scores using a Softmax activation function or estimating the clinical scores using linear activation function.

4. Results and discussions

4.1. Experimental settings

All the experiments are conducted on a machine with Ubuntu 506 Nvidia GTX Titan Xp x2 i76800K and implemented using Keras library with Tensorflow as backend. In our experiments, we applied a grid search for the hyperparameter tuning such that we fixed all the hyperparameters except only the one that is being used as a grid search. Specifically, we find the best batch size within the range of $\{10, 20, \cdots, 100\}$. Also, we search the best number of epochs within the range of $\{40, 50, \cdots, 100\}$ such that the rest of hyperparameters remained constant for different epochs. Furthermore, we searched the best value for learning rate within the range of $\{10^{-6}, 10^{-5}, \cdots, 10^{-2}\}$.

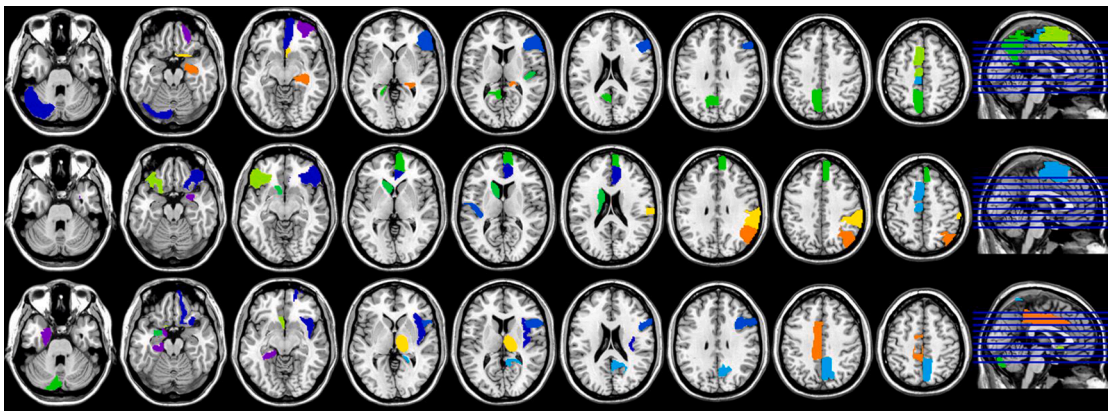

Fig. 13. Top 10 brain regions from MRI data for NC vs. AD, NC vs. sMCI, and NC vs. pMCI shown in top, middle, and bottom rows, respectively.

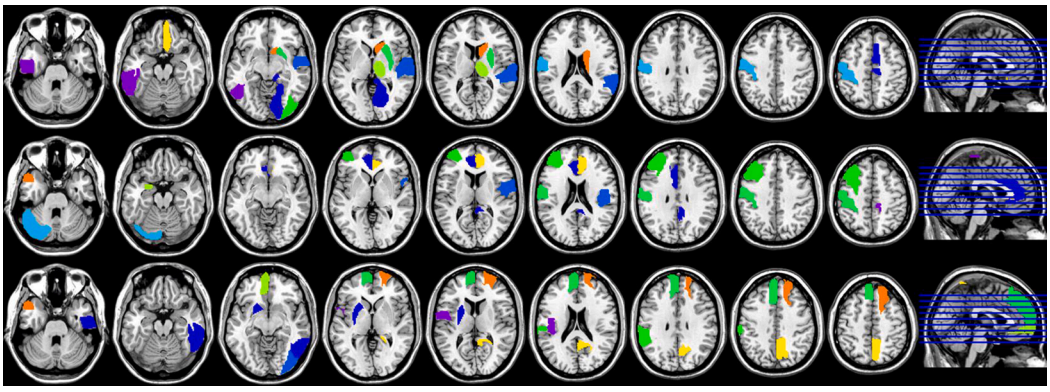

Fig. 14. Top 10 brain from PET data for NC vs. AD, NC vs. sMCI, and NC vs. pMCI shown in top, middle, and bottom rows, respectively.

Table 8
Top 10 MRI ROIs identified by the proposed method for the three tasks.

NC vs. AD		NC vs. sMCI		NC vs. PMCI	
index	Name	index	Name	index	Name
10	Superior Frontal Gyrus Right	42	Parietal Lobe WM Left	6	Lateral Front-Orbital Gyrus Right
91	Thalamus Right	16	Frontal Lobe WM Right	12	Globus Palladus Left
14	Inferior Frontal Gyrus Left	79	Anterior Limb of Internal Capsule Right	68	Entorhinal Cortex Right
69	Hippocampal Formation Left	19	Temporal Pole Right	39	Caudate Nucleus Right
26	Precuneus Right	32	Superior Occipital Gyrus Right	30	Hippocampal Formation Right
80	Middle Temporal Gyrus Right	71	Parietal Lobe WM Right	41	Precuneus Left
19	Temporal Pole Right	15	Putamen Right	21	Nucleus Accumbens Right
67	Lateral Occipitotemporal Gyrus Right	24	Fornix Left	93	Fornix Right
22	Uncus Right	64	Entorhinal Cortex Left	78	Parahippocampal Gyrus Right
38	Superior Parietal Lobule Left	66	Superior Occipital Gyrus Left	33	Caudate Nucleus Left

Table 9
Top 10 PET ROIs identified by the proposed method for the three tasks.

NC vs. AD		NC vs. sMCI		NC vs. PMCI	
index	Name	index	Name	index	Name
89	Cuneus Right	70	Thalamus Left	73	Postcentral Gyrus Right
20	Subthalamic Nucleus Right	36	Occipital Lobe Wm Right	54	Inferior Frontal Gyrus Right
82	Corpus Callosum	18	Angular Gyrus Right	87	Angular Gyrus Left
57	Medial Front-Orbital Gyrus Left	91	Thalamus Right	17	Parahippocampal Gyrus Left
48	Middle Temporal Gyrus Left	31	Inferior Occipital Gyrus Left	90	Lateral Occipitotemporal Gyrus Left
74	Lingual Gyrus Right	57	Medial Front-Orbital Gyrus Left	23	Cingulate Region Left
78	Parahippocampal Gyrus Right	41	Precuneus Left	25	Frontal Lobe WM Left
54	Inferior Frontal Gyrus Right	7	Cingulate Region Right	63	Temporal Pole Left
28	Posterior Limb of Internal Capsule Inc. Cerebral Peduncle Left	32	Superior Occipital Gyrus Right	68	Entorhinal Cortex Right
72	Insula Left	87	Angular Gyrus Left	4	Insula Right

Table 10
Top 10 genes and SNPs identified by the proposed method for the three tasks.

NC vs. AD		NC vs. sMCI		NC vs. pMCI	
Gene	SNPs	Gene	SNPs	Gene	SNPs
GAB2	rs118076932	SORL1	rs1219450	ADAM10	rs73424597
SORL1	rs1791943	SORCS1	rs188635013	APOE	rs438811
PICALM	rs187178331	ADAM10	rs11856657	SORL1	rs1219433
SORL1	rs2850774	SORL1	rs144143061	APOE	rs10119
APOE	rs483082	PICALM	rs138888937	APOE	rs429358
DAPK1	rs4878115	TF	rs1847872	DAPK1	rs2274605
SORL1	rs1448137	ADAM10	rs144083301	SORL1	rs7925712
SORL1	rs1791957	SORCS1	rs142554647	APOE	rs439401
IL33	rs12683567	ADAM10	rs141692019	SORL1	rs1789750
SORCS1	rs7905723	SORCS1	rs149926148	CH25H	rs146757074

Table 11
Comparisons between the proposed method against the state-of-the-art methods on the NC vs. AD classification task.

Algorithm	Subjects	Modalities	ACC	SEN	SPE
Zhang et al. [52]	51AD + 52 NC	PET + MRI + CSF	93.20	93.0	93.3
Liu et al. [53]	93AD + 100 NC	PET	91.20	91.40	91.00
Liu et al. [38]	93AD + 100 NC	PET + MRI	93.26	92.55	93.94
Feng et al. [54]	93AD + 100 NC	PET + MRI	94.29	96.59	92.38
Feng et al. [55]	93 AD + 100 NC	PET + MRI	94.82	97.70	92.45
Ours	186AD + 226 NC	PET + MRI + SNPs	98.20	97.78	98.76

Moreover, we searched the best value of the first order exponential decay rates for the moment estimates, the second order exponential decay rates for the moment estimates, and the epsilon is 10^{-6} within the ranges of $\{0, 0.2, \cdots, 1\}$, $\{0.990, 0.993, \cdots, 0.999\}$, and $\{10^{-8}, 10^{-7}, \cdots, 10^{-1}\}$, respectively. The best found hyperparameters were 40, 80, and 0.2 for the batch size, the number of epochs, the first order exponential decay rate for the moment estimate, respectively. On the other hand, the second order exponential decay rate for the moment estimate was 0.999, the epsilon was 10^{-6}, while the learning rate of the network was set to 10^{-4}. In this work, we used the binary cross entropy as loss function which yielded the best performance compared to other loss function as it matches the nature of our binary classification task. Finally, we used k-fold cross-validation with $k = 10$ for efficient learning of the model's parameters particularly our dataset size is not big enough. Note that, the results reported in our tables are computed by averaging the 10 repetitions of the10-fold cross-validation.

Furthermore, we used the measures: accuracy (ACC), sensitivity (SEN), specificity (SPE), precision (PRE), F1-score (F1), and the area under receive operation curve (AUC) to evaluate the performance of the classification. Whereas, the correlation coefficient (CC) and RMSE were used to evaluate the performance of the regression.

4.2. Comparison methods

For comparison, the proposed method was evaluated against four conventional feature representation, including multi-task feature learning (MTFL) [39], locality preserving projection (LPP) [40], principle component analysis (PCA) [41], and SWDNN [14]. We chose these feature selection methods as they are commonly used in literature for the same purpose or similar tasks. Particularly, the MTFL which is widely used to identify the most informative features using combination of the l2,1-norm regularization and least square loss function [39,42]. This feature selection method has been applied for several association studies to identify the genetic markers causes AD. For MTFL, we found the best regularization parameters within the range of $\{10^{-4}, 10^{-3}, \cdots, 10^{-2}\}$. Similarly, LPP is a well-known method that has been applied for several years because of its promising performance in dimension reduction [40,43,44]. The local manifold structure of the data can be preserved by constructing the adjacency graph among data points, and hence the data points mapped to a subspace by learning the projection matrix such that the local neighborhood structure can be well preserved. Finally, PCA is one of the linear dimensional reduction techniques that have been widely used in scientific research [41,45,46]. The goal of the PCA is to transform the data points into the direction of the maximum variance.

To fairly perform fair comparisons with other methods, the three modality features and clinical scores are fused into a single feature vector and the SVM is adopted for classification and regression tasks. Also, all the comparisons are based on three different tasks (NC vs. AD, NC vs. sMCI, and NC vs. pMCI) and the results of all the techniques are

Table 12
Comparisons between the proposed method against the state-of-the-art methods on the NC vs. sMCI classification task.

Algorithm	Subjects	Modalities	ACC	SEN	SPE
Liu et al.[38]	128sMCI + 100NC	PET + MRI	64.04	63.07	67.05
Feng et al. [54]	128sMCI + 100NC	PET + MRI	64.47	70.43	48.41
Feng et al.[55]	128sMCI + 100NC	PET + MRI	65.35	70.59	69.17
Ours	226sMCI + 226NC	PET + MRI + SNPs	93.11	92.65	93.57

Table 13
Comparisons between the proposed method against the state-of-the-art methods on the NC vs. pMCI classification task.

Algorithm	Subjects	Modalities	ACC	SEN	SPE
Liu et al.[38]	76pMCI + 100NC	PET + MRI	82.95	81.08	84.31
Feng et al [54]	76pMCI + 100NC	PET + MRI	84.66	83.56	85.44
Feng et al.[55]	76pMCI + 100NC	PET + MRI	86.36	83.33	88.78
Ours	167pMCI + 226NC	PET + MRI + SNPs	97.35	97.82	96.71

computed by averaging the 10 repetitions of 10-fold cross-validation. Furthermore, we empirically search the best value for the soft margin parameter C of SVM classifier within the range of $\{10^{-4}, 10^{-3}, \cdots, 10^{-4}\}$.

4.3. Characteristics of the complete dataset

Figs. 3 and 4 show the normalized histogram and estimated probability distribution function (PDF) with number of features between the real and the complete datasets (real + synthetic) of PET and SNPs samples, respectively. It is shown that the normalized histogram and PDF of the real dataset and the complete dataset are close to each other.

To further evaluate the effect of synthetic data on the complete dataset, we performed a statistical *t*-test under risk level ($\alpha = 0.05$) for the mean of the real and complete dataset for PET and SNPs (Table 2). It is shown that the corresponding critical value in the *t*-distribution table under the specified α value is 1.963 and 1.962 for PET and SNPs, respectively. Comparing the *t*-values with the *t*-critical for PET and SNPs, we failed to reject the null hypothesis [47]. Therefore, we could claim that there is no significance difference between the real and the complete dataset for PET and SNPs.

On the other hand, Table 3 shows the mean and SD for fifteen randomly selected features from real and complete dataset (including the synthetic samples). It is clear that, the mean and SD of the complete dataset and the real data are close with nonsignificant variance between them. Hence, the bias in the complete dataset can be ignored. Accordingly, the diagnosis of samples with missing modality can be taken in consideration using our data filling method.

4.4. Classification results

The binary disease classification results for three tasks are given in Table 4 (boldfaces indicate the best performance). In addition, Figs. 5-7 illustrate the boxplot of ACC, SEN, SPEC, PRE, and F1 for the different classification tasks. It is clear that, the proposed method significantly outperforms the competing methods in three classification tasks using five performance metrics. Moreover, our proposed method achieved 98.22%, 93.11%, and 97.35% for NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. Furthermore, our proposed method attained the highest sensitivity in different tasks compared to the conventional methods, which indicate that the discrimination between AD and NC can be effectively increased by our proposed method. Hence, the proposed method can accurately identify AD patients, in comparison to those listed methods. The main reasons of obtaining such performance are the ability of the proposed method to learn the high-level features from all available samples in the ADNI dataset and the effectiveness of combining multimodal data as well.

Table 5 shows the performance comparison of state-of-arts methods and our proposed method on clinical scores of NC vs. AD, NC vs, PMCI, and NC vs. pMCI. It is obvious that, our proposed method achieves the lowest RMSE and the highest CC compared to those provided by SWDNN and other machine learning techniques. On the other hand, Fig. 8 shows the ROC curves of different combinations of the proposed method and SWDNN for NC vs. AD, NC vs. sMCI, and NC vs. pMCI classification tasks. We can notice that, the multi-modal fusion achieves the highest AUC compared with competing methods and different combination of modalities.

Fig. 9 and Tables 6 show the comparisons between the proposed method and different combinations of modalities for three different tasks. It is clear that, the involving of SNPs for AD diagnosis is not always achieving a promising result for classification or regression. For example, combining the SNPs with MRI reduces the performance of classification and regression. However, the combining SNPs with PET achieves better performance than using individual modality. Moreover, the multi-modality fusion achieves higher performance than single modality which verifies the effectiveness of the proposed method for improving the diagnosis of AD. The results of the proposed method are consistent with that achieved in the literature [14]. Besides, our proposed method achieves the best performance in terms of disease classifications.

4.5. Regression results

Figs. 10-12 illustrate the scatter plot of the estimated clinical score vs. the real clinical scores for NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. Furthermore, Table 7 represents the clinical score regression results of the three tasks using different configurations. It is clear that the proposed method can achieve better performance than that obtained by the competing methods in the regression of four clinical scores. Specifically, the MMSE correlation of our method is 84.49%, 37.88%, and 46.30% for the NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. Also, the CDR-SOB correlation of our method is 89.02%, 71.60%, and 77.65% for the NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. In addition, the CDR-GLOB correlation of our method is 89.87%, 85.96%, and 94.45% for the NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. Finally, the ADAS correlation of our method is 81.48%, 60.18%, and 75.19% for the NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. From these results, we can notice that our proposed method attains the highest correlation compared to the state-of-art methods of the four clinical scores by at least 7.41%, 5.15%, and 1.26% for the NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. On the other hand, the proposed method attains the lowest RMSE of 10.23%, 9.85%, and 10.64% for the NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. Also, the CDR-SOB obtained by our method has an improved RMSE of 11.03%, 11.62%, and 10.98% for the NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. Similarly, the RMSE of the CDR-GLOB achieved by our method is better than competing methods by 11.15%, 10.67%, and 10.54% for the NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. Likewise, our method attains the lowest RMSE for ADAS by 10.64%, 10.10%, and 11.06% for the NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively. Finally, it is obvious that our proposed method attains the lowest RMSE compared to the state-of-art methods of the four clinical scores by at least 2.73%, 6.62%, and 7.59% for the NC vs. AD, NC vs. sMCI, and NC vs. pMCI, respectively.

4.6. Discriminative ROIs and SNPs

Figs. 13 and 14 show the top 10 ROIs identified for MRI and PET data for three classification tasks, respectively. We conclude the top 10 ROIs for each modality in Tables 8 and 9. Moreover, these finding are also consistent with the previous studies on AD diagnosis [44,48–50].

Table 10 summarizes the top 10 SNPs and their corresponding genes names identified by our proposed method. These findings are consistent with those reported in many AD diagnostic studies also genome-wide association studies and hence verify the effectiveness of our proposed method to identify the most discriminative brain regions and the most relevant SNPs cause AD [11,51].

4.7. Comparisons with previous studies

Tables 11-13 compare the performance of proposed method with the state-of-the-arts methods on the ADNI database for the three classification tasks. It is shown that, our proposed method achieves better performance than the competing methods in most cases. However, our proposed method has a three major contribution compared to the previous studies. Firstly, our method uses multimodal neuroimaging and genetic data fusion for jointly brain disease classification and estimate the clinical score in one unified framework. Secondly, we propose to utilize the maximum number of the available samples by adopting linear interpolation to fill the incomplete samples. Third, we alleviate the data heterogeneity by concatenating the output features from each modality for brain disease classification and the clinical scores estimation. In

general, our method outperforms the previous methods with single and multimodal data. Furthermore, our proposed method can identify the most discriminative SNPs related to AD.

5. Conclusions and future work

In this work, a robust one-stage joint classification and regression using deep learning framework was proposed for diagnosing AD. Our framework is significantly different from previous works that learned the multimodal data including neuroimaging and genetic data in multiple stages for AD diagnosis. We first introduced a filling mechanism of the missing neuroimaging and genetic data and preserved the local structure of the complete dataset. This was proved to be significant via the statistical *t*-test with almost no bias between the real and the complete datasets. Furthermore, we learned the neuroimaging features from MRI, PET, and SNPs using CNN to alleviate the heterogeneity among genotype and phenotype data. Afterwards, the learned high-level features from each modality are combined for jointly identifying brain diseases and predicting clinical scores. Compared with the traditional methods, our proposed method demonstrated superior performance against competing methods in both tasks of disease classification and clinical scores regression.

Although, the proposed method achieved better performance in both classification and regression tasks than conventional methods in literature, there are few limitations that we will try to address in our future work. First, the proposed method only considered the ADNI 805 dataset. However, including more data shall improve the performance. We will investigate this issue by combining ADNI-2 dataset with ADNI-1 dataset. Second, in this work, we used brain ROIs as features which may induce missing information. This can be addressed by using the original imaging data and use attentive deep neural network models to extract the high-level features.

CRediT authorship contribution statement

Mohammed Abdelaziz: Conceptualization, Methodology, Software, Writing - review & editing. **Tianfu Wang:** Supervision, Guidance, and Research Support. **Ahmed Elazab:** Conceptualization, Data curation, Visualization, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by National Natural Science Foundation of China No. 61950410615.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbi.2021.103863.

References

[1] L. S. Kucmanski, L. Zenevicz, D. S. Geremia, V. S. F. Madureira, T. G. d. Silva, and S. S. d. Souza, "Alzheimer's desease: challenges faced by family caregivers," Revista Brasileira de Geriatria e Gerontologia, vol. 19, no. 6, pp. 1022-1029, 2016.
[2] Z.S. Khachaturian, Diagnosis of Alzheimer's disease, Arch. Neurol. 42 (11) (1985) 1097–1105.
[3] A.M. Anter, et al., A robust swarm intelligence-based feature selection model for neuro-fuzzy recognition of mild cognitive impairment from resting-state fMRI, Inf. Sci. 503 (2019) 670–687.
[4] A. s. Association, "2016 Alzheimer's disease facts and figures," Alzheimer's & Dementia, vol. 12, no. 4, pp. 459-509, 2016.
[5] X. Chen, et al., High-order resting-state functional connectivity network for MCI classification, Hum. Brain Mapp. 37 (9) (2016) 3282–3296.
[6] K.-H. Thung, P.-T. Yap, E. Adeli, S.-W. Lee, D. Shen, and A. s. D. N. Initiative, "Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion," Medical image analysis, vol. 45, pp. 68-82, 2018.
[7] X. Hao, et al., Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease, Med. Image Anal. 60 (2020), 101625.
[8] G. Folego, M. Weiler, R. F. Casseb, R. Pires, and A. Rocha, "Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI," Frontiers in Bioengineering and Biotechnology, vol. 8, 2020.
[9] X. Bi, S. Li, B. Xiao, Y. Li, G. Wang, X. Ma, Computer aided Alzheimer's disease diagnosis by an unsupervised deep learning technology, Neurocomputing 392 (2020) 296–304.
[10] M. Liu, et al., A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease, NeuroImage 208 (2020), 116459.
[11] X. Zhu, H.-I. Suk, H. Huang, D. Shen, Low-rank graph-regularized structured sparse regression for identifying genetic biomarkers, IEEE Trans. Big Data 3 (4) (2017) 405–414.
[12] Y. LeCun, Y. Bengio, and G. J. n. Hinton, "Deep learning," vol. 521, no. 7553, pp. 436-444, 2015.
[13] B.-K. Choi, et al., Convolutional neural network-based mr image analysis for Alzheimer's disease classification, Current Medical Imaging 16 (1) (2020) 27–35.
[14] T. Zhou, K.H. Thung, X. Zhu, D. Shen, Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis, Hum. Brain Mapp. 40 (3) (2019) 1001–1016.
[15] J. Schmidhuber, Deep learning in neural networks: An overview, Neural networks 61 (2015) 85–117.
[16] T. Zhou, K.-H. Thung, X. Zhu, D. Shen, Feature learning and fusion of multimodality neuroimaging and genetic data for multi-status dementia diagnosis, Springer, 2017, pp. 132–140.
[17] B. Jie, D. Zhang, B. Cheng, D. Shen, and A. s. D. N. Initiative, "Manifold regularized multitask feature learning for multimodality disease classification," Human brain mapping, vol. 36, no. 2, pp. 489-507, 2015.
[18] M. Liu, J. Zhang, P.-T. Yap, D. Shen, View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data, Med. Image Anal. 36 (2017) 123–134.
[19] B. J. T. a. i. n. d. Sheehan, "Assessment scales in dementia," vol. 5, no. 6, pp. 349-358, 2012.
[20] S. Balsis, J.F. Benge, D.A. Lowe, L. Geraci, R.S. Doody, How do scores on the ADAS-Cog MMSE, and CDR-SOB correspond? 29 (7) (2015) 1002–1009.
[21] D. Zhang, D. Shen, Multi-modal multi-task learning for joint prediction of clinical scores in Alzheimer's disease, Springer, 2011, pp. 60–67.
[22] M. Liu, D. Zhang, D. Shen, Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment, IEEE Trans. Med. Imaging 35 (6) (2016) 1463–1474.
[23] B. Lei, P. Yang, T. Wang, S. Chen, D. Ni, Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis, IEEE Trans. Cybern. 47 (4) (2017) 1102–1113.
[24] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song, "Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning," Frontiers in neuroscience, vol. 14, 2020.
[25] B. Lei, et al., Deep and joint learning of longitudinal data for Alzheimer's disease prediction, Pattern Recogn. 102 (2020), 107247.
[26] M. Liu, J. Zhang, E. Adeli, D. Shen, Joint Classification and Regression via Deep Multi-Task Multi-Channel Learning for Alzheimer's Disease Diagnosis, IEEE Trans. Biomed. Eng. 66 (5) (2018) 1195–1206.
[27] X. Hong, et al., Predicting Alzheimer's Disease Using LSTM, IEEE Access 7 (2019) 80893–80901.
[28] C.R. Jack Jr, et al., The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods 27 (4) (2008) 685–691.
[29] M.J. McAuliffe, F.M. Lalonde, D. McGarry, W. Gandler, K. Csaky, B.L. Trus, Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001, IEEE, 2001, pp. 381–386.
[30] J. G. Sled, A. P. Zijdenbos, and A. C. J. I. t. o. m. i. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," vol. 17, no. 1, pp. 87-97, 1998.
[31] Y. Wang, et al., Knowledge-guided robust MRI brain extraction for diverse large-scale neuroimaging studies on humans and non-human primates, PLoS ONE 9 (1) (2014), e77810.
[32] Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm, IEEE Trans. Med. Imaging 20 (1) (2001) 45–57.
[33] D. Shen, C. Davatzikos, HAMMER: hierarchical attribute matching mechanism for elastic registration, IEEE Trans. Med. Imaging 21 (11) (2002) 1421–1439.
[34] N.J. Kabani, D.J. MacDonald, C.J. Holmes, A.C. Evans, 3D anatomical atlas of the human brain, Neuroimage 7 (4) (1998) S717.
[35] S. Purcell, et al., PLINK: a tool set for whole-genome association and population-based linkage analyses, The American journal of human genetics 81 (3) (2007) 559–575.
[36] Y. Li, C.J. Willer, J. Ding, P. Scheet, G.R. Abecasis, MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes, Genet. Epidemiol. 34 (8) (2010) 816–834.

[37] X. Hong et al., "Predicting Alzheimer's Disease Using LSTM," vol. 7, pp. 80893-80901, 2019.
[38] M. Liu, D. Cheng, K. Wang, Y. Wang, and A. s. D. N. Initiative, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," Neuroinformatics, vol. 16, no. 3-4, pp. 295-308, 2018.
[39] A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task feature learning," in Advances in neural information processing systems, 2007, pp. 41-48.
[40] X. He and P. Niyogi, "Locality preserving projections," in Advances in neural information processing systems, 2004, pp. 153-160.
[41] J. Yang, D.D. Zhang, A.F. Frangi, J.-Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. (2004).
[42] C. Zu et al., "Label-aligned multi-task feature learning for multimodal classification of Alzheimer's disease and mild cognitive impairment," vol. 10, no. 4, pp. 1148-1159, 2016.
[43] X. Zhu, H.-I. Suk, D. Shen, Sparse discriminative feature selection for multi-class Alzheimer's disease classification, Springer, 2014, pp. 157–164.
[44] X. Zhu, H.-I. Suk, S.-W. Lee, and D. J. I. T. o. B. E. Shen, "Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification," vol. 63, no. 3, pp. 607-618, 2015.
[45] M. López et al., "Principal component analysis-based techniques and supervised classification schemes for the early detection of Alzheimer's disease," vol. 74, no. 8, pp. 1260-1271, 2011.
[46] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, F. Segovia, and A. s. D. N. I. J. Neurocomputing, "Early diagnosis of Alzheimer′ s disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," vol. 151, pp. 139-150, 2015.
[47] N. Abdelaziz, R. T. Abd El-Hakim, S. M. El-Badawy, and H. A. J. I. J. o. P. E. Afify, "International Roughness Index prediction model for flexible pavements," vol. 21, no. 1, pp. 88-99, 2020.
[48] T. M. Nir et al., "Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging," vol. 3, pp. 180-195, 2013.
[49] J. Peng, L. An, X. Zhu, Y. Jin, D. Shen, Structured sparse kernel learning for imaging genetics based Alzheimer's disease diagnosis, Springer, 2016, pp. 70–78.
[50] T. Zhou, M. Liu, K.-H. Thung, and D. J. I. t. o. m. i. Shen, "Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data," vol. 38, no. 10, pp. 2411-2422, 2019.
[51] X. Hao et al., "Identifying Candidate Genetic Associations with MRI-derived AD-related ROI via Tree-guided Sparse Learning," IEEE/ACM transactions on computational biology and bioinformatics, 2018.
[52] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, and A. s. D. N. Initiative, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," Neuroimage, vol. 55, no. 3, pp. 856-867, 2011.
[53] M. Liu, D. Cheng, W. Yan, and A. s. D. N. Initiative, "Classification of Alzheimer's disease by combination of convolutional and recurrent neural networks using FDG-PET images," Frontiers in neuroinformatics, vol. 12, p. 35, 2018.
[54] C. Feng, A. Elazab, P. Yang, T. Wang, B. Lei, and X. Xiao, "3D convolutional neural network and stacked bidirectional recurrent neural network for Alzheimer's disease diagnosis," in International Workshop on PRedictive Intelligence In MEdicine, 2018, pp. 138-146: Springer.
[55] C. Feng, et al., Deep Learning Framework for Alzheimer's Disease Diagnosis via 3D-CNN and FSBi-LSTM, IEEE Access 7 (2019) 63605–63618.