Computer Methods and Programs in Biomedicine 213 (2022) 106503

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Ranking convolutional neural network for Alzheimer's disease mini-mental state examination prediction at multiple time-points

Hezhe Qiao[a,b,1], Lin Chen[a,1], Fan Zhu[a,*]

[a] Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 11 May 2021
Accepted 22 October 2021

Keywords:
Alzheimer's Disease (AD)
Convolutional Neural Network (CNN)
Magnetic Resonance Imaging (MRI)
Mini-Mental State Examination (MMSE)
Ranking learning

ABSTRACT

Background and objective: Alzheimer's disease (AD) is a fatal neurodegenerative disease. Predicting Mini-mental state examination (MMSE) based on magnetic resonance imaging (MRI) plays an important role in monitoring the progress of AD. Existing machine learning based methods cast MMSE prediction as a single metric regression problem simply and ignore the relationship between subjects with various scores. Methods: In this study, we proposed a ranking convolutional neural network (rankCNN) to address the prediction of MMSE through muti-classification. Specifically, we use a 3D convolutional neural network with sharing weights to extract the feature from MRI, followed by multiple sub-networks which transform the cognitive regression into a series of simpler binary classification. In addition, we further use a ranking layer to measure the ranking information between samples to strengthen the ability of the classification by extracting more discriminative features. Results: We evaluated the proposed model on ADNI-1 and ADNI-2 datasets with a total of 1,569 subjects. The Root Mean Squared Error (RMSE) of our proposed model at baseline is 2.238 and 2.434 on ADNI-1 and ADNI-2, respectively. Extensive experimental results on ADNI-1 and ADNI-2 datasets demonstrate that our proposed model is superior to several state-of-the-art methods at both baseline and future MMSE prediction of subjects. Conclusion: This paper provides a new method that can effectively predict the MMSE at baseline and future time points using baseline MRI, making it possible to use MRI for accurate early diagnosis of AD. The source code is freely available at https://github.com/fengduqianhe/ADrankCNN-master.

1. Introduction

Alzheimer's disease (AD) is a irreversible and chronic neurodegenerative disease that accounts for about 60% of all dementia cases [1]. Since the onset of AD is relatively late and the prevalence increases sharply with age, it is listed as the four major health killers along with cardiovascular and cerebrovascular diseases, tumors, and brain diseases [2]. At present, there are about 90 million AD patients in the world, and it is estimated that the number will reach 300 million in future [3]. Cognitive impairment, memory impairment, language impairment and personality changes are the main clinical manifestations of AD [4]. It is precisely because of the concealment and high cost of treatment that the early diagnosis and intervention of AD are very important [5]. As an important material for the early diagnosis of certain brain diseases, magnetic resonance imaging (MRI) has received more attention [6–8] and provides an effective way to determine the stage of the subject including Normal Control (NC), mild cognitive impairment (MCI), and AD [9,10,33,34]. A lot of studies have tried to use the whole image or specific regions of interest (ROI) in MRI to identify potential AD patients and predict the progression of AD for a specific subject [11,12].

Recently, estimating clinical scores like Mini-mental state examination (MMSE) of subjects in current time or future time using MRI also becomes a hot topic in early AD diagnosis. MMSE is a 30-point questionnaire that is widely used to measure the cognitive status of dementia in clinical diagnosis [13]. There are certain differences in the distribution of MMSE among subjects at different stages. Fox example, the MMSE of NC is between 24 and 30 (inclusive), and the MMSE score of AD is between 20-26 [14]. In the Fig. 1, we show the MMSE's distribution of subjects in ADNI-1 at four time points including baseline (BL), month06 (M06), month12 (M12) and month24 (M24). It can be seen the distribution of MMSE for all subjects is relatively concentrated at baseline. At 24 months, the range of MMSE distribution is 1-30,

* Corresponding author.
E-mail addresses: qiaohezhe@cigit.ac.cn (H. Qiao), chenlin@cigit.ac.cn (L. Chen), zhufan@cigit.ac.cn (F. Zhu).
[1] Equal contribution

https://doi.org/10.1016/j.cmpb.2021.106503
0169-2607/

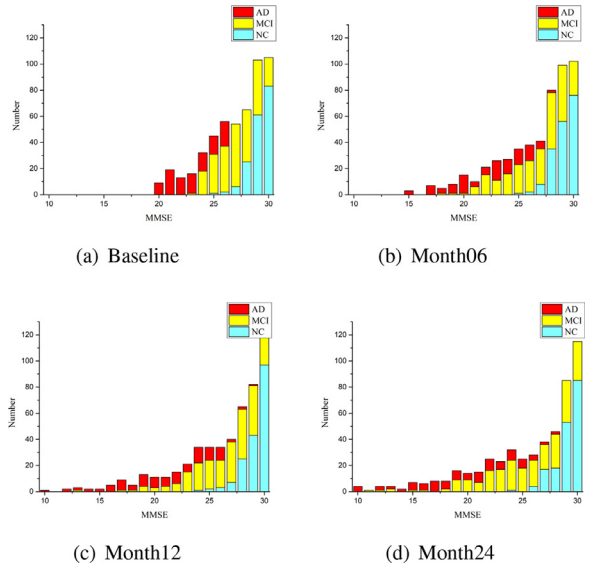


Fig. 1. Distribution of subject with MMSE from ADNI-1 at four time points including Basline, Month06, Month12, and Month24.

which shows that MMSE will change over time, and this variety for a specific subject is very useful for the early diagnosis of AD. The methods based on the structural MRIs for AD diagnosis could be divided into voxel-based morphology (VBM) methods and deep learning methods. We will introduce these two types of methods in the following.

Most of VBM studies focus on the analysis of hippocampus volumes, cortex sickness, subcortical volumes, etc. [15,16]. Conventional statistical methods like SVM, Random Forest were used to building hand-crafted features and explored the difference in the brain between AD and NC. The common method is to split MRI scan into multiple regions of interest and the feature like volume, thickness of ROI was computed as the input of machine learning model [17,18]. Nevertheless, studies based on VBM and machine learning methods required professional software operation and a large amount of expert knowledge [19].

Due to the superiority of deep learning in image classification, convolutional neural network (CNN) [20] was also applied on MRI for AD diagnosis [21]. Several studies use 3D CNN to extract the embedding features from the whole image in a data-driven manner [22]. Duc *et al.* developed a 3-dimensional neural network architecture for joint disease classification and cognitive regression based on functional MRI (fMRI). MMSE scores were achieved using linear square regression, support vector machine (SVM), bagging-based ensemble regression [23]. Liu et al. proposed a weakly supervised model by constructing a weighted loss function based on the incomplete clinical scores [24]. Yang et al. build a model to explore the relationship between MRI data and scores based on the deep polynomial network and SVM[25]. A 3D attention network was proposed by Jin et al. to capture the candidate imaging biomarkers, classification of subject, and regression of clinical scores [26,27]. Liu et al. proposed a multi-task for classification of subject and MMSE prediction simultaneously using MRI data and demographic information [28]. The joint learning of MMSE prediction and subject categories classification can further improve the performance of the model. In order to conduct a more precise analysis of the subjects, many studies have explored multi-modal data [29], combining a variety of image information, clinical information, genetic information, etc.

In addition to the prediction of MMSE directly, numerous studies are devoted to the prediction that whether it will produce a significant decline in the future time [30]. These studies usually use the subject's current visit status or collect multiple historical visit records to establish relationships between multiple visit records to predict future scoring status [37].

Although several methods have been proposed for predicting clinical scores in the future using baseline MRI. The prediction of MMSE is still an undetermined problem, mainly because (1) Analyses of high dimensional MRI data containing a large number of voxels is challenging work, especially the limited available training samples often leads to an over-fitted model. (2) Most of the existing deep learning based methods formulate the MMSE prediction as a one-dimensional regression problem where the precise value sampling from a continuous signal ([0-29]). However, the MMSE values have a nature of the ordinal relationship, which is not exploited in these regression models. (3) MRIs anatomical abnormalities of subjects with similar MMSE scores are relatively small, so it is necessary to find moderate and subtle changes in disease progression.

In this study, we propose a ranking CNN (rankCNN) based on the comparison of cardinality and individual for brain disease prognosis using subjects with MRI. A deep neural network with sharing weights was used to extract the MRI feature, then the regression of MMSE is transformed into the muti-classication tasks with multiple sub-networks with discrete MMSE values, which are rounded into 30 points labels [0,1,..,29], so that the correlation among the discrete MMSE points could be explored by jointly training in an end to end CNN. Furthermore, to portray the subtle changes between individuals, we constructed a ranking layer to force the CNN to explore more discriminative information to improve the classication formulation.

The main contribution of this study can be summarized as follows:

(1) Multiple sub-networks were proposed to transform the direct regression of MMSE into multi-classifications by exploring the relationship between the subject's MMSE and cardinality.

(2) The ranking information was exploited to compare the samples with various MMSE values, which could help the model find out the subtle changes between individuals.

(3) Our model achieves 2.23 of root mean square error (RMSE), and 0.57 of Pearson's correlation coefficient (CC) on ADNI-1, respectively. For the evaluation on ADNI-2, our model achieves 2.43 of RMSE and 0.43 of CC, which indicates that our model is superior to several states of the art methods.

2. Methods

The architecture of this network is shown in Fig. 2. Our model contains three main parts, including the feature extraction module, multiple sub-classifications for the regression module, and subject ranking layer.

2.1. 3D Convolutional neural network

The feature extraction module is composed of multiple blocks and each block unit consists of a 3D convolution layer, a batch normalize (BN) layer for 3D feature map, a rectified linear unit (ReLU) layer, and a 3D max-pooling layer. Although several activation functions have been proposed to obtain high performance in different architectures [31,32], ReLU has been applied in this work due to its computational efficiency. The 3D convolution performed on MRI was defined as following

$$u_j^l(x,y,z) = \sum_{\delta_x}\sum_{\delta_y}\sum_{\delta_z} F_k^{l-1}(x+\delta_x, y+\delta_y, z+\delta_z) \times W_{kj}^l(\delta_x,\delta_y,\delta_z) \quad (1)$$

where x, y, z is a 3D pixel coordinates of MRI, $F(x,y,z)$ represents a specific pixel in MRI at position x, y, z. $\delta_x, \delta_y, \delta_z$ denotes the three

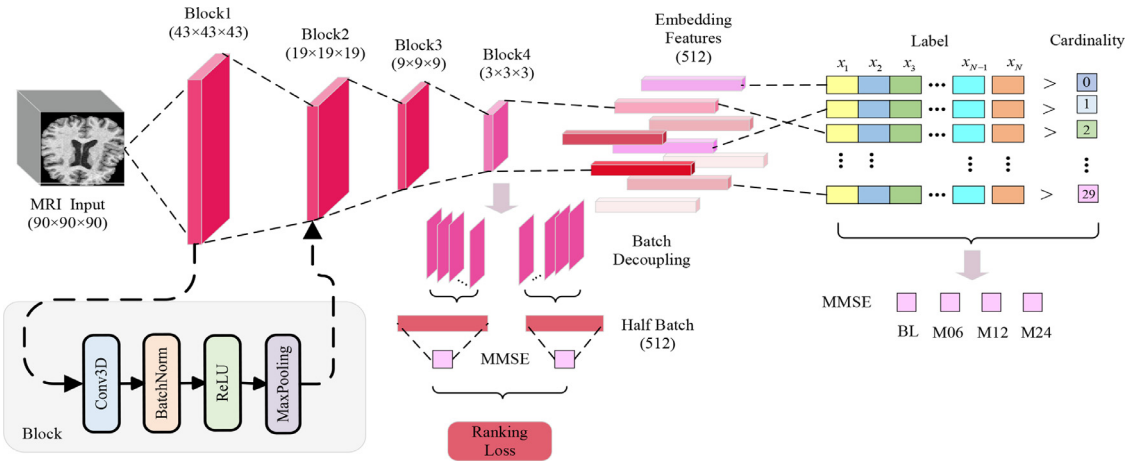


Fig. 2. The overall structure of our proposed model. The model structure consists of three main parts: a feature extraction module with sharing weights, multiple sub-classifiers for regression module, and a ranking layer. The output of the classifier is the distribution probability of the corresponding sequence position. The results of multiple classifiers are combined to form the final MMSE.

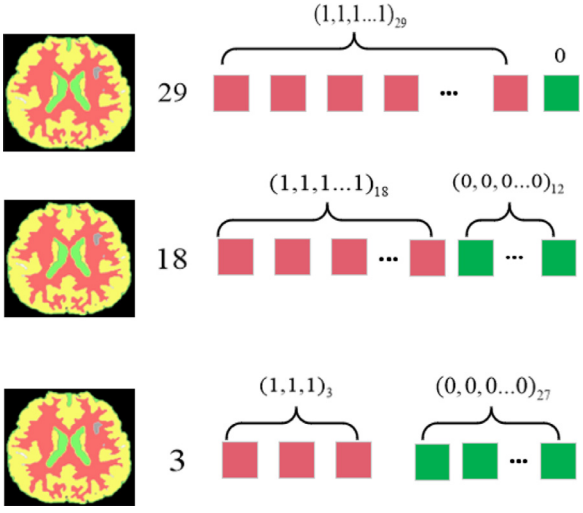


Fig. 3. Given the true value of MMSE corresponding to the MRI of three subjects, and the corresponding sequence label Y. The red block indicates that the current position is 1, and the green block indicates that the current position is 0.

dimensions of the convolution kernel. W_{kj} is the weight between F_k^{l-1}, the kth feature map in $l-1$ layer, and u_j, the jth feature map in the lth layer. The last layer of the feature extraction module is a 3D AvgPooling layer, which could converts the feature maps extracted from MRI into feature vectors. It is worth mentioning that all MRIs share a same feature extraction module.

2.2. Sub-networks for cognitive regression

Before introducing the classification of sub-networks, we first introduce how to construct the vector label based on the real MMSE. Given a labeled training set D with N instances (x_i, l_i), where x_i denotes the ith subject's MRI image and l_i denotes the its MMSE, $i \in [1, N]$. For the MRI image x_i with MMSE l_i, $l_i \in [0, 1, 2, ..., 29]$. As shown in Fig. 3, since MMSE is a discrete sequence, so we can transform it into a binary sequence Y consists of 0 and 1 [38,39], so that the regression of MMSE can be formulated as a series of binary classification problems. The vector label y is defined as the following:

$$y_i^k = \begin{cases} 1 & if(l_i > k) \\ 0 & otherwise \end{cases} \tag{2}$$

where y_i^k is the label of x_i at the k dimension, which denotes whether l_i is greater than the cardinality k. k ranges from 0 to 29. For instance, given a subject's MRI x with MMSE 18, the y consists of $(1, 1, ...1)_n$, $n = 18$ and $(0, 0, .., 0)_m$, $m = 12$.

After the feature extraction module, we got the embedding feature of each MRI. To complete the prediction of MMSE based on classification, we have constructed multiple sub-classifiers, a network composed of multiple fully connected layers. Since the same embedding feature is used as the input of multiple sub-networks, the correlation among these distinct MMSE labels could be explored by jointly training the sub-classifier.

Each binary sub-classifier corresponds to a value in the vector label. The output of k the classifier is the probability of whether the subject's MMSE is greater than k. Each output of the sub-classifier corresponds to a binary classification, the cross-entropy is employed as the loss function L_k defined as

$$\mathcal{L}_k = -\frac{1}{N}\sum_{i=1}^{N} I(o_i = y_i)\log(p(o_i|x_i, W)) \tag{3}$$

where o_i is the output of kth sub-classifier of the x_i, W denotes the parameters of the sub-classifier. Although hybrid loss functions have been proposed in some recent works [35,36], the cross-entropy is nearly a default loss function and preferred due to its efficiency.

Most of the original studies are directly regression MMSE, the L_{MSE} is defined as

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - f(x_i; W))^2 \tag{4}$$

where $f(x_i; W)$ and y_i is the estimated MMSE and ground truth of x_i respectively. Through multiple sub-classifiers, the mean square loss L is transformed into the sum of Multiple sub-classifiers' loss function.

$$\mathcal{L}_{rank-c} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} I(o_i^t = y_i^t)\log(p(o_i^t|x_i, W^t)) \tag{5}$$

where o_i^t and y_i^t is the predicted MMSE and ground truth of tth task forx_i. W^t is the parameters of tth task.

After multiple sub-classifiers, we need to convert the output result into a real MMSE. We adopted two methods to complete the transformation, one is to set the same threshold to decide whether to take 0 or 1 as the output for each classifier. The y_i is the output

of the binary sub-classifier which is defined as

$$y^i = \begin{cases} 0 & \hat{y}^i < \varepsilon \\ 1 & \hat{y}^i \geq \varepsilon \end{cases} \tag{6}$$

where ε is the threshold. The final y is the sum of output. Another way is to directly sum up the output distribution of multiple sub-classifiers without processing. The value obtained in this way is continuous. In this way, the setting of the threshold can be avoided, and the MMSE of the subject can be obtained in a more precise manner. The effects of these two methods will be discussed in the following section.

2.3. Ranking layers

Based on the assumption that the MRI of subjects with similar MMSE scores may have similar characteristics, we designed a ranking layer to measure the similarities or differences between individuals. Based on comparison with cardinality, the ranking layer could explore ranking information between samples to force the CNN extracting more discriminative features. Giving a input batch $(x_0, x_1, ..., x_b)$, where b denotes the batch size, we can split it into two sub-batches $(x_0, ..., x_{b/2})$ and $(x_{b/2}, ..., x_b)$ through batch decoupling. The MRI of each corresponding position in two sub-branches composes a pair (x_i, x_j), which is used as the input of our ranking CNN. We first define Cos_{ij} for each paired MRI [40].

$$\mathrm{Cos}_{ij} = \begin{cases} 0 & if \quad l_i - l_j < -\phi \\ 0.5 & if \quad |l_i - l_j| \leq \phi \\ 1 & if \quad l_i - l_j > \phi \end{cases} \tag{7}$$

where the Cos_{ij} is the target probability of x_i is bigger the x_j. Considering that the MMSE of the subjects may not be strictly equal, and may fluctuate within a certain range, we set ϕ, $\phi \geq 0$, denotes the soft margin for the paired MMSE. When the MMSE difference between the subjects is within ϕ, we think they may be similar in feature representation. $Cos_{ij} = 0.5$ represent that x_i is the same MMSE as x_j. We define the $P(x_i > x_j)$ using a logistic function which denotes the probability the subject x_i has a greater MMSE than x_j.

$$p_{ij} = P(x_i > x_j) = \frac{1}{1 - e^{-(f(\varphi_i) - f(\varphi_j))}} \tag{8}$$

where φ_i is the embedding feature extracted from MRI x_i and $f(\varphi_i)$ denotes the branch output of the ranking layer, which was used to calculate the paired subject rank probability. Instead, it only uses the different information to supervise the training of the model. The loss function of the ranking layer is a binary cross-entropy L_{rank-s} define as

$$\mathcal{L}_{rank-s} = -\mathrm{Cos}_{ij} \log p_{ij} - (1 - \mathrm{Cos}_{ij}) \log(1 - p_{ij}) \tag{9}$$

2.4. Total loss function

We have designed a total loss function to effectively learn our proposed rankCNN, in which the Multiple sub-classifiers loss function and the subject rank layer loss function were combined for modeling. The total loss L_{total} is designed as

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{rank-c} + \beta \mathcal{L}_{rank-s} \tag{10}$$

where α and β are hyperparameters that adjust the weights between these two loss functions. It can be seen that the total loss function is composed of multiple cross-entropy functions, including the comparison with the cardinality and the comparison between individuals.

Table 1
The number of subjects from two databases including ADNI-1, ADNI-2 at four time points BL, M06, M12 and M24.

Data	Diagnosis	BL	M06	M12	M24
ADNI-1	NC	229	215	200	178
	MCI	397	355	317	236
	AD	183	162	137	103
	Subject	809	732	654	517
ADNI-2	NC	184	169	158	127
	MCI	442	310	292	232
	AD	143	98	86	23
	Subject	769	577	536	382

2.5. Materials and image processing

In this section, we will introduce the information of the subject used in our study and the procedure of MR image preprocessing.

The subjects' MRI and MMSE information at different timepoint used in this study was obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI), which is available at http://adni.loni.usc.edu/ [41]. The statistic information of these datasets are available in Table 1. Different numbers of subjects are available at different time points corresponding to the baseline time point. When predicting the MMSE at future time points, we only predict subjects with the baseline MRI.

We first performed anterior joint (AC)-poster joint (PC) correction on all MRIs, then registered all the MRIs into the same space. Due to the large number of invalid background areas in MRI, it may affect the performance of the model. We removed the invalid regions of each MRI and left the only area of the brain. In order to eliminate the influence of the distribution between different MRIs, we resized MRI to the same resolution of $90 \times 90 \times 90$ and adopted intensity normalization.

3. Results

3.1. Experimental setting

To verify the effectiveness of the proposed model, following the experimental setting in [24,28,42,43], we first use the ADNI-1 dataset as the training set and evaluate with ADNI-2. In the second group, we use ADNI-2 as the training set and evaluate with ADNI-1. Our goal is to predict the mmse in BL and future time points using BL MRI. We use RMSE and correlation coefficient (CC) as the criteria of the models. Our model is implemented with Pytorch library and trained on 1 NVIDIA GeForce GTX 1080Ti GPUs with 11G memory. We have also performed t -test (with a significance level of 0.05) on the prediction results achieved by our model rankCNN and the baseline model 3DCNN.

3.2. Comparison methods

(1) SVM: We first use a nonlinear image registration algorithm [44] to spatially normalize all MR images to the template space. Then CAT toolbox (http://dbm.neuro.uni-jena.de/cat/) was used to segment each MRI to gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). The feature vector can be attained by local GM tissue density. All these feature vectors were used as the input of a support vector regression (SVR) for MMSE regression with an RBF kernel.

(2) 3D CNN: A multi-layer 3D convolutional is used to extract embedding features from MRI. After, there is an AvgPooling layer to map the extracted features into a vector for each MRI. Finally, a fully connected layer is used to mapping the vector to the predicted MMSE.

Table 2
Results of RMSE and CC achieved by different models trained on the ADNI-1 dataset and evaluated on the ADNI-2 dataset. The terms denoted by * means that the result of rankCNN are statistically significantly better than tested models ($p < 0.05$) using pairwise t-Test. The best result is highlighted in bold.

Method	RMSE				CC			
	BL	M06	M12	M24	BL	M06	M12	M24
SVM	2.652	2.885	3.553	3.864	0.270	0.247	0.239	0.218
3DCNN	2.748	2.805	3.516	3.623	0.332	0.235	0.255	0.381
rankCNN-C	2.485	2.424	2.795	3.449	0.523	0.408	0.413	0.447
rankCNN-CS	2.316	2.362	2.746	3.346	0.537	0.413	0.423	0.490
rankCNN	**2.238***	**2.299***	**2.578**	**3.300***	**0.578***	**0.435***	**0.427**	**0.504***

(3) rankCNN-C: To investigate the influence of each component, we further compare ranking CNN without ranking layer. rankCNN-C refers to the model that uses only the data where the MMSE is compared with the base and the predicted MMSE is the sum of the output of multiple sub-classifiers with a threshold.

(4) rankCNN-CS: Similar to the rankCNN-C, rankCNN-CS leave the cardinality comparison part while the predicted MMSE is the sum of the output of multiple sub-classifiers without thresholds.

3.3. Results on ADNI-2

Experimental results of model trained on ADNI-1 and evaluated on ADNI-2 were shown in Table 2 and Fig. 5. We also report the statistical analysis using a paired t-Test (with a significance level of 0.05) on prediction results achieved by our rankCNN method and other comparison methods in Table 2. The distribution scatter plots of the predictions generated by our model were illustrated in Fig. 4. The proposed model achieves Pearson's correlation coefficient (CC) of 0.578 at the baseline, higher than other methods. This result suggests that rankCNN could improve the learning performance by exploring the ranking information. The performances of rankCNN at the two time-points of BL and M24 are better than that of M06 and M12. This may be because, at the two time-points of M06 and M12, the MMSE of the subjects is in an unstable stage compared to BL and M24, which makes it difficult for the model to generate more accurate predictions. Result also indicates that rankCNN-CS performs better than rankCNN-C. This is because the sum of multiple sub-classifiers makes the final MMSE value continuous, which brings the predicted value closer to the true value while rankCNN-C needs to manually set the threshold.

In general, our method shows a clear advantage compared to traditional SVM and simple 3D CNN methods in both RMSE and CC. It can be seen rankCNN shows superiority at all four points than rankCNN-CS, which may be due to the additional information from the comparison between individual samples.

3.4. Results on ADNI-1

In this section, the model was trained on ADNI-2 and evaluated on ADNI-1 to further explore the effectiveness. We report the RMSE and CC achieved by four different methods in Table 3 and Fig. 5, and denote the significantly better results of rankCNN with * by using the pairwise t-Test ($p < 0.05$) in Table 3. The distribution scatter plots were shown in Fig. 6.

Similar to the results on ADNI-2, that rankCNN achieves the higher CC of 0.485 and 0.459 respectively than rankCNN-CS at baseline and M24, while rankCNN-CS achieves the highest CC at M06 and M12. These further show the fact that the MMSE changes between subjects at these two time-points are relatively unstable, leading to negative feedback of the ranking layer on the ADNI-2. Compared with other methods, rankCNN achieves the lowest RMSE at four time-points. Similarly, we can see that rankCNN-C performs better than 3D CNN and SVM on ADNI-1, which further demonstrates the effectiveness of our method. For example, 3D CNN has CC of 0.159 and 0.272 at BL and M12. The CC at two time-points has been greatly improved by the performance of rankCNN-C, reaching 0.441 and 0.374.

The results imply that comparing with cardinality is more effective than direct regression for MMSE prediction. Comparing Tables 2 and 3, we find the proposed rankCNN achieves well results on the ADNI-1 and ADNI-2 than other methods. Although all correlations between predicted MMSE and ground truth are not high, this method could provide some help on AD diagnosis to some extent.

Results on ADNI-1 and ADNI-2 are quite different, *i.e.* the model trained with ADNI-2 is lower than the one trained with ADNI-1. This situation is consistent with results from the existing study [24,28]. The major reason is that ADNI-1 and ADNI-2 were acquired by 1.5T and 3.0T scanners, which leads to different imaging quality that directly affects prediction accuracy.

3.5. Comparison with state-of-the-art methods

We compared the proposed methods with state-of-the-art methods in the literature, including conventional methods and deep learning-based models.

(1) Voxel: The Fast algorithm [45] in FSL was used to segment MRI into three tissue types. A support vector regression (SVR) model was construct based on GM tissue density in each voxel for the prediction of MMSE.

(2) ROI: This method registers the MRI to AAL template by using a nonlinear registration algorithm to generate 90 manually labeled ROI. The tissue volume of each ROI in GM is computed as a feature, which is further fed into SVR.

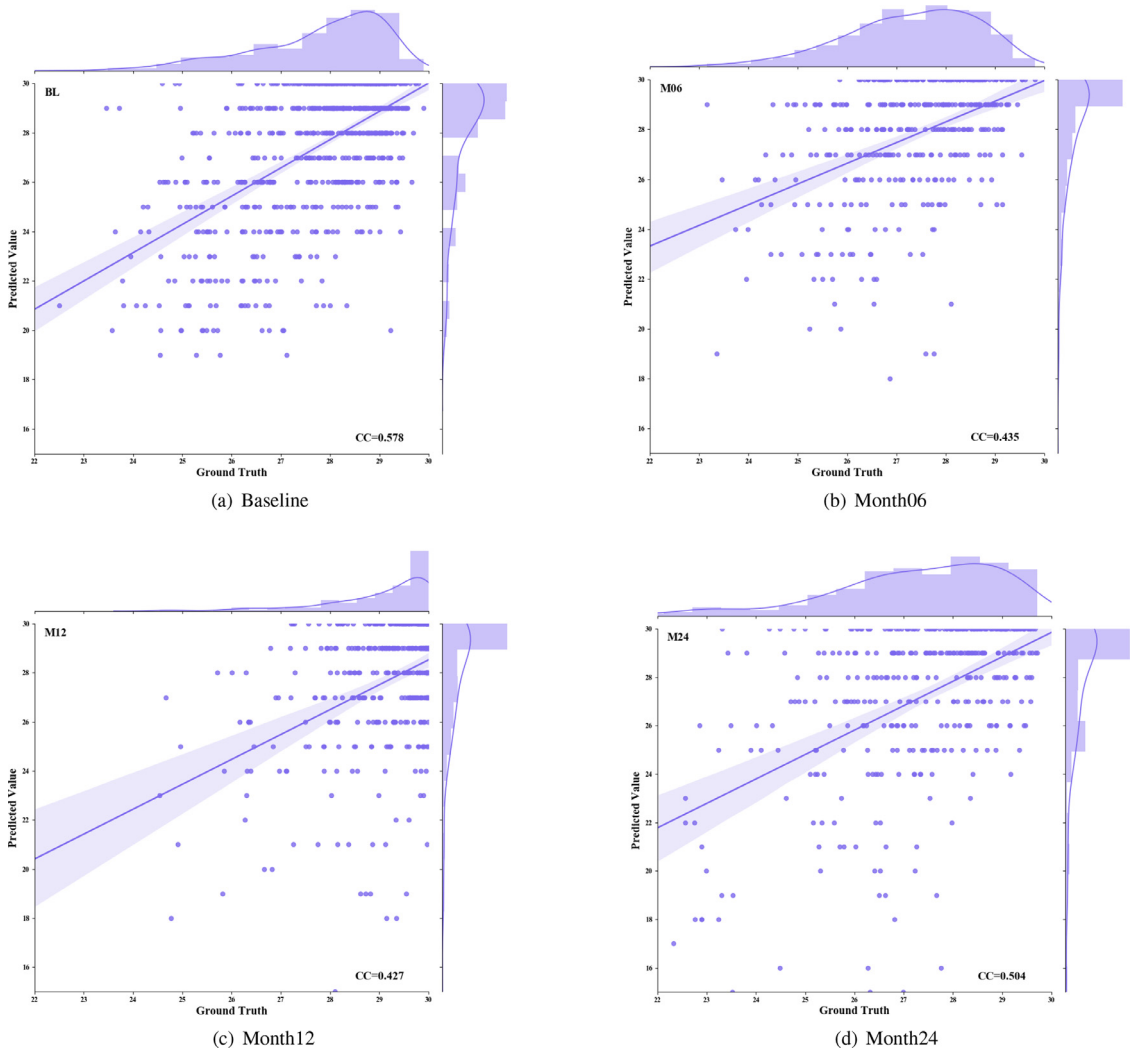
(3) LMF: The method computes the 100-D local energy pattern [46] from the patch extract from K landmarks [47,48]. The vector of the patch was concatenated and used as the input of SVR.

(4) M^{3}TL: The method is a multi-modal multi-task learning model that relies on the hand-crafted feature. A linear SVR was trained for MMSE regression at four time-points [49].

(5) DM^{2}L: DM^{2}L extracts the patches from the discriminative anatomical landmarks and builds a deep multi-task learning model performing the joint regression of MMSE and classification of subjects through CNN [28].

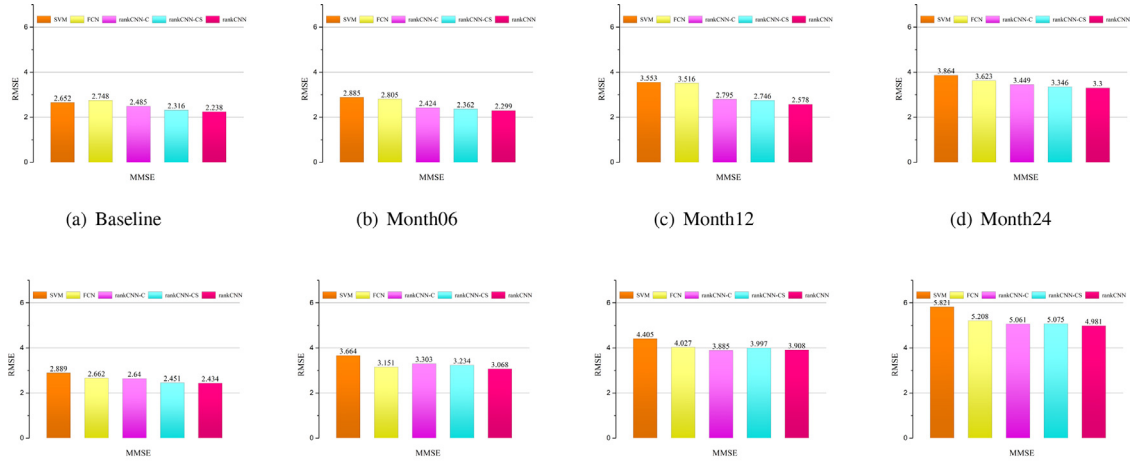
(6) wiseDNN-IS: Similar to the DM^{2}L, wiseDNN-IS further extracts the small scale image patches from landmarks and develops a weakly supervised densely deep model for the prediction of MMSE [24].

We compared the result of proposed method with above state-of-the-art methods (Table 4 and Fig. 7). Results indicate that the proposed rankCNN outperforms the M^{3}TL, DM^{2}L, and wiseDNN in terms of RMSE. For M^{3}TL and DM^{2}L, only results at baseline and M24 are available in their literature. Our method performs well in the evaluation at all time points. It achieves the highest CC at base-



(a) Baseline (b) Month06 (c) Month12 (d) Month24

Fig. 4. The distributions scatter plots of our model trained on the ADNI-1 dataset and evaluated on the ADNI-2 dataset at multiple time points.

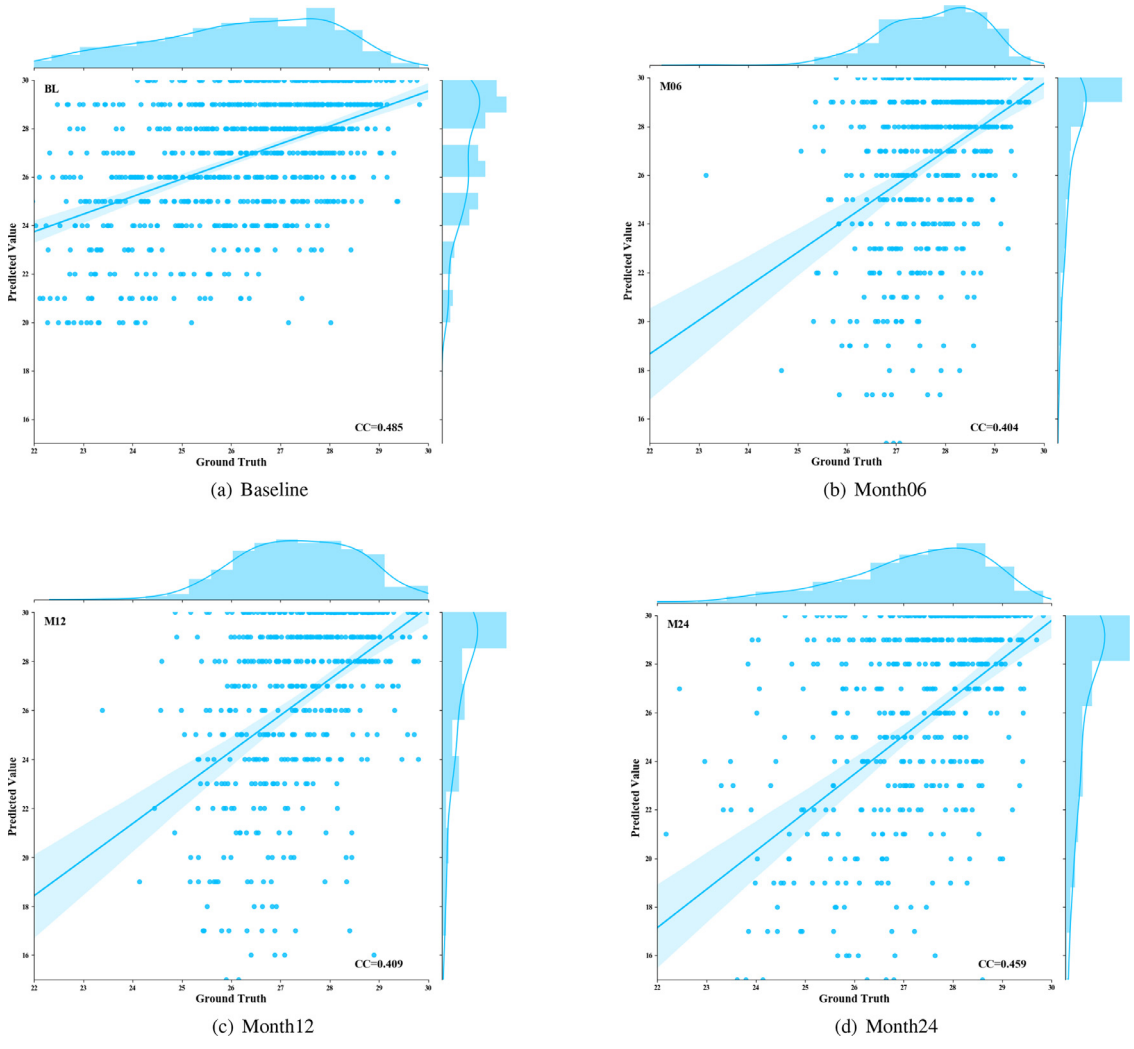


(e) Baseline (f) Month06 (g) Month12 (h) Month24

Fig. 5. Results of RMSE achieve by different methods at multiple time points. (a), (b), (c) and (d) are the results of models trained on ADNI-1 and evaluated on ADNI-2. (e), (f), (g) and (h) are the results of models trained on ADNI-2 and evaluated on ADNI-1.

Table 3
Results of RMSE and CC achieved by different models trained on the ADNI-1 dataset and evaluated on the ADNI-2 dataset. The term denoted by * means that the result of rankCNN are statistically significantly better than tested models ($p < 0.05$) using pairwise t-Test. The best result is highlighted in bold.

Method	RMSE				CC			
	BL	M06	M12	M24	BL	M06	M12	M24
SVM	2.889	3.664	4.405	5.821	0.104	0.219	0.220	0.186
3D CNN	2.622	3.151	4.027	5.208	0.159	0.272	0.315	0.363
rankCNN-C	2.640	3.303	**3.885**	5.061	0.441	0.398	0.405	0.374
rankCNN-CS	2.451	3.234	3.997	5.075	0.437	**0.429**	**0.437**	0.443
rankCNN	**2.434***	**3.068**	3.908	**4.981***	**0.485***	0.404	0.409	**0.459***



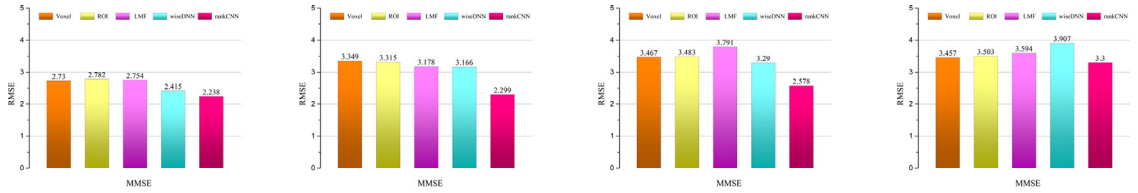
(a) Baseline (b) Month06

(c) Month12 (d) Month24

Fig. 6. The distributions scatter plots of our model trained on the ADNI-2 dataset and evaluated on the ADNI-1 dataset at multiple time points.

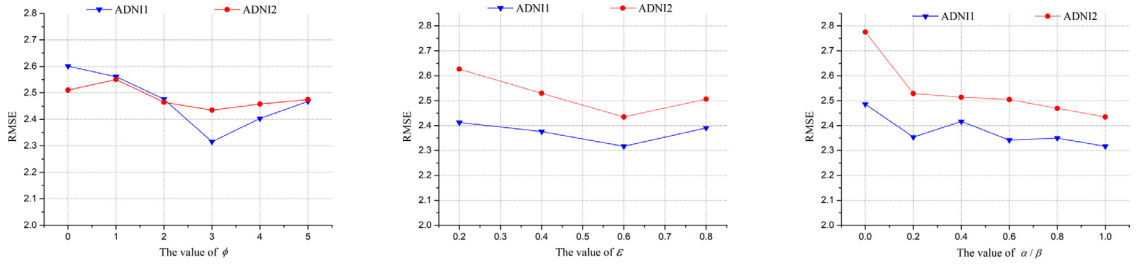
Table 4
A summary table for the comparison of the RMSE and CC achieve by our method and other methods reported in several literature. All models are trained on ADNI-1 and evaluated on ADNI-2. The best result is highlighted in bold.

Reference	Method	RMSE				CC			
		BL	M06	M12	M24	BL	M06	M12	M24
Baron et al. (Voxel) [45]	GM Density +SVR	2.730	3.349	3.467	3.457	0.309	0.254	0.233	0.146
Zhang et al. (ROI) [49]	ROI + SVR	2.782	3.315	3.483	3.503	0.306	0.405	0.423	0.265
Zhang et al.(LMF) [47]	Landmark + SVR	2.754	3.178	3.791	3.594	0.331	0.405	0.423	0.364
Zhang et al. (M^3TL) [49]	Multi-Modal	4.740	–	–	5.853	0.504	–	–	0.445
Liu et al. (DM^2L) [28]	Multi-Channel	3.073	–	–	4.625	0.565	–	–	**0.518**
Liu et al.(wiseDNN) [24]	Weakly Supervised	2.415	3.166	3.290	3.907	0.538	**0.564**	**0.526**	0.477
proposed	MMSE Ranking	**2.238**	**2.299**	**2.578**	**3.300**	0.578	0.435	0.427	0.504



(a) Baseline (b) Month06 (c) Month12 (d) Month24

Fig. 7. Comparison of RMSE achieved by our methods and the methods reported in several literature at multiple time points.



(a) (b) (c)

Fig. 8. Discuss on the effects of different hyperparameters including the threshold of sub-network, the threshold of ranking layer and the hyperparameter of loss function.

line, and is only slightly lower than wiseDNN at M06 and M12. The CC of our model at M24 is lower than DM^{2}L, but slightly higher than wiseDNN.

4. Discussion

In the following, we will discuss the influence of some parameters and then present the limitation and future work of this paper.

4.1. Effects of the threshold of sub-networks

The threshold in each muti-classification of rankCNN-C is generally set to 0.5. To further explore the effectiveness of this threshold, we set different thresholds and report the RMSE changes of the model on ADNI-1 and ADNI-2 in Fig. 8 (a). We find that the effect of the model first gets better and then gets worse, as the threshold increases. If the threshold is too high or too small, it will reduce the performance of the model. The reason may be that when the threshold is set too large, the sub-classifier loses some non-significant samples for the task at a specific location. On the contrary, if the threshold is too small, the sub-classifier will introduce some wrong information. In our experiments, the proposed model yields the best effect when the threshold is set to 0.6.

4.2. Discussion on the threshold of ranking layer

We now evaluate the effects of the threshold in the ranking layer. The threshold needs to be set manually because the MMSE scores between different individuals may not be strictly equal, and there may be certain differences. We adjusted different thresholds and reported the results in Fig. 8 (b). We found that when we set the rank threshold to 3, the RMSE of the model is the smallest. When the threshold is set too small or too large, the performance of the model will decrease. The reason is that a large threshold may impair the ability to mine more similar samples. On the contrary, a small threshold may cause the wrong classification of confused samples.

4.3. Discussion on the hyperparameter of loss function

The loss function of our proposed model consists of two parts, one part is the cross-entropy loss sum of sub-classifiers used to compare the subject's MMSE with cardinality, and the other part is the cross-entropy loss used for comparison between individuals. We adjusted the weights between the two loss functions, and the results were reported in Fig. 8 (c). It can be seen when the weight ratio of the two loss functions is 1, the RMSE of the model reaches the lowest, which further verifies the comparable importance of the two modules and the effectiveness of each module we proposed.

4.4. Limitation and future work

This paper still has several limitations as follows: (1) We extract features from the full MRI image, which inevitably introduces too much useless information and may hamper the performance of the model. (2) We did not consider the different data distribution between ADNI-1 and ADNI-2 datasets due to the different imaging manner of these two data sets, resulting in subtle inconsistent in experimental results. (3) We analyzed each time node separately and did not integrate the historical records of a subject. Analyzing the changes of a subject's MRI over time may be more helpful in the early diagnosis of AD. As future work, we will continue this study in the following: (1) We can build a model that combines local features and global features extracted from MRI, *i.e.* extract features at the ROI, patch, and whole image levels respectively for subject classification and MMSE prediction. (2) Integrate the subject's historical record information to construct a sequence network, *i.e.* build a recurrent neural network model to extract more discriminative features in a multi-sequence MRI for a specific subject. (3) In addition to the prediction of MMSE, we also apply the model to other clinical scores such as ADAS-Cog11, ADAS-Cog13, and CDR-SB. (4) Based on MRI analysis, we will add more clinical information, genetic information, and prior knowledge to construct a comprehensive AD diagnosis method.

5. Conclusion

In this paper, we proposed a rankCNN for MMSE prediction at the baseline and the future time-points for AD diagnosis. In the proposed method, we first use a comparison between MMSE and cardinality to complete the regression of MMSE through muti-classification. We further use the ranking layer to explore the rela-

tionship between the subject with different MMSE. We evaluated the proposed model on ADNI-1 and ADNI-2 datasets with 1569 subjects. Experiment results suggest that the proposed method delivers superior performance compared with other state-of-the-art approaches and can effectively predict the MMSE at baseline and future time points using MRI images collected at baseline. In general, our method could provide the possibility of early AD diagnosis.

Declaration of Competing Interest

All authors declared no conflict of interests. The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the manuscript submitted entitled with: “Ranking convolutional neural network for Alzheimer’s disease mini-mental state examination prediction at multiple time-points”.

CRediT authorship contribution statement

Hezhe Qiao: Conceptualization, Methodology, Software. **Lin Chen:** Conceptualization, Methodology, Software. **Fan Zhu:** Writing – original draft, Writing – review & editing.

Acknowledgements

This research is supported in part by the National Nature Science Foundation of China under grants No. 61902370 and No. 61802360, and in part by the Chongqing Research Program of Technology Innovation and Application under grants No. cstc2019jscx-zdztzxX0019, and in part by the key cooperation project of chongqing municipal education commission(HZ2021008). We obtain the data from ADNI database. Researchers within ADNI have not participated in the analysis or writing of this study. A complete list of ADNI investigators can be found online.

References

[1] C.R. Jack Jr, D.S. Knopman, W.J. Jagust, et al., Tracking pathophysiological processes in Alzheimer’s disease: an updated hypothetical model of dynamic biomarkers, Lancet Neurol. 12 (2) (2013) 207–216, doi:10.1016/S1474-4422(12)70291-0.
[2] M. Tanveer, B. Richhariya, R.U. Khan, et al., Machine learning techniques for the diagnosis of Alzheimer’s disease: a review, ACM Trans. Multimed. Comput.Commun. Appl. 16 (1s) (2020) 1–35, doi:10.1145/3344998.
[3] A. Burns, Alzheimer’s disease: on the verges of treatment and prevention, Lancet Neurol. 8 (1) (2009) 4–5, doi:10.1016/S1474-4422(08)70271-0.
[4] Y.Y. Wang, S.F. Yu, H.Y. Xue, Effectiveness and safety of acupuncture for the treatment of Alzheimer's disease: a systematic review and meta-analysis, Front. Aging Neurosci. 12 (2020) 98, doi:10.3389/fnagi.2020.00098.
[5] B. Dubois, H. Hampel, H.H. Feldman, Preclinical Alzheimer’s disease: definition, natural history, and diagnostic criteria, Alzheimer's Dement. 12 (3) (2016) 292–323, doi:10.1016/j.jalz.2016.02.002.
[6] T.D. Vu, N.H. Ho, H.J. Yang, Non-white matter tissue extraction and deep convolutional neural network for Alzheimer's disease detection, Soft Comput. 22 (20) (2018) 6825–6833, doi:10.1007/s00500-018-3421-5.
[7] U.R. Acharya, S.L. Fernandes, J.E. WeiKoh, Automated detection of Alzheimer's disease using brain MRI images-a study with various feature extraction techniques, J. Med. Syst. 43 (9) (2019) 1–14, doi:10.1007/s10916-019-1428-9.
[8] E. Goceri, C. Songül, Computer-based segmentation, change detection and quantification for lesions in multiple sclerosis, in: 2017 International Conference on Computer Science and Engineering (UBMK), IEEE, 2017, pp. 177–182, doi:10.1109/UBMK.2017.8093371.
[9] N. Yamanakkanavar, J.Y. Choi, B. Lee, MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: a survey, Sensors 20 (11) (2020) 3243, doi:10.3390/s20113243.
[10] H. Amiri, A. de Sitter, K. Bendfeldt, Urgent challenges in quantification and interpretation of brain grey matter atrophy in individual MS patients using MRI, NeuroImage 19 (2018) 466, doi:10.1109/UBMK.2017.8093371.
[11] R. Cui, M. Liu, Alzheimer's disease neuroimaging initiative. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease, Computerized Medical Imaging and Graphics 73 (2019) 1, doi:10.1016/j.compmedimag.2019.01.005.
[12] A. Puente-Castro, E. Fernandez-Blanco, A. Pazos, et al., Automatic assessment of alzheimer's disease diagnosis based on deep learning techniques, Comput. Biol. Med. 120 (2020) 103764, doi:10.1016/j.compbiomed.2020.103764.
[13] P. Kourtesis, E. Margioti, C. Demenega, A comparison of the greek ACE-III, m-ACE, ACE-r, MMSE, and ECAS in the assessment and identification of Alzheimer's disease, J. Int. Neuropsychol. Soc. 26 (8) (2020) 825–834, doi:10.1017/S1355617720000314.
[14] S. Sharma, R. Sharma, S. Gutt, The effect of neurofibrillary tangles on MMSE scores in patients with amyloid plaque: neuroimaging differential diagnosis, Alzheimer's Dement. 16 (2020), doi:10.1002/alz.036316. E036316
[15] S.H. Hojjati, A. Ebrahimzadeh, A. Khazaee, Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI, Comput. Biol. Med. 102 (2018) 30, doi:10.1016/j.compbiomed.2018.09.004.
[16] I. Beheshti, H. Demirel, Alzheimer's disease neuroimaging initiative. Probability distribution function-based classification of structural MRI for the detection of Alzheimer's disease, Comput. Biol. Med. 64 (2015) 208, doi:10.1016/j.compbiomed.2015.07.006.
[17] J. Feng, S. Zhang, L. Chen, Extracting ROI-based contourlet subband energy feature from the sMRI image for Alzheimer's disease classification, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2021, doi:10.1109/TCBB.2021.3051177.
[18] E. Goceri, Automated measurement of changes in cortical thickness from MR images, in: International Conference on Applied Analysis and Mathematical Modeling (ICAAMM2018), 78, 2018.
[19] X. Ji, H. Wang, M. Zhu, Brainstem atrophy in the early stage of Alzheimer's disease: a voxel-based morphometry study, Brain Imaging Behav. 15 (1) (2021) 49–59, doi:10.1007/s11682-019-00231-3.
[20] Y. LeCun, Y. . Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444, doi:10.1038/nature14539.
[21] E. Altinkaya, K. Polat, B. Barakli, Detection of Alzheimer's disease and dementia states based on deep learning from MRI images: a comprehensive review, J. Inst. Electron. Comput. 1 (1) (2020) 39–53, doi:10.33969/JIEC.2019.11005.
[22] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation, Med. Image Anal. 63 (2020) 101694, doi:10.1016/j.media.2020.101694.
[23] N.T. Duc, S. Ryu, M.N.I. Qureshi, 3d-deep learning based automatic diagnosis of Alzheimer's disease with joint MMSE prediction using resting-state fMRI, Neuroinformatics 18 (1) (2020) 71–86, doi:10.1007/s12021-019-09419-w.
[24] M. Liu, J. Zhang, C. Lian, Weakly supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores, IEEE Trans. Cybern. 50 (7) (2020) 3381–3392, doi:10.1109/TCYB.2019.2904186.
[25] M. Yang, P. Yang, A. Elazab, Join and deep ensemble regression of clinical scores for Alzheimer's disease using longitudinal and incomplete data, in: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 1254–1257, doi:10.1109/EMBC.2018.8512549.
[26] D. JinD, B. Zhou, Y. Han, Generalizable, reproducible, and neuroscientifically interpretable imaging biomarkers for Alzheimer’s disease, Adv. Sci. 7 (14) (2020) 2000675, doi:10.1002/advs.202000675.
[27] D. Jin, J. Xu, K. Zhao, Attention-based 3d convolutional network for Alzheimer's disease diagnosis and biomarkers exploration, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, 2019, pp. 1047–1051, doi:10.1109/ISBI.2019.8759455.
[28] M. Liu, J. Zhang, E. Adeli, 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, doi:10.1109/TBME.2018.2869989.
[29] M. Liu, F. Li, H. Yan, A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease, NeuroImage 208 (2020) 116459, doi:10.1016/j.neuroimage.2019.116459.
[30] R. Wang, L. Fratiglioni, G. Kalpouzos, Mixed brain lesions mediate the association between cardiovascular risk burden and cognitive decline in old age: apopulation-based study, Alzheimer's Dement. 13 (3) (2017) 247–256, doi:10.1016/j.jalz.2016.06.2363.
[31] E. Goceri, Analysis of deep networks with residual blocks and different activation functions: classification of skin diseases, in: 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA) IEEE, 2019, pp. 1–6, doi:10.1109/IPTA.2019.8936083.
[32] E. Goceri, Capsnet topology to classify tumours from brain images and comparative evaluation, IET Image Processing 14(5) 882–889. 10.1049/iet-ipr.2019.0312
[33] V.J.A. Verlinden, G.J.N. van der, A. Hofman, Brain MRI-markers associate differentially with cognitive versus functional decline leading to dementia, J. Am. Geriatr. Soc. 65 (6) (2017) 1258–1266, doi:10.1111/jgs.14775.
[34] H. Qiao, L. Chen, Z. Ye, Early Alzheimer's disease diagnosis with the contrastive loss using paired structural MRIs, Comput. Methods Programs Biomed. 208 (2021) 106282, doi:10.1016/j.cmpb.2021.106282.
[35] E. Goceri, Deep learning based classification of facial dermatological disorders, Comput. Biol. Med. 128 (2020) 104118, doi:10.1016/j.compbiomed.2020.104118.
[36] E. Goceri, Diagnosis of skin diseases in the era of deep learning and mobile technology, Comput. Biol. Med. 134 (2021) 104458, doi:10.1016/j.compbiomed.2021.104458.
[37] G.I. Allen, N. Amoroso, C. Anghel, Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease, Alzheimer's Dement. 12 (6) (2016) 645–653, doi:10.1016/j.jalz.2016.02.006.
[38] Z. Niu, M. Zhou, L. Wang, Ordinal regression with multiple output CNN for age estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4920–4928, doi:10.1109/CVPR.2016.532.

[39] T.Y. Yang, Y.H. Huang, Y.Y. Lin, SSR-net: a compact soft stagewise regression network for age estimation[c], in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, 2018, pp. 1078–1084, doi:10.24963/ijcai.2018/150.
[40] K. Li, J. Xing, W. Hu, D2c: deep cumulatively and comparatively learning for human age estimation, Pattern Recognit. 66 (2017) 95, doi:10.1016/j.patcog.2017.01.007.
[41] C.R. Jack Jr, M.A. Bernstein, N.C. Fox, The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods, J. Magn. Reson. Imaging 27 (4) (2008) 685–691, doi:10.1002/jmri.21049.
[42] C. Lian, M. Liu, Y. Pan, Attention-guided hybrid network for dementia diagnosis with structural MR images, IEEE Trans. Cybern. (2020), doi:10.1109/TCYB.2020.3005859.
[43] C. Lian, M. Liu, J. Zhang, Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE Trans. Pattern Anal. Mach.Intell. 42 (4) (2018) 880–893, doi:10.1109/TPAMI.2018.2889096.
[44] Z. Xue, D. Shen, C. Davatzikos, CLASSIC: consistent longitudinal alignment and segmentation for serial image computing, NeuroImage 30 (2) (2006) 388–399, doi:10.1016/j.neuroimage.2005.09.054.
[45] J.C. Baron, G. Chetelat, B. Desgranges, In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease, Neuroimage 14 (2) (2001) 298–309, doi:10.1006/nimg.2001.0848.
[46] J. Zhang, J. Liang, H. Zhao, Local energy pattern for texture classification using self-adaptive quantization thresholds, IEEE Trans. Image Process. 22 (1) (2012) 31–42, doi:10.1109/TIP.2012.2214045.
[47] J. Zhang, Y. Gao, Y. . Gao, Detecting anatomical landmarks for fast Alzheimer's disease diagnosis, IEEE Trans. Med. Imaging 35 (12) (2016) 2524–2533, doi:10.1109/TMI.2016.2582386.
[48] M. Liu, J. Zhang, E. Adeli, Landmark-based deep multi-instance learning for brain disease diagnosis, Med. Image Anal. 43 (2018) 157, doi:10.1016/j.media.2017.10.005.
[49] D. Zhang, D. Shen, Alzheimer's disease neuroimaging initiative. multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease, NeuroImage 59 (2) (2012) 895–907, doi:10.1016/j.neuroimage.2012.03.006.