CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/150037/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Zheng, Guowei, Zhang, Yu, Zhao, Ziyang, Wang, Yin, Liu, Xia, Shang, Yingying, Cong, Zhaoyang, Dimitriadis, Stavros I. , Yao, Zhijun and Hu, Bin 2022. A transformer-based multi-features fusion model for prediction of conversion in mild cognitive impairment. Methods 205 , pp. 241-248. 10.1016/j.ymeth.2022.04.015

Publishers page: http://dx.doi.org/10.1016/j.ymeth.2022.04.015

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

A Transformer-based Multi-features Fusion Model for Prediction of Conversion in Mild Cognitive Impairment

Guowei Zheng[a], Yu Zhang[a], Ziyang Zhao[a], Yin Wang[a], Xia Liu[e], Yingying Shang[a], Zhaoyang Cong[a], Stavros Dimitriadis[f, g, h, i, j, k, l, m, n, *], Zhijun Yao[a, *] and Bin Hu[a, b, c, d, *]

[a] Gansu Provincial Key Laboratory of Wearable Computing, School of Information Science and Engineering, Lanzhou University, Lanzhou, China

[b] CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

[c] Joint Research Center for Cognitive Neurosensor Technology of Lanzhou University & Institute of Semiconductors, Chinese Academy of Sciences, Lanzhou, China

[d] Engineering Research Center of Open Source Software and Real-Time System (Lanzhou University), Ministry of Education, Lanzhou, China

[e] School of Computer Science, Qinghai Normal University, Xining, China.

[f] Integrative Neuroimaging Lab, 55133, Thessaloniki (Makedonia), Greece.

[g] Neuroinformatics Group, Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, -College of Biomedical and Life Sciences, Cardiff, Wales, United Kingdom.

[h] 1st Department of Neurology, G.H. "AHEPA " School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece.

[i] Greek Association of Alzheimer's Disease and Related Disorders, Thessaloniki, Makedonia, Greece.

[j] Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, College of Biomedical and Life Sciences, Cardiff University, Cardiff, Wales, United Kingdom.

[k] Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, College of Biomedical and Life Sciences, Cardiff University, Cardiff, Wales, United Kingdom.

[l] School of Psychology, College of Biomedical and Life Sciences, Cardiff University, Cardiff, Wales, United Kingdom.

[m] Neuroscience and Mental Health Research Institute, School of Medicine, College of Biomedical and Life Sciences, Cardiff University, Cardiff, Wales, United Kingdom.

[n] MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, College of Biomedical and Life Sciences, Cardiff University, Cardiff, Wales, United Kingdom.

[*] Correspondence: DimitriadisS@cardiff.ac.uk (Stavros Dimitriadis), yaozj@lzu.edu.cn (Zhijun Yao), bh@lzu.edu.cn (Bin Hu)

Email: zhenggw20@lzu.edu.cn (Guowei Zheng), yzhang20@lzu.edu.cn (Yu Zhang), zhaozy2021@lzu.edu.cn (Ziyang Zhao), wangyin20@lzu.edu.cn (Yin Wang), liux2016@lzu.edu.cn (Xia Liu), shangyy20@lzu.edu.cn (Yingying Shang), congchy19@lzu.edu.cn (Zhaoyang Cong), DimitriadisS@cardiff.ac.uk (Stavros Dimitriadis), yaozj@lzu.edu.cn (Zhijun Yao), bh@lzu.edu.cn (Bin Hu)

Abstract

Mild cognitive impairment (MCI) is usually considered the early stage of Alzheimer's disease (AD). Therefore, the accurate identification of MCI individuals with high risk in converting to AD is essential for the potential prevention and treatment of AD. Recently, the great success of deep learning has sparked interest in applying deep learning to neuroimaging field. However, deep learning techniques are prone to overfitting since available neuroimaging datasets are not sufficiently large. Therefore, we proposed a deep learning model fusing cortical features to address the issue of fusion and classification blocks. To validate the effectiveness of the proposed model, we compared seven different models on the same dataset in the literature. The results show that our proposed model outperformed the competing models in the prediction of MCI conversion with an accuracy of 83.3% in the testing dataset. Subsequently, we used deep learning to characterize the contribution of brain regions and different cortical features to MCI progression. The results revealed that the caudal anterior cingulate and pars orbitalis contributed most to the classification task, and our model pays more attention to volume features and cortical thickness features.

Keywords: Mild cognitive impairment, Magnetic resonance imaging, Deep learning, Transformer

1 Introduction

Alzheimer's disease (AD) is a common degenerative disease in aging populations. Cognitive

impairment and progressive memory loss are the fundamental characteristics of AD [1]. More than 30 million people worldwide are suffering from AD cause of the extending life expectancy, and this number is estimated to be tripled by 2050 [2]. Despite the dramatic increase in the prevalence of AD, no treatment can completely cure it currently. Thus, early diagnosis is crucial to developing treatments for AD [3, 4]. Mild cognitive impairment (MCI) is generally considered a transitional stage between normal aging and AD [5]. Studies have shown that approximately 5% to 15% of persons with MCI will progress to AD each year [6, 7]. MCI can be divided into two subtypes, progressive mild cognitive impairment (pMCI) and stable mild cognitive impairment (sMCI). Subjects classified as pMCI were those with a higher risk of conversion to AD in a short period, while subjects in the sMCI group remained stable for a certain period and had a lower risk of progression to AD than the former [8]. Therefore, classifying the two different types of MCI can predict the conversion from MCI to AD as early as possible, which is beneficial for AD prevention and therapy.

Neuroimaging is widely used to understand the pathology of MCI and AD [9]. In previous studies on the mechanism of AD, structural magnetic resonance imaging (MRI) is one of the most extensively utilized imaging modalities in AD detection and prediction for its wide practicality, non-invasion, high resolution, and moderate cost [10]. Applying machine learning techniques to neuroimaging diagnosis is a developing field. In terms of MCI conversion prediction, numerous studies are using different methods, including network features constructed based on graph theory [11, 12],voxel-based morphometry (VBM) based on the segmentation of grey matter [13, 14], multiple methods of hippocampal segmentation [15], etc. However, research using traditional machine learning methods still suffers from inadequacies. The performance of traditional machine learning methods largely depends on data representation [16], and it is challenging to learn high-level information from poorly hand-picked features.

Recently, with the development of deep learning technology, many researchers have achieved outstanding achievements in neuroscience [17-19]. Deep learning network models also progressed in predicting AD conversion in advance from MCI [20-23]. Nevertheless, most of these studies used 3D subject‑level features as input to deep learning network models, which suffer from overfitting issues, since the sample size of available neuroimaging data sets is not significant compared with millions of features in each image [24, 25]. Freesurfer is a powerful tool to reliably extract cortical

features such as volume, surface area, cortical thickness, and curvature index [26-33] through an automated pipeline without any user interaction. The dimension of cortical features is significantly lower than the original neuroimage but contains rich ROI-level brain morphological information, which can effectively alleviate the overfitting problem. In 2017, the transformer was first proposed by Vaswani et al. [34] and successfully applied to natural language processing (NLP) tasks. Researchers have recently extended it to other tasks such as computer vision (CV) with great success [35]. Its strong global perception capability makes it possible to find differences in brain morphology of the cortex between pMCI and sMCI from fused cortical features for classification.

Based on the above considerations, in this work, we proposed a transformer-based multi-features fusion model to predict conversion in MCI by using MRI. Specifically, our architecture was designed to fuse the multiple cortical features and automatically learn high-level information from the fused features. To validate the proposed model, we perform the classification on the MRI datasets from the Alzheimer's Disease Neuroimaging Initiative (ADNI, http://adni.loni.usc.edu/) [36], and achieved better performance over other models. Furthermore, with occlusion analysis, we investigated the contribution of different brain regions and different cortical features to the classifying progression and stability of MCI.

The rest of the paper has been organized as follows. In Section 2, we mainly introduce the architecture of the proposed model and the details of its construction and validation and the implementation of occlusion analysis. Section 3 gives the analysis of the results followed by further discussion in Section 4. Finally, Section 5 summarizes the full text.

2 Materials and methods

2.1 Experimental Data

Data used in our study were obtained from the ADNI database. The ADNI is an ongoing and multicenter study that aims to develop imaging, clinical, genetic, and biochemical biomarkers for AD's early detection and tracking [37]. 249 MCI participants with baseline T1-weighted structural MRI were selected from ADNI in this work. All MCI subjects were divided into two groups: (1) stable mild cognitive impairment (sMCI) who did not convert to AD within three years. In addition, the

subjects who were diagnosed as MCI at least twice, but reverse to a standard control, at last, are also considered as sMCI [23]; (2) progressive mild cognitive impairment (pMCI) who were diagnosed as MCI at the first visit, but converted to AD at longitudinal visits within three years. The detailed demographic information is given in Table 1.

TABLE 1 The demographic information.

	sMCI	**pMCI**
Subjects' number	104	145
Age range	55-88	55-88
Males/Females	67/37	90/55

Abbreviations: pMCI = progressive mild cognitive impairment, sMCI = stable mild cognitive impairment.

2.2 Image Pre-processing

T1-weighted structural images were processed using the Freesurfer software (v6.0; http://surfer.nmr.mgh.harvard.edu/) [38]. The preprocessing steps are described below. Firstly, the correction for non-uniformity artifacts was performed on the images [39], followed by the coordinate transformation [40] and the brain tissue segmentation (including gray matter, white matter, cerebrospinal fluid, and other background categories). Subsequently, the surface of white/gray matter boundaries was reconstructed [40]. After completing the construction of boundary models, surface expansion and registration were performed [30, 38]. Finally, we extracted multiple cortical measurements including volume (VOL), cortical thickness (CT), curvature index (CD), folding index (FD), and surface area (SA) for 62 brain regions (31 regions in each hemisphere of the brain) using the Desikan-Killiany-Tourville (DKT) atlas [41].

2.3 The Transformer-based Multi-features Fusion Model

Here we proposed a multi-features fusion model to predict conversion in MCI, which is based on the transformer model [34]. Our model was designed to input a cortical feature matrix (extracted from the preprocessed image) and output the classification result. The model consists of a fusion block and a classification block. For an overview, refer to Fig.1.

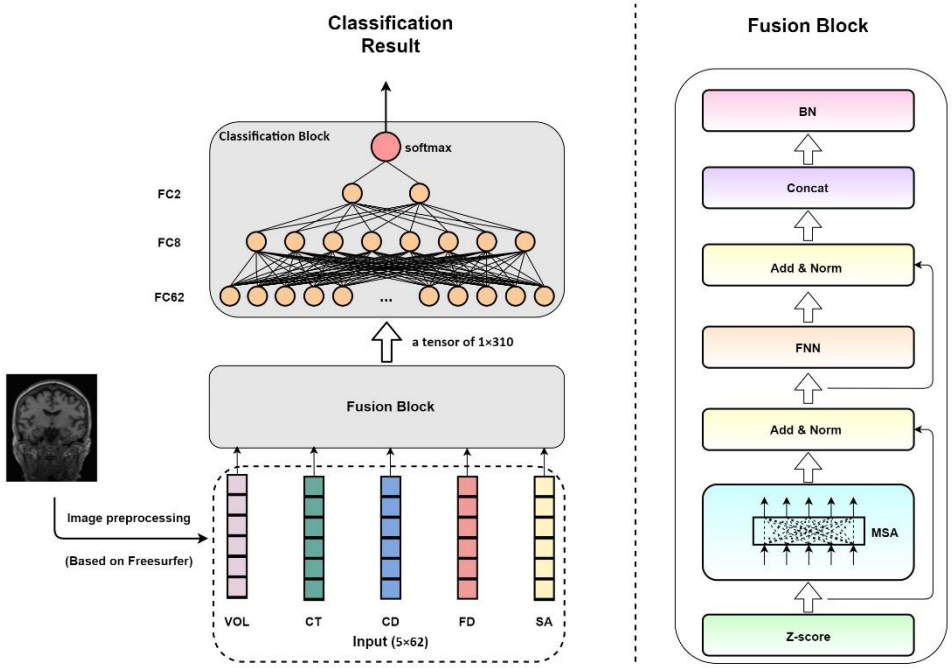

Fig. 1. Illustration of proposed deep learning model. Abbreviations: VOL = volume, CT = cortical thickness, CD = curvature index, FD = folding index, SA = surface area, MSA = multi-head self-attention, FFN = feed-forward network, Concat = concatenate, BN = Batch Normalization, FC62 = 62-units fully connected layer, FC8 = 8-units fully connected layer, FC2 = 2-units fully connected layer.

The fusion block consists of five different sub-layers. Firstly, the Z-score method was applied to the input features to remove the effect of different feature sizes (Eq. (1)). The second is a multi-head self-attention (MSA) [34] and the third is a feed-forward network (FFN), the residual connections [42] were employed after the MSA and FFN, followed by layer normalization (LN) [43] (Eqs. (2) and (3)). To improve the model's efficiency, we set the number of heads in the MSA to 2, which could reduce the number of model parameters. The dimension of outputs for MSA and FFN is 62, which matches the model's input and enables these residual connections. The fourth is a concatenate (Concat) layer (Eq. (4)), which reshapes the input data (cortical feature matrix, 5×62) to a tensor of 1×310 for later classification. Finally, the Batch Normalization (BN) layers were applied to accelerate convergence. The output of fusion block f is calculated using Eq. (5) (x is the input of the model). Then, took f as the input to the classification block.

$$l_1 = Z - score(x) \tag{1}$$

$$l_2 = LN\big(l_1 + MSA(l_1)\big) \tag{2}$$

$$l_3 = LN\big(l_2 + FNN(l_2)\big) \tag{3}$$

$$l_4 = Concat(l_3) \quad (4)$$

$$f = BN(l_4) \quad (5)$$

The classification block consists of three fully connected (FC) layers with 62, 8, and 2 units respectively. Later, the softmax activation function was used to predict the results.

2.4 Implementation

We implemented our model with Pytorch 1.8.0. Model training and testing were performed on the Ubuntu 18.04 operating system. During training, we used the Binary Cross-Entropy (BCE) loss function and set the number of epochs to 15, with a mini-batch size of 32. The optimizer was Adam [44] with a learning rate of 1e-4 and weight decay of 1e-8.

2.5 Validation Framework

To validate the efficacy of the proposed model, we split our 249 subjects randomly into three groups, including the training dataset (n=200), the validation dataset (n=25), and the testing dataset (n=24). The training dataset was used for training models, while the validation dataset was used for parameter tuning and the testing dataset for evaluating model performance.

2.6 Model Comparison

The proposed model was compared with four traditional machine learning methods: support vector machine [45], decision tree [46], random forest [47], and logistic regression [48]. Compared to traditional machine learning methods with feature engineering, deep learning models aim to extract features automatically. Therefore, deep learning methods including Recurrent Neural Network (RNN) [49], Long Short-Term Memory (LSTM) [50], and Gated Recurrent Unit (GRU) [51] were also employed in this study for comparison with the proposed model.

To verify the performance of the above models, we randomly divide the entire dataset into a training dataset, a validation dataset, and a testing dataset in a ratio of 8:1:1 (the division of the dataset is the same as our proposed model). All the traditional machine learning models were implemented using sklearn (https://scikit-learn.org/stable/) library in python3 (used the default settings) and were trained on the training dataset then tested on the testing dataset. In addition, all deep learning models were implemented with Pytorch 1.8.0, and for these deep learning

models, we defined the optimal hyperparameters of the classifiers by using the training and validating datasets. Subsequently, when the model achieved the best performance (the optimized hyperparameters were listed in *Supplementary Materials* Table S1) in the validation dataset, the model was validated using the testing dataset. We also performed 10-fold cross-validation on the entire dataset to compare the performance between our proposed and other models to ensure generalizability.

Furthermore, to validate the performance of all the models, we employed four measurements including classification accuracy (ACC), sensitivity (SEN), specificity (SPE) (shown in Eq. (6) to Eq. (8)), and the area under the receiver operating characteristic (ROC) curve (AUC). For these measurements, higher values demonstrate better performance.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$SEN = \frac{TP}{TP + FN} \tag{7}$$

$$SPE = \frac{TN}{TN + FP} \tag{8}$$

Where TP, TN, FP, and FN are abbreviations for True Positive, True Negative, False Positive, and False Negative, respectively.

2.7 Implementation of Occlusion Analysis

Occlusion analysis was employed to investigate the contribution of each brain region and each cortical feature to the performance of the proposed model. First, we set the value of five cortical features of each brain region (both left and right) to 0 from the cortical feature matrix of the test stage and retested the trained proposed model. The input corresponding to brain region m is x_m :

$$BrainRegionOcc_n = \begin{cases} 0 & \text{if m=n} \\ x_m & \text{otherwise} \end{cases} \tag{9}$$

Where $BrainRegionOcc_n$ represents the occlusion of brain region n. If the m is equal to n, the value of x_m is set to 0. Then, we masked different features to explore their impact on the model. See Eq. (10) for details (the $FeatureOcc_i$ means the occlusion of i-th feature and x_j means the input corresponding j-th feature).

$$FeatureOcc_i = \begin{cases} 0 & \text{if j=i} \\ x_j & \text{otherwise} \end{cases} \tag{10}$$

3 Results

3.1 Classification Performance

We test the trained model on the testing dataset to verify the proposed model. The proposed model achieved 83.3% accuracy and an AUC of 0.888 (Fig.2), with a sensitivity of 0.727 and a specificity of 0.923(Fig.3).

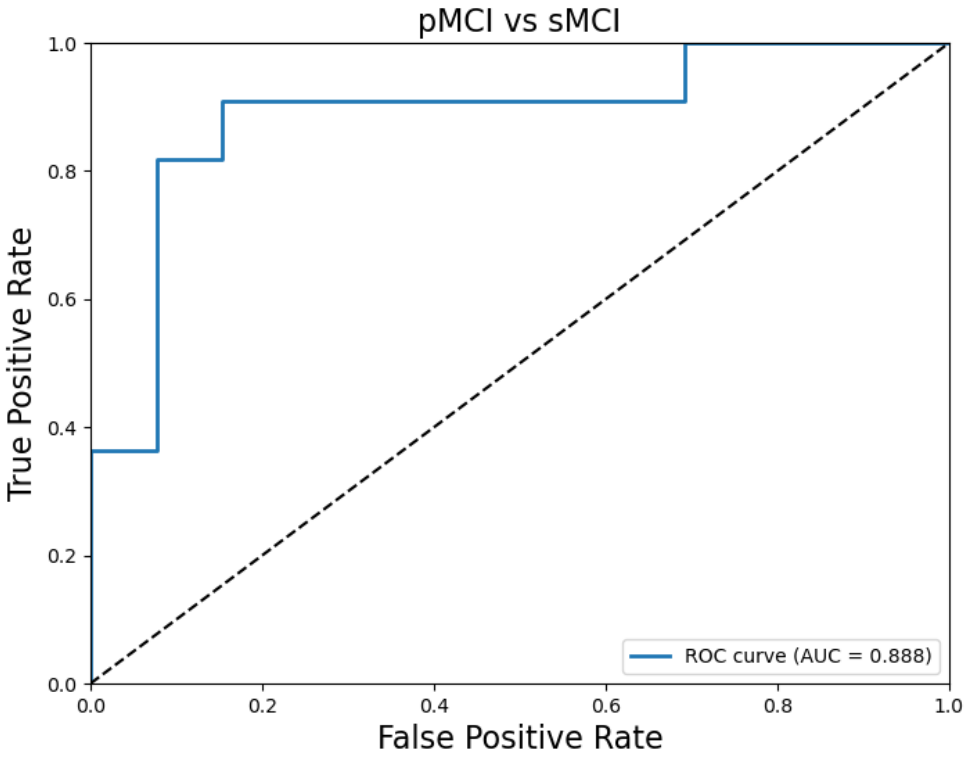

Fig. 2. ROC curve for classifying pMCI versus sMCI. Abbreviations: pMCI = progressive mild cognitive impairment, sMCI = stable mild cognitive impairment, ROC = receiver operating characteristic, AUC = area under the curve.

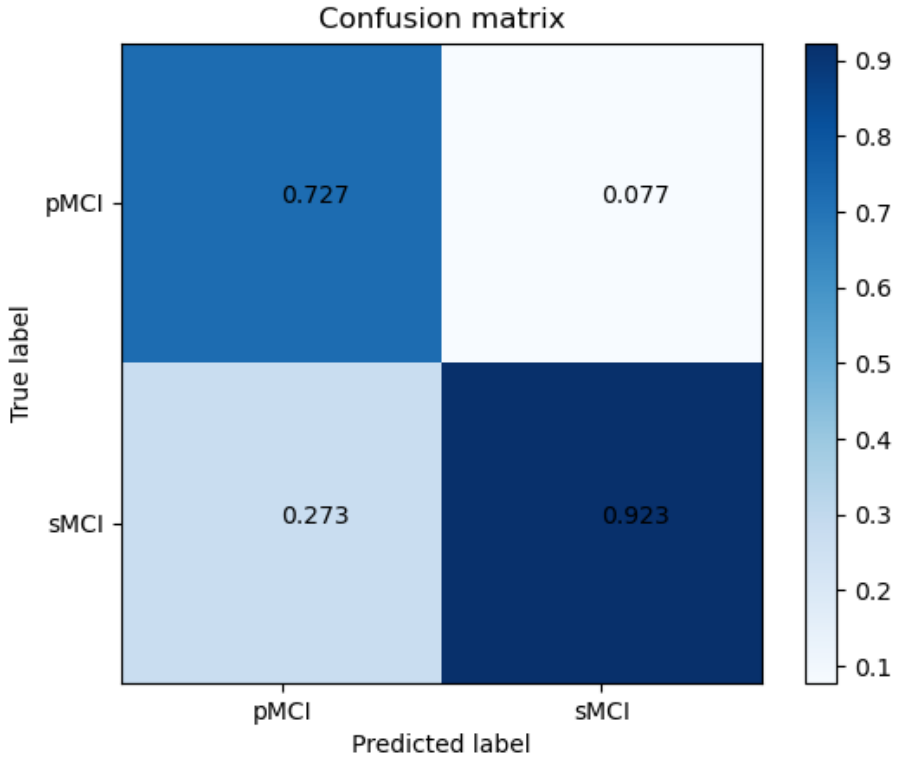

Fig. 3. Confusion matrix, evaluating the SEN and SPE obtained in pMCI versus sMCI. The matrix values were rescaled to the scope of [0,1]. Abbreviations: pMCI = progressive mild cognitive impairment, sMCI = stable mild cognitive impairment.

In addition, we compared our proposed method with three different deep learning methods (RNN, LSTM, GRU) and four different machine learning methods (random forest, decision tree, logistic regression, and support vector machine). As shown in Table 2, our proposed model showed better performance than the other models. The results show that GRU (AUC = 0.853) performs better than LSTM (AUC = 0.839) and RNN (AUC = 0.790) among the three deep learning models. Furthermore, the random forest has the best performance (AUC=0.678) among four machine learning models. See Table 2 for more detailed information. The results of 10-fold cross-validation also showed that our model can predict MCI conversion more accurately (see *Supplementary Materials* Table S2).

TABLE 2 The performance of different models.

Model	ACC	SEN	SPE	AUC
Proposed model	**83.3%**	0.727	**0.923**	**0.888**
RNN	66.7%	**0. 818**	0.538	0.790
LSTM	70.8%	0. 727	0.692	0.839
GRU	70.8%	0. 727	0.692	0.853

Random Forest	66.7%	**0.818**	0.538	0.678
Decision Tree	54.2%	0.545	0.538	0.542
Logistic Regression	66.7%	0.727	0.615	0.671
Support vector machine	54.2%	**0.818**	0.308	0.563

The best results for each column are shown in boldface. Abbreviations: RNN = Recurrent Neural Network, LSTM = Long Short-Term Memory, GRU = Gated Recurrent Unit, ACC = accuracy, SEN = sensitivity, SPE = specificity, AUC = area under the receiver operating characteristic (ROC) curve.

3.2 Occlusion Analysis

After extracting five features based on the DKT template for each subject, the relevant contribution of different brain regions and different features to the classification performance was evaluated using computer vision's commonly used occlusion analysis method [52].

As shown from Fig.4, the masking of most brain regions causes a decrease in model accuracy, and the masking of a small number of brain regions does not affect model accuracy (Fig.5). Notably, masking of the caudal anterior cingulate and the pars orbitalis (Fig.6) resulted in a dramatic decrease in model performance. Then, we performed occlusion analysis for different features. It can be seen from Fig.7 that the occlusion of different features all caused a significant decrease in model accuracy.

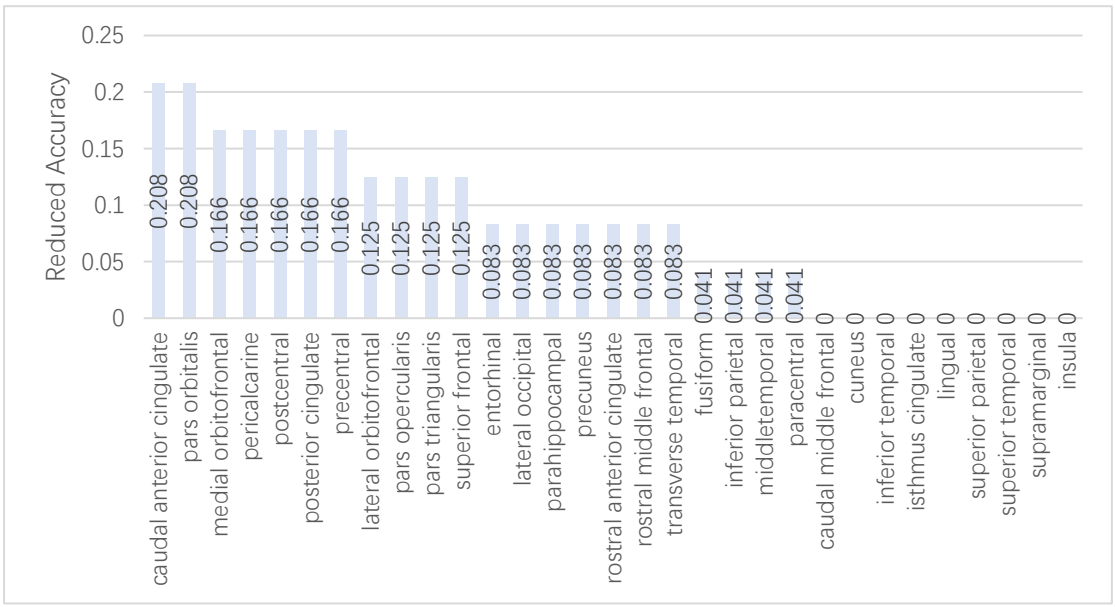

Fig. 4. The reduced accuracies with each brain region occluded compared to the original intact model.

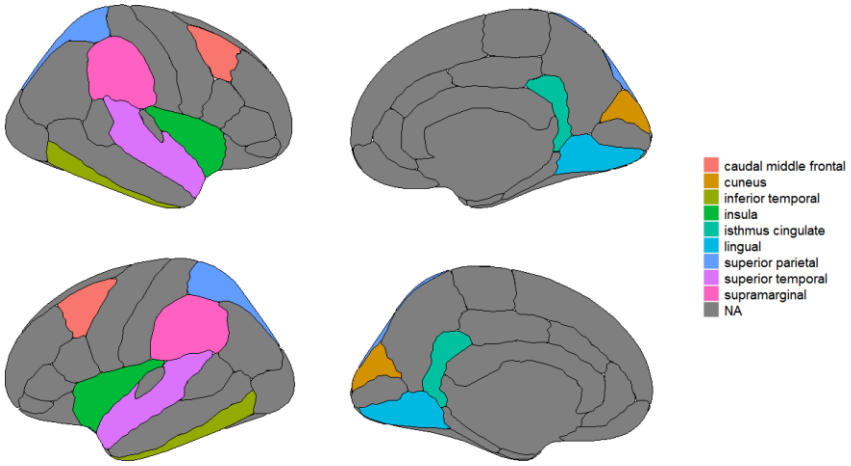

Fig. 5. The location of brain regions that do not affect classification accuracy.

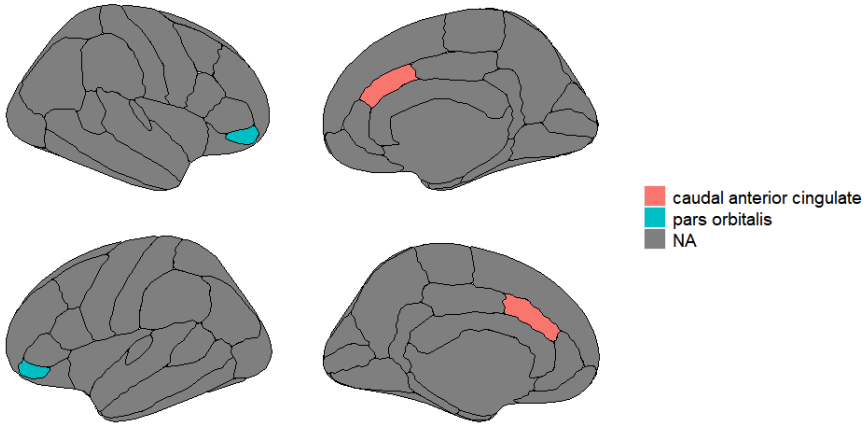

Fig. 6. The location of the brain regions that contributed most to the classification task.

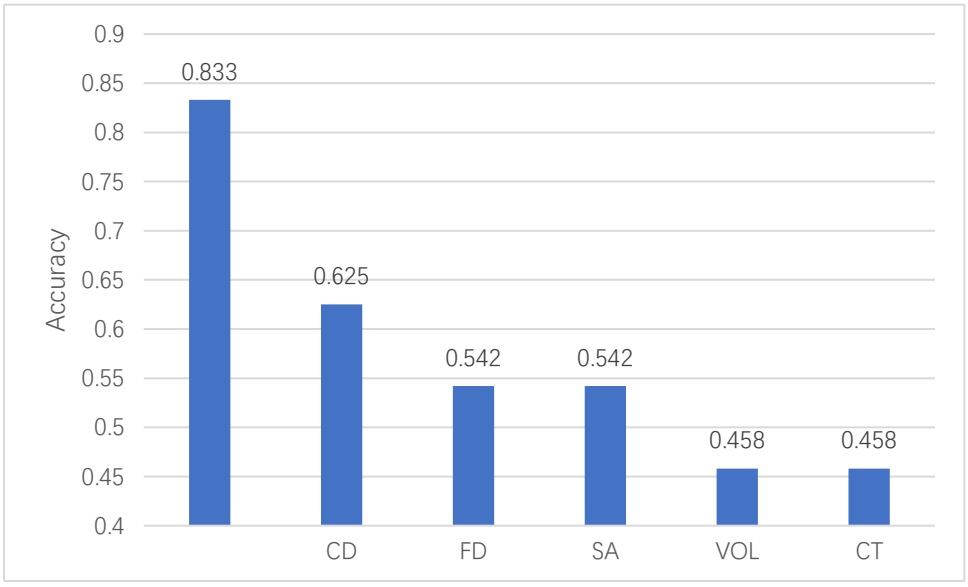

Fig. 7. Results for each feature occlusion (the first column is the model's accuracy with all features input). Abbreviations: CD = curvature index, FD = folding index, SA = surface area, VOL = volume, CT = cortical thickness.

4 Discussion

Patients with MCI show a strong variable trajectory of symptoms, with some individuals finally diagnosed with AD, while others show a more stable cognitive ability pattern for a certain period. Identifying these two different types of MCI is crucial and essential to preventing and treating AD. Therefore, many researchers are committed to developing computer-aided systems to diagnose AD early. To solve the overfitting problem of most previous methods, we proposed a transformer-based model that predicts conversion in MCI using multiple ROI-level cortical features and achieved an accuracy of 83.3% on the testing dataset.

The model comparison results demonstrated that the proposed model performs better than other traditional machine learning models (random forest, decision tree, logistic regression, and support vector machine) and deep learning models (RNN, LSTM, and GRU). The traditional machine learning methods rely on the manual selection of features. For features that have not been carefully selected, the traditional machine learning methods are challenging to thoroughly learn sufficient information in cortical features. In addition, compared with other deep learning methods, the proposed model includes a fusion block with MSA, which takes into account the features themselves and fully considers the correlation between different cortical features to achieve better performance. Furthermore, the classification performance of the proposed model also outperformed previously developed deep learning models for classifying pMCI versus sMCI based on MRI data [22, 53-55], which ranged from 73.95% to 78.79%.

The occlusion analysis results both extend and support prior reports by describing the contribution of different brain regions and different cortical features to the progression of MCI. On the one hand, the results revealed significant differences between the brain regions differentiating pMCI from sMCI. Notably, the results have shown that the caudal anterior cingulate and pars orbitalis (Fig.6) were most important for the classification task than any other brain region. Previous studies have shown that neuronal loss in the caudal anterior cingulate begins in the early stage of AD [56], and this timing may need to be advanced. This brain region contributed the most

to the model, possibly indicating that some neurons have been lost in pMCI. Given that the caudal anterior cingulate is important to cognitive control of behavior [57], it suggests that pMCI may show more severe cognitive impairment than sMCI. In addition, a previous study found that the pars orbitalis, as well as some other brain regions, contributed to good classification performance in this task [58], but the central role of the pars orbitalis should be highlighted. On the other hand, the occlusion of different features all caused a significant decrease in model accuracy, this finding demonstrates the existence of important complementary information in all five features. Furthermore, the occlusion analysis on different features showed that VOL and CT had the strongest impact on model performance, this may be related to the different volumes and atrophy rates between sMCI and pMCI [59] and the significantly thinner cortical thicknesses in many brain regions in pMCI [60]. The results also indicated that VOL and CT were more distinct in sMCI and pMCI brains than CD, FD, and SA and were more reliable biomarkers in the progression of MCI.

Our study has some limitations. Firstly, our work only used MRI images, while researchers have continuously disclosed the strength of multimodal features in computer-aided diagnosis models [61-63]. Therefore, the model performance is expected to be improved by incorporating data from multiple modalities, such as functional MRI. In addition, the cross-sectional nature is another limitation of our study. Therefore, longitudinal data should be employed in our future research.

5 Conclusion

This study proposed a transformer-based multi-features fusion model to predict the MCI-to-AD conversion only using MRI data. Results show that our model can fuse the cortical features extracted by Freesurfer. Compared with other models in the literature, our proposed model achieves higher accuracy and AUC. In addition, our study reveals the contribution of brain regions in differentiating between pMCI and sMCI, highlighting the central role of the caudal anterior cingulate and pars orbitalis. Finally, the occlusion analysis results demonstrate that VOL and CT may be more reliable biomarkers in MCI progression.

Authors' Contributions:

In this paper, Zhijun Yao and Bin Hu conceived the project. Guowei Zheng completed all experiments in this work. Guowei Zheng wrote the manuscript, Stavros Dimitriadis, Yu Zhang, Ziyang Zhao, Yin Wang, Xia Liu, Yingying Shang, and Zhaoyang Cong revised the manuscript.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFA0706200), in part by the National Natural Science Foundation of China (Grant No.61632014, No.61627808), in part by the National Basic Research Program of China (973 Program, Grant No. 2014CB744600), in part by the Natural Science Foundation of Gansu Province of China (Grant No.20JR5RA292), in part by the Fundamental Research Funds for the Central Universities (Grant No. lzujbky-2018-125), and in part by the Department of education of Gansu Province: "Innovation Star" project for excellent postgraduates (2021CXZX-121).

References

[1] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper-González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, O. Colliot, Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation, Medical image analysis 63 (2020) 101694.

[2] D.E. Barnes, K. Yaffe, The projected effect of risk factor reduction on Alzheimer's disease prevalence, The Lancet Neurology 10(9) (2011) 819-828.

[3] C. Samaey, A. Schreurs, S. Stroobants, D. Balschun, Early cognitive and behavioral deficits in mouse models for tauopathy and Alzheimer’s disease, Frontiers in aging neuroscience 11 (2019) 335.

[4] W. Shao, S. Xiang, Z. Zhang, K. Huang, J. Zhang, Hyper-graph based sparse canonical correlation analysis for the diagnosis of Alzheimer’s disease from multi-dimensional genomic data, Methods 189 (2021) 86-94.

[5] R.C. Petersen, Mild cognitive impairment as a diagnostic entity, Journal of internal medicine 256(3) (2004) 183-194.

[6] A.J. Mitchell, M. Shiri‐Feshki, Rate of progression of mild cognitive impairment to dementia–meta‐analysis of 41 robust inception cohort studies, Acta psychiatrica scandinavica 119(4) (2009) 252-265.

[7] R. Roberts, D.S. Knopman, Classification and epidemiology of MCI, Clinics in geriatric medicine 29(4) (2013) 753-772.

[8] A.M. Anter, Y. Wei, J. Su, Y. Yuan, B. Lei, G. Duan, W. Mai, X. Nong, B. Yu, C. Li, A robust swarm intelligence-based feature selection model for neuro-fuzzy recognition of mild cognitive impairment from resting-state fMRI, Information Sciences 503 (2019) 670-687.

[9] Z. Li, H.-I. Suk, D. Shen, L. Li, Sparse multi-response tensor regression for Alzheimer's disease study

with multivariate clinical assessments, IEEE transactions on medical imaging 35(8) (2016) 1927-1936.
[10] T. Tong, Q. Gao, R. Guerrero, C. Ledig, L. Chen, D. Rueckert, A.s.D.N. Initiative, A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease, IEEE Transactions on Biomedical Engineering 64(1) (2016) 155-165.
[11] S.H. Hojjati, A. Ebrahimzadeh, A. Khazaee, A. Babajani-Feremi, A.s.D.N. Initiative, Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM, Journal of neuroscience methods 282 (2017) 69-80.
[12] R. Wei, C. Li, N. Fogelson, L. Li, Prediction of conversion from mild cognitive impairment to Alzheimer's Disease using MRI and structural network features, Frontiers in aging neuroscience 8 (2016) 76.
[13] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, A.s.D.N. Initiative, Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects, Neuroimage 104 (2015) 398-412.
[14] G. Chételat, B. Landeau, F. Eustache, F. Mézenge, F. Viader, V. de La Sayette, B. Desgranges, J.-C. Baron, Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study, Neuroimage 27(4) (2005) 934-946.
[15] C. Platero, M.C. Tobar, A fast approach for hippocampal segmentation from T1-MRI for predicting progression in Alzheimer's disease from elderly controls, Journal of neuroscience methods 270 (2016) 61-75.
[16] H. Li, L. Chen, Z. Huang, X. Luo, H. Li, J. Ren, Y. Xie, DeepOMe: a web server for the prediction of 2′ -O-Me sites based on the hybrid CNN and BLSTM architecture, Frontiers in cell and developmental biology 9 (2021) 1244.
[17] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, J. Li, A robust deep model for improved classification of AD/MCI patients, IEEE journal of biomedical and health informatics 19(5) (2015) 1610-1616.
[18] N. Zeng, Z. Wang, H. Zhang, W. Liu, F.E. Alsaadi, Deep belief networks for quantitative analysis of a gold immunochromatographic strip, Cognitive Computation 8(4) (2016) 684-692.
[19] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, A.M. Dobaie, Facial expression recognition via learning deep sparse autoencoders, Neurocomputing 273 (2018) 643-649.
[20] 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 Transactions on Biomedical Engineering 66(5) (2018) 1195-1206.
[21] E. Ocasio, T.Q. Duong, Deep learning prediction of mild cognitive impairment conversion to Alzheimer’s disease at 3 years after diagnosis using longitudinal and whole-brain 3D MRI, PeerJ Computer Science 7 (2021) e560.
[22] J. Zhang, B. Zheng, A. Gao, X. Feng, D. Liang, X. Long, A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification, Magnetic Resonance Imaging 78 (2021) 119-126.
[23] W. Lin, T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, X. Qu, Convolutional neural networks-based MRI image analysis for the Alzheimer’s disease prediction from mild cognitive impairment, Frontiers in neuroscience 12 (2018) 777.
[24] N. Goenka, S. Tiwari, Deep learning for Alzheimer prediction using brain biomarkers, Artificial Intelligence Review (2021) 1-45.
[25] X. Zhao, X.-M. Zhao, Deep learning of brain magnetic resonance images: A brief review, Methods 192 (2021) 131-140.
[26] A.M. Dale, B. Fischl, M.I. Sereno, Cortical surface-based analysis: I. Segmentation and surface

reconstruction, Neuroimage 9(2) (1999) 179-194.
[27] B. Fischl, A.M. Dale, Measuring the thickness of the human cerebral cortex from magnetic resonance images, Proceedings of the National Academy of Sciences 97(20) (2000) 11050-11055.
[28] B. Fischl, A. Liu, A.M. Dale, Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex, IEEE transactions on medical imaging 20(1) (2001) 70-80.
[29] B. Fischl, D.H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain, Neuron 33(3) (2002) 341-355.
[30] B. Fischl, M.I. Sereno, A.M. Dale, Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system, Neuroimage 9(2) (1999) 195-207.
[31] B. Fischl, M.I. Sereno, R.B. Tootell, A.M. Dale, High‐resolution intersubject averaging and a coordinate system for the cortical surface, Human brain mapping 8(4) (1999) 272-284.
[32] B. Fischl, D.H. Salat, A.J. Van Der Kouwe, N. Makris, F. Ségonne, B.T. Quinn, A.M. Dale, Sequence-independent segmentation of magnetic resonance images, Neuroimage 23 (2004) S69-S84.
[33] B. Fischl, A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D.H. Salat, E. Busa, L.J. Seidman, J. Goldstein, D. Kennedy, Automatically parcellating the human cerebral cortex, Cerebral cortex 14(1) (2004) 11-22.
[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems, 2017, pp. 5998-6008.
[35] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, An image is worth 16x16 words: Transformers for image recognition at scale, arXiv preprint arXiv:2010.11929 (2020).
[36] B.T. Wyman, D.J. Harvey, K. Crawford, M.A. Bernstein, O. Carmichael, P.E. Cole, P.K. Crane, C. DeCarli, N.C. Fox, J.L. Gunter, Standardization of analysis sets for reporting results from ADNI MRI data, Alzheimer's & Dementia 9(3) (2013) 332-337.
[37] V.J. Lowe, P.J. Peller, S.D. Weigand, C.M. Quintero, N. Tosakulwong, P. Vemuri, M.L. Senjem, L. Jordan, C.R. Jack, D. Knopman, Application of the National Institute on Aging–Alzheimer's Association AD criteria to ADNI, Neurology 80(23) (2013) 2130-2137.
[38] B. Fischl, FreeSurfer, Neuroimage 62(2) (2012) 774-781.
[39] J.G. Sled, A.P. Zijdenbos, A.C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, IEEE transactions on medical imaging 17(1) (1998) 87-97.
[40] W. Zheng, Z. Yao, Y. Xie, J. Fan, B. Hu, Identification of Alzheimer's disease and mild cognitive impairment using networks constructed based on multiple morphological brain features, Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 3(10) (2018) 887-897.
[41] A. Klein, J. Tourville, 101 labeled brain images and a consistent human cortical labeling protocol, Frontiers in neuroscience 6 (2012) 171.
[42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
[43] J.L. Ba, J.R. Kiros, G.E. Hinton, Layer normalization, arXiv preprint arXiv:1607.06450 (2016).
[44] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
[45] V. Vapnik, The nature of statistical learning theory, Springer science & business media1999.
[46] Y.-Y. Song, L. Ying, Decision tree methods: applications for classification and prediction, Shanghai

archives of psychiatry 27(2) (2015) 130.
[47] A. Liaw, M. Wiener, Classification and regression by randomForest, R news 2(3) (2002) 18-22.
[48] D.R. Cox, The regression analysis of binary sequences, Journal of the Royal Statistical Society: Series B (Methodological) 20(2) (1958) 215-232.
[49] W. Zaremba, I. Sutskever, O. Vinyals, Recurrent neural network regularization, arXiv preprint arXiv:1409.2329 (2014).
[50] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9(8) (1997) 1735-1780.
[51] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, arXiv preprint arXiv:1412.3555 (2014).
[52] J. Sadr, I. Jarudi, P. Sinha, The role of eyebrows in face recognition, Perception 32(3) (2003) 285-293.
[53] H.-I. Suk, S.-W. Lee, D. Shen, A.s.D.N. Initiative, Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis, NeuroImage 101 (2014) 569-582.
[54] K. Oh, Y.-C. Chung, K.W. Kim, W.-S. Kim, I.-S. Oh, Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning, Scientific Reports 9(1) (2019) 1-16.
[55] K. Kwak, M. Niethammer, K.S. Giovanello, M. Styner, E. Dayan, A.s.D.N. Initiative, Differential Role for Hippocampal Subfields in Alzheimer's Disease Progression Revealed with Deep Learning, Cerebral Cortex (2021).
[56] R.J. Killiany, T. Gomez-Isla, M. Moss, R. Kikinis, T. Sandor, F. Jolesz, R. Tanzi, K. Jones, B.T. Hyman, M.S. Albert, Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease, Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society 47(4) (2000) 430-439.
[57] J.R. Gray, T.S. Braver, Personality predicts working-memory—related activation in the caudal anterior cingulate cortex, Cognitive, Affective, & Behavioral Neuroscience 2(1) (2002) 64-75.
[58] C.Y. Wee, P.T. Yap, D. Shen, A.s.D.N. Initiative, Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns, Human brain mapping 34(12) (2013) 3411-3425.
[59] H. Tabatabaei-Jafari, M.E. Shaw, E. Walsh, N. Cherbuin, A.s.D.N. Initiative, Regional brain atrophy predicts time to conversion to Alzheimer's disease, dependent on baseline volume, Neurobiology of aging 83 (2019) 86-94.
[60] V. Julkunen, E. Niskanen, S. Muehlboeck, M. Pihlajamäki, M. Könönen, M. Hallikainen, M. Kivipelto, S. Tervo, R. Vanninen, A. Evans, Cortical thickness analysis to detect progressive mild cognitive impairment: a reference to Alzheimer's disease, Dementia and geriatric cognitive disorders 28(5) (2009) 389-397.
[61] T. Zhang, M. Shi, Multi-modal neuroimaging feature fusion for diagnosis of Alzheimer's disease, Journal of Neuroscience Methods 341 (2020) 108795.
[62] L. Xu, X. Wu, R. Li, K. Chen, Z. Long, J. Zhang, X. Guo, L. Yao, A.s.D.N. Initiative, Prediction of progressive mild cognitive impairment by multi-modal neuroimaging biomarkers, Journal of Alzheimer's Disease 51(4) (2016) 1045-1056.
[63] F. Liu, C.-Y. Wee, H. Chen, D. Shen, Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification, NeuroImage 84 (2014) 466-475.

Supplementary Materials

TABLE S1 The hyperparameters of GRU, LSTM, and GRU.

Hyperparameter	RNN	LSTM	GRU
Epoch	100	50	50
Batch size	64	32	64
Optimizer	Adam	Adam	Adam
Learning rate	1e-3	1e-3	1e-3
Weight decay	1e-8	1e-8	1e-8

Abbreviations: RNN = Recurrent Neural Network, LSTM = Long Short-Term Memory, GRU = Gated Recurrent Unit.

TABLE S2 Results of the 10-fold cross-validation. All metrics are reported as mean ± SD across folds.

Model	ACC	SEN	SPE	AUC
Proposed model	**0.719 ± 0.084**	0.797 ± 0.109	**0.545 ± 0.214**	**0.668 ± 0.092**
RNN	0.603 ± 0.108	0.866 ± 0.095	0.272 ± 0.182	0.600 ± 0.105
LSTM	0.615 ± 0.105	**0.928 ± 0.101**	0.203 ± 0.139	0.624 ± 0.072
GRU	0.611 ± 0.106	0.898 ± 0.110	0.264 ± 0.207	0.603 ± 0.100
Random Forest	0.562 ± 0.063	0.638 ± 0.150	0.478 ± 0.128	0.621 ± 0.108
Decision Tree	0.618 ± 0.119	0.693 ± 0.143	0.533 ± 0.157	0.613 ± 0.122
Logistic Regression	0.619 ± 0.062	0.682 ± 0.126	0.529 ± 0.164	0.619 ± 0.071
Support vector machine	0.518 ± 0.117	0.583 ± 0.159	0.449 ± 0.227	0.516 ± 0.109

The best results for each column are shown in boldface. Abbreviations: SD = standard deviation, RNN = Recurrent Neural Network, LSTM = Long Short-Term Memory, GRU = Gated Recurrent Unit, ACC = accuracy, SEN = sensitivity, SPE = specificity, AUC = area under the receiver operating characteristic (ROC) curve.