Research Article

Classification and Interpretability of Mild Cognitive Impairment Based on Resting-State Functional Magnetic Resonance and Ensemble Learning

Mengjie Hu,1,2 Yang Yu,1 Fangping He,1 Yujie Su,1 Kan Zhang,1 Xiaoyan Liu,1 Ping Liu,1 Ying Liu,2 Guoping Peng,1,2 and Benyan Luo1

1Department of Neurology, First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China
2Department of General Practice, First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China

Correspondence should be addressed to Guoping Peng; guopingpeng@zju.edu.cn and Benyan Luo; luobenyan@zju.edu.cn

Received 8 February 2022; Revised 12 June 2022; Accepted 6 July 2022; Published 19 August 2022

Academic Editor: Paolo Crippa

Copyright © 2022 Mengjie Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The combination and integration of multimodal imaging and clinical markers have introduced numerous classifiers to improve diagnostic accuracy in detecting and predicting AD; however, many studies cannot ensure the homogeneity of data sets and consistency of results. In our study, the XGBoost algorithm was used to classify mild cognitive impairment (MCI) and normal control (NC) populations through five rs-fMRI analysis datasets. Shapley Additive exPlanations (SHAP) is used to analyze the interpretability of the model. The highest accuracy for diagnosing MCI was 65.14% (using the mPerAF dataset). The characteristics of the left insula, right middle frontal gyrus, and right cuneus correlated positively with the output value using DC datasets. The characteristics of left cerebellum 6, right inferior frontal gyrus, opercular part, and vermis 6 correlated positively with the output value using fALFF datasets. The characteristics of the right middle temporal gyrus, left middle temporal gyrus, left temporal pole, and middle temporal gyrus correlated positively with the output value using PerAF datasets. The characteristics of the right middle temporal gyrus, left middle temporal gyrus, and left hippocampus correlated positively with the output value using Wavelet-ALFF datasets. We found that the XGBoost algorithm constructed from rs-fMRI data is effective for the diagnosis and classification of MCI. The accuracy rates obtained by different rs-fMRI data analysis methods are similar, but the important features are different and involve multiple brain regions, which suggests that MCI may have a negative impact on brain function.

1. Introduction

Mild cognitive impairment (MCI) is a heterogeneous syndrome that causes little or no impairment of daily living activities and thus does not meet the criteria for dementia [1, 2]. Among the aging population (60 years of age and above) in China, the prevalence of MCI is 14.71%; besides, females of older age or living in rural areas of western China have a higher prevalence of MCI [3]. The currently available diagnoses of MCI are based on subjective indicators, including observation, clinical history, and neuropsychological assessment; moreover, its reliable diagnosis is challenging [4]. Approximately 30% of MCI patients progress to AD [5]. Early diagnosis and intervention delay the transformation of MCI to AD and improve its prognosis [6].

Machine learning is extensively used in the clinical and early diagnosis of diseases. Dichotomous and tripartite diagnosis is the most basic application in Alzheimer’s Disease (AD), i.e., diagnosis of AD and normal control (NC), as well as that of AD, MCI, and NC. Notably, classification diagnosis based on these two types is still in use today [7].

Several types of data commonly used in machine learning include structure magnetic resonance imaging...
Computational Intelligence and Neuroscience

channel for task information [11].
nection in the resting state provides a communication
itionally, accumulating studies indicate that internal con-
analyze changes in brain networks in AD patients. Addi-
[9, 10]. Rs-fMRI is highly sensitive to AD and is used to
alyze changes in brain networks in AD patients. Addi-
itionally, accumulating studies indicate that internal con-
nection in the resting state provides a communication channel for task information [11].

Rs-fMRI is a noninvasive imaging method with a high
spatial and temporal resolution, continually adopted in
scientific research and clinical work. Rs-fMRI primarily
reflects neuronal activity by observing the blood oxygen
level–dependent (BOLD) signal changes. The spontaneous
activity of neurons may trigger low-frequency fluctuation
(LFF). Studies integrating neuron electrophysiology and rs-
fMRI reveal that many cognitive and behavioral processes
are related to LFF [12–14]. Biswal et al. [15] discovered a
highly synchronous spontaneous LFF between motor cor-
tices, and the LFF of a BOLD signal is closely related to
neuronal spontaneous activity and is used to reflect changes
in brain functional activities. Changes in the amount of LFF
in different brain regions may be related to the interruption
of automatic regulation of the cerebral microvascular system
[16]. Numerous studies on LFF, including low-frequency
fluctuation (ALFF) [17], fractional ALFF (fALFF) [13],
percent amplitude of fluctuation (PerAF) [18], Wavelet-
ALFF [19], have been documented.

At present, studies on machine learning- (ML-) based
diagnosis studies with rs-fMRI have reached maturity (Ta-
ble 1). Most of the studies focuses on the interpretability of
models and the improvement of feature extraction methods
and classification algorithm. The accuracy of some predictive
models has reached more than 90%. Nevertheless, because of
the small sample size, the credibility of these studies is at
stake.

NC: normal controls; eMCI: early MCI; IMCI: late MCI;
aMCI: amnestic MCI; MCI-C: MCI converter; MCI-NC:
MCI nonconverter; SCD: subjective cognitive decline; VD:
vascular dementia; MDX: “mixed VAD dementia”; CNN:
convolutional neural network; SVM: support vector ma-
chine; LDA: linear discriminant analysis; RF: random forest;
ANFIS: adaptive neurofuzzy inference system; ELM: ex-
treme learning machine; DAG: directed acyclic graph; AE:
autoencoder.

Herein, we established a database of rs-fMRI studies
involving MCI and NC based on the local population; this
increased the applicability of the findings. Besides, the
combination and integration of multimodal imaging and
clinical markers have elicited numerous classifiers that
improve diagnostic accuracy in detecting and predicting AD
or MCI. Although the accuracy obtained is significantly
attractive, numerous studies cannot guarantee data
homogeneity and consistent results [45]. We used the
XGBoost algorithm to classify MCI and NC populations,
and the results were explained.

2. Materials and Methods

2.1. Participants. Between January 2017 and December
2020, patients were recruited from the Memory Clinic of the
First Affiliated Hospital, Zhejiang University School of
Medicine. Eligible participants were aged 55 years or older,
with primary school education or above. Peterson’s criteria
were used to select the MCI patients [46]. Individuals were
excluded if they had evidence of other diseases potentially
causing dementia other than AD; a history of stroke and
focal signs of nervous system; other neurological diseases
that potentially cause brain dysfunction (including schizo-
phrenia, severe anxiety, depression, frontotemporal de-
mentia, Huntington’s disease, brain tumors, Parkinson’s
disease, metabolic encephalopathy, encephalitis, multiple
sclerosis, epilepsy, and brain trauma); other systemic dis-
ases that potentially cause cognitive impairment including
hypothyroidism, folic acid and vitamin B12 deficiency,
specific infections (e.g., syphilis and HIV), and alcohol and
drug abuse; severe liver, kidney and lung insufficiency; se-
vere anemia, gastrointestinal disease and arrhythmia, and
myocardial infarction within 6 months; contraindications
including metal implantation in vivo; aphasia, consciousness
disorders, and other diseases that potentially hinder the
completion of cognitive examination; did not sign informed
cconsent. This study was authorized and approved by the
Ethics Committee of First Affiliated Hospital, Zhejiang
University School of Medicine, and conducted based on the
principles of the Helsinki Declaration. After obtaining in-
formed consent, participants were subjected to initial tests,
including clinical evaluation, neuropsychological tests,
laboratory examination, and MRI scanning.

2.2. Data Acquisition. All rs-fMRI data were collected from
the Second Affiliated Hospital of Hangzhou Normal Uni-
versity from a Discovery MR750 3.0 T scanner of General
Electric Company. Rs-fMRI scans were acquired based on
the following parameters: 43 slices, TR = 2000 ms, IE = 30 ms,
FA = 90, FOV = 64 mm × 64 mm, matrix = 200 × 200, scanning time = 8 min.

2.3. Rs-fMRI Data Preprocessing. Data preprocessing was
performed using the RESTplus V1.2 tool [47] (http://www.
restfmri.net/Forum/RESTplusV1.2) in the SPM12 (Statistical
Parametric Mapping 12) (http://www.fil.ion.ucl.ac.uk/spm).
The rs-fMRI data preprocessing steps included the following:
(1) removing volumes, i.e., the first ten volumes of each
subject were removed to ensure a steady condition; (2) slice
timing, i.e., data scanning was performed in intervals, with
odd-numbered layers having priority; (3) realignment, i.e.,
subjects with a maximum translation of more than 3.0 mm
or maximum rotation of more than 3.0° were excluded; (4)
normalization, i.e., the rs-fMRI scans were registered to
correspond sMRI and split using the Diffeomorphic

Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) and new segment, which were spatially normalized to the Montreal Neurological Institute (MNI) space; (5) detrend, i.e., the offset generated during data acquisition may have an impact on the later calculation process. Detrend eliminates such an impact when the data is acquired; (6) Nuisance Covariates Regression, i.e., factors affecting the results were removed, including Friston24 rotation parameter, white matter, cerebrospinal fluid, and global mean signal [48, 49].

2.4. Rs-fMRI Features Extraction. Further, features extraction of rs-fMRI was performed using RESTplus V1.2; consequently, 116 features based on Anatomical Automatic Labeling (AAL) were extracted in each type of calculation method. The features included fractional amplitude of low-frequency fluctuations (fALFF). The preprocessed data results were registered into the MNI space; then, each voxel was resampled using a sampling template of 3 mm × 3 mm × 3 mm. In RESTplus, BOLD was transformed from a time domain to a frequency domain by the fast Fourier transform formula (FFT), and the power spectrum of the BOLD signal in the frequency domain was obtained. The power spectrum obtained was calculated via square root, and the result obtained by calculating the mean value of the effective frequency band divided by a mean value of the amplitude of the whole frequency band was fALFF. Subsequently, the spatial fALFF maps were divided by the mean value of the whole brain (mfALFF). This study calculated fALFF in three frequency bands, i.e., Norm-1 (0.01–0.08 Hz), Slow-4 (0.027–0.073 Hz), and Slow-5 (0.01–0.027 Hz) frequency bands. A Gaussian smoothing kernel of 4 mm full width-

<table>
<thead>
<tr>
<th>Year</th>
<th>ML method</th>
<th>Subjects</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2022 [21]</td>
<td>AdaBoost</td>
<td>eMCI: 34, IMCI: 32</td>
<td>Accuracy 70%</td>
</tr>
<tr>
<td>2022 [22]</td>
<td>SVM</td>
<td>NC: 20, AD: 27</td>
<td>Accuracy (fMRI) 78.72%</td>
</tr>
<tr>
<td>2022 [23]</td>
<td>SVM</td>
<td>NC: 41, aMCI: 30, AD: 36</td>
<td>Accuracy (sMRI + rs-fMRI) 91.49%</td>
</tr>
<tr>
<td>2022 [24]</td>
<td>LDA</td>
<td>NC: 30, AD: 28</td>
<td>Accuracy (NC vs. AMCI) 68%</td>
</tr>
<tr>
<td>2021 [25]</td>
<td>SVM</td>
<td>MCI-C: 14, MCI-NC: 41</td>
<td>Accuracy (NC vs. AD) 71%</td>
</tr>
<tr>
<td>2021 [26]</td>
<td>SVM</td>
<td>MCI-C: 30, MCI-NC: 55, AD: 19</td>
<td>Accuracy (MCI-C vs. MCI-NC) 83.5%</td>
</tr>
<tr>
<td>2020 [27]</td>
<td>SVM</td>
<td>NC: 51, MCI: 66</td>
<td>Accuracy (MCI-C vs. MCI-NC) 83.5%</td>
</tr>
<tr>
<td>2020 [28]</td>
<td>SVM</td>
<td>NC: 20, SCD: 22</td>
<td>Accuracy (MCI-C vs. MCI-NC) 67.6%</td>
</tr>
<tr>
<td>2020 [29]</td>
<td>RF</td>
<td>NC: 93, MCI: 82</td>
<td>Accuracy (MCI-C vs. MCI-NC) 66%</td>
</tr>
<tr>
<td>2020 [31]</td>
<td>ANFIS</td>
<td>AD: 33, VD: 27, MDX: 15</td>
<td>Accuracy (MCI-C vs. MCI-NC) 98.58%</td>
</tr>
<tr>
<td>2020 [32]</td>
<td>SVM</td>
<td>eMCI: 77, IMCI: 64</td>
<td>Accuracy (MCI-C vs. MCI-NC) 97.76%</td>
</tr>
<tr>
<td>2020 [33]</td>
<td>SVM</td>
<td>NC: 60, MCI: 39</td>
<td>Accuracy (MCI-C vs. MCI-NC) 93.8%</td>
</tr>
<tr>
<td>2019 [35]</td>
<td>SVM</td>
<td>NC: 45, AD: 45</td>
<td>Accuracy (MCI-C vs. MCI-NC) 97.76%</td>
</tr>
<tr>
<td>2019 [36]</td>
<td>CNN</td>
<td>NC: 172, eMCI: 179</td>
<td>Accuracy (MCI-C vs. MCI-NC) 90.24%</td>
</tr>
<tr>
<td>2020 [37]</td>
<td>CNN</td>
<td>NC: 198, AD: 133</td>
<td>Accuracy (MCI-C vs. MCI-NC) 93.8%</td>
</tr>
<tr>
<td>2019 [38]</td>
<td>SVM</td>
<td>NC: 45, SCD: 39, aMCI: 45, AD: 38</td>
<td>Accuracy (MCI-C vs. MCI-NC) 95.58%</td>
</tr>
<tr>
<td>2019 [39]</td>
<td>SVM</td>
<td>NC: 24, eMCI: 24, IMCI: 24, AD: 24</td>
<td>Accuracy (MCI-C vs. MCI-NC) 95.8%</td>
</tr>
<tr>
<td>2019 [40]</td>
<td>ELM</td>
<td>NC: 31 + 152, MCI: 31 + 132, AD: 33 + 81</td>
<td>Accuracy (MCI-C vs. MCI-NC) 97.76%</td>
</tr>
<tr>
<td>2018 [41]</td>
<td>DAG network</td>
<td>NC: 34, AD: 34</td>
<td>Accuracy (MCI-C vs. MCI-NC) 95.5%</td>
</tr>
<tr>
<td>2018 [42]</td>
<td>SVM</td>
<td>MCI-C: 18, MCI-NC: 62</td>
<td>Accuracy (MCI-C vs. MCI-NC) 96.52%</td>
</tr>
<tr>
<td>2019 [43]</td>
<td>AE</td>
<td>NC: 79, MCI: 91</td>
<td>Accuracy (MCI-C vs. MCI-NC) 95.5%</td>
</tr>
<tr>
<td>2018 [44]</td>
<td>SVM</td>
<td>NC: 35, AD: 25</td>
<td>Accuracy (MCI-C vs. MCI-NC) 95.5%</td>
</tr>
</tbody>
</table>
half maximum (FWHM) was selected to improve the signal-to-noise ratio of the data.

2.4.1. PerAF and mPerAF. Based on the formula, PerAF was calculated by subtracting the BOLD signal intensity of each voxel from the mean-time series value of the voxel and then dividing by the mean-time series value. Then the sum of absolute values of each voxel in the time series was divided by the number of time points to obtain the percentage of the fluctuation relative to the mean BOLD signal intensity, namely, the PerAF value of each time series. Unlike ALFF, fALFF, and PerAF, the results are directly used for comparison or can be compared after averaging (mPerAF). This study calculated mPerAF and PerAF in three frequency bands, i.e., Norm-1, Slow-4, and Slow-5. A Gaussian smoothing kernel of 4 mm FWHM was selected to improve the signal-to-noise ratio of data.

2.4.2. Wavelet-ALFF. The continuous wavelet transform was performed on data, and the convolution of scaling and translation form of the mother wavelet function was calculated. Then, the coefficients of each frequency point at all-time points were added for calculation, then the average coefficient of a given frequency band was obtained. This study calculated Wavelet-ALFF in three frequency bands, i.e., Norm-1, Slow-4, and Slow-5. A Gaussian smoothing kernel of 4 mm FWHM was selected to improve the signal-to-noise ratio of data.

2.4.3. Degree Centrality (DC). Other nodes with significant functional connection \( r > 0.25 \) with each node in each brain functional connection group were calculated to obtain the sum DC value of the significant correlation weight of each node, then divided by the average DC value of the whole brain to obtain the standardized DC value. This study calculated DC in three frequency bands, including Norm-1 (0.01–0.08 Hz), Slow-4 (0.027–0.073 Hz), and Slow-5 (0.01–0.027 Hz) frequency bands. A Gaussian smoothing core of 4 mm full width-half maximum (FWHM) was selected to improve the signal-to-noise ratio of data.

2.5. Statistical Analysis. SPSS 23.0 software was used for statistical analysis in the demographic statistics part of this study. Categorical variables, including gender, were marked with the number of each group for direct description. Age, education level, scale score, and other continuous variables were described as mean ± standard deviation (SD) for the MCI group and NC group. An independent sample t-test or Chi-square test was used for comparison between the two groups.

2.6. Extreme Gradient Boosting (XGBoost) Classifier. XGBoost is a type of composite tree model comprising a series of regression and classification trees. As an open source package, XGBoost is widely recognized in many machine learning and data mining challenges, for example, 17 out of 29 challenge solutions posted on the Kaggle blog in 2015 used XGBoost, and the top 10 winning teams in the 2015 KDD Cup used XGBoost [50]. PyCaret 2.1 in Jupyter Notebook was used to train and validate the XGBoost classifier.

3. Results

3.1. Demographics Differences among NC and MCI Groups. The demographic characteristics of study participants are shown in Table 2. The MMSE score (NC: 28.53 ± 1.248, MCI: 25.47 ± 2.506, \( p < 0.001 \)) and MoCA score (NC: 26.23 ± 1.820, MCI: 19.60 ± 2.768, \( p < 0.001 \)) were significantly different among groups, while no significant differences were noted in age, gender ratio, and education level. Independent-samples t-test was used to examine the differences in the characteristics of NC and MCI groups and categorical data were compared using \( \chi^2 \) tests. *Statistically significant differences \( (p < 0.05) \).

3.2. Classification Performance. The XGBoost classifier was trained and validated using 10-fold cross-validation to estimate out-of-sample performance. AUC, recall rate, precision, F1-score, Kappa value, and accuracy were reported. Table 3 shows binary classification performances of the XGBoost classifier in feature datasets. The results revealed that lower levels of accuracy were achieved in all comparisons. The highest accuracy (65.14%) was observed in the mPerAF datasets. Highest AUC (0.6608), recall rate (53.33%), and F1-score (0.5285) were obtained in the fALFF datasets. The highest precision (60.00%) was obtained in the DC datasets. The highest Kappa value (0.2191) was obtained in Wavelet-ALFF datasets.

The receiver operating characteristic (ROC) curves of the XGBoost classifier trained on 90% of datasets and tested on the remaining 10% of datasets are shown in Figure 1. The AUC of the micro-average ROC curve and macro-average ROC for prediction using DC datasets were 0.61 and 0.63 (Figure 1(a)). The AUC of the micro-average ROC curve and macro-average ROC for prediction using fALFF datasets were 0.61 and 0.64 (Figure 1(b)). The AUC of the micro-average ROC curve and macro-average ROC for prediction using mPerAF datasets were 0.58 and 0.62 (Figure 1(c)). The AUC of the micro-average ROC curve and macro-average ROC for prediction using PerAF datasets were 0.58 and 0.61 (Figure 1(d)). The AUC of the micro-average ROC curve and macro-average ROC for prediction using Wavelet-ALFF datasets were 0.66 and 0.65 (Figure 1(e)).

3.3. Model Interpretation: Shapley Additive exPlanations (SHAP). Anatomical Automatic Labeling (AAL) is provided by Montreal Neurological Institute (MNI), with a total of 116 regions. A total of 90 regions belong to the brain, while the remaining 26 regions belong to the cerebellum. Each region has the MNIcro number from 1 to 116. Based on the SHAP algorithm, the feature ranking interpretation of the XGBoost classifier shows the top 20 great characteristics in
### Table 2: Demographics of datasets.

<table>
<thead>
<tr>
<th></th>
<th>NC ($n = 47$)</th>
<th>MCI ($n = 70$)</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>64.91 ± 9.029</td>
<td>66.47 ± 8.007</td>
<td>0.72</td>
</tr>
<tr>
<td>Gender (M/F)</td>
<td>17/30</td>
<td>36/34</td>
<td>0.131</td>
</tr>
<tr>
<td>Education (years)</td>
<td>12.53 ± 2.869</td>
<td>9.10 ± 3.046</td>
<td>0.582</td>
</tr>
<tr>
<td>MMSE</td>
<td>28.53 ± 1.248</td>
<td>25.47 ± 2.506</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>MoCA</td>
<td>26.23 ± 1.820</td>
<td>19.60 ± 2.768</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

### Table 3: Classification performance of XGBoost classifier.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy (%)</th>
<th>AUC</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
<th>F1-score</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>DC</td>
<td>62.78</td>
<td>0.6558</td>
<td>40.00</td>
<td>60.00</td>
<td>0.4538</td>
<td>0.1803</td>
</tr>
<tr>
<td>fALFF</td>
<td>61.94</td>
<td>0.6608</td>
<td>53.33</td>
<td>58.33</td>
<td>0.5285</td>
<td>0.2086</td>
</tr>
<tr>
<td>mPerAF</td>
<td>65.14</td>
<td>0.6333</td>
<td>40.00</td>
<td>54.00</td>
<td>0.4243</td>
<td>0.2077</td>
</tr>
<tr>
<td>PerAF</td>
<td>57.78</td>
<td>0.5867</td>
<td>34.17</td>
<td>46.67</td>
<td>0.3802</td>
<td>0.0742</td>
</tr>
<tr>
<td>Wavelet-ALFF</td>
<td>63.33</td>
<td>0.6142</td>
<td>48.33</td>
<td>55.83</td>
<td>0.4833</td>
<td><strong>0.2191</strong></td>
</tr>
</tbody>
</table>

**Figure 1**: Continued.
predicting outcomes with different datasets (Figure 2). Generally, the characteristics of the superior parietal gyrus (59), parahippocampal gyrus (40), right cerebellum 7b (102), inferior temporal gyrus (90), and right superior frontal gyrus, orbital part (6) positively correlated with the outcomes using DC datasets. The characteristics of left cerebellum 6 (99), left postcentral gyrus (57), right inferior frontal gyrus, opercular part (12), right cerebellum 6 (100), and vermis 6 (112) positively correlated with outcomes using fALFF datasets. The characteristics of the right middle temporal gyrus (86), left middle temporal gyrus (85), left temporal pole, middle temporal gyrus (87), right cerebellum 7b (102), and right olfactory cortex (22) positively correlated with outcomes using mPerAF datasets. The features of the left cerebellum 10 (107), right middle temporal gyrus (86), right inferior frontal gyrus, opercular part (12), vermis 6 (112), and right amygdala (42) positively correlated with outcomes using PerAF datasets. The characteristics of left cerebellum 9 (105), right amygdala (42), left supramarginal gyrus (63), left posterior cingulate gyrus (35), and right precentral gyrus (2) positively correlated with outcomes using Wavelet-ALFF datasets.

SHAP force plot (Figure 3) shows the interpretability of a single model prediction used for error analysis to identify an interpretation for a particular instance prediction. The output value of the XGBoost classifier using DC datasets was −0.10, and the characteristics of the left insula (29), right middle frontal gyrus (8), and right cuneus (46) positively correlated with the output value, whereas the characteristics of the right cerebellum 7b (102), left superior parietal gyrus (59), right superior frontal gyrus, and orbital part (6) negatively correlated with the output value. The output value of the XGBoost classifier using fALFF datasets was 3.85 and features of left cerebellum 6 (99), right inferior frontal gyrus, opercular part (12), and vermis 6 (112) positively correlated with the output value, whereas the characteristics of vermis 7 (113), left middle frontal gyrus, orbital part (9), and left postcentral gyrus (57) negatively correlated with the output value. The output value of the XGBoost classifier using mPerAF datasets was −0.43; the characteristics of the right middle temporal gyrus (86), left middle temporal gyrus (85), left temporal pole, and middle temporal gyrus (87) positively correlated with the output value, whereas the characteristics of right thalamus (78), right cerebellum 7b (102), and left hippocampus (37) negatively correlated with the output value. The output value of the XGBoost classifier using PerAF datasets was −2.04; the characteristics of the right middle temporal gyrus (86), left middle temporal gyrus (85), and left hippocampus (37) positively correlated with the output value, whereas the characteristics of vermis 6 (112), left cerebellum 10 (107), and right amygdala (42) negatively correlated with the output value. The output value of the XGBoost classifier using Wavelet-ALFF datasets was −0.02 and the characteristics of left cerebellum 9 (105), vermis 9 (115), and right precentral gyrus (2) correlated positively with the output value, whereas the characteristics of the left supramarginal gyrus (63), right amygdala (42), and left middle occipital gyrus (51) negatively correlated with the output value.

4. Discussion

A total of 15 machine learning models were used in each of the five datasets (see Tables S1–S5 in the Supplementary Materials for classification performance on the five datasets), and eventually, XGBoost algorithm was selected for the classification diagnosis of MCI and NC based on the overall performance and interpretability of the model. Besides, we used 116 features from rs-fMRI analysis in model classification diagnosis. Based on the analysis of model performance, it was difficult to classify MCI and NC using rs-fMRI features alone, and the highest accuracy was only 65.14% (using the mPerAF dataset).
In contrast with the classification of AD and NC, the classification of MCI and NC is more difficult but appears meaningful because although AD cannot be cured, intervention in MCI patients effectively delays recognition and decreases cognitive capacity [51]. In previous studies, unlike the classification diagnosis of AD and NC, the diagnostic accuracy of MCI and NC is lower. For instance, Lama and Kwon [52] used functional magnetic resonance image features based on graph theory for classification. In the classification diagnosis of MCI and NC, although the accuracy rate of 97.80% is obtained using Lasso regression, only 80%–86% accuracy rate is obtained when other algorithms, including support vector machine based on feature elimination, adaptive structure learning, feature learning based on pairwise correlation are used. Bergeron et al. [53] used the MemTrax test combined with the MoCA score to make model predictions and obtained a prediction accuracy of approximately 90%. Unlike the accuracy rates reported in other studies, our accuracy rate is lower; this is potentially attributed to the overfitting of the XGBoost algorithm model because of the small sample size.

Based on our XGBoost algorithm model, the results are quite different when applied to different datasets. The important features correspond to the following AAL regions, i.e., superior parietal gyrus (59), parahippocampal gyrus (40), right cerebellum 7b (102), left cerebellum 6 (99), left post-central gyrus (57), right inferior frontal gyrus, opercular part (12), right middle temporal gyrus (86), left middle temporal gyrus (85), right amygdala (42), and left supramarginal gyrus (63). We counted the top 20 features in five SHAP algorithm graphs. Among the 116 features, 66 features appeared in the graph, most of which only appeared in SHAP of a certain dataset, among which right cerebellum 7b (102), right superior frontal gyrus, orbital part (6), right middle temporal gyrus (86), right amygdala (42), and vermis 6 (112) appeared with high frequency (≥3). This suggests that the effect of disease state on brain function is extensive, and comprehensive analysis combined with multiple indicators may be beneficial to further analyze the mechanism of cognitive impairment.
Previous studies on MCI indicate that patients with MCI have posterior cingulate gyrus, cuneus, superior marginal gyrus, hippocampus (belonging to the default network), insula (belonging to the prominence network), the lingual gyrus, middle occipital gyrus, and inferior temporal gyrus (belonging to the visual network), which are different from NC [54, 55]. Lenka [56] et al. applied psychophysiological interaction (PPI) analysis to detect specific alterations in PCC connectivity associated with visual processing while controlling brain atrophy. This approach separated the MCI from NC with 77% sensitivity and 89% specificity. Suh et al. [57] developed a 2-step algorithm using a convolutional neural network to perform brain parcellation followed by 3 classifier techniques, including XGBoost for disease prediction. Compared with SVM and logistic regression, XGBoost had a sensitivity of 68% and a specificity of 70% in terms of differentiating AD from MCI group. In terms of MCI from the NC group, XGBoost had a sensitivity of 79% and a specificity of 80%. Shmulev et al. [58] used brain MRI and clinical data to predict MCI conversion to the AD. The resulting accuracy by the XGBoost algorithm is 0.76 ± 0.01, and the AUC of this model is 0.82. A machine learning framework proposed in this paper for MCI detection achieved an accuracy of 65.14% when using the mPerAF dataset. These findings provide novel insights into the understanding of pathological changes in the brain functional network organization of MCI and show the potential of the PerAF analysis-related features for MCI detection.

In the present study, we found abnormal cerebellar activation in all five datasets. In the past, the cerebellum was primarily associated with voluntary movement and postural balance. However, clinical and anatomical work suggests that the cerebellum may also play a role in cognition [60]. A significant number of fMRI research reports further provide supporting evidence that the cerebellum is activated to varying degrees in cognitive tasks (including language, working memory, and spatial processing) [61, 62]. Studies have proposed a connection system between the cerebellum and thalamus; the cerebellum is connected with the thalamus through the brain stem and participates in related functions of the frontal lobe cognitive circuit [63]. This work has compelling shortcomings. First, as a single-center study, only 117 subjects were included; this is a relatively small sample size; thus, a larger sample is required in subsequent studies to further verify the stability of the results.
Secondly, the study lacks long-term follow-up, which makes it impossible to track the conversion of MCI. In contrast with the classification diagnosis of MCI and NC, predicting the conversion of MCI in practical application is necessary. Lastly, this study lacks PET-CT/MR radionuclide (such as AV-45) labeling results of subjects, thus limiting its credibility and conviction. Therefore, it is necessary to cooperate with multiple hospitals to perform multicenter researches. Besides, reliable pathological diagnosis, including the improvement of Aβ protein examination of cerebrospinal fluid or AV-45 PET, improves the value of the results in clinical application.

5. Conclusion

In conclusion, our findings demonstrate that the XGBoost algorithm constructed from rs-fMRI data is effective in classifying and diagnosing MCI. Using mPerAF dataset, we obtained the highest accuracy for diagnosing MCI. This suggests that the outcomes of rs-fMRI analysis may be useful as imaging markers for MCI diagnosis. The accuracy rates obtained by different rs-fMRI data analysis methods are similar, but the important features are different and involve multiple brain regions, which suggests that MCI may have a negative impact on brain function.

Data Availability

The rs-fMRI data used to support the findings of this study are restricted by the Ethics Committee of First Affiliated Hospital, Zhejiang University School of Medicine, in order to protect patient privacy. Data are available from the corresponding author for researchers who meet the criteria for access to confidential data.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors’ Contributions

GP and BL conceived and designed the project. MH wrote the manuscript with inputs from other authors. YY and FH provided guidance on data analysis. Other authors helped to revise the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Acknowledgments

This work was supported by the Key R & D Program of Zhejiang (2022C03064), the Zhejiang Science and Technology Project (WKJ-ZJ-2024), the National Natural Science Foundation of China (82101251), and the Ministry of Science and Technology of the People’s Republic of China (2019YFC0118203 and 2016YFC1306402).

Supplementary Materials

Tables S1–S5 in the Supplementary Material show comparison of the performance of 15 machine learning models using the DC/fALFF/mPerAF/PerAF/Wavelet-ALFF dataset. (Supplementary Materials)

References

Computational Intelligence and Neuroscience


