



OPEN Identification of texture MRI brain abnormalities on Fibromyalgia syndrome using interpretable machine learning models

Hongyang Jiang^{1,3,4}, Aihui Liu^{2,4} & Zhenhua Ying^{1,2}✉

To provide objective diagnostic markers for fibromyalgia symptoms (FMS) diagnosis, we have created interpretable extreme gradient boosting (XGBoost) models using radiomics to aid in the diagnosis of chronic pain (CP) and to develop nomogram models for diagnosing subgroups of FMS. A group of 54 patients with CP and 71 healthy controls was randomly separated into training and validation groups, using a 7:3 ratio. Radiomics features were extracted from grey-matter and white-matter in the filtered mwp0* image. The Mann-Whitney U test, Spearman's rank correlation test, and least absolute shrinkage and selection operator (LASSO) were utilized to select features. An XGBoost model was created based on these features, and Shapley Additive exPlanations (SHAP) was used for personalization and visual interpretation. A nomogram was developed for the diagnosis of FMS subgroups, utilizing radiomics scores and clinical predictors. The efficacy of the nomogram was evaluated using the area under the receiver operating characteristic curve, while decision curve analysis was employed to evaluate its clinical efficacy. The XGBoost model displays stability in the training validation group, indicating lower overfitting of CP model. The nomogram model combined with the rad-score has a greater ability to distinguish between typical and sub-clinical than the clinical factor model alone. We developed and validated a CP diagnosis model by XGBoost and realized model visualization through SHAP. The rad-score obtained by machine learning was used to build a nomogram model that combines clinical scales to distinguish patients with typical and sub-clinical fibromyalgia.

Keywords Fibromyalgia, XGBoost, Radiomics, Chronic pain, MRI

Fibromyalgia syndrome (FMS) is a chronic disorder characterized by widespread musculoskeletal pain, fatigue, and tenderness in localized areas¹. Identifying the underlying causes of FMS can be challenging due to its complex nature. The diagnosis can be difficult because there are no specific diagnostic tests. It is usually made based on the patient's symptoms and a physical examination². The diagnostic proposal criteria of 2011 and 2016 eliminated the previous 1990 tender point exam and defined FMS as a multi-symptom disorder. Even with the guidance of criteria, clinician-based FMS diagnosis may lead to bias leading to overdiagnosis or underdiagnosis of patients, due to the subjective nature of the symptoms assessed³. Some articles use the terms sub-clinical, pre-fibromyalgia, and incomplete fibromyalgia to describe patients who do not fully meet the guideline criteria^{4,5}. Although controversial, the proper identification and management of these patients need to be addressed.

Since the diagnosis of FMS lacks objective biomarkers, a large number of neuroimaging methods have been used to explore its central mechanism⁶. The study found that the resting state blood oxygen dependent level signal of these brain regions such as the inferior parietal lobule, insular cortex, medial prefrontal cortex, amygdala, and medial hypothalamic nucleus decreased in FMS patients. Resting-state functional networks such as default mode network, dorsal attention network, salience network, and related functional connectivity changes have also been confirmed to be related to the occurrence and progression of FMS⁷⁻⁹. Grey-matter (GM) analysis

¹Medical College of Soochow University, Suzhou, China. ²Center for General Practice Medicine, Department of Rheumatology and Immunology, Zhejiang Provincial People's Hospital (Affiliated People's Hospital), Hangzhou Medical College, Hangzhou, Zhejiang, China. ³Center for Rehabilitation Medicine, Department of Radiology, Zhejiang Provincial People's Hospital (Affiliated People's Hospital), Hangzhou Medical College, Hangzhou, Zhejiang, China. ⁴Hongyang Jiang and Aihui Liu co-first author. Contributed equally to the study. ✉email: yingzh2021@163.com

achieved abundant results through voxel-based and surface-based morphometry analyses. META analysis and reviews summarized GM volume changes in the amygdala, thalamus, hippocampus, insula, anterior cingulate cortex, and inferior frontal gyrus, which indicate that chronic pain is related to subtle brain structures and spatial distribution¹⁰. Microstructural changes in white-matter (WM) connectivity were also demonstrated by DTI in the medial lemniscus, corpus callosum, and peri-thalamic and connective tracts in FMS patients^{11,12}. Several research studies, including META analysis and reviews, have been unable to consistently identify which brain regions undergo specific sequential changes and become the central biological marker of FMS¹³.

Most of the above studies have focused on regional volume or intensity measures and have not fully exploited the rich information contained in brain MRI¹⁴. To solve this problem, we decided to take the whole GM and WM as the region of interest (ROI) and use radiomics to extract characteristic parameters to reflect the changes in microstructure. Radiomics reflected tumor heterogeneity in initial studies, in which most features were processed by comparing each voxel with its neighboring voxels and saving the results as binary numbers^{15,16}. In recent years, radiomics texture features with their potential as image-based biomarkers have been widely used across several central nervous systems studies, for Alzheimer's¹⁷ and Parkinson's disease¹⁸ as neurodegenerative diseases, major depression¹⁹, and schizophrenia^{20,21}. The main advantage of applying radiomics texture features is their potential to capture microscopic alterations in tissue characteristics of the brain, however, the reproducibility and visualization of the method need to be emphasized.

Currently, there has been no research that has utilized radiomics in the study of central nervous systems mechanisms related to FMS. Our study aimed to identify any differences between individuals with FMS and healthy controls (HC) by analyzing 12 texture signatures of GM and WM using non-segmented MR images. We then took these signatures and input them into a machine-learning binary classification model to determine how each feature impacted the model, which was presented visually.

Materials and methods

Participants

A total of 131 subjects were included in this study, 6 of them were excluded by the image quality control. 71 subjects were in the HC group, 46 of them are cases in our hospital, and 25 of them were the normal health from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. 54 subjects were in the CP group, 35 of them were fibromyalgia-ness chronic pain, and 19 of them were other over 3 months chronic pain which includes rheumatoid arthritis, chronic low back pain, and ankylosing spondylitis. All FMS-ness patients had been examined by a rheumatologist and a neurologist, and divided into the typical FMS group and the sub-clinical FMS group. The typical group fulfilled the diagnostic criteria for FMS according to the guidelines released by the American College of Rheumatology 2016 reversion²²: (1) Presence of generalized pain (four quadrants and axial of five regions). (2) Symptoms have been present at a similar level for at least 3 months. (3) Widespread Pain Index (WPI) ≥ 7 and Symptom Severity (SS) ≥ 5 , or WPI of 4–6 and SS ≥ 9 . The sub-clinical groups exhibited similar fatigue or sleep disturbance for at least 3 months, with the presence of at least 3 regional pain points (WPI ≥ 3). Our hospital conducted comprehensive examinations on both groups to rule out other possible diagnoses. Additionally, both groups completed the self-rating anxiety scale (SAS) and self-rating depression scale (SDS). The Institutional Review Board (IRB) of the Zhejiang Provincial People's Hospital (IRB No.2023KY053) reviewed and approved the study protocol. All patients gave their written informed consent, following the principles of the Declaration of Helsinki.

Image acquisition

High-resolution 3D-FSPGR images were acquired by GE Discovery 750 3.0T with the following parameters: TE = 2.9 ms, TR = 6.7 ms, flip angle = 15 degree, FOV = 216 × 216 mm, matrix = 256 × 256, slice thickness = 1 mm, 192 scanning images. The data from ADNI were also acquired by the same machine and similar parameters: TE = 3.1 ms, TR = 7.4 ms, flip Angle = 11 degree, FOV = 216 × 216 mm, matrix = 256 × 256, slice thickness = 1 mm, 192 scanning images.

Image preprocessing and segmentation

Data were preprocessed with the Computational Anatomy Toolbox (CAT12.8.2 r2130, <http://www.neuro.uni-jena.de/cat/>) ran under Statistical Parametric Mapping, Version 12 (SPM12, <http://www.fil.ion.ucl.ac.uk/spm/software/spm12/>)²³. To ensure quality, we visually inspected all raw images for artifacts and statistically controlled all segmented images for inter-subject homogeneity by the ratio between weighted overall image quality and quartic mean Z-score. All the structural images were segmented into GM, WM, and cerebrospinal fluid (CSF), and were normalized to Montreal Neurological Institute (MNI) standard space by Geodesic Shooting templates. To preserve the volume of GM and WM, we applied “modulation” during the normalization step and then resampled the images to 1.5 × 1.5 × 1.5 mm³. Finally, we smoothed the modulated images with an isotropic 8 mm full-width half maximum Gaussian kernel. The graphical representation delineates the steps involved in the image processing pipeline, as illustrated in Fig. 1.

Radiomics signatures extraction

Radiomics features were extracted using the PyRadiomics open-source Python package (version 2.1.0; <https://pyradiomics.readthedocs.io/>)²⁴. The width of discretization bins for feature extraction was fixed to 25. We extracted the standard feature classes, which included shape, first-order, texture, wavelet, exponential, and square transform features²⁵. ComBaTool and z-scores were then used to normalize the features. ComBaTool, a free online application (https://forlhac.shinyapps.io/Shiny_ComBat/)²⁶, was used to pool features and minimize inter-scanner variability. Principal component analysis was utilized to visualize the effects of Combat on feature uniformization²⁷. Finally, all radiomics features were standardized using z-scores. All features were extracted

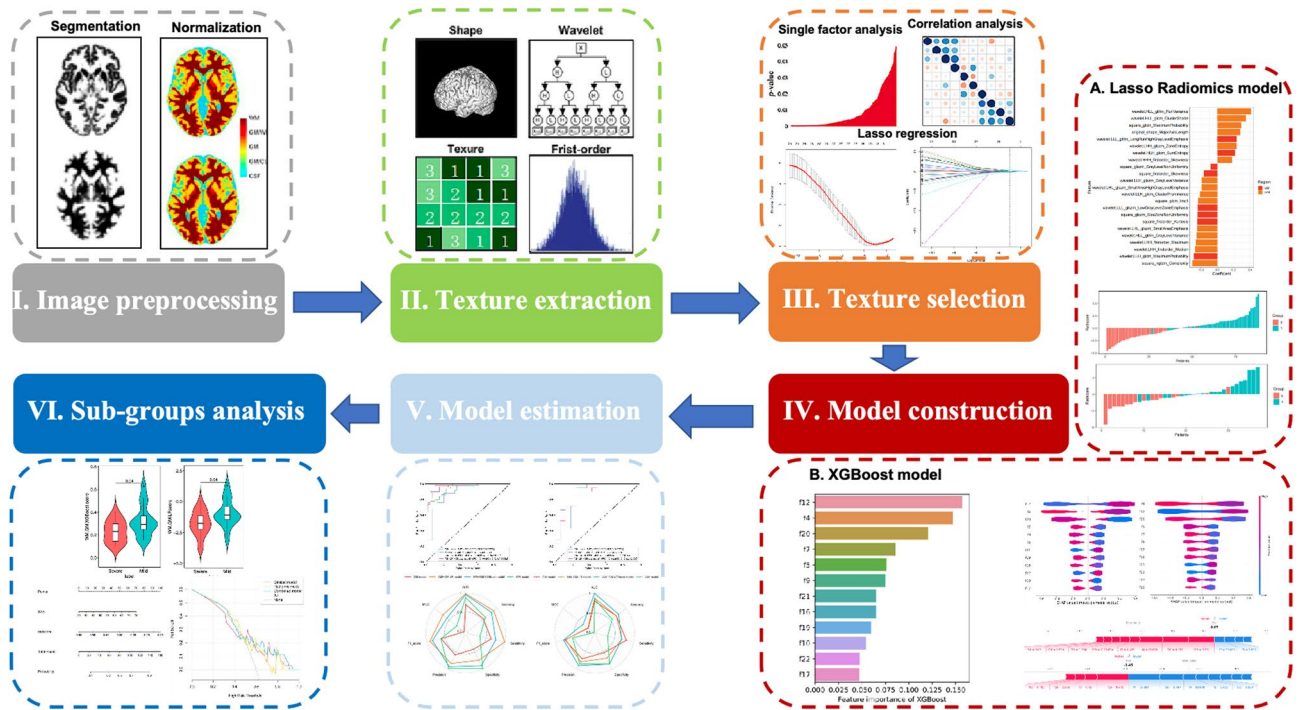


Figure 1. Flowchart of the study.

from transformed $mwp0^*$ images with the individual masks of GM and WM. Details can be found in the software package and the source code²⁸.

Feature selection and binary logistic regression (LR) model construction

The patients were selected using random stratified sampling, with consistent control of the CP/HC ratio in the training and validation groups at 7:3. The model was constructed and features were selected from both GM and WM to distinguish between the two groups. The selected features were screened through three steps: (1) include features with significant differences using an independent-sample t-test or Mann-Whitney U test ($p < 0.05$). (2) Elimination of internal redundant features using Spearman's rank correlation test (threshold of 0.8). (3) Reduction of inter-sequence redundancy and selection of the best predictive features using the least absolute shrinkage and selection operator (LASSO)- binary logistic regression model. The best lambda to determine non-zero regression coefficient variables was determined through a 10-fold cross-validation process. Important WM and GM features were used to construct respective models through logistic analysis. After eliminating a feature, logistic modeling analysis is conducted by merging the characteristics of both regions.

Establishment and validation of the XGBoost model

To classify CPs and HCs in the validation dataset, an XGBoost model was built using the training dataset. GridSearchCV was used to optimize the model parameters. The radiomics signatures were arranged and combined, and cross-validation was used to return evaluation index scores for all parameter combinations. The optimal value was selected as the parameter corresponding to the combination with the highest score, and the model prediction probability is taken as rad-score.

To comprehend the reasons for the complexity of XGBoost models, Shapley Additive exPlanations (SHAP) is utilized for analyzing the correlation between features and outputs in the XGBoost “black box”²⁹. With SHAP analysis, the contribution of each feature to changes in the model output is represented by its SHAP value. The prediction results are linearly decomposed into the influence of individual features, which enables the calculation of feature importance and visualization of the role of different features in the model based on their sensitivity to changes in output.

Combined nomogram model building

The combined model was established using the rad-score of the radiomics model with the best prediction performance and the highly related clinical indicators of FMS by logistic regression analysis, which was presented in the form of a nomogram³⁰. The accuracy of quantitative prediction was evaluated using the area under curve (AUC). The consistency between the predicted results and the actual results was evaluated using a calibration curve, and its clinical effectiveness was evaluated using a decision curve analysis (DCA).

Statistical analysis

The study used two software programs, SPSS 26.0 and R software (4.1.3, <http://www.r-project.org>), to perform statistical analysis. A statistically significant difference was considered when $P < 0.05$. Clinical data were analyzed using the SPSS software. The chi-square test was used for classified variable analysis, the t-test for continuous variables of normal distribution, and the Mann–Whitney U test for abnormal or unknown distribution. Univariate and multivariate logistic analyses were performed to identify clinical indicators that have a high correlation with FMS. The R software was used to establish and evaluate the nomogram. The software packages “car”, “rms”, “pROC”, and “DecisionCurve” were used to analyze the nomogram, receiver operating characteristic (ROC) curve, calibration curve, and decision curve analysis (DCA)³¹.

Results

Patient characteristics

The statistical analysis results for the baseline data of the HC and CP groups are shown in Table 1. FMS is widely recognized as a condition that is more prevalent in women, with a reported incidence ratio of women to men ranging from 2:1 to 7:1. The ratio of females (37 cases) to males (17 cases) in chronic pain cohort is also close to 2:1. Additionally, to mitigate gender as a potential confounding factor affecting brain structure, we controlled for the gender ratio in the normal control group by randomly sampling more female cases from the ADNI database. This approach ensured that there was no statistically significant difference in the gender composition between the CP group and the HC group. There were significant differences in CSF ($P = 0.02$) and related CSF ($P = 0.027$). However, other baseline characteristics were not significantly different between the two groups. The statistical analysis results for the sub-groups are shown in Table 2. In addition to WPI ($P < 0.001$), SS ($P < 0.001$), and SAS ($P = 0.052$), there were no significant differences in typical and sub-clinical groups.

Radiomics features selection and LR model

A total of 944 features were obtained from GM and WM in the filtered mwp0* image. 242 features of WM and 121 features of GM were found significant for chronic pain in the Mann–Whitney U test. Spearman’s rank correlation test further screened the features and obtained 42 WM features and 32 GM features. In the LASSO model, λ was chosen by ten-fold cross-validation, and $\log(\lambda)$ of -2.987 and -2.727 were the optimal subset for radiomics features of WM and GM. 15 WM features and 9 GM features were finally selected. Through Spearman correlation analysis (with a threshold value of 0.8), a redundant feature was identified and removed. The remaining 23 features (15 WM and 8 GM features) were used to construct the LR model, with each feature’s respective coefficient factored into the linear combination. Waterfall plots showed the rad-score for individuals in the training cohort and validation cohort ([Supplementary Material](#)).

Establishment and validation of the XGBoost model

GridSearchCV was used to further optimize the modeling parameters from the features selected by LASSO dimensionality reduction. The classified bar chart of the SHAP summary plots was obtained by extracting the average absolute value of SHAP for 12 radiomics signatures to show the global significance (Fig. 2A). The Beeswarm graph for both training and validation groups showed features such as wavelet.HLL_glcm_ClusterShade, square_ngtdm_Complexity of WM, and wavelet.LLH_glcm_MaximumProbability of GM was

	Chronic pain (n=54)	Healthy control (n=71)	p
Age	59	54	0.502
	[44.25, 66.00]	[45.50, 65.50]	
Gender			0.683
Female	37 (68.5)	45 (63.4)	
Male	17 (31.5)	26 (36.6)	
TIV	1387.5	1444	0.219
	[1337.00, 1501.50]	[1330.00, 1524.50]	
CSF	279	318	0.02*
	[258.00, 362.50]	[286.50, 359.50]	
Related.CSF	0.2	0.23	0.027*
	[0.19, 0.25]	[0.21, 0.25]	
GM	624.5	631	0.737
	[603.50, 654.50]	[594.50, 681.00]	
Related.GM	0.45	0.44	0.164
	[0.43, 0.47]	[0.43, 0.46]	
WM	462	478	0.608
	[439.25, 504.50]	[436.00, 508.00]	
Related.WM	0.33	0.33	0.216
	[0.32, 0.35]	[0.32, 0.35]	

Table 1. The baseline data of CP and HC groups. *Significant level at $p < 0.05$.

	Sub-clinical FMS (n=14)	Typical FMS (n=21)	p
Age	49.5	45	0.4
	[43.25, 58.75]	[38.00, 58.00]	
Gender			0.233
Female	14 (100.0)	17 (81.0)	
Male	0 (0.0)	4 (19.0)	
SDS	45	50	0.296
	[42.25, 53.12]	[45.00, 58.75]	
SAS	47.5	52.5	0.052
	[45.50, 52.19]	[48.75, 57.00]	
WPI	4	9	<0.001*
	[3.00, 6.00]	[7.00, 14.00]	
SS	3	6	<0.001*
	[2.25, 3.75]	[5.00, 8.00]	

Table 2. The baseline of patients with typical and sub-clinical FMS. *Significant level at $p < 0.05$.

consistently among the top 3, indicating that they are highly representative (Fig. 2B). SHAP value plot is a heatmap used to interpret the output of a machine learning model, which visually shows how the top 3 features and the remaining 9 features affect model predictions. The plot visualizes feature importance on a color scale, with each row representing different features and columns representing individual instances. The color intensity correlates with the SHAP value; red indicates a high positive impact on the model's output prediction while blue indicates a negative impact (Fig. 2C).

Performance comparison

The AUC of training group are (Fig. 3A), respectively, GM 0.880 (95% CI, 0.808–0.943), WM 0.940 (95% CI, 0.884–0.979), rad-score 0.974 (95% CI, 0.939–0.996), and XGBoost model 0.966 (95% CI, 0.927–0.992). The AUC of validation groups are, respectively, GM 0.854 (95% CI, 0.705–0.966), WM 0.875 (95% CI, 0.740–0.974), rad-score 0.884 (95% CI, 0.773–0.974), and XGBoost model 0.932 (95% CI, 0.821–1.000). The radar charts (Fig. 3B) show the performance of models in six dimensions accuracy, sensitivity, specificity, precision, F1_score, and Matthew's correlation coefficient (MCC). In the training group, the rad-score model has higher accuracy, sensitivity, and specificity. However, the XGBoost model displays stability in both the training group and validation group, indicating lower model overfitting.

Sub-group analyses and nomogram validation

We used the radiomics model of chronic pain for the subgroup analysis of FMS. We found that there were significant differences in SAS ($P = 0.052$), and rad-score ($p = 0.04$) had an excellent ability to distinguish typical and sub-clinical groups. Combined with the conventional clinical factors and radiomics signature, a clinical and radiomics nomogram was built, as presented in Fig. 4A. The AUC value of the ROC curve and DCA (Fig. 4B) were used to evaluate the ability of the model to classify typical and sub-clinical groups. The nomogram model with SAS as the clinical factor has more accuracy and sensitivity (AUC 0.697 (0.514–0.863), accuracy 0.714, sensitivity 0.81, specificity 0.571), while the nomogram model combined with the rad-score has extremely high specificity (AUC 0.724 (0.544–0.874), accuracy 0.657, sensitivity 0.476, specificity 0.929).

Discussion

We utilized machine learning techniques to choose radiomics features from the GM and WM of MR images. Then we proceeded with constructing two diagnostic models to differentiate CP patients from normal populations. The XGBoost model exhibits the best discrimination power among the two models and was further confirmed to demonstrate superior calibration and clinical utility. To further explore the biomarkers of FMS, we assumed the diagnostic model of CP also could be applied to the differentiation of symptom severity in FMS; this hypothesis was verified. The radiomics nomogram model incorporating the radiomics signature and clinical psychological scale exhibited a higher power for classifying the typical and sub-clinical groups of FMS patients. At last, SHAP and nomogram provided reasonable visual interpretations of the classification model, including positive and negative effects.

Pain is an unpleasant sensory and emotional experience associated with actual or potential tissue damage or an experience similar thereto, which is defined by the International Classification of Diseases (ICD-11). CP refers to persistent or recurrent episodes of more than 3 months, involving multiple factors such as biological, psychological, and social. Absence of objective and quantitative tools for diagnosis, measurements of brain structure or activity would be potential biomarkers of disease, like the amygdala, hippocampus, striatum, anterior insula, and prefrontal cortex, which were confirmed related to CP³². There are differences in gender, age, pain distribution, and intensity of each individual, which may be the reason why many studies on the central mechanism have no consistent results³³. Inspired by some literature^{12,28}, we separately extract texture information of gray-matter and white-matter from unsegmented brain T1WI. And using the method of machine learning, the rad-score score is generated through ten-fold cross-validation LASSO and Logistic regression to

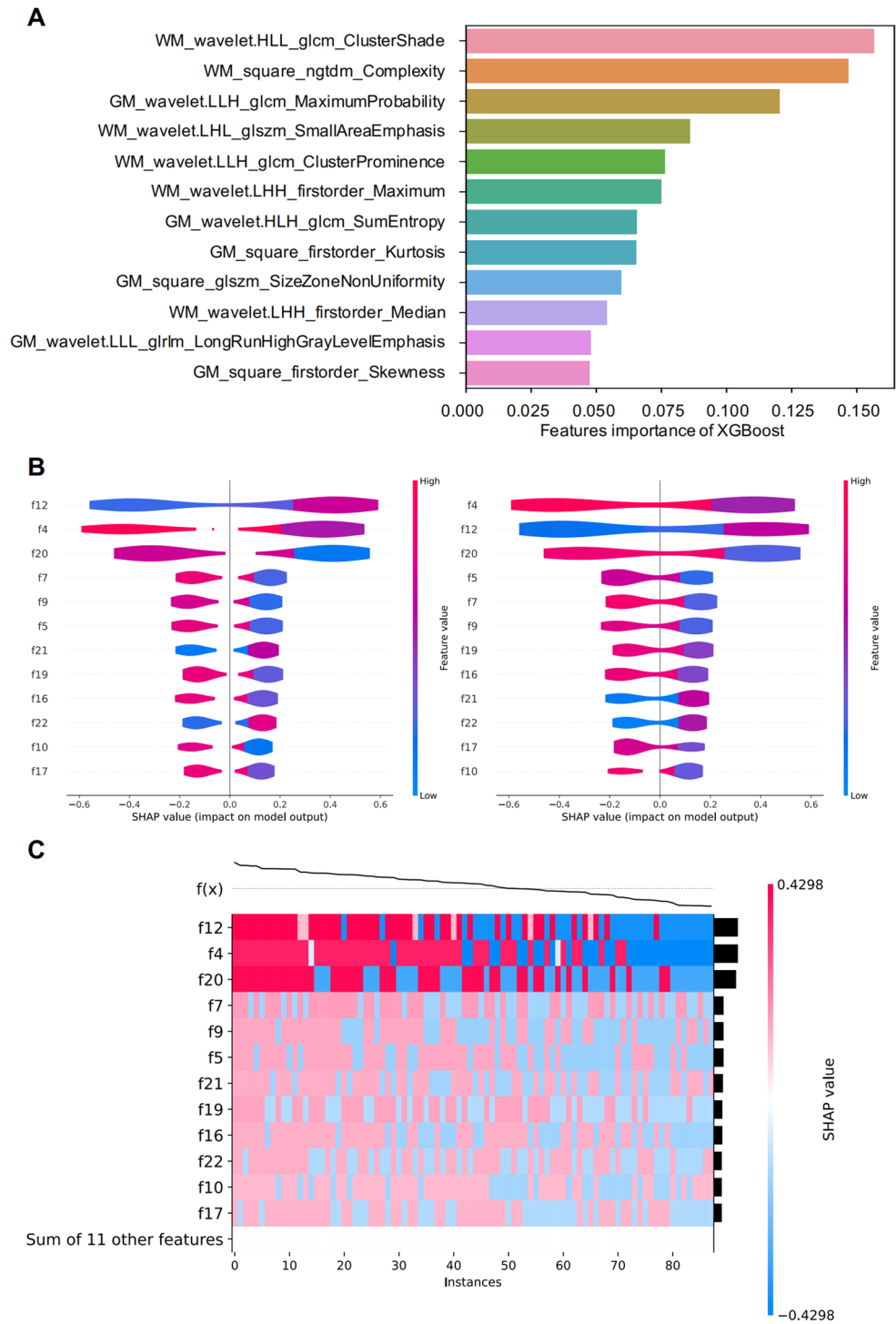


Figure 2. SHAP plots of the XGBoost model. (A) The classified bar charts of the SHAP summary plots show the influence of each parameter on the XGBoost model. (B) The top 3 features remain ahead in the Beeswarm map, which shows the relationship between the characteristic value and the predicted probability through colors, including positive (f12) and negative (f4 and f20) predictive effects (The corresponding table of feature names and numbers is shown in Supplementary Table 2). (C) The $f(x)$ curve at the top is the model prediction for the instance, the x-axis is for each instance, the feature importance is displayed in descending order in the y-axis, and the color describes the direction and strength of the influence of the feature on the instance (red indicates a high positive impact on the model's output prediction while blue indicates a negative impact).

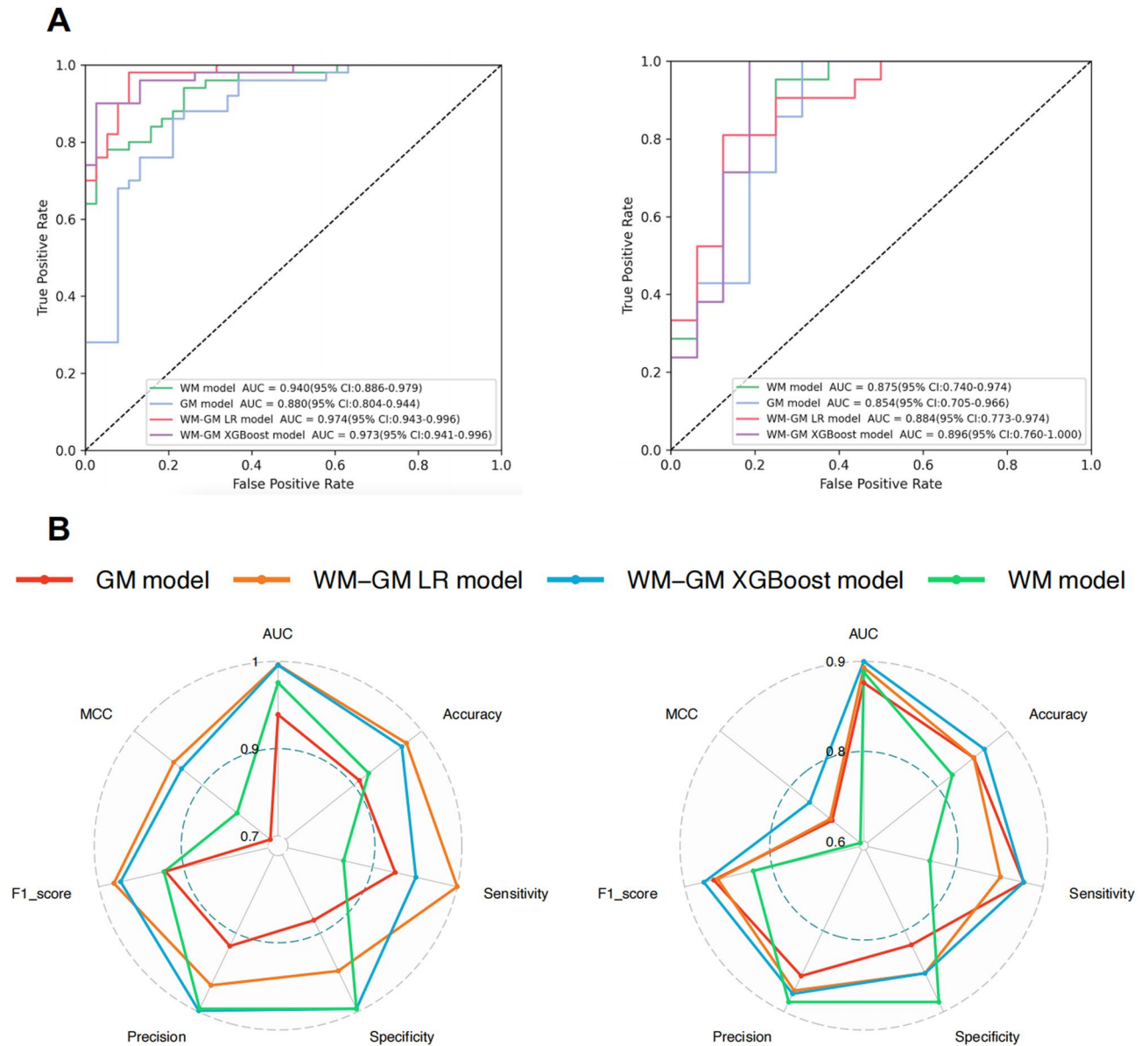


Figure 3. Both the ROC curve and the radar chart show the excellent diagnostic performance of the LR model and the XGBoost model, while the XGBoost model has a lower degree of overfitting.

establish a binary classification model of CP. The AUC of the WM model in the training group was 0.940 (95%CI 0.886–0.979), and the AUC of the test group was 0.875 (95%CI 0.740–0.974), which performed better than the gray matter model. This suggests that microstructural damage of white-matter can be more significantly reflected by radiomics features, compared with the changes of gray matter (training AUC 0.880, 95%CI 0.804–0.944; validation AUC 0.854, 95%CI 0.705–0.966).

Important requirements for machine learning models are their repeatability and interpretability. XGBoost is a massively parallel Boosting Tree algorithm that can prevent overfitting by selecting an appropriate model complexity. According to GridSearchCV optimization parameters, we calculated the importance of the 23 features selected by LASSO and found that 12 of them had strong model contributions. The model established by XGBoost has similar excellent performances in both the training group (AUC 0.966) and the validation group (AUC 0.932), indicating that it has a lower degree of overfitting than the Logistic model. Despite the ‘black box’ nature of machine learning, XGBoost showed better interpretability than other machine learning methods. SHAP analysis was performed to show 3 features have the top three contributions in both the training and validation group. Two of them are Gray Level Co-occurrence Matrix (GLCM) Features and one of them is Neighbouring Gray Tone Difference Matrix (NGTDM) Feature, which all describe the gray relationship with neighborhood pixels in the image. The heterogeneity in GW and WM image texture was quantitatively evaluated from the aspects of ray-level distribution symmetry, uniformity, and complexity, reflecting gray matter atrophy³⁴ and white-matter microstructural damage³³.

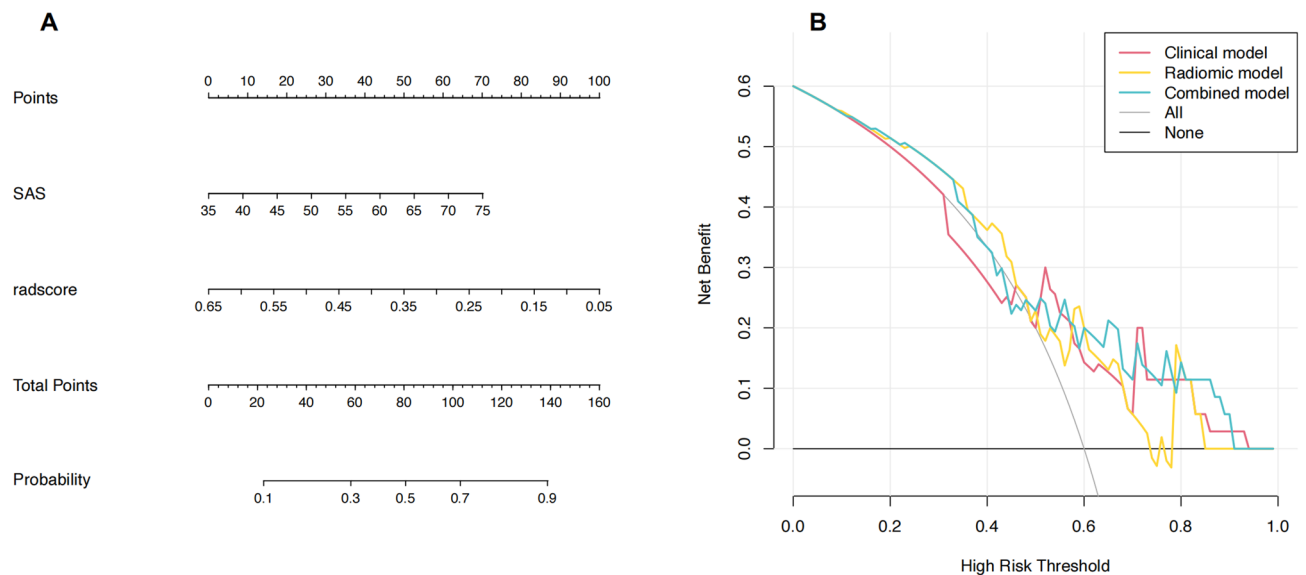


Figure 4. The Combined nomogram incorporated conventional clinical factors (SAS) and rad-score. The nomogram is valued to obtain the probability of severe symptoms by adding up the points identified on the points scale for each variable.

Our findings suggest that the diagnostic markers identified through radiomics features from MR images could potentially transform the way FMS subgroups are diagnosed. We use the CP diagnostic model to distinguish between typical and sub-clinical groups of FMSs and found that a high score on the psychological scale SAS was a clinical predictor of severe symptoms. A number of studies have reported that combining radiomics signatures with clinical risk factors can improve the accuracy of predictive models^{30,35}. We established a nomogram model including clinical features and rad-score for FMS subgroup differentiation, and its diagnostic efficiency was high, indicating that the degree of discrimination of the nomogram model was good. The individual clinical feature model has more accuracy and sensitivity, while the combined model has better performance and higher specificity. The radiomics nomogram model, which combines radiomics signatures with clinical psychological scales, could serve as a predictive tool to assess the severity of symptoms and guide personalized therapy. We hope these markers could complement existing clinical assessment methods by providing a more objective and quantitative approach, enabling clinicians to make more informed decisions regarding treatment plans and monitor disease progression.

We acknowledge several potential limitations and challenges in their implementation within a clinical setting. First, this study did not test the model's generalization using external validation due to the lack of data from public chronic pain databases and other centers. Future studies should aim to validate these models across different cohorts and healthcare settings in multicenter and large sample sizes. Second, this study did not include other fMRI sequences commonly used in clinical practice, such as DTI and fMRI. We expect to further explore whether this information can be used to establish diagnostic models in subsequent studies. Finally, the efficacy evaluation of chronic pain is also highly subjective, and objective evaluation methods are also required. Due to factors such as patient compliance, the data volume of longitudinal studies is insufficient.

Conclusion

We developed and validated a chronic pain diagnosis model by XGBoost and realize model visualization through SHAP. A radiomics nomogram model was created using the rad-score from machine learning. This model combines clinical scales to differentiate between patients with typical and subclinical fibromyalgia. It is a powerful tool for clinical diagnosis and assessment of FMS patients.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 2 January 2024; Accepted: 25 September 2024

Published online: 09 October 2024

References

1. Yang, S. & Chang, M. C. Chronic Pain: Structural and Functional Changes in Brain Structures and Associated Negative Affective States. *Int. J. Mol. Sci.* **20**(13), <https://doi.org/10.3390/ijms20133130> (2019).
2. Häuser, W. et al. Fibromyalgia syndrome: classification, diagnosis, and treatment. *Dtsch. Arztebl Int.* **106**(23), 383–391. <https://doi.org/10.3238/arztebl.2009.0383> (2009).

3. Galvez-Sánchez, C. M. & Del Reyes, G. A. Diagnostic criteria for fibromyalgia: critical review and future perspectives. *J. Clin. Med.* **9**(4), <https://doi.org/10.3390/jcm9041219> (2020).
4. Brummett, C. M. et al. Characteristics of fibromyalgia independently predict poorer long-term analgesic outcomes following total knee and hip arthroplasty. *Arthritis Rheumatol. (Hoboken NJ)*. **67**(5), 1386–1394. <https://doi.org/10.1002/art.39051> (2015).
5. Yunus, M. B. & Aldag, J. C. The concept of incomplete fibromyalgia syndrome: comparison of incomplete fibromyalgia syndrome with fibromyalgia syndrome by 1990 ACR classification criteria and its implications for newer criteria and clinical practice. *J. Clin. Rheumatol.* **18**(2), 71–75. <https://doi.org/10.1097/RHU.0b013e318247b7da> (2012).
6. Staud, R. Brain imaging in fibromyalgia syndrome. *Clin. Exp. Rheumatol.* **29**(6 Suppl 69), S109–S117 (2011).
7. Schrepf, A. et al. A multi-modal MRI study of the central response to inflammation in rheumatoid arthritis. *Nat. Commun.* **9**(1), 2243. <https://doi.org/10.1038/s41467-018-04648-0> (2018).
8. Basu, N. et al. Neurobiologic features of Fibromyalgia are also Present among Rheumatoid Arthritis patients. *Arthritis Rheumatol.* **70**(7), 1000–1007. <https://doi.org/10.1002/art.40451> (2018).
9. Kong, J. et al. Altered functional connectivity between hypothalamus and limbic system in fibromyalgia. *Mol. Brain.* **14** (1), 17. <https://doi.org/10.1186/s13041-020-00705-2> (2021).
10. Henn, A. T. et al. Structural imaging studies of patients with chronic pain: an anatomical likelihood estimate meta-analysis. *Pain* **164**(1), e10–e24. <https://doi.org/10.1097/j.pain.0000000000002681> (2023).
11. Mosch, B., Hagen, V., Herpertz, S. & Diers, M. Brain morphometric changes in fibromyalgia and the impact of psychometric and clinical factors: a volumetric and diffusion-tensor imaging study. *Arthritis Res. Ther.* **25**(1), 81. <https://doi.org/10.1186/s13075-023-03064-0> (2023).
12. Cagnie, B. et al. Central sensitization in fibromyalgia? A systematic review on structural and functional brain MRI. *Semin Arthritis Rheum.* **44**(1), 68–75. <https://doi.org/10.1016/j.semarthrit.2014.01.001> (2014).
13. Xin, M., Qu, Y., Peng, X., Zhu, D. & Cheng, S. A systematic review and meta-analysis of Voxel-based morphometric studies of fibromyalgia. *Front. Neurosci.* **17**, 1164145. <https://doi.org/10.3389/fnins.2023.1164145> (2023).
14. Chaddad, A., Desrosiers, C. & Toews, M. Multi-scale radiomic analysis of sub-cortical regions in MRI related to autism, gender and age. *Sci. Rep.* **7**, 45639. <https://doi.org/10.1038/srep45639> (2017).
15. Rimola, J. Heterogeneity of hepatocellular carcinoma on Imaging. *Semin Liver Dis.* **40**(1), 61–69. <https://doi.org/10.1055/s-0039-1693512> (2020).
16. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: images are more than pictures, they are data. *Radiology* **278**(2), 563–577. <https://doi.org/10.1148/radiol.2015151169> (2016).
17. Li, T.-R. et al. Radiomics analysis of magnetic resonance imaging facilitates the identification of preclinical Alzheimer's disease: an exploratory study. *Front. Cell. Dev. Biol.* **8**, 605734. <https://doi.org/10.3389/fcell.2020.605734> (2020).
18. Hu, X. et al. Multivariate radiomics models based on 18F-FDG hybrid PET/MRI for distinguishing between Parkinson's disease and multiple system atrophy. *Eur. J. Nucl. Med. Mol. Imaging.* **48** (11), 3469–3481. <https://doi.org/10.1007/s00259-021-05325-z> (2021).
19. Korda, A. I. et al. Identification of voxel-based texture abnormalities as new biomarkers for schizophrenia and major depressive patients using layer-wise relevance propagation on deep learning decisions. *Psychiatry Res. Neuroimaging.* **313**, 111303. <https://doi.org/10.1016/j.pscychres.2021.111303> (2021).
20. Park, Y. W. et al. Differentiating patients with schizophrenia from healthy controls by hippocampal subfields using radiomics. *Schizophr Res.* **223**, 337–344. <https://doi.org/10.1016/j.schres.2020.09.009> (2020).
21. Bang, M. et al. An interpretable multiparametric radiomics model for the diagnosis of schizophrenia using magnetic resonance imaging of the corpus callosum. *Transl Psychiatry.* **11** (1), 462. <https://doi.org/10.1038/s41398-021-01586-2> (2021).
22. Wolfe, F. et al. 2016 revisions to the 2010/2011 fibromyalgia diagnostic criteria. *Semin Arthritis Rheum.* **46** (3), 319–329. <https://doi.org/10.1016/j.semarthrit.2016.08.012> (2016).
23. Gaser, C. et al. CAT – a computational anatomy toolbox for the analysis of Structural MRI Data. *bioRxiv* **2022**, 2006.2011.495736. <https://doi.org/10.1101/2022.06.11.495736> (2022).
24. van Griethuysen, J. J. M. et al. Computational Radiomics System to Decode the Radiographic phenotype. *Cancer Res.* **77** (21), e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339> (2017).
25. Zwanenburg, A. et al. The Image Biomarker Standardization Initiative: standardized quantitative Radiomics for High-Throughput Image-based phenotyping. *Radiology.* **295** (2), 328–338. <https://doi.org/10.1148/radiol.2020191145> (2020).
26. Orlhac, F. et al. How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *Eur. Radiol.* **31** (4), 2272–2280. <https://doi.org/10.1007/s00330-020-07284-9> (2021).
27. Da-Ano, R. et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci. Rep.* **10** (1), 10248. <https://doi.org/10.1038/s41598-020-66110-w> (2020).
28. Korda, A. I. et al. Identification of texture MRI brain abnormalities on first-episode psychosis and clinical high-risk subjects using explainable artificial intelligence. *Transl Psychiatry.* **12** (1), 481. <https://doi.org/10.1038/s41398-022-02242-z> (2022).
29. Shi, Y. et al. Ultrasound-based radiomics XGBoost model to assess the risk of central cervical lymph node metastasis in patients with papillary thyroid carcinoma: individual application of SHAP. *Front. Oncol.* **12**, 897596. <https://doi.org/10.3389/fonc.2022.897596> (2022).
30. Zhang, X. et al. Radiomics Nomogram based on multi-parametric magnetic resonance imaging for predicting early recurrence in small hepatocellular carcinoma after radiofrequency ablation. *Front. Oncol.* **12**, 1013770. <https://doi.org/10.3389/fonc.2022.1013770> (2022).
31. R Core Team R. R: A language and environment for statistical computing. (2013).
32. Treede, R.-D. et al. Chronic pain as a symptom or a disease: the IASP classification of Chronic Pain for the International classification of diseases (ICD-11). *Pain.* **160** (1), 19–27. <https://doi.org/10.1097/j.pain.0000000000001384> (2019).
33. Rahimi, R. et al. Microstructural white matter alterations associated with migraine headaches: a systematic review of diffusion tensor imaging studies. *Brain Imaging Behav.* **16** (5), 2375–2401. <https://doi.org/10.1007/s11682-022-00690-1> (2022).
34. Cao, X. et al. A Radiomics Approach to Predicting Parkinson's disease by incorporating whole-brain functional activity and Gray Matter structure. *Front. Neurosci.* **14**, 751. <https://doi.org/10.3389/fnins.2020.00751> (2020).
35. Huang, K. et al. A multipredictor model to predict the conversion of mild cognitive impairment to Alzheimer's disease by using a predictive nomogram. *Neuropsychopharmacology.* **45** (2), 358–366. <https://doi.org/10.1038/s41386-019-0551-0> (2020).

Acknowledgements

This work was supported by Zhejiang Provincial Key Laboratory of Traditional Chinese Medicine Cultivation for Arthritis Diagnosis and Treatment (grant number: C-2023-W2008). We also would like to express our sincere gratitude to all those who have contributed to this research. We would like to thank AH for their invaluable assistance in collecting and analyzing the data, reviewing the literature, and drafting the manuscript. We would also like to thank ZH for critically revising important intellectual content in the manuscript. Finally, we would like to acknowledge the contributions of all authors who have read and approved the final manuscript.

Author contributions

HY and AH collected and analyzed the data, reviewed the literature, and drafted the manuscript. ZH has critically revised important intellectual content in the manuscript. Final manuscript read and approved by all authors.

Declarations

Competing interests

The authors declare that they have no conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-74418-0>.

Correspondence and requests for materials should be addressed to Z.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024, corrected publication 2025