BMC Medical Imaging

RESEARCH Open Access

Enhancing classification of active and non-active lesions in multiple sclerosis: machine learning models and feature selection techniques

Atefeh Rostami[1,2], Mostafa Robatjazi[1,2*], Amir Dareyni[3], Ali Ramezan Ghorbani[4], Omid Ganji[5], Mahdiye Siyami[6] and Amir Reza Raoofi[7]

Abstract

Introduction Gadolinium-based T1-weighted MRI sequence is the gold standard for the detection of active multiple sclerosis (MS) lesions. The performance of machine learning (ML) and deep learning (DL) models in the classification of active and non-active MS lesions from the T2-weighted MRI images has been investigated in this study.

Methods 107 Features of 75 active and 100 non-active MS lesions were extracted by using SegmentEditor and Radiomics modules of 3D slicer software. Sixteen ML and one sequential DL models were created using the 5-fold cross-validation method and each model with its special optimized parameters trained using the training-validation datasets. Models' performances in test data set were evaluated by metric parameters of accuracy, precision, sensitivity, specificity, AUC, and F1 score.

Results The sequential DL model achieved the highest AUC of 95.60% on the test dataset, demonstrating its superior ability to distinguish between active and non-active plaques. Among traditional ML models, the Hybrid Gradient Boosting Classifier (HGBC) demonstrated a commendable test AUC of 86.75%, while the Gradient Boosting Classifier (GBC) excelled in cross-validation with an AUC of 87.92%.

Conclusion The performance of sixteen ML and one sequential DL models in the classification of active and non-active MS lesions was evaluated. The results of the study highlight the effectiveness of sequential DL approach and ensemble methods in achieving robust predictive performance, underscoring their potential applications in classifying MS plaques.

Keywords Multiple Sclerosis, Active lesions, Machine Learning, Deep Learning, Feature importance

*Correspondence:
Mostafa Robatjazi
Robatjazi1361@gmail.com
[1]Department of Medical Physics and Radiological Sciences, Sabzevar University of Medical Sciences, Sabzevar, Iran
[2]Non-communicable Disease Research Center, Sabzevar University of Medical Sciences, Sabzevar, Iran
[3]Department of Medical Physics and Biomedical Engineering, Tehran University of Medical Sciences, Tehran, Iran
[4]Department of Radiology, Rasoul Akram Hospital, Iran University of Medical Sciences, Tehran, Iran
[5]Department of MRI, Sina Hospital, Tehran University of Medical Sceinces, Tehran, Iran
[6]Student Research Committee, Sabzevar University of Medical Sciences, Sabzevar, Iran
[7]Department of Anatomy, Sabzevar University of Medical Sciences, Sabzevar, Iran

Introduction

Multiple Sclerosis (MS) is a chronic autoimmune disease where the immune system wrongly attacks the central nervous system (CNS) [1, 2]. This immune-mediated disease is one of the most common causes of neurological disability in young adults and affects the quality of life of 2.5 million people worldwide [3]. Weakness, dizziness, spasticity, cognitive changes, difficulty walking, emotional changes, vision problems, and fatigue are the Common symptoms of MS [4–6]. While scientists believe that a combination of genetic and environmental factors, such as vitamin D deficiency, obesity, geography, and smoking, plays crucial roles in the development of multiple sclerosis (MS), the exact causes of the disease remain unknown [7–11].

Magnetic resonance imaging (MRI) is the most important and available paraclinical tool that serves as the best technique to support the diagnosis, differential detection, monitoring, and evaluation of treatment response in MS disease so that, this imaging technique has been considered the gold standard for the accurate detection of active and non-active MS lesions [12–14]. All MS lesions can be detected based on hyperintense areas in T2-weighted, fluid-attenuated inversion recovery (FLAIR), and proton density MR images. However, not all detected lesions based on these images are active [15]. Early prognosis and diagnosis of active MS lesions can improve the outcome of MS disease treatment [16, 17]. Conventionally, T1-weighted gadolinium-based MR images are used to detect active MS lesions which is related to acute focal inflammation and blood-brain barrier (BBB) disruption [18]. However, intravenous administration of contrast agents may lead to side effects, including nephrogenic systemic fibrosis, deposition of gadolinium in various tissues, dizziness, headache, and nausea [19–22]. Furthermore, the detection of MS lesions in various pre-contrast and post-contrast MR images is challenging and time-consuming for physicians [2]. Due to the constraints in the identification of active MS lesions, alternative approaches, such as Artificial Intelligence (AI) models, can be leveraged to streamline the procedures and avoid the need for gadolinium administration in these patients [6, 23–26]. Machine learning (ML) and deep learning (DL), which are two branches of AI, have great potential in this area [27].

Potential applications of ML and DL models with supervised and unsupervised learning in MS disease have been investigated in some studies. The detection of MS using ML and DL models has been evaluated in T2 weighted, FLAIR, diffusion-weighted, and functional MR images [28–31]. Some studies have reported the capability of DL models to differentiate MS diseases from similar conditions observed in MRI images, such as neuromyelitis optica spectrum disorders, migraine, and CNS vasculitis [32–34]. The segmentation of MS lesion and the measurement of brain volumes that is necessary for MS detection have been performed by DL models based on 3D MR images [7, 27, 35]. Another key area of focus in the application of ML and DL models for MS is the classification and differentiation of active (acute or contrast-enhanced) and non-active (chronic or non-enhanced) MS lesions. While there are limited studies in this field, Narayana et al. evaluated the potential of a DL method using the VGG16 network to predict active lesions based on pre-contrast T1-weighted, T2-weighted, and FLAIR MRI data from 1,970 scans with 519 enhanced lesions. They reported a slice-wise prediction sensitivity and specificity of 78% and 73%, respectively [26].

Radiomics is high-throughput extraction of numerous quantitative features from texture of medical images which is mineable for quantitative analysis alongside of ML algorithms [36, 37]. In recent years, radiomics has been increasingly applied to the study of pathological changes, diagnosis, differential diagnosis, and prognosis in MS [31, 38–40]. In the current study, we aim to investigate the application of radiomics data in various AI models for predicting MS active lesions using T2-weighted MRI images.

Table 1 Imaging parameters of MR protocols

Protocol Parameters	T1	T2	FLAIR	T1-enhanced
TE (ms)	10	121	113	10
TR (ms)	485	5340	9000	491
Flip angle (°)	90	150	150	90
Slice thickness (mm)	4.5	4.5	4.5	4.5
Nex	2	2	1	2
Gap (mm)	0.45	0.45	0.45	0.45
Received band width (kHz)	150	160	190	150

Materials and methods

Study population or study participants

A dataset of T1-weighted, T2-weighted, post-contrast T1-weighted, and FLAIR MRI images was utilized in this study. Details of the imaging parameters are presented in Table 1. At the beginning of the study, during the data preparation phase, patients who did not have the specified four image categories were excluded from the analysis. Additionally, lesions in patients who suffered from ischemia were also excluded, as ischemic lesions can be identified in MRI images. In the lesion labeling phase, any lesions that were uncertain in terms of whether they were active or inactive were removed from the segmented areas to ensure accuracy in our evaluations.

All plaques, including 75 active lesions and 100 non-active lesions from 134 patients, were labeled by two experienced radiologists, who worked collaboratively throughout the assessment process. Each radiologist

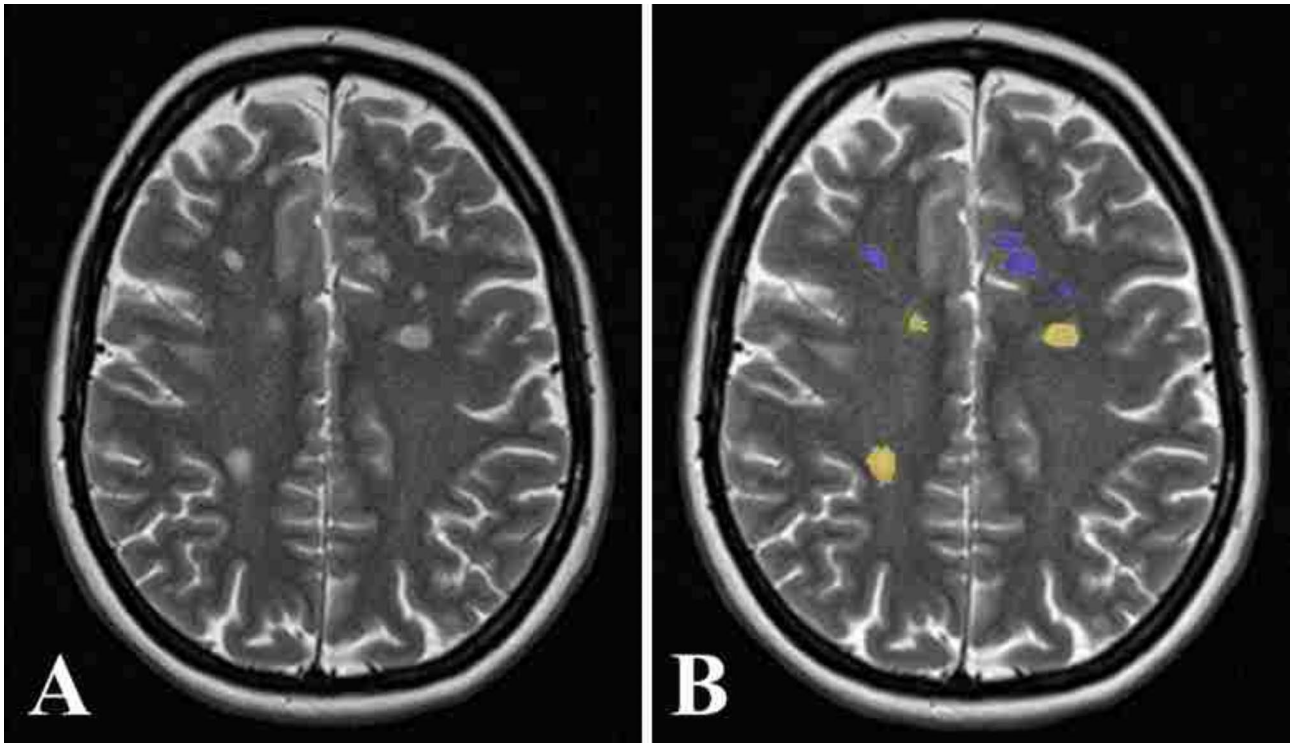

Fig. 1 The segmented active (Yellow) and non-active (Blue) plaques on T2-weighted MRI images using Slicer

independently reviewed the post-contrast T1-weighted images alongside the other non-contrast MR images to identify and label the lesions.

Features extraction

The feature extraction stage is a crucial phase in ML and DL studies, as the quality of the optimized models is heavily dependent on the quality of the input data (extracted features). In this study, T2-weighted MRI images were selected for feature extraction due to their superior quality, characterized by lower noise and a higher signal-to-noise ratio compared to the other available sequences. The T2-weighted MRI volumes were normalized and resampled to a standardized axial orientation using linear interpolation for feature extraction. Quantitative radiomics features were extracted from the active and non-active lesions using the 3D Slicer software (version 4.13). The radiologists identified active and non-active MS lesions based on post-contrast T1-weighted and other non-contrast MR images. The labeled regions were segmented using the SegmentEditor module of the 3D Slicer software on the T2-weighted MR images (As illustrated in Fig. 1). Subsequently, the segmented region of each lesion was converted to a binary label map. The Radiomics module of the 3D Slicer software was then used to extract 107 features using solely the T2-weighted MR images and the corresponding binary label maps. The results of the extracted features for each lesion were saved in separate tables. By combining all the extracted tables, the raw data for training the models was provided. The types of extracted features are reported in Table 2.

Table 2 Extracted features of Radiomics Module of 3D slicer software

PyRadiomics features	Number of extracted features
First Order Statistics	18
Shape-based	14
Gray Level Co-occurence Matrix	24
Gray Level Run Length Matrix	19
Gray Level Size Zone Matrix	13
Neighbouring Gray Tone Difference Matrix	5
Gray Level Dependence Matrix	14
Total	107

Machine learning models

Raw data was checked for null features. Statistical analysis and correlations of various features were evaluated. Finally, all quantitative features scaled for training of ML models.

In this study, the performance of 16 various ML models of LogisticRegression (LR), SGDclassifier (SGD), PassiveAggressiveClassifier (PAC), Supportvectorclassifier (SVC), KNeighborsClassifier (KNN), DecisionTreeClassifier (DTC), RandomForestClassifier (RFC), GradientBoostingClassifier (GBC),

HistGradientBoostingClassifier (HGBC), AdaBoostClassifier (ABC), BaggingClassifier (BC), ExtraTreesClassifier (ETC), GaussianNB (GNB), and LinearDiscriminantAnalysis (LDA), QuadraticDiscriminantAnalysis (QAD) and MLPClassifier (MLPC) were investigated in the prediction of active and non-active MS lesions. Models were developed on randomly drawn 75% of the training-validation data set and its performance was evaluated on the 25% of test data set. First, models have been made using 5-fold cross-validation method to control the learning process of models and optimized their parameters based on the highest accuracy on training data set via GridSearchCV technique. In the second step, each model was trained with its special optimized parameters using training-validation data set. After evaluating the ML performance on the validation dataset the final results were reported based on the testing dataset.

Deep learning model (sequential model with dense layers)

Deep learning studies can be conducted in two primary ways. In traditional neural network approaches, extracted features are input into the model, which is then optimized based on the architecture defined by the number of layers and neurons. Conversely, in image-based deep learning models, feature extraction occurs automatically within the model, without human intervention. In this study, we employed a model utilizing dense layers for the deep learning component, ensuring that the features extracted are of comparable quality to those used in the ML models. In the DL model, the sequential model with 5 dense layers was trained. Based on the numbers of extracted features, 107 neurons were set for the first layer as the dimensional of input data. Details of optimized sequential model are reported in Table 3.

Evaluation of ML and DL models

Designed models were evaluated using independent test data set. Confusion matrix was calculated for each model and metric parameters of accuracy, precision, sensitivity, specificity, AUC, and F1 score were reported. Feature importance for all models was investigated by three methods of coef_, feature_importances_, and permutation_importance.

Results

Descriptive analysis of extracted features

In our analysis, we evaluated a total of 107 radiomic features across two classes. The summary statistics for these features, including minimum, maximum, mean, standard deviation, and correlation with the target variable, are provided in *Appendix A*. Additionally, Fig. 2 presents the correlated heatmaps for these features, highlighting significant relationships and patterns that may inform further analysis.

Based on the evaluation of the correlation of various features with the target variable, the features exhibiting the highest correlations are summarized in Table 4. Notably, the feature 'Median' from the first-order statistics category demonstrated the maximum correlation of 0.42. According to the guidelines for selecting and reporting intraclass correlation coefficients, correlations below 0.5 are considered poor [41]. Consequently, all evaluated features in this study fall into the poor correlation category with respect to the target variable.

Evaluation of ML and DL models

The performance of ML and DL models was assessed using a variety of evaluation metrics, including accuracy, precision, sensitivity, specificity, AUC, and F1 score. Both the test and cross-validation evaluation scores are summarized in Table 5, providing a comprehensive comparison of model effectiveness. The reported evaluation metrics highlights the strengths and weaknesses of each approach, allowing for a comprehensive understanding of their performance in the context of the specific tasks addressed in this study. The results indicate that while DL models generally demonstrate superior performance in handling complex data patterns, certain ML models also achieve competitive results, particularly in scenarios with limited data.

Evaluation of feature importance in the best ML models

In this analysis, we evaluated the significance of various features in predicting outcomes across six ML models that achieved an AUC greater than 90% using test data. Figure 3 shows the AUC curves of these models and the ten most important features for each model are presented in Table 6.

The table highlights the relative importance scores assigned to each feature, indicating their contribution to

Table 3 Parameters of optimized sequential model

Number of layers	Type of layer	Neurons	Trainable parameters	Activation function
1	Dense input layer	107	11,556	selu
2	Dense internal layer	150	16,200	selus
3	Dense internal layer	100	15,100	selu
4	Dense internal layer	50	5050	selu
5	Dense output layer	1	51	sigmoid

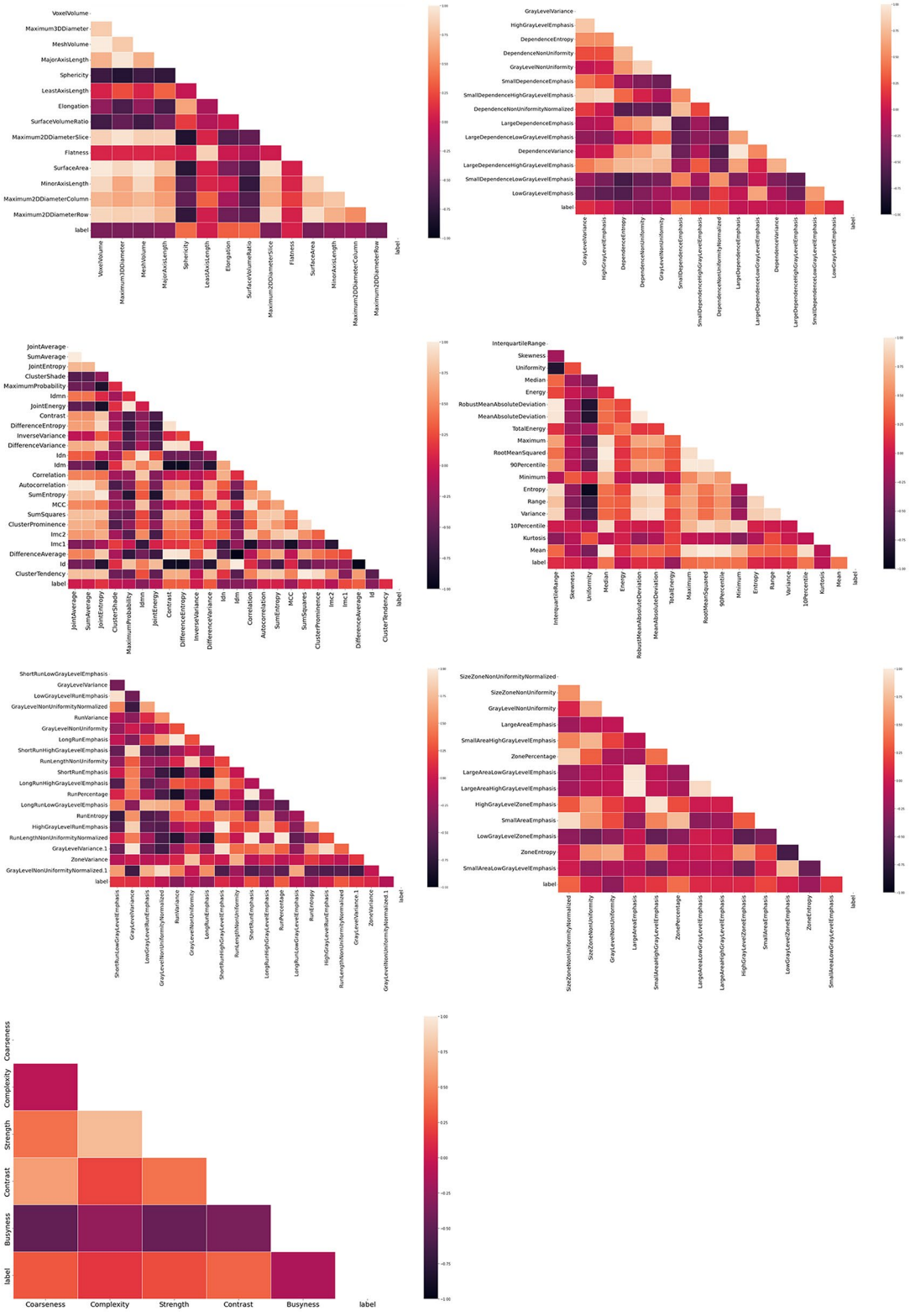

Fig. 2 The correlation heatmaps for each group of features with the target

Table 4 Features with maximum absolute value of correlation with target

Kind of feature	Name of feature	Correlation with target
Shape based	SurfaceArea	-0.39
Gray Level Dependence Matrix	SmallDependenceEmphasis	0.37
Gray Level Co-occurence Matrix	DifferenceAverage	0.33
First order statistics	Median	0.42
Gray Level Run Length Matrix	RunPercentage	0.31
	RunVariance	-0.31
Gray Level Size Zone Matrix	Zone%	0.39
Neighbouring Gray Tone Difference Matrix	contrast	0.33

the model's predictive performance. For instance, in the LR model, the feature "Surface Volume Ratio" emerged as the most influential, with a score of 13.86, followed closely by "Size Zone Non-Uniformity" and "Sphericity." Similarly, the PAC model identified "Elongation" as its top feature, with an importance score of 1.98.

Other models, such as the SGD and SVM models, also exhibited distinct feature priorities. The SGD model ranked "Elongation" and "Sphericity" as the most critical features, while the SVM model highlighted "Dependence Non-Uniformity Normalized" and "Surface Volume Ratio" as key predictors.

This comprehensive evaluation not only underscores the variability in feature importance across different models but also emphasizes the need for tailored feature selection strategies in ML applications. By identifying and leveraging the most impactful features, we can enhance the predictive accuracy and interpretability of these models, ultimately leading to more informed decision-making in the respective domain.

Discussion

Early diagnosis of MS lesions can significantly enhance treatment procedures. Traditionally, gadolinium-based T1-weighted MR images have been utilized for detecting active MS lesions; however, the use of contrast agents presents several limitations and potential side effects in certain cases. Therefore, detection using non-contrast images can improve patient safety and reduce clinical costs. ML and DL algorithms have been widely applied in the early detection of various diseases. In this study, we explored the potential of utilizing radiomics-based AI models to evaluate the classification of MS lesions on T2-wighted MRI images. We developed and assessed 16 ML models and one sequential DL model.

Texture analysis plays a crucial role in the field of radiomics, which focuses on extracting a large number of quantitative features from medical images. Texture describes the spatial arrangement of basic attributes on a surface, such as the structure of a brick wall or the grains within a brick. In medical imaging, these attributes are represented by image pixels, where texture reflects the visual characteristics of the image, which can range from coarse to fine and uniform to irregular.

In practice, texture analysis produces a set of parameters that encapsulate the image's content, aiding in detection and classification, which can be performed on radiomics data analysis using ML and DL algorithms. In MRI of MS lesions, it is believed that the gray level distribution within lesions reflects the ultra-structural properties of affected tissues, influenced by disease processes and treatments. This concept has been supported by recent histopathological studies of hyper-intense white matter lesions on T2-weighted MR images [42–45]. In this way, Nicolas Michoux et al. evaluated texture analysis of T2-weighted MR images as an alternative method for assessing acute inflammation in MS lesions. The study involved 21 MS patients who underwent various MR imaging sequences. Significant differences in texture parameters were found between MS lesions and normal appearing white matter, indicating texture analysis's effectiveness in differentiation. Multi-parametric models using classifiers like LR achieved a sensitivity of 86% and specificity of 84% in identifying contrast-enhanced lesions. Their findings suggest that T2-weighted texture parameters can accurately assess brain inflammatory activity, offering a promising alternative to traditional contrast-enhanced imaging methods [42].

Peng Yuling et al. investigated the use of radiomics-based ML models to predict the evolution of unenhanced lesions in MS using follow-up MRI scans. The research involved 36 MS patients and analyzed 45 pairs of baseline and follow-up MR images, extracting 972 radiomic features from FLAIR images. The lesions were categorized based on volume changes into "interval activity" (enlarged lesions) and "interval inactivity" (shrunk lesions). Various ML classifiers, including LR, random forest, and SVM, were employed, with feature selection methods like ReliefF, LASSO, and recursive feature elimination (RFE). In their study, the SVM classifier with the ReliefF algorithm demonstrated the best predictive performance, achieving an accuracy of 82.7% and an AUC of 86%. Their findings suggest that radiomics can effectively assess lesion activity in MS, providing valuable insights

Table 5 Evaluation metrics of 16 ML and one DL optimized models using all 107 features for test and cross-validation data

		Accuracy		Precision		Sensitivity		Specificity		AUC		F1 score	
Category of models	**Models**	**Test**	**CV**	**Test**	**CV**	**Test**	**CV**	**Test**	**CV**	**Test**	**CV**	**Test**	**CV**
Linear	LR	90.90	80.19	94.44	87.66	85.00	63.64	95.83	92.08	94.16	77.85	89.47	72.21
	PAC	84.09	79.45	88.24	83.21	75.00	69.09	91.67	86.67	93.33	77.88	81.08	71.86
	SGD	86.36	75.58	85	85.28	85.00	52.73	87.50	92.17	95.00	72.45	85.00	62.15
SVM	SVC	81.82	80.14	92.86	84.40	65.00	67.27	95.83	89.33	93.54	78.30	76.47	73.75
Neighbors	KNN	75.00	75.64	90.91	0.91	50.00	47.27	95.83	96.08	88.65	71.68	64.51	59.31
Tree	DTC	77.27	76.35	81.25	71.33	65.00	74.55	87.5	77.67	77.40	76.11	72.22	72.80
Ensemble	RFC	81.82	68.83	87.5	69.20	70.00	54.55	91.67	79.17	85.94	66.86	77.78	59.45
	GBC	81.82	75.53	87.5	76.87	70.00	61.828.	91.67	85.42	87.92	73.62	67.78	67.81
	HGBC	92.94	79.40	80.00	80.53	26.67	69.09	99.35	86.92	86.75	78.00	40.00	73.51
	ABC	79.55	76.27	86.67	76.22	65.00	64.45	91.67	84.03	87.29	74.77	74.29	69.17
	BC	81.82	77.07	87.5	81.39	70.00	0.60	91.67	89.72	87.19	74.71	77.78	67.92
	ETC	81.82	76.35	85.5	75.11	70.00	64.45	91.67	84.25	86.88	74.85	77.78	69.61
Neural Network	MLPC	84.09	79.40	93.33	79.78	70.00	70.90	95.83	85.42	95.00	78.16	80.00	74.21
Naive Bayes	GNB	45.45	62.70	91.67	56.25	55.00	63.64	95.83	62	85.62	62.82	68.74	58.40
Discriminant Analysis	LDA	86.36	82.48	93.75	89.06	75.00	67.27	95.83	93.33	92.70	80.30	83.33	75.24
	QDA	63.64	52.73	70.00	42.93	35.00	41.82	87.50	61.5	61.25	51.16	46.67	42.21
Deep Learning Sequential model	Dense	91.00	91.26	94.44	91.30	85.00	75.82	95.83	95.69	95.60	84.16	89.47	85.80

Abbreviation: Logistic Regression (LR), Passive Aggressive Classifier (PAC), Stochastic Gradient Decent (SGD), Support Vector Classifier (SVC), K-Nearest Neighbors (KNN), Decision Tree Classifier (DTC), Random Forest Classifier (RFC), Gradient Boosting Classifier (GBC), Hist Gradient Boosting Classifier (HGBC), AdaBoost Classifier (ABC), Bagging Classifier (BC), Extra Trees Classifier (ETC), Multi-Layer Perception Classifier (MLPC), Gaussian Naive Bayes (GNB), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA).

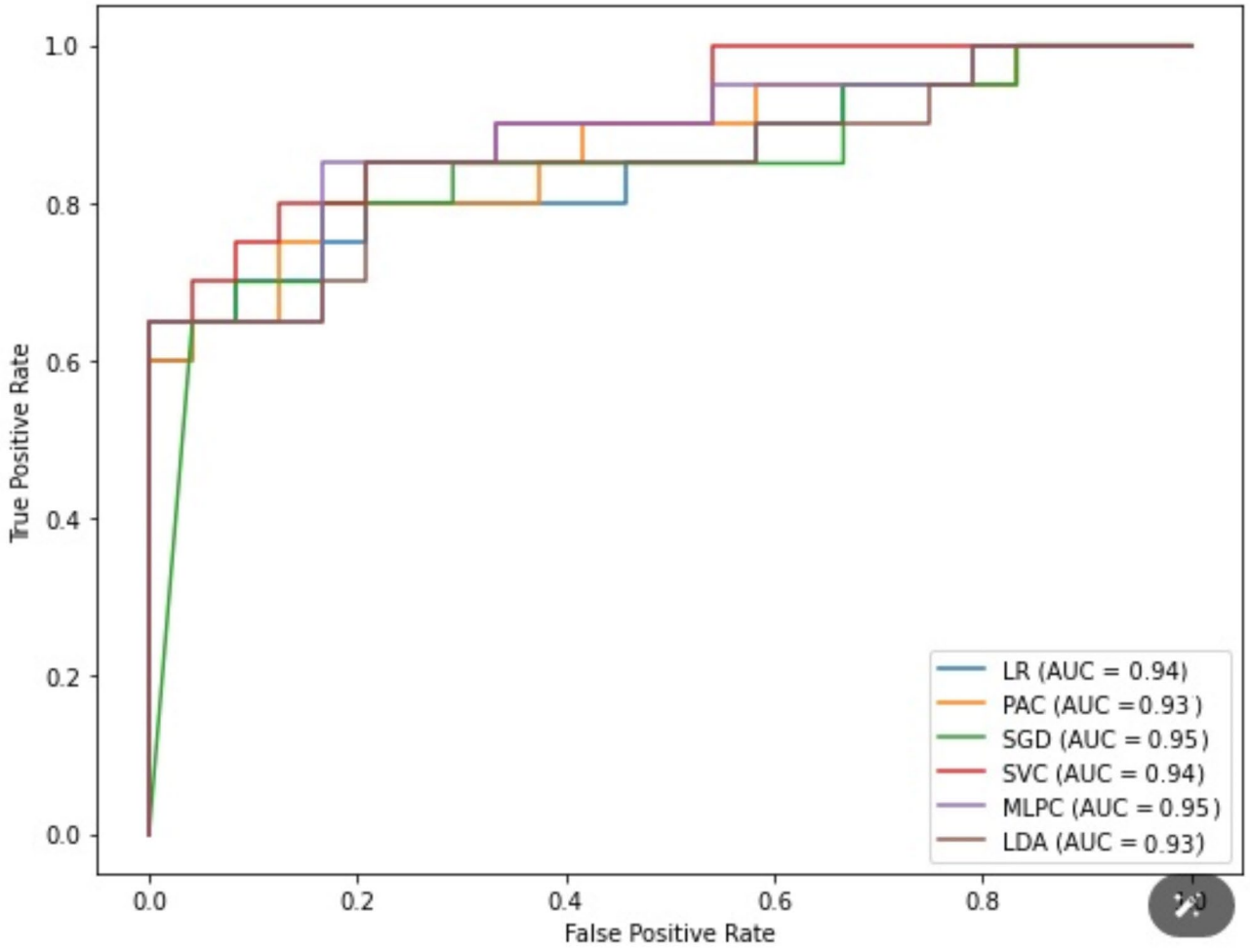

Fig. 3 AUC curve for the six models which was greater than 90%

for monitoring disease progression and guiding clinical treatment [46].

In our study, we analyzed approximately 175 MS plaques, extracting 107 radiomic features to evaluate their predictive capabilities. According to the metrics presented in Table 5 by using test data, the highest accuracy of 92.94% was achieved with the HGBC ML model. For precision, the LR and DL models attained a maximum value of 94.44%. Sensitivity was maximized at 85.35% in the LR, SGD ML, and DL models. Notably, the HGBC model demonstrated an impressive specificity of 99.35%. The DL model excelled with a maximum AUC of 95.60% and an F1 score of 89.47%.

Our findings indicate that the sequential DL model outperformed others in predicting active and non-active MS lesions based on the 107 extracted features, showcasing superior accuracy, precision, specificity, and AUC values [47]. The sensitivity and F1 score for the DL model were also robust, falling within the range of 80–90%, which reflects its reliability in clinical applications.

When comparing our results to previous studies, such as those utilizing texture analysis as a quantitative approach, our findings align with the notion that advanced ML techniques can enhance diagnostic accuracy [42, 46]. Similar to the literature, which highlights the significance of texture features in differentiating MS lesions from other conditions, our study reinforces the idea that a comprehensive analysis of radiomic features can lead to improved detection and characterization of MS lesions. The DL model's performance further supports the growing body of evidence advocating for the integration of radiomics and AI in the clinical assessment of MS.

One notable study by Narayana et al. explored the efficacy of a DL method using the VGG16 network to predict active lesions based on pre-contrast T1-weighted, T2-weighted, and FLAIR MRI images. They reported a sensitivity of 78% and a specificity of 73% using a slice-wise prediction approach [26]. In contrast, our study achieved superior accuracy across all evaluation metrics, with results exceeding 85%.

Table 6 Feature importance of ML models

LR model		**PAC model**	
SurfaceVolumeRatio	13.86	Elongation	1.98
SizeZoneNonUniformity	9.40	SurfaceVolumeRatio	1.74
Sphericity	8.91	Skewness	1.48
SmallAreaLowGrayLevelEmphasis	7.21	Sphericity	1.42
Median	7.17	SizeZoneNonUniformity	1.40
Range	7.08	Maximum	1.25
GrayLevelVariance	6.32	SmallAreaLowGrayLevelEmphasis	1.22
SmallAreaHighGrayLevelEmphasis	5.35	10Percentile	1.11
GrayLevelNonUniformity	5.29	SmallDependenceLowGrayLevelEmphasis	1.06
Maximum	5.00	SmallAreaHighGrayLevelEmphasis	1.00
SGD model		**SVM model**	
Elongation	43.47	DependenceNonUniformityNormalized	0.016
Sphericity	40.30	SurfaceVolumeRatio	0.012
10Percentile	26.93	ShortRunEmphasis	0.010
Median	26.90	RunLengthNonUniformityNormalized	0.010
Mean	25.08	Maximum3DDiameter	0.007
Minimum	24.81	Elongation	0.007
RootMeanSquared	24.05	RunEntropy	0.007
Zone%	23.15	Sphericity	0.006
SurfaceVolumeRatio	20.60	Maximum2DDiameterSlice	0.006
LargeDependenceEmphasis	19.62	SurfaceArea	0.004
MPLC model		**DA model**	
SmallAreaLowGrayLevelEmphasis	0.024	TotalEnergy	0.02290
Elongation	0.021	10Percentile	0.02290
Coarseness	0.019	LongRunHighGrayLevelEmphasis	0.02137
SurfaceVolumeRatio	0.018	SumAverage	0.01985
SmallDependenceLowGrayLevelEmphasis	0.016	LargeDependenceEmphasis	0.01527
Median	0.015	SurfaceArea	0.01374
Maximum	0.015	Minimum	0.00916
DependenceNonUniformityNormalized	0.013	SizeZoneNonUniformity	0.00916
Correlation	0.013	Maximum	0.00763
10Percentile	0.013	Maximum3DDiameter	0.00611

Table 6 highlights the ten most important features identified across six ML models, all of which achieved an AUC greater than 90%. Notably, features such as Maximum, Elongation, Sphericity, Surface-to-Volume Ratio, and Median consistently appeared among the key features across different models. This consistency underscores the relevance of these features in effectively characterizing MS lesions and enhances the robustness of our findings.

In our analysis, we observed that all 107 evaluated features exhibited correlation coefficients below 0.5 with the target variable, suggesting limited predictive ability. However, it is crucial to understand that ML and DL models do not solely rely on correlation coefficients to establish feature relationships. These models can leverage combinations of features to enhance predictive performance, as demonstrated by the results presented in this study. While Table 4 outlines the correlation coefficients, Table 6 highlights the importance of features from the ML perspective, revealing that high correlation does not necessarily imply greater significance in model performance.

There are limitations in the current study. The relatively small number of MRI images and primary data may impact the generalizability of our findings. Our analysis was conducted using images from a single MRI center, which employed specific imaging protocols. While this approach ensures consistency, it may limit the applicability of our results to broader clinical settings. To address concerns regarding overfitting associated with using data from a single center, we implemented rigorous cross-validation techniques throughout our analysis. However, increasing the dataset size and utilizing multicenter data could further enhance the performance of our models in predicting active and non-active MS lesions. Furthermore, the use of 2D images for analysis instead of 3D images can affect the results and is also considered as a limitation in our study. The labeling and feature extraction were conducted on 2D slices, which may restrict the contextual information available for accurate lesion

characterization. Employing advanced MRI sequences and 3D imaging protocols could significantly improve the detection capabilities for active MS lesions. Future research should focus on expanding the dataset and incorporating multicenter data to validate our findings and enhance their generalizability.

Conclusion

In conclusion, our study demonstrates the promising potential of radiomics and AI models in classifying of active and non-active MS lesions using T2-weighted MRI images. By extracting 107 quantitative features from a dataset of 175 MS plaques, we achieved significant accuracy and robustness in our predictions, with the DL model outperforming traditional methods. The findings highlight the importance of texture analysis and the role of specific features, such as Maximum, Elongation, and Sphericity, in effectively distinguishing between lesion types.

The results underscore the feasibility of utilizing non-contrast imaging techniques, which can enhance patient safety and reduce reliance on gadolinium-based contrast agents. Ultimately, the integration of AI in the assessment of MS lesions holds significant promise for improving diagnostic accuracy, informing treatment decisions, and enhancing patient outcomes in the management of MS.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-024-01528-6.

Supplementary Material 1

Acknowledgements
The authors sincerely thank from Vasei Clinical Research Development Unit at Sabzevar University of Medical Sciences, for providing advice and guidance in conducting this research.

Author contributions
Atefeh Rostami: Conceptualization, Methodology, Formal analysis, Resource, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administrationMostafa Robatjazi: Conceptualization, Methodology, Resource, Writing - Review & Editing, Supervision, Project administrationAmir Dareyni: Methodology, Software, Validation, Investigation, Data CurationAli Ramezan Ghorbani: Resource, Writing - Review & Editing, Supervision, Project administrationOmid Ganji: Resource, Writing - Review & EditingMahdiye Siyami: Methodology, Software, Investigation, Data CurationAmir Reza Raoofi: Methodology, Resource, Writing - Review & Editing.

Funding
This work was fully supported by Sabzevar University of Medical Sciences (grant number 401220).

Data availability
Due to ethical and legal concerns, the data and materials used in this study will not be publicly available. However, requests for access to the data and materials may be made by the corresponding author. We will make every effort to provide access to the data and materials in a manner that is consistent with ethical and legal standards.

Declarations

Ethics approval and consent to participate
This study was approved by the Ethics Committee of Sabzevar University of Medical Sciences (IR.MEDSAB.REC.1402.032). Obtaining any informed consent was waived by the Ethics Committee of Sabzevar University of Medical Sciences. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication
Not Applicable.

Competing interests
The authors declare no competing interests.

Received: 22 August 2024 / Accepted: 11 December 2024
Published online: 20 December 2024

References

1. Bonacchi R, Filippi M, Rocca MA. Role of artificial intelligence in MS clinical practice. NeuroImage: Clin. 2022;35:103065.
2. Shoeibi A, Khodatars M, Jafari M, Moridian P, Rezaei M, Alizadehsani R, et al. Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review. Comput Biol Med. 2021;136:104697.
3. Tullman MJ. Overview of the epidemiology, diagnosis, and disease progression associated with multiple sclerosis. Am J Manag Care. 2013;19(2 Suppl):S15–20.
4. Calabresi PA. Diagnosis and management of multiple sclerosis. Am Family Phys. 2004;70(10):1935–44.
5. Zhao Y, Healy BC, Rotstein D, Guttmann CR, Bakshi R, Weiner HL, et al. Exploration of machine learning techniques in predicting multiple sclerosis disease course. PLoS ONE. 2017;12(4):e0174866.
6. Eshaghi A, Young AL, Wijeratne PA, Prados F, Arnold DL, Narayanan S, et al. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. Nat Commun. 2021;12(1):2078.
7. García-Lorenzo D, Francis S, Narayanan S, Arnold DL, Collins DL. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. Med Image Anal. 2013;17(1):1–18.
8. Hartmann M, Fenton N, Dobson R. Current review and next steps for artificial intelligence in multiple sclerosis risk research. Comput Biol Med. 2021;132:104337.
9. Oksenberg JR, Begovich AB, Erlich HA, Steinman L. Genetic factors in multiple sclerosis. JAMA. 1993;270(19):2362–9.
10. Ascherio A. Environmental factors in multiple sclerosis. Expert Rev Neurother. 2013;13(sup2):3–9.
11. Milo R, Kahana E. Multiple sclerosis: geoepidemiology, genetics and the environment. Autoimmun rev. 2010;9(5):A387–94.
12. Filippi M, Preziosa P, Barkhof F, Chard DT, De Stefano N, Fox RJ, et al. Diagnosis of progressive multiple sclerosis from the imaging perspective: a review. JAMA Neurol. 2021;78(3):351–64.
13. Sastre-Garriga J, Pareto D, Battaglini M, Rocca MA, Ciccarelli O, Enzinger C, et al. MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. Nat Reviews Neurol. 2020;16(3):171–82.
14. Wattjes MP, Ciccarelli O, Reich DS, Banwell B, de Stefano N, Enzinger C, et al. 2021 MAGNIMS–CMSC–NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. Lancet Neurol. 2021;20(8):653–70.
15. Lohrke J, Frenzel T, Endrikat J, Alves FC, Grist TM, Law M, et al. 25 years of contrast-enhanced MRI: developments, current challenges and future perspectives. Adv therapy. 2016;33:1–28.
16. Gajofatto A, Benedetti MD. Treatment strategies for multiple sclerosis: when to start, when to change, when to stop? World J Clin Cases: WJCC. 2015;3(7):545.
17. Scalfari A, Knappertz V, Cutter G, Goodin DS, Ashton R, Ebers GC. Mortality in patients with multiple sclerosis. Neurology. 2013;81(2):184–92.
18. Absinta M, Sati P, Reich DS. Advanced MRI and staging of multiple sclerosis lesions. Nat Reviews Neurol. 2016;12(6):358–68.

19. Fraum TJ, Ludwig DR, Bashir MR, Fowler KJ. Gadolinium-based contrast agents: a comprehensive risk assessment. J Magn Reson Imaging. 2017;46(2):338–53.
20. Kanal E. Gadolinium-based contrast agents: the plot thickens. Radiological Society of North America; 2017. pp. 340–2.
21. Kanda T, Ishii K, Kawaguchi H, Kitajima K, Takenaka D. High signal intensity in the dentate nucleus and globus pallidus on unenhanced T1-weighted MR images: relationship with increasing cumulative dose of a gadolinium-based contrast material. Radiology. 2014;270(3):834–41.
22. McDonald RJ, McDonald JS, Dai D, Schroeder D, Jentoft ME, Murray DL, et al. Comparison of gadolinium concentrations within multiple rat organs after intravenous administration of linear versus macrocyclic gadolinium chelates. Radiology. 2017;285(2):536–45.
23. Storelli L, Azzimonti M, Gueye M, Vizzino C, Preziosa P, Tedeschi G, et al. A deep learning approach to predicting disease progression in multiple sclerosis using magnetic resonance imaging. Invest Radiol. 2022;57(7):423–32.
24. Al Jannat S, Hoque T, Supti NA, Alam MA, editors. Detection of multiple sclerosis using deep learning. 2021 Asian conference on innovation in technology (ASIANCON); 2021: IEEE.
25. Acar ZY, Başçiftçi F, Ekmekci AH. A Convolutional Neural Network model for identifying Multiple Sclerosis on brain FLAIR MRI. Sustainable Computing: Inf Syst. 2022;35:100706.
26. Narayana PA, Coronado I, Sujit SJ, Wolinsky JS, Lublin FD, Gabr RE. Deep learning for predicting enhancing lesions in multiple sclerosis from noncontrast MRI. Radiology. 2020;294(2):398–404.
27. Afzal HR, Luo S, Ramadan S, Lechner-Scott J. The emerging role of artificial intelligence in multiple sclerosis imaging. Multiple Scler J. 2022;28(6):849–58.
28. Wang S-H, Tang C, Sun J, Yang J, Huang C, Phillips P, et al. Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. Front NeuroSci. 2018;12:818.
29. Zhang Y-D, Pan C, Sun J, Tang C. Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU. J Comput Sci. 2018;28:1–10.
30. Eitel F, Soehler E, Bellmann-Strobl J, Brandt AU, Ruprecht K, Giess RM, et al. Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. NeuroImage: Clin. 2019;24:102003.
31. Zurita M, Montalba C, Labbé T, Cruz JP, da Rocha JD, Tejos C, et al. Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data. NeuroImage: Clin. 2018;20:724–30.
32. Kim H, Lee Y, Kim Y-H, Lim Y-M, Lee JS, Woo J, et al. Deep learning-based method to differentiate neuromyelitis optica spectrum disorder from multiple sclerosis. Front Neurol. 2020;11:599042.
33. Roca P, Attye A, Colas L, Tucholka A, Rubini P, Cackowski S, et al. Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI. Diagn Interv Imaging. 2020;101(12):795–802.
34. Cacciaguerra L, Storelli L, Rocca MA, Filippi M. Current and future applications of artificial intelligence in multiple sclerosis. Augmenting Neurological Disorder Prediction and Rehabilitation Using Artificial Intelligence. Elsevier; 2022. pp. 107–44.
35. Danelakis A, Theoharis T, Verganelakis DA. Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. Comput Med Imaging Graph. 2018;70:83–100.
36. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5(1):4006.
37. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563–77.
38. Caruana G, Pessini LM, Cannella R, Salvaggio G, de Barros A, Salerno A, et al. Texture analysis in susceptibility-weighted imaging may be useful to differentiate acute from chronic multiple sclerosis lesions. Eur Radiol. 2020;30:6348–56.
39. Luo T, Oladosu O, Rawji KS, Zhai P, Pridham G, Hossain S, et al. Characterizing structural changes with evolving remyelination following experimental demyelination using high angular resolution diffusion MRI and texture analysis. J Magn Reson Imaging. 2019;49(6):1750–9.
40. Verma RK, Wiest R, Locher C, Heldner MR, Schucht P, Raabe A, et al. Differentiating enhancing multiple sclerosis lesions, glioblastoma, and lymphoma with dynamic texture parameters analysis (DTPA): a feasibility study. Med Phys. 2017;44(8):4000–8.
41. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63.
42. Michoux N, Guillet A, Rommel D, Mazzamuto G, Sindic C, Duprez T. Texture analysis of T2-weighted MR images to assess acute inflammation in brain MS lesions. PLoS ONE. 2015;10(12):e0145497.
43. Rovira À, Wattjes MP. Gadolinium should always be used to assess disease activity in MS–No. Multiple Scler J. 2020;26(7):767–9.
44. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Trans Syst man cybernetics. 1973;6:610–21.
45. Zhang Y, Moore GW, Laule C, Bjarnason TA, Kozlowski P, Traboulsee A, et al. Pathological correlates of magnetic resonance imaging texture heterogeneity in multiple sclerosis. Ann Neurol. 2013;74(1):91–9.
46. Peng Y, Zheng Y, Tan Z, Liu J, Xiang Y, Liu H, et al. Prediction of unenhanced lesion evolution in multiple sclerosis using radiomics-based models: a machine learning approach. Multiple Scler Relat disorders. 2021;53:102989.
47. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med. 2013;4(2):627.

Publisher's note