Investigating the Ability of Radiomics Features for Diagnosis of the Active Plaque of Multiple Sclerosis Patients

Hassan Tavakoli (PhD)[1,2,3*], Gila Pirzad Jahromi (PhD)[1], Abdolrasoul Sedaghat (MD)[4]

[1]Neuroscience Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran
[2]Radiation Injuries Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran
[3]Department of Physiology and Biophysics, Baqiyatallah University of Medical Sciences, Tehran, Iran
[4]Department of Radiology, Karaj Central Medical Imaging Institute, Karaj, Alborz, Iran

ABSTRACT

Background: Multiple sclerosis (MS) is the most common non-traumatic disabling disease.

Objective: The aim of this study is to investigate the ability of radiomics features for diagnosing active plaques in patients with MS from T_2 Fluid Attenuated Inversion Recovery (FLAIR) images.

Material and Methods: In this experimental study, images of 82 patients with 122 MS lesions were investigated. Boruta and Relief algorithms were used for feature selection on the train data set (70%). Four different classifier algorithms, including Multi-Layer Perceptron (MLP), Gradient Boosting (GB), Decision Tree (DT), and Extreme Gradient Boosting (XGB) were used as classifiers for modeling. Finally, Performance metrics were obtained on the test data set (30%) with 1000 bootstrap and 95% confidence intervals (95% CIs).

Results: A total of 107 radiomics features were extracted for each lesion, of which 7 and 8 features were selected by the Relief method and Boruta method, respectively. DT classifier had the best performance in the two feature selection algorithms. The best performance on the test data set was related to Boruta-DT with an average accuracy of 0.86, sensitivity of 1.00, specificity of 0.84, and Area Under the Curve (AUC) of 0.92 (95% CI: 0.92-0.92).

Conclusion: Radiomics features have the potential for diagnosing MS active plaque by T_2 FLAIR image features. Additionally, choosing the feature selection and classifier algorithms plays an important role in the diagnosis of active plaque in MS patients. The radiomics-based predictive models predict active lesions accurately and non-invasively.

Citation: Tavakoli H, Pirzad Jahromi G, Sedaghat A. Investigating the Ability of Radiomics Features for Diagnosis of the Active Plaque of Multiple Sclerosis Patients. *J Biomed Phys Eng*. 2023;13(5):421-432. doi: 10.31661/jbpe.v0i0.2302-1597.

Keywords

Multiple Sclerosis; Machine Learning; Magnetic Resonance Imaging; Radiomics; Flair

Introduction

Multiple sclerosis (MS) is the most common non-traumatic disabling disease [1], showing a growing prevalence in developing and advanced countries [2]. MS is affected by various factors, such as genetics and environmental factors, vitamin D deficiency, Epstein-Barr virus infection, obesity, and smoking [3].

Magnetic resonance imaging (MRI) technique is increasingly used in research and clinical fields of MS [4]. This imaging method is used as the gold standard in the diagnosis of MS plaques in the clinic. MRI-

*Corresponding author:
Hassan Tavakoli
Department of Physiology and Biophysics, Baqiyatallah University of Medical Sciences, Tehran, Iran
E-mail: tavakoli6402@gmail.com

Received: 13 February 2023
Accepted: 5 March 2023

based studies in MS patients with central nervous system atrophy and plaques have provided valuable information about this disease for physicians in the last two decades [5, 6]. This information has not only led to the creation of a framework for the diagnosis and management of this disease in the clinic, but has also been widely used in studies on the mechanism of this disease [5]. Special protocols with specific sequences are used to image the nervous system in patients with MS, including T_2 weight images. In these images, the location of the plaque on the white matter is marked as points with high signals relative to the surrounding points. Fluid Attenuated Inversion Recovery (FLAIR) sequences are also used to examine and view periventricular plaques that may not be visible in T_2 images due to the high cerebrospinal fluid signal. Other images that are valuable in MS studies are T_1 weight images collected after intravenous injection of a contrast agent. Violation of the blood–brain barrier, which is associated with active inflammation, is assessed by injecting a contrast agent 5 minutes before T_1 imaging. In these images, active plaques show an increased signal by drug uptake. Other sequences that are uncommon and requested in special cases include Diffusion-weighted Imaging (DWI), Magnetization Transfer Imaging (MTI), and Magnetic Resonance Spectroscopy (MRS) sequences [7, 8].

Radiomics is an emerging branch of image processing to link the qualitative and quantitative information extracted from medical images [9]. This science is also based on the fact that all anatomical and functional images in medicine contain qualitative and quantitative information related to pathophysiology [10]. Quantitative radiomics information can be extracted from images using special software. Therefore, the quality of images and the accuracy in selecting parameters when processing images play an important role in the quality and acceptability of the extracted data. Obviously, the quality of the quantitatively extracted features, the relationship between these features and clinical observations, as well as the model obtained from this data, will be affected by the type of image collection, image processing, and segmentation of the target areas [10, 11].

Michoux et al. [12] examined the possibility of detecting active MS plaque in patient's T_2-weighted images. In that study, they used 21 MS patients who were imaged with the same protocol and the same device to extract quantitative features of plaque area and normal white tissue area. The results showed that the use of different models, such as Linear Discriminant Analysis (LDA), Partial Least Squares (PLS), and Logistic Regression (LR) identifies the quantitative parameters that distinguish normal brain white tissue from MS plaques with 76 to 88% accuracy.

Zhang et al. [13] examined the difference between the texture of MRI images in the area of plaques with absorption of contrast material (active plaque) and chronic plaques (inactive) and normal areas of brain tissue. The group's research showed that the texture of the images in the areas of active plaque is more coarseness than in the areas of inactive plaques and normal brain tissue. In the conclusions of that study, it is stated that one of the signs of absorption of the contrast agent by the plaque is the change of texture of the images from soft to coarse mode.

Yulinga et al. [14] showed that the radiomics-based machine learning model has the potential to predict MS lesions and investigated 135 inactive and 110 active plaques. In that study, they used Support-vector Machine (SVM), Logistic Regression (LR), and Random Forest (RF) machine learning classifiers. The results showed that the SVM classifier had the best performance with Area Under the Curve (AUC) equal to 0.85.

To the best of our knowledge, MRI with Gadolinium contrast injection is the only method for distinguishing between active and inactive MS plaques in the clinic. On the other

hand, studies have shown that an increased dose of gadolinium (up to three times the normal dose) should be used to maximize the appearance of new MS plaques; additionally, a 20- to 30-minute interval should be between the injection time and the start of collecting T_1 weight sequences [15]. Apart from discussing the cost of preparing the contrast material as well as increasing the imaging time, the interpretation of the images is visual and can be prone to error under different conditions. Therefore, it is important to find a quantitative method to accurately distinguish plaque type from T_2 images in terms of activity.

The aim of this study is to investigate the possibility of diagnosing active plaques in patients with MS from T_2 FLAIR images before contrast injection using radiomics features of lesions and machine learning algorithms.

Material and Methods

This study is an experimental study. The workflow of the current study is presented in Figure 1.

Patients' specification

In this study, images of 82 MS patients were used for periodic MRI imaging examinations. Out of all the studied patients, 16 (13.1%) and 106 (86.9%) patients had active and inactive plaques, respectively. The standard imaging protocol for patients included T_2, T_2 FLAIR, and T_1 with and without contrast agents. According to the standard protocol, there were at least 5 minutes between the injection of the contrast agent and the imaging of T_1 sequences. Images of all patients were taken with a 1.5 T GE signa explorer scanner (GE Healthcare, Milwaukee, USA) with the same protocol. T_2 FLAIR images were obtained with the following condition: slice thickness: 1.2 mm, pixel size: 1×1 mm^2, matrix size: 256×256 mm^2, NEX: 1, TE: 124 ms, and TR: 6500 ms. T_1 images were obtained with the following condition: slice thickness: 1.2 mm, pixel size: 1×1 mm^2, matrix size: 256×256 mm^2, NEX: 1, TE: 12 ms, and TR: 500.

Lesion segmentation

Region of Interest (ROI) must first be specified in all slices to extract the quantitative features of the images. For this purpose, the segment editor tool was used in the 3D Slicer software (version 4.10.0, Harvard University, National Institutes of Health). In the T_2 FLAIR images, the areas of the MS plaque had a completely defined margin relative to the surrounding texture and were segmented in consecutive slices. It is worth noting that T_2 FLAIR images were collected in volume (3D), leading to examining the segmentation of the desired areas in different anatomical

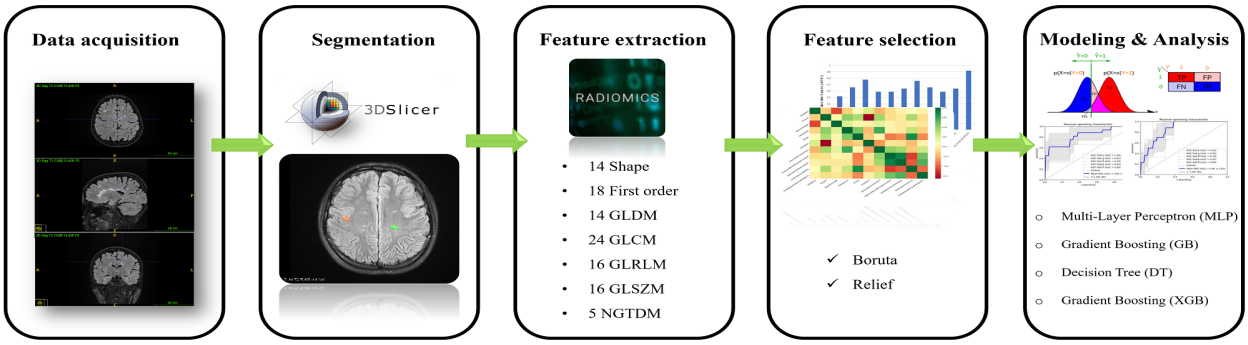


Figure 1: Schematic design of workflow in this study
Gray Level Dependence Matrix (GLDM), Gray Level Co-occurrence Matrix (GLCM), first order, Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Neighborhood Gray Tone Difference Matrix (NGTDM).

views (coronal, sagittal, and axial). In this section, 122 lesions were segmented by a specialist radiologist with over twenty years of experience.

Feature extraction

For the feature extraction, the software used in this study was 3D Slicer (version 4.10.0). A total of 107 features were extracted using the PyRadiomics module of this software [16]. The features type included shape, Gray Level Dependence Matrix (GLDM), Gray Level Co-occurrence Matrix (GLCM), first order, Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Neighborhood Gray Tone Difference Matrix (NGTDM).

Features selection

The number of features extracted from the images was very large, given that many of these features were redundant and should be removed. There are different algorithms for feature selection and in this study Boruta, and Relief algorithms were used. Boruta's algorithm is a powerful and recently introduced feature selection method, which trained a random forest classifier on a duplicate dataset (including original features and shadows) and mark an object as important by comparing its Z-score with the duplicate score [17, 18]. As a method of selecting individual rating filtering features, Relief calculates proxy statistics for each feature that can be used to estimate the "quality" or "relevance" of that feature to the concept target (i.e., predict the value of the endpoint). These feature statistics are called feature weights (W [A]=feature weight "A"), or more simply "feature score" can range from −1 (worst) to +1 (best) [19]. Feature selection was performed by R software (version 3.5.1, R Core Team, Vienna, Austria).

Radiomics modeling

In this step, according to the clinical observations (absorption or non-absorption of contrast agent by MS plaque in the T_1 images), the relationship between the extracted quantitative data and these observations was examined. The dataset was divided into training and testing cohorts with a proportion of 70% to 30%. The Z-score method was used to normalize the training data, and the mean and standard deviation were applied to the testing data. Decision Tree (DT), Gradient Boosting (GB), Extreme Gradient Boosting (XGB), and Multi-Layer Perceptron (MLP) classifiers were implemented in the R software (version 3.5.1, R Core Team, Vienna, Austria) for diagnosing the active plaque from the features of T_2 FLAR images because it was proven to be at the top of prediction performance [20-22]. According to the conducted studies, the advantages and disadvantages of these classifiers can be seen in Table 1.

Hyperparameters were optimized by 3-fold cross-validation in the training data and the best hyperparameters were selected to train the model. Finally, the mean±standard deviation values of sensitivity, specificity, accuracy, and Area Under the Curve (AUC) were calculated.

Validation of models

There are two internal and external methods for the validation of the model. The best validation method is the external method, but it was not possible due to the limitations and the need for a large sample size. Synthetic Minority Oversampling Technique (SMOTE) [28] was implemented on the training data to overcome imbalance followed by applying the trained model on the testing data by 1000 bootstraps.

Results

In the current study, 82 patients and 122 lesions were studied. As it can be seen from the data in Table 2, 16 (13.1%) active plaques and 106 (86.9%) inactive plaques were examined in this study. This study included 65 (79.2%) women and 17 (20.8%) men with an

Table 1: Advantages and disadvantages of the classifiers used in the present study

Classifiers	Advantages	Disadvantages	Reference
GB	1. By providing optimization through functions rather than parameters, it makes it easier to use custom functions. 2. It can handle large and very different datasets well since it is a step-by-step algorithm.	1. The trees are built sequentially, so training takes a long time. 2. Noisy data produces errors in estimation or classification.	[21, 23, 24]
XGB	1. In the case of clean data, overfitting can be prevented. 2. It is capable of handling missing values. 3. As a consequence, it optimizes the number of iterations by allowing cross-validation at each stage of the process.	1. Unlike other linear algorithms, it is more challenging to understand. 2. In the case of noisy data, overfitting may occur.	[21, 25, 26]
DT	1. Formulates rules that can be understood and interpreted without any statistical knowledge. 2. Both numerical and categorical variables can be used to make classifications. 3. This algorithm performs the classification with less computational complexity. 4.Non-parametric	1. DT is commonly flawed by overfitting. 2. Applied less to estimation tasks, such as predicting the value of continuous features. 3. Applying DT to continuous values results in a loss of information. 4. Non-optimal solution	[21]
MLP	1. Can be applied to complex non-linear problems. 2. Works well with large input data. 3. Provides quick predictions after training. 4. It is possible to achieve the same accuracy ratio even when the data is smaller.	1. It is not known to what extent each independent variable is affected by the dependent variable. 2. Computations are difficult and time consuming. 3. The proper functioning of the model depends on the quality of the training data. If the model does not work properly, generalization problems arise.	[21, 27]

MLP: Multi-Layer Perceptron, GB: Gradient Boosting, DT: Decision Tree, XGB: Extreme Gradient Boosting

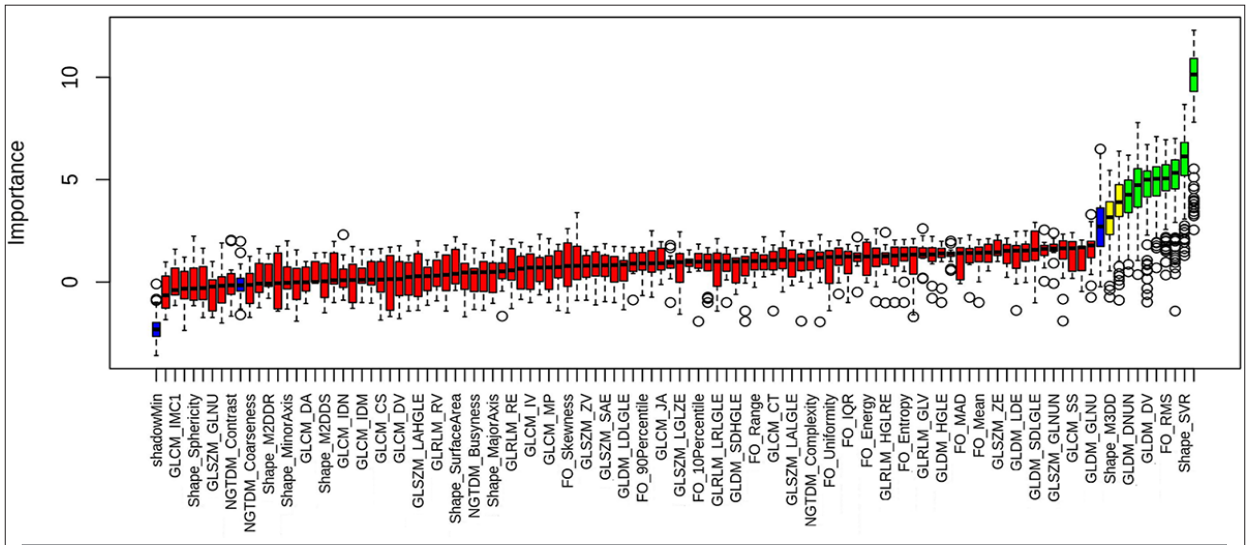
Table 2: Characteristics of multiple sclerosis patients in the present study

Parameter	Value (percentage)
Total number of patients	82
Sex	
Female	65 (79.2%)
Male	17 (20.8%)
Age (Year)	
Mean±standard deviation	38.59±9.46
Range	16-63
Total number of lesions	122
Active	16 (13.1%)
Inactive	106 (86.9%)

average age of 38.59 years, ranging from 16 to 63 years.

Comparison of feature selection methods

As it was mentioned, Boruta and Relief algorithms were used to select the robust features. Figures 2 and 3 show the results of feature selection utilizing Boruta and Relief algorithms, respectively. As it is shown in Figure 2, out of the total of 107 extracted features, only 8 features were selected as robust features in the Boruta method, including one demographic feature (age), three shape features (mesh volume, surface volume ratio, voxel volume),



GLCM_IMC1: Gray Level Co-occurrence Matrix_ Informational Measure of Correlation 1, GLSZM_GLNU: Gray Level Size Zone Matrix_ Gray Level Non-Uniformity, NGTDM: Neighbouring Gray Tone Difference Matrix, M2DDR: Maximum 2D Diameter Row, GLCM_DA: Gray Level Co-occurrence Matrix_ Difference Average, M2DDS: Maximum 2D Diameter Slice, GLCM_IDN: Gray Level Co-occurrence Matrix_ Inverse Difference Moment, GLCM_IDNM: Gray Level Co-occurrence Matrix_ Inverse Difference Moment Normalized, GLCM_CS: Gray Level Co-occurrence Matrix_ Cluster Shade, GLCM_DV: Gray Level Co-occurrence Matrix_ Difference Variance, GLSZM_LAHGLE: Gray Level Size Zone Matrix_ Large Area High Gray Level Emphasis, GLRLM_RV: Gray Level Run Length Matrix_ Run Variance, GLRLM_RE: Gray Level Run Length Matrix_ Run Entropy, GLCM_IV: Gray Level Co-Occurrence Matrix_ Inverse-Variance, GLCM_MP: Gray Level Co-Occurrence Matrix_ Maximum Probability, FO: First Order, GLSZM_ZV: Gray Level Size Zone Matrix_ Zone Variance, GLSZM_SAE: Gray Level Size Zone Matrix_ Small Area Emphasis, GLDM_LDLGLE: Gray Level Dependence Matrix_ Large Dependence Low Gray Level Emphasis, GLCM_JA: Gray Level Co-Occurrence Matrix_ Joint Average, GLSZM_LGLZE: Gray Level Size Zone Matrix_ Low Gray Level Zone Emphasis, GLRML_LRLGLE: Gray Level Run Length Matrix_ Long Run Low Gray Level Emphasis, GLDM_SDHGLE: Gray Level Dependence Matrix_ Small Dependence High Gray Level Emphasis, GLCM_CT: Gray Level Co-Occurrence Matrix_ Cluster Tendency, GLSZM_LALGLE: Gray Level Size Zone Matrix_ Large Area Low Gray Level Emphasis, FO_IQR: First Order_ Interquartile Range, GLRLM_HGLRE: Gray Level Run Length Matrix_ High Gray Level Run Emphasis, GLRLM_GLV: Gray Level Run Length Matrix_ Gray Level Variance, GLDM_HGLE: Gray Level Dependence Matrix_ High Gray Level Emphasis, FO_MAD: First Order_ Mean Absolute Deviation, GLSZM_ZE: Gray Level Size Zone Matrix_ Zone Entropy, GLDM_LDE: Gray Level Dependence Matrix_ Large Dependence Emphasis, GLDM_SDLGLE: Gray Level Dependence Matrix_ Small Dependence Low Gray Level Emphasis, GLSZM_GLNUN: Gray Level Size Zone Matrix_ Gray Level Non-Uniformity Normalized, GLCM_SS: Gray Level Co-occurrence Matrix_ Sum Squares, GLDM_GLNU: Gray Level Dependence Matrix_ Gray Level Non-Uniformity, M3DD: Maximum 3D Diameter, GLDM_DNUN: Gray Level Dependence Matrix_ Dependence Non-Uniformity Normalized, GLDM_DV: Gray Level Dependence Matrix_ Dependence Variance, FO_RMS: First Order_ Root Mean Squared, SVR: Surface Volume Ratio

Figure 2: Result of Boruta feature selection method: blue boxplots=shadow attributes (minimum, mean and maximum attribute), red boxplots: unimportant features, green boxplots: important features, yellow boxplots: tentative features

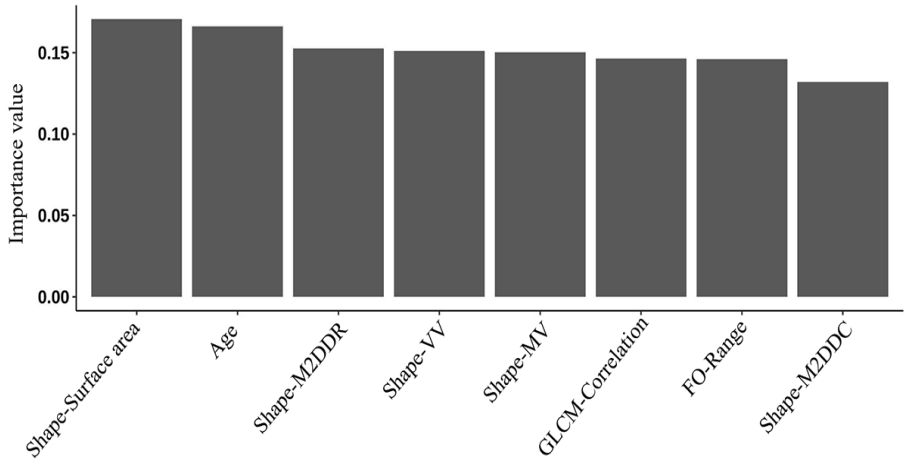


Figure 3: Results of Relief feature selection method
M2DDR: Maximum 2D Diameter Row, VV: Voxel Volume, MV: Mesh Volume, GLCM: Gray Level Co-occurrence Matrix, FO: First Order

two first-order features (root mean squared, total energy), and two GLDM features (dependence non-uniformity normalized, dependence variance). As seen in Figure 3, in the Relief method, eight features were selected, including one demographic feature (age), five shape features (surface area, maximum 2D diameter row, voxel volume, mesh volume, maximum 2D diameter column), one GLCM feature (correlation), and one first-order feature (range).

Comparison of performance metrics of models

In this study, four classifiers were used, including GB, XBG, DT, and MLP. From the combination of these 4 classifiers with 2 feature selection methods, 8 models were obtained. Table 3 shows the performance metrics of eight models including area under the curve with a 95% confidence interval, accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV). The bar chart and heat map of these metrics are shown in Figures 4 and 5, respectively.

As seen in Table 3, in the Boruta method, GB, XGB, DT, and MLP classifiers had AUC values of 0.84, 0.70, 0.92, and 0.89, respectively. Furthermore, in the Relief method, GB, XGB, DT, and MLP classifiers had AUC values of 0.81, 0.84, 0.90, and 0.83, respectively. The corresponding Receiver Operating Characteristic (ROC) diagrams are presented in Figure 6.

Discussion

In this study, the ability of radiomic features was investigated in predicting active plaques of patients with MS from T_2 FLAIR images. As seen in Table 2, based on the examinations and observations performed by a specialist radiologist, 13.1% of the lesions were active. The data collected from patients in this study are not balanced and machine-learning algorithms may be adversely affected by imbalanced datasets.

Misclassifications of prediction are often caused by an uneven distribution of major and minor classes [29]. In the current study, minority class events can include active lesions, which are examples of adverse incidents of low occurrence. The clinical impact of a wrong prediction within the minority group is greater than that of a wrong prediction within the majority group. A patient incorrectly classified as being low or normal risk could inadvertently be treated as such, leading to

Table 3: Performance metrics of nine models used to diagnose active plaque (mean±standard deviation)

Model	AUC	ACC	SEN	SPE	PPV	NPV
Boruta-GB	0.84±0.04 (95% CI: 0.84-0.84)	0.72±0.07	1.00±0.00	0.68±0.08	0.30±0.11	1.00±0.00
Boruta-XGB	0.70±0.14 (95% CI: 0.69-0.71)	0.85±0.06	0.50±0.27	0.90±0.06	0.42±0.24	0.93±0.05
Boruta-DT	0.92±0.03 (95% CI: 0.92-0.92)	0.86±0.06	1.00±0.00	0.84±0.07	0.48±0.15	1.00±0.00
Boruta-MLP	0.89±0.04 (95% CI: 0.89-0.89)	0.8±0.07	1.00±0.00	0.78±0.08	0.38±0.14	1.00±0.00
Relief -GB	0.81±0.04 (95% CI: 0.8-0.81)	0.66±0.08	1.00±0.00	0.61±0.09	0.27±0.10	1.00±0.00
Relief -XGB	0.84±0.11 (95% CI: 0.83-0.84)	0.91±0.05	0.74±0.22	0.94±0.05	0.63±0.23	0.96±0.03
Relief -DT	0.90±0.04 (95% CI: 0.90-0.90)	0.83±0.06	1.00±0.00	0.80±0.07	0.41±0.14	1.00±0.00
Relief -MLP	0.83±0.12 (95% CI: 0.82-0.84)	0.89±0.05	0.75±0.23	0.91±0.05	0.54±0.20	0.96±0.04

AUC: Area Under the Curve, ACC: Accuracy, SEN: Sensitivity, SPE: Specificity, PPV: Positive Predictive Value, NPV: Negative Predictive Value, MLP: Multi-Layer Perceptron, GB: Gradient Boosting, DT: Decision Tree, XGB: Extreme Gradient Boosting, CI: Confidence Interval

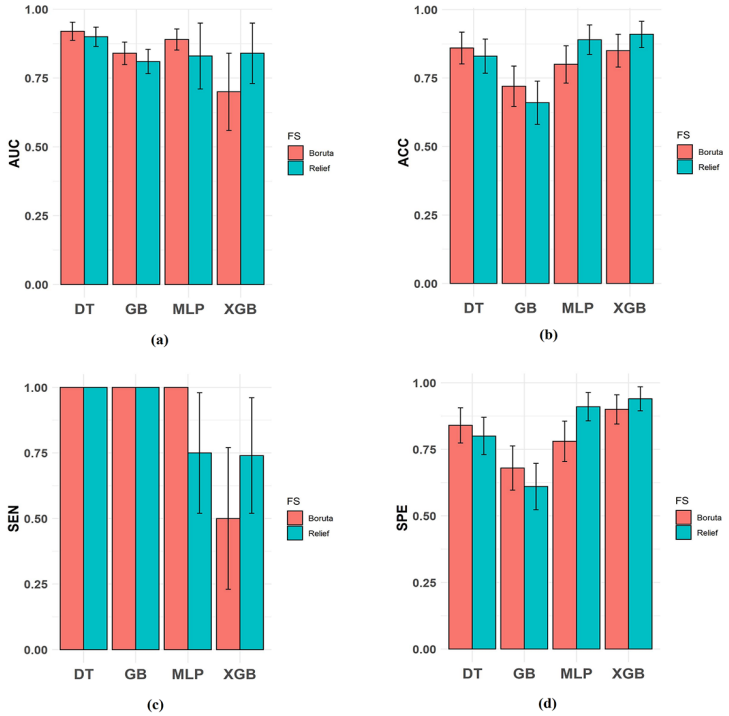


Figure 4: Bar chart of different feature selection and classifiers: Area under the curve (a); Accuracy (b); Sensitivity (c) and Specificity (d)
MLP: Multi-Layer Perceptron, GB: Gradient Boosting, DT: Decision Tree, XGB: Extreme Gradient Boosting, AUC: Area Under the Curve, ACC: Accuracy, SEN: Sensitivity, SPE: Specificity, NPV: Negative Predictive Value

adverse outcomes [30, 31]. Re-sampling has been shown to be an effective pre-processing method for resolving imbalanced problems in previous studies [32, 33]. Prior to classifier training, resampling approaches modify the imbalance distribution between minority and majority classes. Under-sampling methods may eliminate potentially useful information inherently related to their selection process, and oversampling methods can cause overfitting by simply creating repetitions of the minority class [34, 35].

There are different resampling methods, such as 1-random oversampling (ROS), 2- adaptive synthetic (ADASYN) [36], 3- SMOTE [28], and 4- borderline-SMOTE (bSMOTE) [37]. In a study conducted on 10 different resampling methods by Xie et al. [38], resampling techniques have shown a significant positive effect on prediction performance in unbalanced data sets. Recently, several new resampling techniques, such as the Synthetic Minority Oversampling Technique (SMOTE) have been proposed and shown to be more appropriate in some studies [28, 39]. This study applied SMOTE to oversample the minorities and balance the distribution of classes.

As mentioned earlier, two feature selection methods and four classifiers were used in this study. The results presented in Table 3, showed that in both feature selection methods, the decision tree classifier has the best performance. As seen in Table 3, Figure 4 (part a), and Figure 5 (part a), the area under the curve for Boruta-DT and Relief-DT were 0.92±0.03 (95% CI: 0.92-0.92) and 0.90±0.04 (95% CI: 0.90-0.90), respectively. The corresponding ROC diagrams are shown in Figure 6 (part c) and (part g). Additionally, the results show that in all classifiers, except for XGB, the perfor-

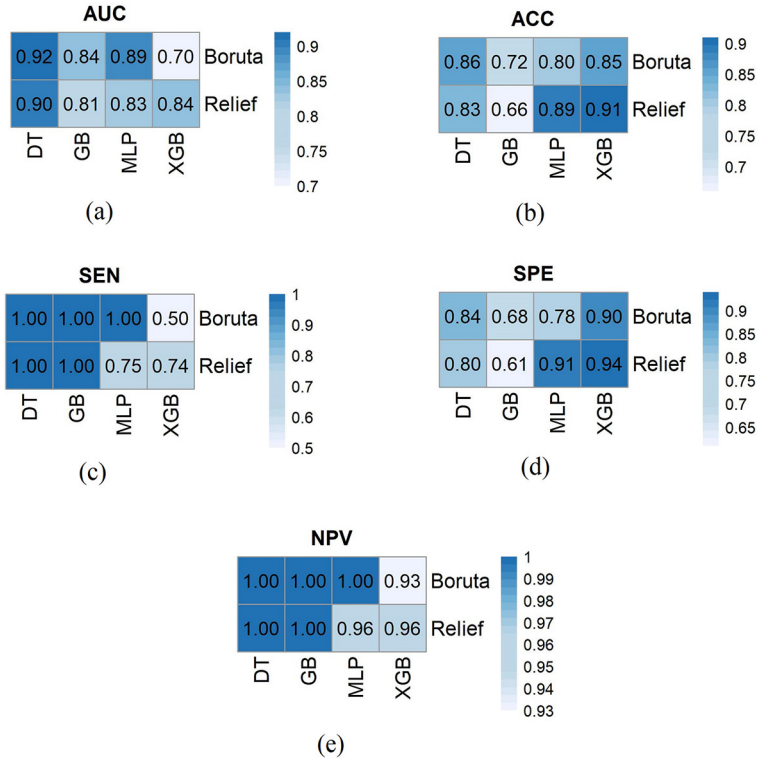


Figure 5: Heat map of different feature selection and classifiers: Area under the curve (a); Accuracy (b); Sensitivity (c); Specificity (d) and Negative predictive value (e)
MLP: Multi-Layer Perceptron, GB: Gradient Boosting, DT: Decision Tree, XGB: Extreme Gradient Boosting, AUC: Area Under the Curve, ACC: Accuracy, SEN: Sensitivity, SPE: Specificity, NPV: Negative Predictive Value

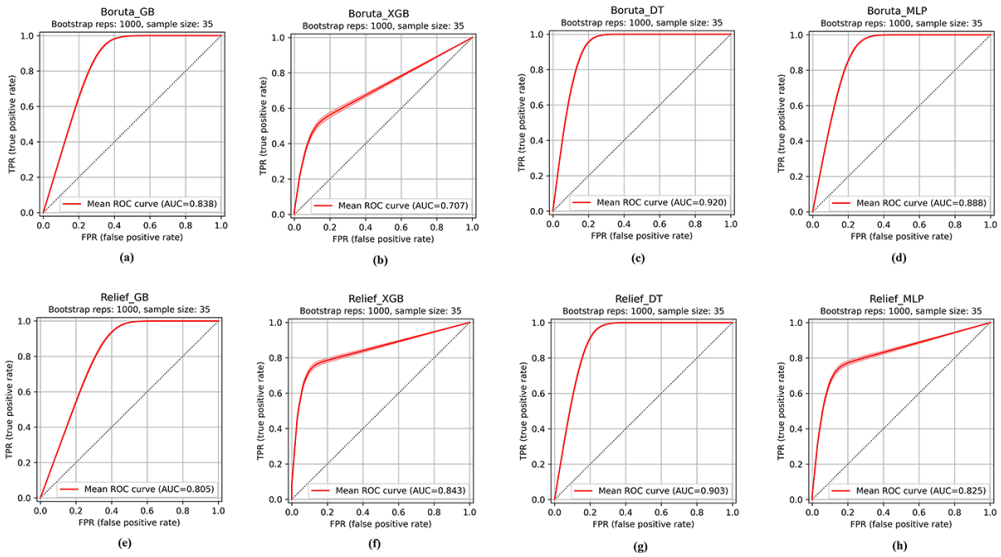


Figure 6: ROC diagrams of different feature selection and classifiers: Boruta-GB (a); Boruta-XGB (b); Boruta-DT (c); Boruta-MLP (d); Relief-GB (e); Relief-XGB (f); Relief-DT (g) and Relief-MLP (h)
ROC: Receiver Operating Characteristic, MLP: Multi-Layer Perceptron, GB: Gradient Boosting, DT: Decision Tree, XGB: Extreme Gradient Boosting, AUC: Area Under the Curve

mance of the active lesion predictor model is better in Boruta feature selection method. According to the results obtained from this study, the Boruta is a more efficient feature selection method than the Relief method (Figure 6).

Some studies have been conducted on the possibility of using quantitative image- processing methods to detect plaque in MS [12, 14]. But one of the limitations of these studies is the low number of their sample sizes. Therefore, in order to have reliable results from radiomics analysis, the sample size should be as large as possible. Furthermore, the validation methods used in these studies were based on the sample sizes of the studies, and the use of these validation methods, due to the use of learning data for testing data, can lead to false high sensitivity. In addition to these, the images of the patients in those studies were obtained in two-dimensional, while the analysis was done on volumetric (three-dimensional) images in this study.

In a study by Yulinga et al. [14], three feature selection methods were investigated along with three different classifiers. They also evaluated Recursive Feature Elimination (RFE), the Relief algorithm, and Least Sbsolute Shrinkage and Selection Operator (LASSO) as feature selection methods and logistic regression, random forest, and Support Vector Machine (SVM) as classifiers to build the models. Although the selection of algorithms in the present study was different from the study of Yulinga et al. [14], the final results of the present study are in good agreement with those of that study, showing that the radiomics features of T_2 FLAIR images can be used as a non-invasive method to help clinicians in the diagnosis of active plaques in MS patients in the future.

However, the clinical application of this research field still faces many challenges. Radiomics pipelines are challenged by the lack of a uniform methodology, the small sample size, and the lack of external datasets [40, 41].

Similar to other radiomics studies, the present study faced some limitations in that the small sample size and lack of external validation are significantly important. Considering that the accurate contouring of the lesion affects the extracted radiomc features, the use of auto-segmentation methods is suggested in future research. It is also necessary to use a larger sample size and examine other sequences of MRI images in future studies.

Conclusion

Radiomics features can lead to predicting active plaques of MS and in the future, radiomics -based models could be viable alternatives to a non-invasive diagnosis of active plaques of MS.

Acknowledgment

The authors would like to extend thanks to Karaj Central Medical Imaging Institute (Karaj, Iran) for its support and for providing data for this research. The authors also thank Dr. Dadafarin for his tireless assistance in providing the patient's data.

Authors' Contribution

H. Tavakoli conceived the idea. Introduction of the paper was written by H. Tavakoli and G. Pirzad Jahromi. A. Sedaghat and G. Pirzad Jahromi gather the images and the related literature and also help with writing of the related works. The method implementation was carried out by H. Tavakoli and G. Pirzad Jahromi. Results and Analysis was carried out by A. Sedaghat and H. Tavakoli. The research work was proofread and supervised by H. Tavakoli and G. Pirzad Jahromi. All the authors read, modified, and approved the final version of the manuscript.

Ethical Approval

All data were collected from Karaj Central Medical Imaging Institute (Karaj, Iran). Investigations on patient data and images were performed based on the ethical guidelines of Baqiyatallah University of Medical Sciences (Tehran, Iran). This work was approved by the Ethics Committee of Baqiyatallah University of Medical Sciences with ethical code of IR.BMSU.REC.1399.412.

Informed Consent

Informed consent was obtained from the patients.

Funding

This study was supported by Baqiyatallah University of Medical Sciences (Grant number of 98000578 (Dr Tavakoli)).

Conflict of Interest

None

References

1. Kobelt G, Thompson A, Berg J, Gannedahl M, Eriksson J. New insights into the burden and costs of multiple sclerosis in Europe. *Mult Scler*. 2017;**23**(8):1123-36. doi: 10.1177/1352458517694432. PubMed PMID: 28273775. PubMed PMCID: PMC5476197.
2. Browne P, Chandraratna D, Angood C, Tremlett H, Baker C, Taylor BV, Thompson AJ. Atlas of Multiple Sclerosis 2013: A growing global problem with widespread inequity. *Neurology*. 2014;**83**(11):1022-4. doi: 10.1212/WNL.0000000000000768. PubMed PMID: 25200713. PubMed PMCID: PMC4162299.
3. Ascherio A. Environmental factors in multiple sclerosis. *Expert Rev Neurother*. 2013;**13**(12 Suppl):3-9. doi: 10.1586/14737175.2013.865866. PubMed PMID: 24289836.
4. Bakshi R, Thompson AJ, Rocca MA, Pelletier D, Dousset V, Barkhof F, et al. MRI in multiple sclerosis: current status and future prospects. *Lancet Neurol.* 2008;**7**(7):615-25. doi: 10.1016/S1474-4422(08)70137-6. PubMed PMID: 18565455. PubMed PMCID: PMC2586926.
5. Bakshi R, Minagar A, Jaisani Z, Wolinsky JS. Imaging of multiple sclerosis: role in neurotherapeutics. *NeuroRx.* 2005;**2**(2):277-303. doi: 10.1602/neurorx.2.2.277. PubMed PMID: 15897951. PubMed PMCID: PMC1064992.
6. Filippi M, Rocca MA. Conventional MRI in multiple sclerosis. *J Neuroimaging*. 2007;**17**(Suppl 1):3S-9. doi: 10.1111/j.1552-6569.2007.00129.x. PubMed PMID: 17425727.
7. Lennon VA, Wingerchuk DM, Kryzer TJ, Pittock SJ, Lucchinetti CF, Fujihara K, et al. A serum autoantibody marker of neuromyelitis optica: distinction from multiple sclerosis. *Lancet.* 2004;**364**(9451):2106-12. doi: 10.1016/S0140-6736(04)17551-X. PubMed PMID: 15589308.
8. Pretorius PM, Quaghebeur G. The role of MRI in the diagnosis of MS. *Clin Radiol.* 2003;**58**(6):434-48. doi: 10.1016/s0009-9260(03)00089-8. PubMed PMID: 12788312.
9. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;**278**(2):563-77. doi: 10.1148/radiol.2015151169. PubMed PMID: 26579733. PubMed PMCID: PMC4734157.
10. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, De Jong EEC, Van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;**14**(12):749-62. doi: 10.1038/nrclinonc.2017.141. PubMed PMID: 28975929.
11. Larue RTHM, Van Timmeren JE, De Jong EEC, Feliciani G, Leijenaar RTH, Schreurs WMJ, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncol.* 2017;**56**(11):1544-53. doi: 10.1080/0284186X.2017.1351624. PubMed PMID: 28885084.
12. 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. doi: 10.1371/journal.pone.0145497. PubMed PMID: 26693908. PubMed PMCID: PMC4687842.
13. Zhang Y. MRI texture analysis in multiple sclerosis. *Int J Biomed Imaging.* 2012;**2012**:762804. doi: 10.1155/2012/762804. PubMed PMID: 22144983. PubMed PMCID: PMC3227516.
14. 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. *Mult Scler Relat Disord*. 2021;**53**:102989. doi: 10.1016/j.msard.2021.102989. PubMed PMID: 34052741.
15. Uysal E, Erturk SM, Yildirim H, Seleker F, Basak M. Sensitivity of immediate and delayed gadolinium-enhanced MRI after injection of 0.5 M and 1.0 M gadolinium chelates for detecting multiple sclerosis lesions. *Am J Roentgenol.* 2007;**188**(3):697-702. doi: 10.2214/AJR.05.2212. PubMed PMID: 17312056.
16. Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res*. 2017;**77**(21):e104-7. doi: 10.1158/0008-5472.CAN-17-0339. PubMed PMID: 29092951. PubMed PMCID: PMC5672828.
17. Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Softw*. 2010;**36**(11):1-13. doi: 10.18637/jss.v036.i11.
18. Kursa MB, Jankowski A, Rudnicki WR. Boruta–a system for feature selection. *Fundamenta Informaticae.* 2010;**101**(4):271-85. doi: 10.3233/FI-2010-288.

19. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief-based feature selection: Introduction and review. *J Biomed Inform.* 2018;**85**:189-203. doi: 10.1016/j.jbi.2018.07.014. PubMed PMID: 30031057. PubMed PMCID: PMC6299836.
20. Wan S, Liang Y, Zhang Y, Guizani M. Deep multilayer perceptron classifier for behavior analysis to estimate Parkinson's disease severity using smartphones. *IEEE Access.* 2018;**6**:36825-33. doi: 10.1109/ACCESS.2018.2851382.
21. Jafarzadeh H, Mahdianpari M, Gill E, Mohammadimanesh F, Homayouni S. Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: a comparative evaluation. *Remote Sens.* 2021;**13**(21):4405. doi: 10.3390/rs13214405.
22. Akkaya B, Çolakoğlu N. Comparison of multi-class classification algorithms on early diagnosis of heart diseases. y-BIS Conference 2019: Recent Advances in Data Science and Business Analytics; Istanbul, Turkey: Mimar Sinan Fine Arts University Publications; 2019. p. 162-72.
23. Ravanshad A. Gradient Boosting Vs Random Forest. 2018 [Accessed 2019 August 12]. Available from: https://medium.com/@aravanshad/gradient-boosting-versus-random-forest-cfa3fa8f0d80.
24. Rao H, Shi X, Rodrigue A K, Feng J, Xia Y, Elhoseny M, et al. Feature selection based on artificial bee colony and gradient boosting decision tree. *Applied Soft Computing.* 2019;**74**:634-42. doi: 10.1016/j.asoc.2018.10.036.
25. Kumar N. Advantages of XGBoost algorithm in machine learning. 2019 [Accessed 2019 August 13]. Available from: http://theprofessionalspoint.blogspot.com/2019/03/advantages-of-xgboostalgorithm-in.html.
26. Nielsen D. Why Does XGBoost Win "Every" Machine Learning Competition? [Dissertation]. Norwegian University of Science and Technology Department of Mathematical Sciences; 2016.
27. Venkateswaran B, Ciaburro G. Neural Networks with R. Packt Publishing; 2017.
28. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: synthetic minority over sampling technique. *J Artif Intell Res*. 2002;**16**:321-57. doi: 10.1613/jair.953.
29. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng.* 2009;**21**(9):1263-84. doi: 10.1109/TKDE.2008.239.
30. Kabir MF, Ludwig S. Classification of breast cancer risk factors using several resampling approaches. 17th IEEE International Conference on Machine Learning and Applications (ICMLA); Orlando, FL, USA: IEEE; 2018.
31. Fotouhi S, Asadi S, Kattan MW. A comprehensive data level analysis for cancer diagnosis on imbalanced data. *J Biomed Inform.* 2019;**90**:103089. doi: 10.1016/j.jbi.2018.12.003. PubMed PMID: 30611011.
32. Batuwita R, Palade V. Efficient resampling methods for training support vector machines with imbalanced datasets. The 2010 International Joint Conference on Neural Networks (IJCNN); Barcelona, Spain: IEEE; 2010.
33. Loyola-González O, Martínez-Trinidad JF, Carrasco-Ochoa JA, García-Borroto M. Study of the impact of resampling methods for contrast pattern based classifiers in imbalanced databases. *Neurocomputing*. 2016;**175**:935-47. doi: 10.1016/j.neucom.2015.04.120.
34. Chawla NV. Data mining for imbalanced datasets: An overview. In: Data mining and knowledge discovery handbook. Springer; 2009. p. 875-86.
35. Nitesh VC, Nathalie J, Aleksander K. Editorial: special issue on learning from imbalanced data sets. *Sigkdd Explorations*. 2004;**6**(1):1-6.
36. He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence); Hong Kong: IEEE; 2008.
37. Han H, Wang WY, Mao BH. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In International Conference on Intelligent Computing; Berlin, Heidelberg: Springer; 2005. p. 878-87.
38. Xie C, Du R, Ho JW, Pang HH, Chiu KW, Lee EY, Vardhanabhuti V. Effect of machine learning resampling techniques for imbalanced datasets in 18F-FDG PET-based radiomics model on prognostication performance in cohorts of head and neck cancer patients. *Eur J Nucl Med Mol Imaging*. 2020;**47**(12):2826-35. doi: 10.1007/s00259-020-04756-4. PubMed PMID: 32253486.
39. Elreedy D, Atiya AF. A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. *Inf Sci.* 2019;**505**:32-64. doi: 10.1016/j.ins.2019.07.070.
40. Sotoudeh H, Sarrami AH, Roberson GH, Shafaat O, Sadaatpour Z, Rezaei A, et al. Emerging Applications of Radiomics in Neurological Disorders: A Review. *Cureus.* 2021;**13**(12):e20080. doi: 10.7759/cureus.20080. PubMed PMID: 34987940. PubMed PMCID: PMC8719529.
41. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Phys Med Biol.* 2016;**61**(13):R150-66. doi: 10.1088/0031-9155/61/13/R150. PubMed PMID: 27269645. PubMed PMCID: PMC4927328.