Machine Learning-Based Analysis of MR Multiparametric Radiomics for the Subtype Classification of Breast Cancer

Tianwen Xie [1,2†], Zhe Wang [3,4†], Qiufeng Zhao [5], Qianming Bai [6], Xiaoyan Zhou [6], Yajia Gu [1,2], Weijun Peng [1,2*] and He Wang [3,7*]

[1] Department of Radiology, Fudan University Shanghai Cancer Center, Shanghai, China, [2] Department of Oncology, Fudan University Shanghai Cancer Center, Shanghai, China, [3] Human Phenome Institute, Fudan University, Shanghai, China, [4] Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China, [5] Department of Radiology, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China, [6] Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, China, [7] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

Objective: To investigate whether machine learning analysis of multiparametric MR radiomics can help classify immunohistochemical (IHC) subtypes of breast cancer.

Study design: One hundred and thirty-four consecutive patients with pathologically-proven invasive ductal carcinoma were retrospectively analyzed. A total of 2,498 features were extracted from the DCE and DWI images, together with the new calculated images, including DCE images changing over six time points (DCE$_{sequential}$) and DWI images changing over three b-values (DWI$_{sequential}$). We proposed a novel two-stage feature selection method combining traditional statistics and machine learning-based methods. The accuracies of the 4-IHC classification and triple negative (TN) vs. non-TN cancers were assessed.

Results: For the 4-IHC classification task, the best accuracy of 72.4% was achieved based on linear discriminant analysis (LDA) or subspace discrimination of assembled learning in conjunction with 20 selected features, and only small dependent emphasis of Kendall-tau-b for sequential features, based on the DWI$_{sequential}$ with the LDA model, yielding an accuracy of 53.7%. The linear support vector machine (SVM) and medium k-nearest neighbor using eight features yielded the highest accuracy of 91.0% for comparing TN to non-TN cancers, and the maximum variance for DWI$_{sequential}$ alone, together with a linear SVM model, achieved an accuracy of 83.6%.

Conclusions: Whole-tumor radiomics on MR multiparametric images, DCE images changing over time points, and DWI images changing over different b-values provide a non-invasive analytical approach for breast cancer subtype classification and TN cancer identification.

Keywords: radiomics, breast cancer, immunohistochemical subtypes, dynamic contrast-enhanced imaging, diffusion-weighted imaging, machine learning

INTRODUCTION

Breast cancer is a heterogeneous disease with diverse clinical manifestations, treatment responses, and survival outcomes (1). The immunohistochemical (IHC) subtypes based on the expression of the estrogen and progesterone receptor, the detection of overexpression of the human epidermal growth factor receptor 2 (HER2) oncogene, and Ki-67 labeling index, are routinely used to identify tumor subtypes with different clinical outcomes and responses to therapy, including Luminal A cancer, Luminal B cancer, HER2-positive cancer, and triple negative (TN) cancer (2). HER2-positive breast cancers are more likely to have a pathologic complete response (pCR) to neoadjuvant chemotherapy, whereas reduced pCR rates are demonstrated in Luminal-type breast cancers (2, 3). Patients with TN breast cancer have a relatively high death rate due to the aggressive features of this subtype and lack of effective targeted therapy (4). Therefore, it is important to identify the subtypes to select the appropriate therapy and predict the therapeutic response (5).

Multiparametric MR imaging, including dynamic contrast-enhanced (DCE) imaging and diffusion-weighted imaging (DWI), has excellent sensitivity and good specificity for breast cancer diagnosis and plays an important role in the preoperative staging and subsequent choice of appropriate therapy. If DCE imaging and DWI can be used to differentiate breast cancer subtypes, this would provide a complementary method for constructing the IHC profile and therefore assist in treatment planning. Several authors have described the breast cancer subtypes with the BI-RADS lexicon on MRI (6–8), but inter- and intra-observer variability continues to exist among radiologists (9, 10).

Beyond visual interpretation by radiologists, there is more quantitative diagnostic information about the tumor hidden in thousands of acquired images. Radiomics (referring to computational algorithms) used extracted imaging texture features to serve as noninvasive biomarkers that could predict breast cancer subtypes (10). Multiple studies have used the texture analysis of DCE images for the subtype classification of breast cancer or the identification of TN cancer (11, 12). Grimm et al. demonstrated a correlation between Luminal subtypes of breast cancer and texture features on DCE images (13). Agner et al. used the morphologic and texture features extracted from the whole tumor on early postcontrast images in conjunction with an SVM classifier to identify TN cancer (14). Several studies have proposed the histogram or textural features extracted from DWI for subtype differentiation (15, 16). However, no study has attempted to investigate the texture analysis based radiomics of combining DCE imaging and DWI in the subtype classification of breast cancer.

The purpose of this present study was to evaluate the performance of the MR multiparametric radiomics model to differentiate among Luminal A cancer, Luminal B cancer, HER2-positive cancer, and TN breast cancer using DCE imaging and DWI. Furthermore, we investigated whether the radiomic model could identify the subtype of the worst clinical outcome (TN breast cancer) from other subtypes.

MATERIALS AND METHODS

Study Population

Between February 2016 and May 2017, 190 consecutive patients with core needle biopsy-proven invasive ductal carcinoma (IDC) were enrolled in the study. All the patients underwent preoperative MRI examinations with DCE and DWI sequences. Patients with a prior history of malignancy ($N = 5$), those treated with neoadjuvant chemotherapy before MR examination ($N = 31$), and those with lesions smaller than 1 cm ($N = 16$) were excluded. Our database also excluded cases that had poor fat saturation on DWI ($N = 3$) or discordance in the number of slices among the postcontrast sequences ($N = 1$). Two radiologists with 2- and 8-years' experience in breast MR imaging, respectively, who were blinded to the pathologic results but were aware of the IDC diagnosis, reviewed the MR images. For 22 patients with multicentric or multifocal tumors, the largest tumor was selected for analysis based on the first post-contrast DCE images. A total of 134 tumors from 134 women (mean age, 51.2 years; age range, 24–84 years) were ultimately evaluated.

The immunohistochemical subtype of breast cancer was classified as Luminal A (ER- and/or PR-positive, HER2-negative, and Ki-67 < 14%), Luminal B (ER- and/or PR-positive, HER2-negative, and Ki-67 ≥ 14%, or ER- and/or PR-positive, HER2-positive, irrespective of Ki-67 expression), HER2-positive (ER- and PR-negative, HER2-positive), and TN (ER-negative, PR-negative, and HER2-negative) (5).

This retrospective study was approved by our institutional review board, which waived informed consent.

MR Imaging

All the MR scans were performed using a 3.0T MAGNETOM Skyra system (Siemens Healthcare, Erlangen, Germany) with a 16-channel phased-array breast coil, with patients in the prone position. The breast MRI examinations included a transverse fat-suppressed T2-weighted (TR/TE, 3,570/69 ms) sequence and a transverse T1-weighted (TR/TE, 5.4/2.4 ms) sequence.

Before contrast injection, DWI was performed in the transverse plane covering both breasts at the position of the tumor using a single-shot echo-planar imaging sequence with the following parameters: TR/TE, 3,000/54 ms; flip angle, 90°; field of view, 340 × 150 – 280 mm^2; matrix, 220 × 220; slice thickness, 6 mm; 3 b-values, 50, 400, and 800 s/mm^2, with the number of averages 3, 4, and 5, respectively; rate 3 GRAPPA acceleration. The total acquisition time was 2:09 min.

DCE-MRI was performed using a 3D T1-weighted fat-suppressed, fast spoiled gradient-echo sequence (TR/TE 4.5/1.6 ms; flip angle, 10°; bandwidth, 380 Hz/Pixel) with one pre-contrast and five consecutive post-contrast dynamic series after a bolus injection of 0.1 mmol/L of gadopentetate dimeglumine (Magnevist; Bayer Schering Pharma, Berlin, Germany) per kilogram of body weight, injected at a rate of 1.5 mL/s. Image acquisition in the transverse plane lasted for 60 s per volumetric acquisition with slice thickness and was 1.5 mm with no gap and a matrix of 384 × 384.

Multiparametric MR Radiomics

The semi-automated segmentation of whole tumor on the DWI and DCE images was conducted by the radiologist using the prototype MR Multiparameter Analysis software (Siemens Healthcare, Erlangen, Germany). The segmentation process was composed of three steps: (a) The seed points were manually drawn inside the tumor (including the necrotic regions) and outside the tumor on DWI images with a b-value of 800 s/mm^2 (DWI_{b800}) and the first postcontrast of DCE images, respectively; (b) The 3D segmentation of the whole tumor was executed (based on these seed points) with a random-walker algorithm (17). We performed manual adjustments in 9 of 134 women for DWI_{b800} and in 46 of 134 women for the first postcontrast of DCE images; and (c) For DWI, 3D segmented contouring created on DWI_{b800} was propagated to the DWI images with other b-values. For DCE, 3D segmented contouring created on the first postcontrast
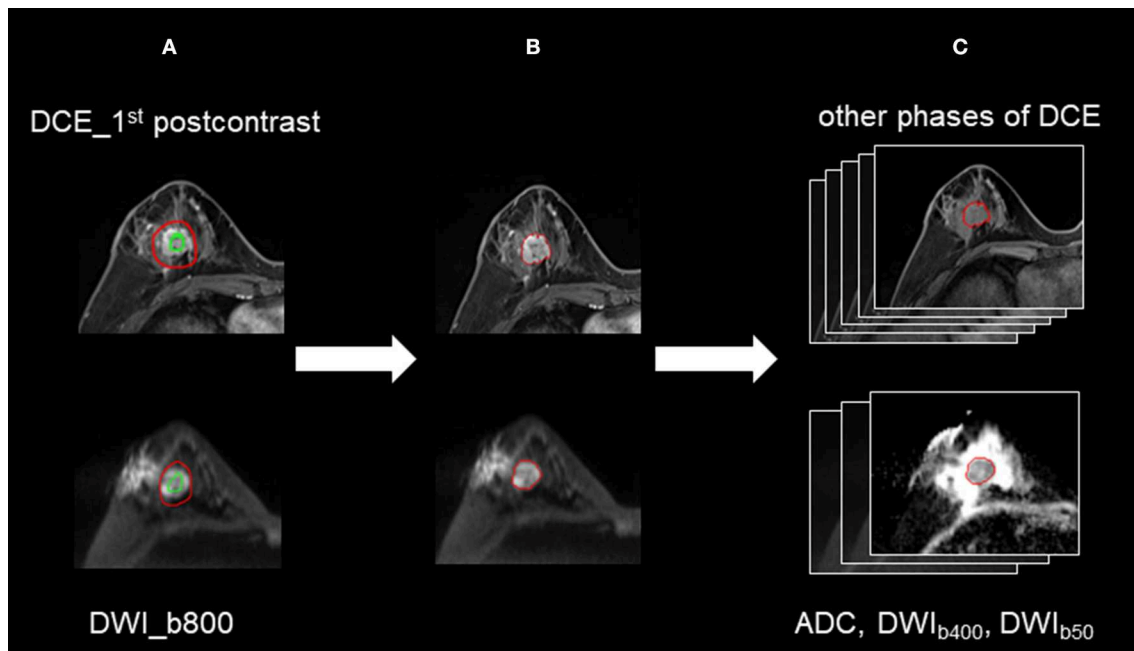

FIGURE 1 | The process for 3D segmentation. The segmentation process comprised three steps: **(A)** The foreground and background seed points were manually drawn inside and outside the tumor on the DWI images with a b-value of 800 s/mm^2 (DWI_{b800}) and the first postcontrast of the DCE images, respectively. **(B)** The whole-tumor segmentations on DWI_{b800} and the first postcontrast of the DCE images were executed with a semiautomatic algorithm. **(C)** For DWI, 3D segmented contouring that was created on DWI_{b800} was propagated to the DWI images with other b-values. For DCE, 3D segmented contouring that was created on the first postcontrast images was propagated to the precontrast and four postcontrast phases of the DCE images.
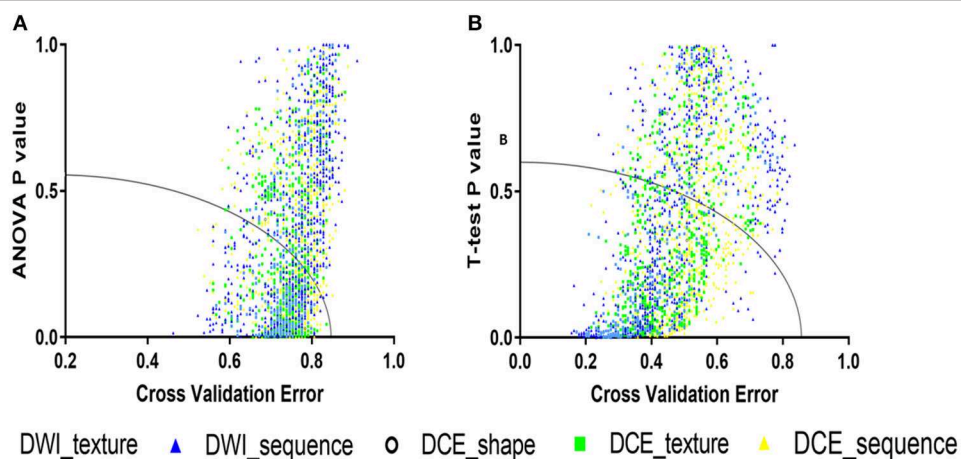

FIGURE 2 | The accepted domains for coarse feature selection. **(A)** For the 4-IHC classification, the p-value for the analysis of variance (ANOVA) and cross-validation error were set to 0.6 and 0.85, respectively. **(B)** For the TN vs. non-TN cancers, the t-test p-value and cross-validation error were set to 0.54 and 0.76, respectively.

images were propagated to pre-contrast and the other four post-contrast phases of DCE images. Finally, 3D segmentation of the whole tumor from nine different imaging series (DWI MR images with three different b-values, and six DCE images) were saved. The progress for 3D segmentation is described in **Figure 1**.

The 3D segmentation from the nine-image series were loaded onto a personal laptop for further texture analysis. The radiomics features were comprised of five parts: (a) Shape features on DCE (DCE_{shape}): The fourteen shape-based features were calculated using the first postcontrast DCE images; (b) Texture features based on DCE images ($DCE_{texture}$): The 92 texture features (18) features, were calculated on six time series individually, yielding 552 features; (c) Sequential features based on DCE time series ($DCE_{sequential}$): To characterize the textural changes on DCE images over time serials, we measured ten new sequential features for each texture feature described in group b (**Supplemental Material 1**). The first six features, including mean, variance, kurtosis, skewness, energy, and entropy, were extracted from each individual subject. The other four features, including Kendall-tau-b, conservation, stability, and dispersion, were calculated from the interactive information between the current subject and the remainder of the subjects. Therefore, a total of 920 $DCE_{sequential}$ features were extracted from 92 texture features; (d) Texture features based on DWI images ($DWI_{texture}$): The same 92 texture features with group b were calculated on DWI images with three b-values set to yield 276 features; and (e) Sequential features based on DWI images ($DWI_{sequential}$): Like group c, to characterize the textural changes on DWI images over different b-values, we measured eight sequential features for each texture feature. Kurtosis and entropy were not measured because of the limited number of b-values. Therefore, 736 $DWI_{sequential}$ features were extracted. A total of 2,498 features were extracted to obtain a classification model. Texture extraction was applied to the multiparametric images using PyRadiomics package (18) in the Python software (v. 3.6, Python Software Foundation, https://www.python.org/).

Feature Selection

To reduce the dimensionality of the feature space, we proposed a two-staged method combining traditional statistics and machine learning-based feature selection.

For coarse feature selection, the features that were considered to be noisy were excluded. The purpose of the coarse feature selection was to prevent the resulting model from becoming a linear combination of noises. Regarding the 4-IHC classification tasks of breast cancer, the one-way analysis of variance (ANOVA) and cross-validation (CV) were performed for every radiomics feature. The t-test and CV for each radiomics feature were performed for TN vs. non-TN cancers. The feature that had both a small p-value and a small cross-validation error rate was informative. The feasible domain for excluding noisy features was a strong convex set, which reduced a high dimensionality of complex data sets to a two-dimensional space (19). In addition, the ellipse was chosen as the acceptance domain. Based on this

idea, the accepted domain equation was set as follows:

$$\frac{(p\_value)^2}{a^2} + \frac{(CVerrorrate)^2}{b^2} < 1 \qquad (1)$$

where a and b are acceptance factors. In order to reduce approximately half of the features, we set a and b to 0.6 and 0.85, respectively, for the 4-IHC classification after an ergodic process by an interval of 0.01. For TN vs. non-TN cancers, a and b were set to 0.54 and 0.76, respectively (**Figure 2**).

For fine feature selection, we proposed a novel model by combing the lasso regression, ridge regression, and elastic net to select the features. For the 4-IHC classification, lasso regression, ridge regression, and elastic net were utilized separately to rank the selected features from the coarse feature selection. The ranks from these three methods were summed up as a score for each feature. The top 40 features with the lowest scores were selected, and 20 features that were correlated with other features (the absolute value of the Pearson correlation index > 0.4) were subsequently removed manually to reduce the collinearity of feature combinations. Using the same selection method as the 4-IHC classification, a total of eight features were selected for TN vs. non-TN.

Machine Learning-Based Classification

The best models for the 4-IHC classification and TN vs. non-TN cancers were selected from three decision tree classifiers, two discriminant analysis classifiers, six SVM classifiers, four k-nearest neighbor (KNN) classifiers, and five ensemble learning classifiers (20–24). The three decision tree classifiers were fine tree, medium tree, and coarse tree. The two discriminant analysis classifiers were LDA and quadratic discriminant analysis (QDA). The six SVM classifiers included the linear, quadratic, cubic, fine gaussian, medium gaussian, and coarse gaussian. The four KNN classifiers were fine KNN, medium KNN, coarse KNN, and weighted KNN. The five ensemble learning classifiers included subspace discriminant, subspace KNN, the AdaBoost algorithm with decision tree, the bootstrap-aggregated (bagged) tree algorithm with decision tree, and the RUSBoost algorithm with decision tree.

All the machine learning models were conducted using the 5-fold cross validation, whereby 20% of the data were used to test the model created by the other 80% of the data. The procedure was repeated for ten rounds to average the estimates of performance. The accuracy for two classification tasks was assessed for the models. To prove the superiority of our fine feature selection method, the classification accuracy combining lasso regression, ridge regression, and elastic net, and the classification accuracy using any single one of these three methods, were also assessed. The feature selection and machine learning-based classification was achieved using the Statistics and Machine Learning Toolbox in MatLab (v. R2018a; MathWorks, Natick, MA). The flowchart for multiparametric MR radiomics is described in **Figure 3**.
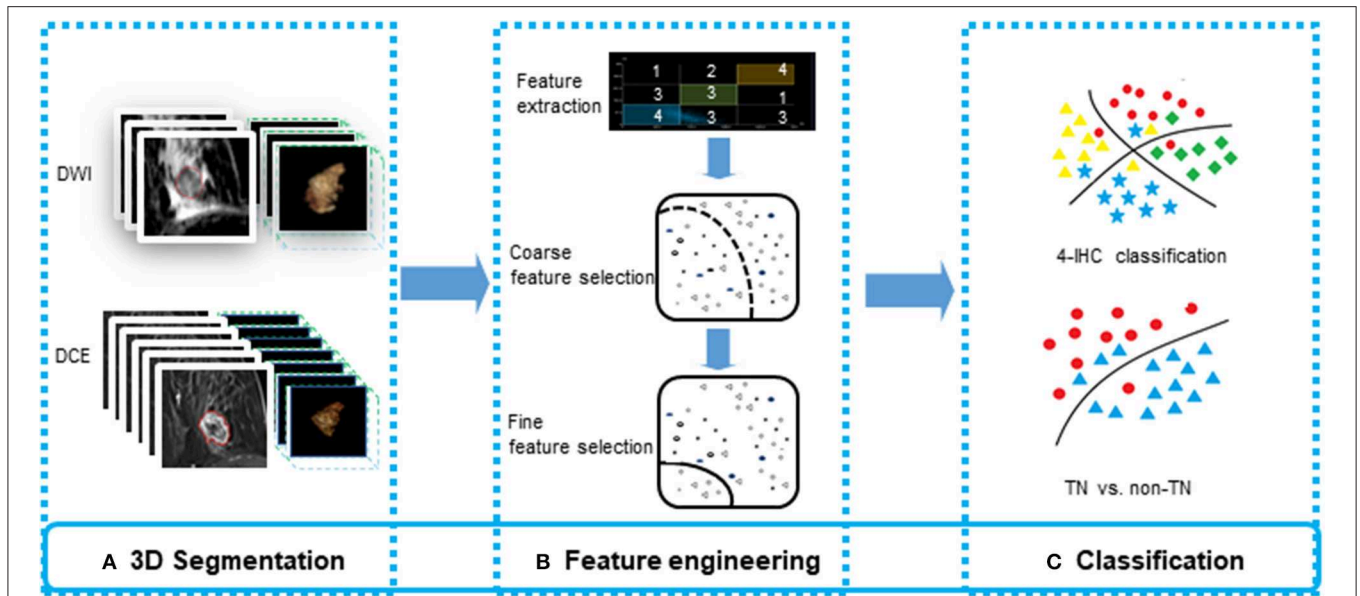
FIGURE 3 | The flowchart for feature engineering of multiparametric MR radiomics. **(A)** The whole-tumor segmentations from a total of 10 sequence images were executed. **(B)** A total of 2,498 features were extracted, and a two-stage feature selection method was subsequently performed. **(C)** Machine learning-based classifiers were used for the 4-IHC classification and TN vs. non-TN cancers.

TABLE 1 | Clinicopathologic characteristics of patients.

Parameter	Luminal A	Luminal B	HER2-positive	TN
Age (y)*	26	68	18	22
Menopausal status				
Peri- or postmenopausal	11	29	8	14
Premenopausal	15	38	10	8
Unknown	0	1	0	0
Tumor size (mm^3)*	18 (11–47)	25 (10–80)	21 (10–43)	24 (13–39)
Axial lymph nodes				
Negative	21	30	7	13
Positive	5	38	11	9

HER2, human epidermal growth factor receptor 2; TN, triple negative.
Data for continuous variables are means ± standard deviation, with medians in parentheses.

RESULTS

Clinical Data

The patient demographic and cancer characteristics are shown in **Table 1**. Of the 134 cancers, the numbers of Luminal A, Luminal B, HER2-positive, and TN cancers were 26, 68, 18, and 22, respectively.

Performance of the Machine Learning-Based Classification

Of 2,498 features, a total of 1,292 for the 4-IHC classification and 1,555 for the TN vs. non-TN cancers were chosen after coarse feature selection. Fine feature selection was conducted to reduce the feature sets to 20 features for the 4-IHC
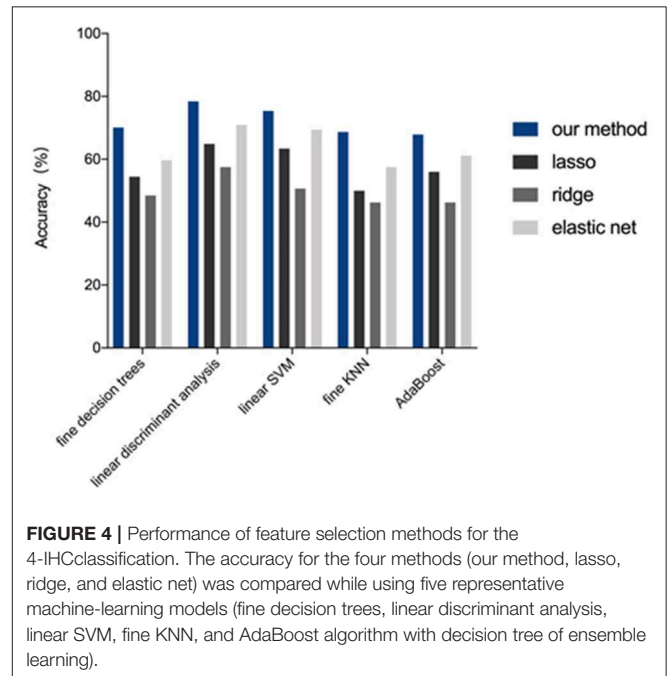

FIGURE 4 | Performance of feature selection methods for the 4-IHCclassification. The accuracy for the four methods (our method, lasso, ridge, and elastic net) was compared while using five representative machine-learning models (fine decision trees, linear discriminant analysis, linear SVM, fine KNN, and AdaBoost algorithm with decision tree of ensemble learning).

classification and eight for the TN vs. non-TN cancers (**Supplementary Materials 2, 3**). Z-score value distribution of every selected feature for all the patients in two classification tasks is displayed in **Figure 4**. As shown in **Figure 5**, our fine feature selection combining the lasso regression, ridge regression, and elastic net had higher accuracy compared with the use of any single one of these three methods. The Rad-score
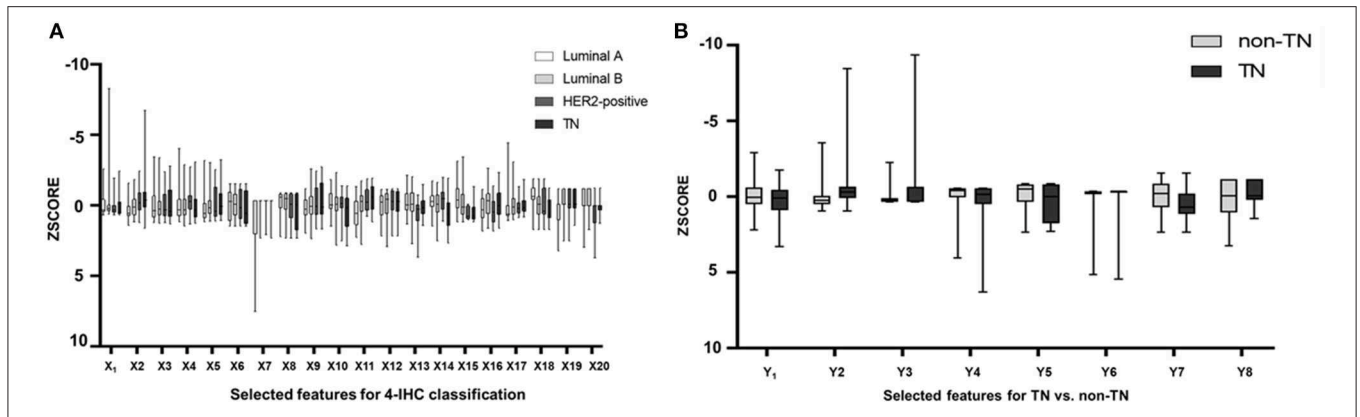
FIGURE 5 | Box plots showed z-score value distribution of every selected feature for all patients in the 4-IHC classification **(A)** and TN vs. non-TN **(B)**.
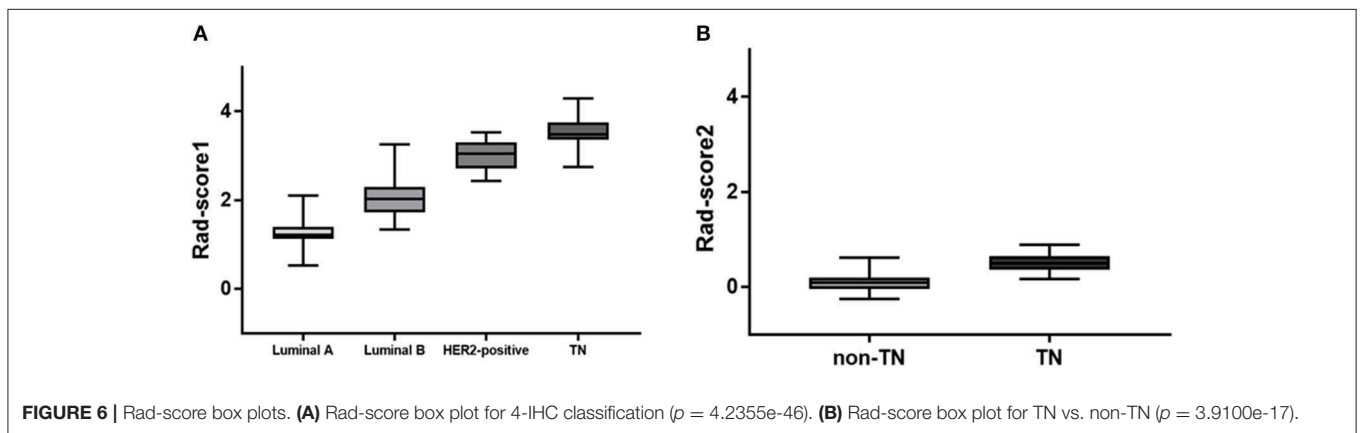

FIGURE 6 | Rad-score box plots. **(A)** Rad-score box plot for 4-IHC classification ($p = 4.2355e-46$). **(B)** Rad-score box plot for TN vs. non-TN ($p = 3.9100e-17$).

formulas for the two classification tasks were calculated via a linear combination of selected features that were weighted by the respective coefficients (**Supplementary Materials 4, 5**, and **Figure 6**).

The results for model performance of the two classification tasks are listed in **Table 2**. For the 4-IHC classification task, the accuracy of the dataset for different classifiers ranged from 50.7 to 72.4%. The best models with the highest accuracy were the LDA and subspace discriminant of ensemble learning. Notably, only a small dependence emphasis of Kendall-tau-b feature for $DWI_{sequential}$ with the LDA model yielded the highest accuracy of 53.7%.

When comparing the TN to the non-TN cancers, the accuracy of the test dataset ranged from 72.4 to 91.0%. The linear SVM and medium KNN each yielded the highest accuracy of 91.0%. With only the maximum of variance feature for $DWI_{sequential}$ with the linear SVM model, we obtained an accuracy of 83.6%.

DISCUSSION

We investigated whether different breast cancer subtypes could be differentiated on multiparametric MR radiomics. Texture features extracted from DCE- and DWI-related original images, particularly those sequential features changing over time

points or several *b*-values, could quantify the heterogenous differences between the IHC subtypes. Although this is early work, multiparametric MR radiomics has the potential to provide a non-invasive approach and further insight into tumor imaging phenotypes. Moreover, the methodology may assist in the differentiation of details of subtypes that are imperceptible to the human eye and is also free of inter- or intra-observer variability (9).

Multiparametric imaging assessing tumor heterogeneity has recently been extended to help identify the subtype of breast cancer (25–28). The combined MR texture features could make the process of subtype classification quantitative and yield increased accuracy. Waugh et al. extracted 220 GLCM features in 72 IDC patients and found that the accuracy for subtype classification was 57.2% using all features and 43.6% using only the entropy feature (29). Sutton et al. extracted the first order and GLCM texture features from 178 IDC cancers to distinguish three subtypes, including luminal-like, HER2+, and TN (30). Their study analyzed only one slice showing the largest lesion diameter on the DCE images, whereas the necrotic tissue was excluded in Waugh's study. Agner et al. used the morphologic and texture features extracted from the whole tumor on the early postcontrast images in conjunction with an SVM classifier, to yield an area under the curve (AUC) of 0.74 for the TN

TABLE 2 | Model performance for the test datasets of two classification tasks.

	Classifiers	Accuracy (4-IHC classification)	Accuracy (TN vs. non-TN)
Decision tree	Fine tree	69.4%	87.3%
	Medium tree	69.4%	87.3%
	Coarse tree	64.9%	88.8%
Discriminant analysis	LDA	72.4%	90.3%
	QDA	70.9%	90.3%
SVM	Linear	67.9%	91.0%
	Quadratic	69.4%	89.6%
	Cubic	59.7%	72.4%
	Fine gaussian	68.7%	87.3%
	Median gaussian	68.7%	89.6%
	Coarse gaussian	70.1%	88.1%
KNN	Fine	69.4%	87.3%
	Medium	64.9%	91.0%
	Coarse	50.7%	83.6%
	Weighted	70.1%	88.1%
Ensemble learning	Subspace discriminant	72.4%	90.3%
	Subspace KNN	69.4%	87.3%
	AdaBoost tree	67.2%	83.6%
	Bagged tree	70.1%	88.1%
	RUSBoost tree	58.2%	85.8%

LDA, linear discriminant analysis; QDA, quadratic discriminant analysis; SVM, support vector machines; KNN, k-nearest neighbor; Bagged, bootstrap-aggregated; IHC, Immunohistochemical; TN, triple negative.

vs. non-TN cancers (14). In our study, whole-tumor texture features extracted from DCE and DWI-related original images and those changing over six time points or three b-values yielded the best accuracy of 72.4% for the 4-IHC classification task, and the best accuracy of 91.0% for the TN vs. non-TN cancers. One possible interpretation for our model performance is that texture features capture the information of whole-tumor about perfusion, diffusion, and heterogeneity.

Compared with DCE, DWI was relatively less frequently performed in the radiomics workflow due to a lower resolution and more distortion. DWI reflecting the microenvironment of tumor structures was introduced to be a complementary tool that provided additional tissue information (31). Most studies used the histogram or texture analysis on ADC maps with a mono-exponential fit (15, 32, 33). In addition, there is one study for which the authors performed histogram analysis on maps from mono-exponential DWI and biexponential intravoxel incoherent motion (IVIM) to predict the HER2 status in ER-positive breast cancer (33). These parametric maps were calculated based on different diffusion models. The applicability of several diffusion models in clinical practice remains controversial. Instead of maps, the original DWI images were used to extracted features in our study, and classification using only minor dependence emphasis on Kendall-tau-b for DWI$_{sequential}$ yielded an accuracy of 53.7% for the 4-IHC classification. Furthermore, an accuracy

of 83.6% for the TN vs. non-TN cancers was obtained using the maximum of variance for DWI$_{sequential}$ alone. Such a finding may indicate that TN breast cancer was more heterogeneous compared with other subtypes. A possible explanation for the findings was that TN demonstrated more necrosis (34). In our study, the DWI images with three b-values were obtained using a 16-channel breast coil and relatively accurate b-values (35, 36). The features extracting for DWI may therefore offer a new strategy for reconsidering the role of DWI in radiomics. Examples of the four IHC subtypes are shown in **Figure 7**.

Multiple studies performed texture analysis on images from different systems and institutions, including varied protocols (14, 29, 37). The breast MR imaging protocol has a variety of acquisition parameters, such as spatial resolution and temporal resolution, dynamic time points, repetition time and echo time, b-values, numbers of average and acceleration factors. All these variables affect texture features, and consequently the robustness of the radiomics classification model. For the widespread use of texture analysis in clinical applications, the approach to overcome the limitation that texture features are sensitive to the imaging parameters and modalities should be considered (38). In this investigation, we extracted radiomic features from patients on both the DCE and DWI images with two new sequential images, including DCE images changing over several time points, and DWI images changing over different b-values, which have not been used or described before in the domain of breast radiomics. Further studies on whether these normalized features minimized the variations introduced by different reconstruction algorithms and scanning parameters between different MR vendors and institutions will be validated.

Furthermore, a novel two-stage method of feature selection was performed by combining traditional statistics with machine learning. The coarse feature selection aimed to choose a relatively small subset of the features by removing the present features of redundant, noisy, and irrelevant dimensions (39, 40). Because of the differing performance of lasso regression, ridge regression, and elastic net, we combined the three methods to select intersecting features for robustness and stability during the fine feature selection step (41). The number of fine features selected in models depends on the difficulty of the classification problem (42). In addition, the classification accuracy obtained by combining the three methods was higher compared with that obtained by using any single one of these three methods. Similarly, Li et al. combined the Mann-Whitney U-test and SVM to select informative genes (43).

This study had several limitations. First, it was a retrospective, single-institution study. Second, the limited sample size and the imbalance distribution of breast subtypes may not have trained the classification models sufficiently in such a retrospective design. Third, the same data were used for training and testing. The model may perform differently if multicenter datasets were used. Further prospective study using a larger and more balanced population from multiple centers and a subset of the data as a validation set is required. Fourth, the number of b-values for DWI to assess the texture features based on DWI image change over multiple b-values, was relatively low. Further validation on original DWI with more b-values in the IVIM (44) or
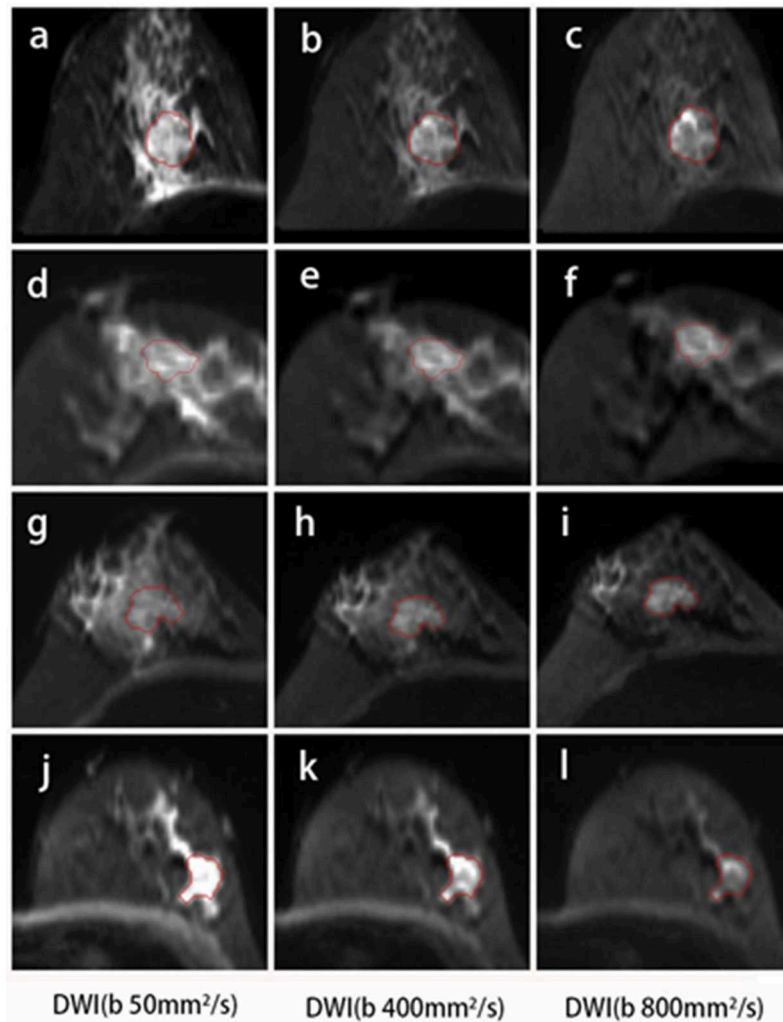
FIGURE 7 | DWI images for the four subtypes of breast cancer. A 46-year-old female with Luminal A breast cancer **(A–C)**; a 60-year-old female with Luminal B breast cancer **(D–F)**; a 47-year-old female with HER2-positive breast cancer **(G–I)**; and a 57-year-old female with TN breast cancer **(J–L)**. The small dependence emphasis of Kendall-tau-b for DWI$_{sequential}$ for Luminal A, Luminal B, HER2-positive, and TN breast cancers were −0.915, 0.358, 0.915, and 0.299, respectively.

diffusion kurtosis imaging (45) are needed. Finally, portions of the current work were performed offline. To implement this method into routine clinical practice, the automatic registration and segmentation tools, as well as the radiomics tools, will need to be integrated in the same platform.

In conclusion, we have shown that machine learning-based analysis of multiparametric MR radiomics assessing biological characteristics and heterogeneity of the whole tumor can effectively differentiate different subtypes of breast cancer and identify TN cancers. This approach could serve as a more convenient and non-invasive biomarker for the prediction of breast cancer subtypes.

DATA AVAILABILITY

The datasets for this manuscript are not publicly available because the data of texture features and subtypes for all the patients in our study was uploded in the **Supplementary Table 1**.

The raw images for patients can not allowed to upload for the public access. Requests to access the datasets should be directed to pengweijun2017@163.com.

ETHICS STATEMENT

This retrospective study was approved by the review board of Fudan University Shanghai 96 Cancer Center, which waived informed consent.

AUTHOR CONTRIBUTIONS

TX, ZW, WP, and HW conceived and designed the experiments. TX, ZW, QZ, YG, and HW analyzed the imaging data. QB and XZ reviewed the molecular data. TX

wrote the manuscript. All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.00505/full#supplementary-material

REFERENCES

1. Martelotto LG, Ng CK, Piscuoglio S, Weigelt B, Reis-Filho JS. Breast cancer intra-tumor heterogeneity. *Breast Cancer Res.* (2014) 16:210. doi: 10.1186/bcr3658
2. Zambetti M, Mansutti M, Gomez P, Lluch A, Dittrich C, Zamagni C. Pathological complete response rates following different neoadjuvant chemotherapy regimens for operable breast cancer according to ER status, in two parallel, randomized phase II trials with an adaptive study design (ECTO II) *Breast Cancer Res Treat.* (2012) 132:843–51. doi: 10.1007/s10549-011-1660-6
3. Bhargava R, Beriwal S, Dabbs DJ, Ozbek U, Soran A, Johnson RR. Immunohistochemical surrogate markers of breast cancer molecular classes predicts response to neoadjuvant chemotherapy: a single institutional experience with 359 cases. *Cancer.* (2010) 116:1431–9. doi: 10.1002/cncr.24876
4. Cleator S, Heller W, Coombes RC. Triple-negative breast cancer: therapeutic options. *Lancet Oncol.* (2007) 8:235–44. doi: 10.1016/S1470-2045(07)70074-8
5. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes–dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol.* (2011) 22:1736–47. doi: 10.1093/annonc/mdr304
6. Koo HR, Cho N, Song IC, Kim H, Chang JM, Yi A. Correlation of perfusion parameters on dynamic contrast-enhanced MRI with prognostic factors and subtypes of breast cancers. *J Magn Reson Imaging.* (2012) 36:145–51. doi: 10.1002/jmri.23635
7. Youk JH, Son EJ, Chung J, Kim JA, Kim EK. Triple-negative invasive breast cancer on dynamic contrast-enhanced and diffusion-weighted MR imaging: comparison with other breast cancer subtypes. *Eur Radiol.* (2012) 22:1724–34. doi: 10.1007/s00330-012-2425-2
8. Schmitz AM, Loo CE, Wesseling J, Pijnappel RM, Gilhuijs KG. Association between rim enhancement of breast cancer on dynamic contrast-enhanced MRI and patient outcome: impact of subtype. *Breast Cancer Res Treat.* (2014) 148:541–51. doi: 10.1007/s10549-014-3170-9
9. Mazurowski MA. Radiogenomics: what it is and why it is important. *J Am Coll Radiol.* (2015) 12:862–6. doi: 10.1016/j.jacr.2015.04.019
10. Grimm LJ. Breast MRI radiogenomics: Current status and research implications. *J Magn Reson Imaging.* (2016) 43:1269–78. doi: 10.1002/jmri.25116
11. Blaschke E, Abe H. MRI phenotype of breast cancer: kinetic assessment for molecular subtypes. *J Magn Reson Imaging.* (2015) 42:920–4. doi: 10.1002/jmri.24884
12. Chang RF, Chen HH, Chang YC, Huang CS, Chen JH, Lo CM. Quantification of breast tumor heterogeneity for ER status, HER2 status, and TN molecular subtype evaluation on DCE-MRI. *Magn Reson Imaging.* (2016) 34:809–19. doi: 10.1016/j.mri.2016.03.001
13. Grimm LJ, Zhang J, Mazurowski MA. Computational approach to radiogenomics of breast cancer: luminal A and luminal B molecular subtypes are associated with imaging features on routine breast MRI extracted using computer vision algorithms. *J Magn Reson Imaging.* (2015) 42:902–7. doi: 10.1002/jmri.24879

14. Agner SC, Rosen MA, Englander S, Tomaszewski JE, Feldman MD, Zhang P, et al. Computerized image analysis for identifying triple-negative breast cancers and differentiating them from other molecular subtypes of breast cancer on dynamic contrast-enhanced MR images: a feasibility study. *Radiology.* (2014) 272:91–9. doi: 10.1148/radiol.14121031
15. Choi Y, Kim SH, Youn IK, Kang BJ, Park WC, Lee A. Rim sign and histogram analysis of apparent diffusion coefficient values on diffusion-weighted MRI in triple-negative breast cancer: Comparison with ER-positive subtype. *PLoS ONE.* (2017) 12:e0177903. doi: 10.1371/journal.pone.0177903
16. Fan M, He T, Zhang P, Cheng H, Zhang J, Gao X. Diffusion-weighted imaging features of breast tumours and the surrounding stroma reflect intrinsic heterogeneous characteristics of molecular subtypes in breast cancer. *NMR Biomed.* (2018) 31:e3869. doi: 10.1002/nbm.3869
17. Grady L. Random walks for image segmentation. *IEEE Trans Pattern Anal Mach Intell.* (2006) 28:1768–83. doi: 10.1109/TPAMI.2006.233
18. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
19. Soyster AL. Convex programming with set-inclusive constraints and applications to inexact linear-programming. *Oper Res.* (1973) 21:1154–7. doi: 10.1287/opre.21.5.1154
20. Cristianini N, Scholkopf B. Support vector machines and kernel methods - The new generation of learning machines. *AI Mag.* (2002) 23:31–41. doi: 10.1609/aimag.v23i3.1655
21. Ressom HW, Varghese RS, Zhang Z, Xuan J, Clarke R. Classification algorithms for phenotype prediction in genomics and proteomics. *Front Biosci.* (2008) 13:691–708. doi: 10.2741/2712
22. Rokach L. Ensemble-based classifiers. *Artif Intell Rev.* (2010) 33:1–39. doi: 10.1007/s10462-009-9124-7
23. Samworth RJ. Optimal weighted nearest neighbour classifiers. *Ann Stat.* (2012) 40:2733–63. doi: 10.1214/12-AOS1049
24. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J.* (2015) 13:8–17. doi: 10.1016/j.csbj.2014.11.005
25. Chen JH, Agrawal G, Feig B, Baek HM, Carpenter PM, Mehta RS. Triple-negative breast cancer: MRI features in 29 patients. *Ann Oncol.* (2007) 18:2042–3. doi: 10.1093/annonc/mdm504
26. Schrading S, Kuhl CK. Mammographic, US, and MR imaging phenotypes of familial breast cancer. *Radiology.* (2008) 246:58–70. doi: 10.1148/radiol.2461062173
27. Loo CE, Straver ME, Rodenhuis S, Muller SH, Wesseling J, Vrancken Peeters MJ. Magnetic resonance imaging response monitoring of breast cancer during neoadjuvant chemotherapy: relevance of breast cancer subtype. *J Clin Oncol.* (2011) 29:660–6. doi: 10.1200/JCO.2010.31.1258
28. Martincich L, Deantoni V, Bertotto I, Redana S, Kubatzki F, Sarotto I. Correlations between diffusion-weighted imaging and breast cancer biomarkers. *Eur Radiol.* (2012) 22:1519–28. doi: 10.1007/s00330-012-2403-8
29. Waugh SA, Purdie CA, Jordan LB, Vinnicombe S, Lerski RA, Martin P. Magnetic resonance imaging texture analysis classification of primary breast cancer. *Eur Radiol.* (2016) 26:322–30. doi: 10.1007/s00330-015-3845-6
30. Sutton EJ, Dashevsky BZ, Oh JH, Veeraraghavan H, Apte AP, Thakur SB. Breast cancer molecular subtype classifier that incorporates MRI

features. *J Magn Reson Imaging*. (2016) 44:122–9. doi: 10.1002/jmri.25119

31. Marino MA, Helbich T, Baltzer P, Pinker-Domenig K. Multiparametric MRI of the breast: a review. *J Magn Reson Imaging*. (2018) 47:301–15. doi: 10.1002/jmri.25790

32. Mori N, Ota H, Mugikura S, Takasawa C, Ishida T, Watanabe G. Luminal-type breast cancer: correlation of apparent diffusion coefficients with the Ki-67 labeling index. *Radiology*. (2015) 274:66–73. doi: 10.1148/radiol.14140283

33. Vidic I, Egnell L, Jerome NP, Teruel JR, Sjobakk TE, Ostlie A. Support vector machine for breast cancer classification using diffusion-weighted MRI histogram features: preliminary study. *J Magn Reson Imaging*. (2018) 47:1205–16. doi: 10.1002/jmri.25873

34. Uematsu T, Kasami M, Yuen S. Triple-negative breast cancer: correlation between MR imaging and pathologic findings. *Radiology*. (2009) 250:638–47. doi: 10.1148/radiol.2503081054

35. Bogner W, Gruber S, Pinker K, Grabner G, Stadlbauer A, Weber M. Diffusion-weighted MR for differentiation of breast lesions at 3.0 T: how does selection of diffusion protocols affect diagnosis? *Radiology*. (2009) 253:341–51. doi: 10.1148/radiol.2532081718

36. Thomassin-Naggara I, De Bazelaire C, Chopier J, Bazot M, Marsault C, Trop I. Diffusion-weighted MR imaging of the breast: advantages and pitfalls. *Eur J Radiol*. (2013) 82:435–43. doi: 10.1016/j.ejrad.2012.03.002

37. Mazurowski MA, Zhang J, Grimm LJ, Yoon SC, Silber JI. Radiogenomic analysis of breast cancer: luminal B molecular subtype is associated with enhancement dynamics at MR imaging. *Radiology*. (2014) 273:365–72. doi: 10.1148/radiol.14132641

38. Mayerhoefer ME, Szomolanyi P, Jirak D, Materka A, Trattnig S. Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study. *Med Phys*. (2009) 36:1236–43. doi: 10.1118/1.3081408

39. Bolón-Canedo V, Sánchez-Maroño N, Alonso-Betanzos A, Benítez JM, Herrera F. A review of microarray datasets and applied feature selection methods. *Inf Sci*. (2014) 282:111–35. doi: 10.1016/j.ins.2014.05.042

40. Wang L, Wang Y, Chang Q. Feature selection methods for big data bioinformatics: a survey from the search perspective. *Methods*. (2016) 111:21–31. doi: 10.1016/j.ymeth.2016.08.014

41. Ogutu JO, Schulz-Streeck T, Piepho HP. Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proc*. (2012) 6(Suppl. 2):S10. doi: 10.1186/1753-6561-6-S2-S10

42. Lu CF, Hsu FT, Hsieh KL, Kao YJ, Cheng SJ, Hsu JB. Machine learning-based radiomics for molecular subtyping of gliomas. *Clin Cancer Res*. (2018) 24:4429–36. doi: 10.1158/1078-0432.CCR-17-3445

43. Li HD, Liang YZ, Xu QS, Cao DS, Tan BB, Deng BC. Recipe for uncovering predictive genes using support vector machines based on model population analysis. *IEEE/ACM Trans Comput Biol Bioinform*. (2011) 8:1633–41. doi: 10.1109/TCBB.2011.36

44. Cho GY, Moy L, Kim SG, Baete SH, Moccaldi M, Babb JS. Evaluation of breast cancer using intravoxel incoherent motion (IVIM) histogram analysis: comparison with malignant status, histological subtype, and molecular prognostic factors. *Eur Radiol*. (2016) 26:2547–58. doi: 10.1007/s00330-015-4087-3

45. Sun K, Chen X, Chai W, Fei X, Fu C, Yan X. Breast cancer: diffusion kurtosis MR imaging-diagnostic accuracy and correlation with clinical-pathologic factors. *Radiology*. (2015) 277:46–55. doi: 10.1148/radiol.15141625