Journal of Medical and Biological Engineering (2024) 44:749–762
https://doi.org/10.1007/s40846-024-00900-9

ORIGINAL ARTICLE

# Predicting Immunohistochemical Biomarkers of Breast Cancer Using $^{18}$F-FDG PET/CT Radiomics: A Multicenter Study

**Sahar Faraji[1] · Farshad Emami[2] · Zahra Vosoughi[3] · Ghasem Hajianfar[4] · Shahrokh Naseri[1] · Rezvan Samimi[5] · Habibeh Vosoughi[6] · Parham Geramifar[6] · Habib Zaidi[4,7,8,9]**

Received: 20 July 2024 / Accepted: 3 September 2024 / Published online: 26 September 2024


## Abstract

**Purpose** This study aimed at predicting four important immunohistochemical biomarkers, including estrogen receptor, progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki67 (cell proliferation rate index)) in breast cancer using radiomic features derived from multicentric $^{18}$F-FDG PET/CT images.

**Methods** Sixty-two patients with locally advanced breast cancer who underwent $^{18}$F-FDG PET/CT imaging before any treatment were included. Three different PET/CT scanner models were used to acquire the images. After tumor segmentation, radiomic features from PET and CT images were extracted using the Python PyRadiomics package. Fusion features were created at the feature level, including concatenation (Con) and averaging (Avg). Combat was applied for features harmonization. The area under the curve (AUC), sensitivity, and specificity were used to evaluate the performance of predictive models.

**Results** Random Forest (RF) model in Con features with mean AUC of $0.69 \pm 0.11$, Support Vector Machine (SVC) model in radiomic features from CT with a mean AUC of $0.74 \pm 0.02$ were outstanding in predicting ER and PR, respectively. The best models for predicting HER2 were RF and SVC using CT images, with mean AUC of $0.72 \pm 0.04$ and $0.73 \pm 0.03$. Respectively. In addition, Ki67 was predicted on radiomic features derived from PET images by RF and SVC models with mean AUC of $0.8 \pm 0.09$ and $0.83 \pm 0.03$, respectively.

**Conclusion** Machine learning classifiers based on PET, CT, and PET/CT radiomic features could be correlated with the immunohistochemical biomarkers in breast cancer.

**Keywords** Breast Cancer · Immunohistochemical Biomarkers · PET/CT · Radiomics · Machine Learning

✉ Habibeh Vosoughi
vosoughi64@gmail.com

✉ Parham Geramifar
pgeramifar@tums.ac.ir

1 Department of Medical Physics, Mashhad University of Medical Science, Mashhad, Iran

2 Nuclear Medicine and Molecular Imaging Department, Imam Reza International University, Razavi Hospital, Mashhad, Iran

3 Department of Medical Physics, School of Medicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

4 Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, 1211 Geneva 4, Switzerland

5 Khatam PET-CT Center, Khatam Hospital, Tehran, Iran

6 Research Center for Nuclear Medicine, Tehran University of Medical Science, Shariati Hospital, North Kargar Ave.1411713135, Tehran, Iran

7 Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, 9700 RB Groningen, Netherlands

8 Department of Nuclear Medicine, University of Southern Denmark, 500 Odense, Denmark

9 University Research and Innovation Center, Óbuda University, Budapest, Hungary

## 1 Introduction

Breast Cancer (BC) ranks second compared to other cancers in terms of disease prevalence. In addition, BC is known as the most frequent cancer all over the world. This disease is still considered the fifth most common cause of death in women, despite being investigated in its early stages, thanks to improvements in diagnostic tests [1]. BC is an extremely heterogeneous disease [2], and this heterogeneity is believed to be a significant agent leading to cancer therapy defeat and weak prognosis [3, 4]. In addition to the fact that the heterogeneity of breast tumors varies between different patients, heterogeneity in different parts of a tumor has also been reported for one patient with BC [5]. The evaluation of heterogeneity depends on distinguishing different biomarkers, utilizing tissue biopsy-based immunohistochemistry [6, 7]. A careful biological assessment of BC is critical since every subtype has a unique biological and genetic profile, resulting in different prognoses and treatment choices. Because subtypes have a potential impact on clinical management and are determined by different molecular specifications, propagation rates, tumor receptors, and grades [8]. BC biopsies consistently assess four immunohistochemical (IHC) biomarkers: estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki67 (cell proliferation rate index) [8, 9]. Based on IHC studies, BC is currently classified into four molecular subtypes: Luminal A(ER +, PR +, HER2–, Ki67 $<$ 20%), Luminal B/HER2– (ER +, PR +, HER2–, Ki67 $\geq$ 20%), Luminal B/HER2 + (ER +, PR +, HER2 +, any Ki67), HER2 + /non-luminal(ER–, PR–, HER2 +), Triple-negative(ER–, PR–, HER2–) [10]. Hence, biomarkers are of great importance for tailoring treatment procedures to each patient as part of the personalized medicine approach [11]. However, it is impossible to sample an area to accurately determine tumor heterogeneity using invasive biopsy-based methods. Accurate diagnosis of heterogeneity requires sampling from multiple areas which leads to patient complications [1, 2, 12]. In order to determine the type of treatment, a non-invasive examination of the heterogeneity of all parts of the tumor region seems necessary [2]. In medical image analysis, radiomics is a popular topic because it can extract many hidden features in image voxels that are normally invisible to the human eye. The relationship between these extracted features and different tumor phenotypes is identified helping data mining [3, 13]. Non-invasiveness and evaluation of all parts of the tumor are important advantages of radiomics-based methods [14, 15].

$^{18}$F-FDG PET/CT imaging is a valuable modality for staging, detecting recurrence, and treatment response evaluation of BC patients [16]. The Standardized uptake value (SUV) is a biological marker used for quantitative analysis of PET images [17]. However, the maximum SUV ($SUV_{max}$) is a single voxel value representing the maximum $^{18}$F-FDG uptake in the tumor and as such, it doesn't reflect whole tumor metabolism, particularly in heterogeneous BC lesions [18]. Therefore, newer and better imaging features that more accurately evaluate tumor heterogeneity are needed to support individualized treatment [19]. Radiomics methods allow for the extraction of more quantitative features pertaining to tumor intensity, shape, and tissue parameters from different medical imaging modalities. Machine learning-based methods currently widely used for data analysis, use features extracted by radiomics-based methods to quantify pixel intensity heterogeneity and provide more detailed information about tumors [20].

Most previous studies applied machine learning techniques to determine the molecular subtypes or the status of IHC factors using various medical imaging modalities, including CT [21], MRI [22], ultrasound [23], mammography [24], with a small number of studies using PET/CT images. For instance, Liu et al. [25] utilized PET and CT radiomic features to prognosticate the molecular subtypes of BC. In addition to CT and PET alone, Chen et al. [26] trained machine learning models by combining PET and CT features (fusion features), which were rarely investigated in previous studies. It is worth noting that all previous PET/CT studies were single-centered. Therefore, this research aims to assess the correlation between multi-centric radiomic features (PET only, CT only, PET/CT fusion) extracted from $^{18}$F-FDG PET/CT images acquired on different scanners and IHC factors (ER, PR, HER2, and Ki67) in patients with BC.

## 2 Materials and Methods

Figure 1 summarizes the different steps followed in our study. Initially, PET/CT images and pathology reports related to IHC factors (ER, PR, HER2, and Ki67) of eligible patients were collected from three nuclear medicine departments. In the next step, radiomic features were extracted from regions of interests (ROIs) that were segmented from SUV PET and CT images. Subsequently, to regulate the batch effect because of the multicentric data set, ComBat harmonization was implemented on every feature set. Then, machine learning models were assessed to predict the status of IHC factors.

### 2.1 Patients

A total of 62 BC patients among 335 were retrospectively collected from three different PET/CT centers equipped with three various models, namely a Biograph 6 TrueV,

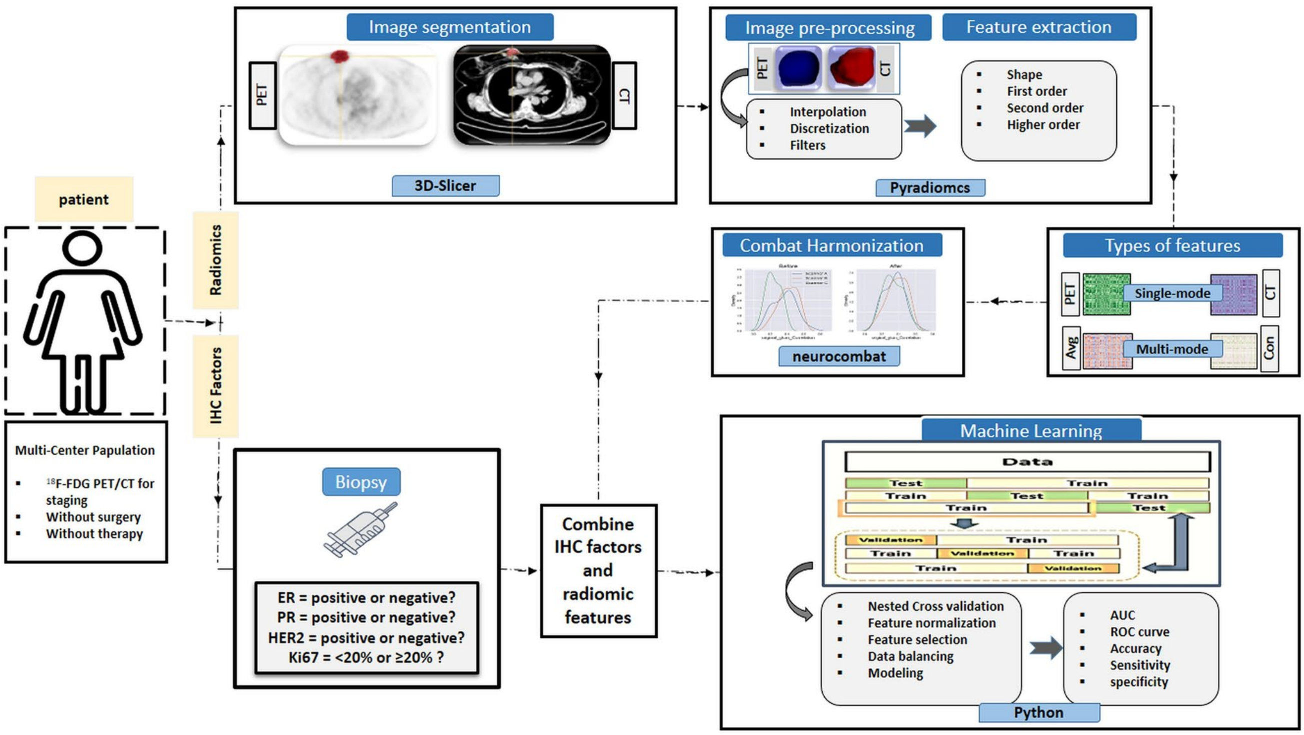


**Fig. 1** Schematic workflow adopted in the present study which encompasses data collection, image segmentation, feature extraction, and use of machine learning algorithms to find the best predictive model

Biograph mCT, and Biograph Horizon (Siemens Healthineers, Knoxville, USA). Inclusion/exclusion criteria used to select included patients are: (1) Women with proven BC who underwent $^{18}$F-FDG PET/CT for staging without any surgery or therapy, (2) patients having pathological reports related to the expression of ER, PR, HER2, and Ki67, (3) omitting patients with prior surgery or treatment and unknown IHC factors status from the study. Figure 2 shows the steps followed to select the patients enrolled in this study.

## 2.2 Image Acquisition and Reconstruction

Whole-body PET/CT imaging was performed 60 min after intravenous injection of $^{18}$F-FDG**.** All PET/CT images were acquired according to each scanner's acquisition parameters. CT images were reconstructed using filtered backprojection (FBP) with 512×512 matrix size. The iterative ordered subset expectation maximization (OSEM) algorithm was utilized to reconstruct PET images with predefined numbers of iterations and subsets followed by Gaussian post-reconstruction smoothing filter. Routine reconstruction settings in all three scanners included 2 iterations, 21 subsets and 3 mm FWHM Gaussian filter on Biograph 6 TrueV, 2 iterations, 24 subsets with 4 mm of Gaussian filter for Biograph mCT, and 4 iterations, 10subsets, 4 mm FWHM Gaussian filter on the Biograph Horizon. In addition, normalization, scatter, attenuation, decay corrections, and resolution recovery were applied.

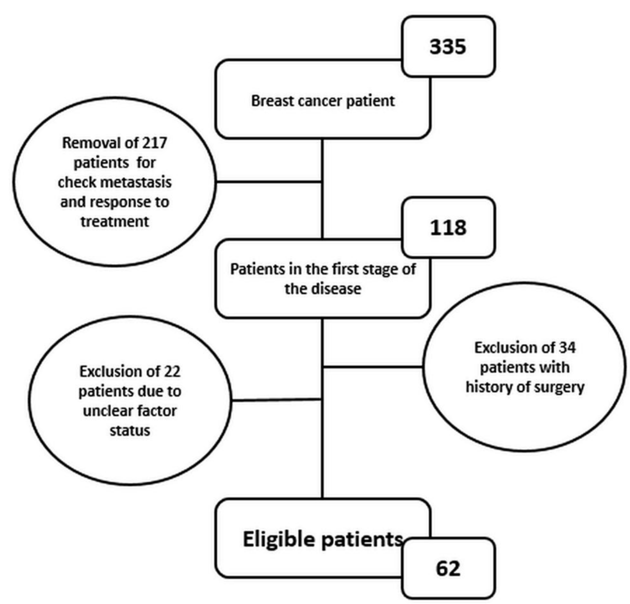


**Fig. 2** Eligible patients according to inclusion/exclusion criteria in this study protocol

## 2.3 Image Segmentation

Tumor delineation and segmentation was performed manually by an experienced nuclear medicine physician using the 3D-Slicer software (version 5.2.1). This was done separately on CT and PET images. All PET images (in Bq/ml) were converted to SUV PET images and the resulting images adopted for radiomic features extraction.

## 2.4 Image Pre-Processing and Feature Extraction

Image pre-processing was performed using the Python PyRadiomics package (version 3.7) [27] for PET and CT images separately based on the Biomarker Imaging Standardization Initiative (IBSI) guidelines [28]. Interpolation to isotropic voxel size, discretization of intensity values, and two Laplacian of Gaussian (LOG) and Wavelet filters were applied to CT and PET images. The voxel spacing was interpolated to $4 \times 4 \times 4$ mm$^3$ for PET images from three different scanners. Image intensity was discretized into 0.1 SUV as fixed bin width. All CT images were interpolated into $1 \times 1 \times 1$ isotropic voxel and were discretized into 64 bins. For each PET and CT image, a LOG filter was used with different sigma values (0.5, 1, 1.5, 2, and 2.5) to create five filtered images. In addition, after applying the wavelet filter, PET and CT images were generated in eight frequency ranges; HHH, HHL, HLH, LHH, HLL, LHL, LLH, and LLL. Indeed, low-frequency and high-frequency signals are abbreviated as L and H, respectively. Hence, for instance, LHL is an image volume made using a low-pass filter (L) on the X-axis, a high-pass filter (H) on the Y-axis, and a low-pass filter (L) on the Z-axis [29].

A total of 1288 radiomic features were extracted from the segmented volumes of interest (VOIs) of CT and PET images, including shape, first-, second-, and higher-order statistics. Among the 1288 features, 105 features were associated with the original images without filters, 728 and 455 features were related to images with Wavelet and LOG filters, respectively. Following the extraction of radiomic features from only PET and only CT images, for feature-level fusion, radiomic features from CT and PET images were averaged (Avg) and concatenated (Con). The fusion radiomic features of Con were calculated from the combination of all PET and CT radiomic features, and the fusion radiomic features of Avg were calculated from the average of CT radiomic features and PET radiomic features. Finally, 1288 Avg features and 2576 Con features were created.

## 2.5 Harmonization of Multicenter Data

ComBat was used to reduce the variability of radiomic features. The ComBat model, initially proposed in expression

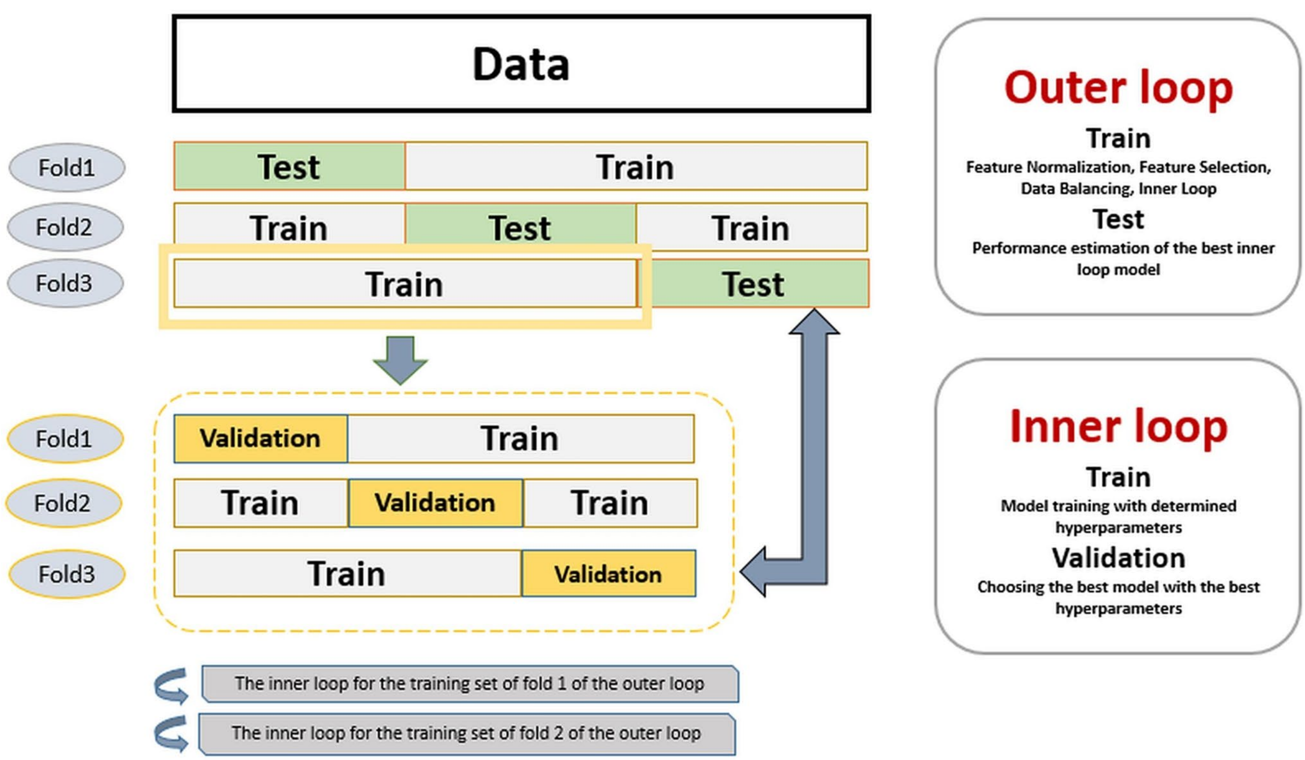


**Fig. 3** Splitting data using nested cross-validation techniques and applying different stages of machine learning consisting of feature normalization, feature selection, data balancing, and modeling to different parts of the data

**Table 1** Clinical characteristics of BC patients

| Characteristics | Biograph6 TrueV (Scanner A) | Biograph Horizon (Scanner B) | Biograph mCT (Scanner C) | Total |
|---|---|---|---|---|
| Median age (range) | 50 (40–80) | 54 (33–73) | 57 (33–79) | 53 |
| ER | | | | |
| Positive | 24 | 12 | 13 | 49 |
| Negative | 6 | 3 | 4 | 13 |
| PR | | | | |
| Positive | 21 | 9 | 13 | 44 |
| Negative | 9 | 6 | 4 | 18 |
| HER2 | | | | |
| Positive | 14 | 3 | 3 | 20 |
| Negative | 11 | 11 | 10 | 32 |
| Ki67 | | | | |
| $<20\%$ | 7 | 4 | 5 | 16 |
| $\geq 20\%$ | 16 | 10 | 12 | 38 |
| Location | | | | |
| Right breast mass | 17 | 8 | 5 | 30 |
| Left breast mass | 13 | 7 | 12 | 32 |
| Histological type | | | | |
| Invasive Ductal Carcinoma | 17 | 3 | 14 | 34 |
| Other | 13 | 12 | 3 | 28 |

analysis by Johnson et al. can reduce the influence of multicentric data [30, 31]. Batch or site effects in radiomic studies can be considered when there is a desire to pool data emanating from different scanner models at various sites, collected using different acquisition and reconstruction protocols [32]. Consequently, it is highly recommended to harmonize features from different datasets before aggregating them to generate consistent and robust models using multicenter datasets [33]. To effectively remove multicenter effects from radiomic features, ComBat creates batch-specific transformations to align all data in a common space, eliminating center effects [34, 35]. This approach ensures that the generated models are not biased by differences in data acquisition or reconstruction parameters across the multiple centers involved in the study.

### 2.6 Machine Learning Models

PET, CT, and PET/CT models were constructed to predict IHC markers. These models were compared based on their performance in anticipating the IHC factors. All machine learning steps were implemented in Python. According to the sample size, using the Nested Cross-Validation (NCV) method, all data were divided to train and test for model building. The NCV method includes an inner cross-validation loop that is nested in an outer cross-validation loop [36, 37]. In this study, NCV with threefold cross-validation for the inner and outer loop was used to split the data into training, validation, and test datasets. The MinMaxScaler method was used to standardize the features. Feature standardization was performed to change the scope of each feature; because many machine learning algorithms do not perform well when trained with widely varying ranges of input features [38]. Two feature selection methods were considered to select optimal features including Recursive Feature Elimination Cross Validation (RFEcv), and Minimum Redundancy and Maximum Relevance (MRMR). Twelve optimal features were selected by MRMR and at least 4 optimal features were chosen using the RFEcv method. Balancing the distribution of samples was done by the synthetic minority oversampling technique (SMOTE). Imbalance issues can affect model learning performance [39]. After these steps in the outer training set, the GridSearchCV method was employed in the inner loop to select the best hyperparameters of the models. Five machine learning classification algorithms were invoked in each outer loop, including Support Vector Classifier (SVC), Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbors (KNN). After the selection of the best hyperparameters for all five classification models in the inner loop, it was evaluated on the outer loop test data. In each outer loop, the performance of five algorithms with their best hyperparameters was checked with the receiver operating characteristic (ROC) curve, the area under the curve (AUC), accuracy, sensitivity, and specificity. Finally, the three best algorithms of each outer loop for each five

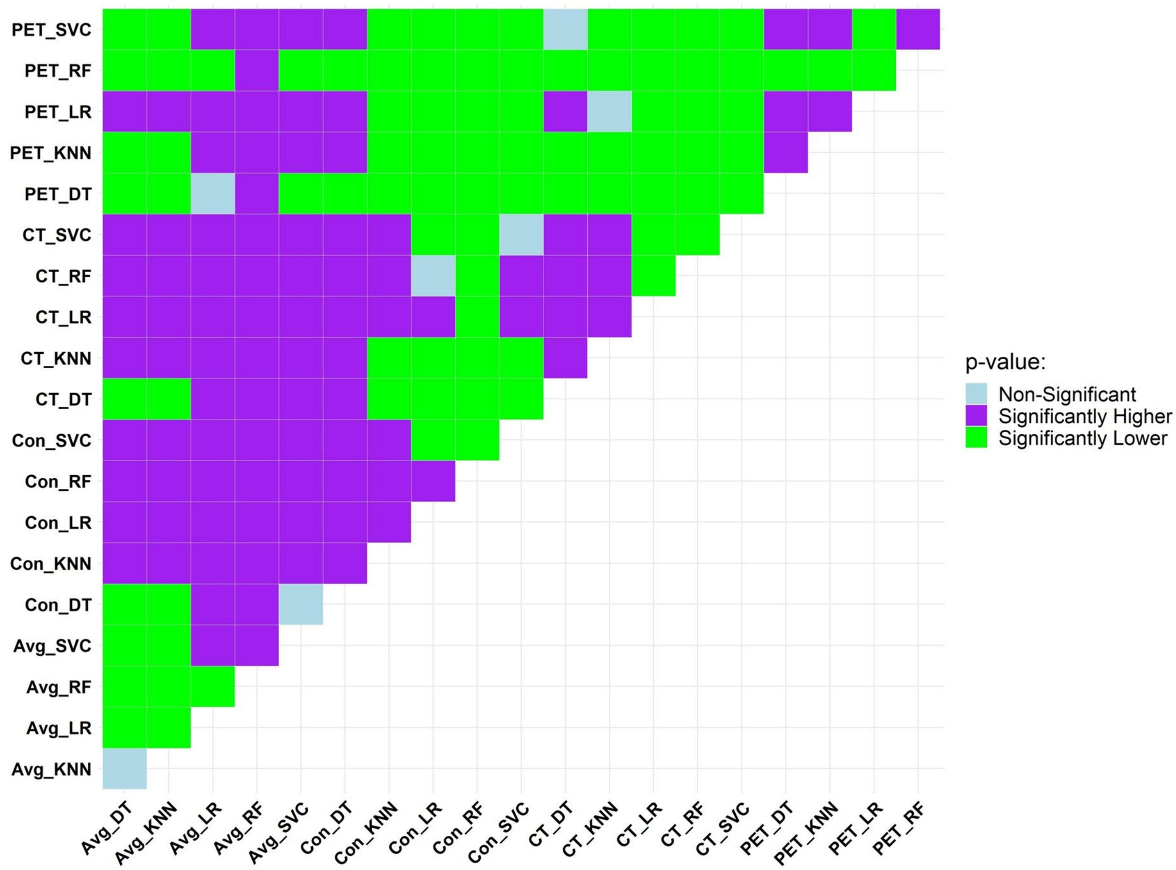


**Fig. 4** Wilcoxon test on predictive models of ER. The model's AUCs were appraised against one another in columns and rows. Light blue: If the row model did not significantly outperform the column model in terms of p-values. Purple: the row model performed significantly higher than the column model in terms of p-values. Green: the performance of the row model is significantly lower than that of the column model in terms of p-values

models were compared with each other. Hyperparameters were opted according to scikit-learn. All steps of machine learning in this study are displayed in Fig. 3.

### 2.7 Statistical Analysis

The AUCs obtained from the prediction algorithms were compared using the Wilcoxon statistical test. The statistical significance level was determined considering a threshold of 0.05. According to the P-value, three modes including significantly lower, significantly higher, and non-significant differences were defined. These results represent which models are the best predictive models for IHC factors. All analyses were accomplished using R statistical software (version 3.6).

## 3 Results

### 3.1 Clinical Characteristics

Among 62 patients referred for PET/CT, 30 images from Biograph 6 TrueV (scanner A), 15 images from Biograph Horizon (scanner B), and 17 images from Biograph mCT (scanner C) were included. Patients' clinical information is summarized in Table 1.

### 3.2 Efficacy of Predictive Models and Related Radiomic Features for ER, PR, HER2, and Ki67 Factors

The results related to the best predictive models of ER, PR, HER2, and Ki67 factors were separately obtained. Since several machine learning models were used to predict these factors, the highest-performance model should have been recognized. Therefore, the Wilcoxon test was carried out on the AUC of models to identify the superior model with the highest certainty. The comparison of all models for ER,

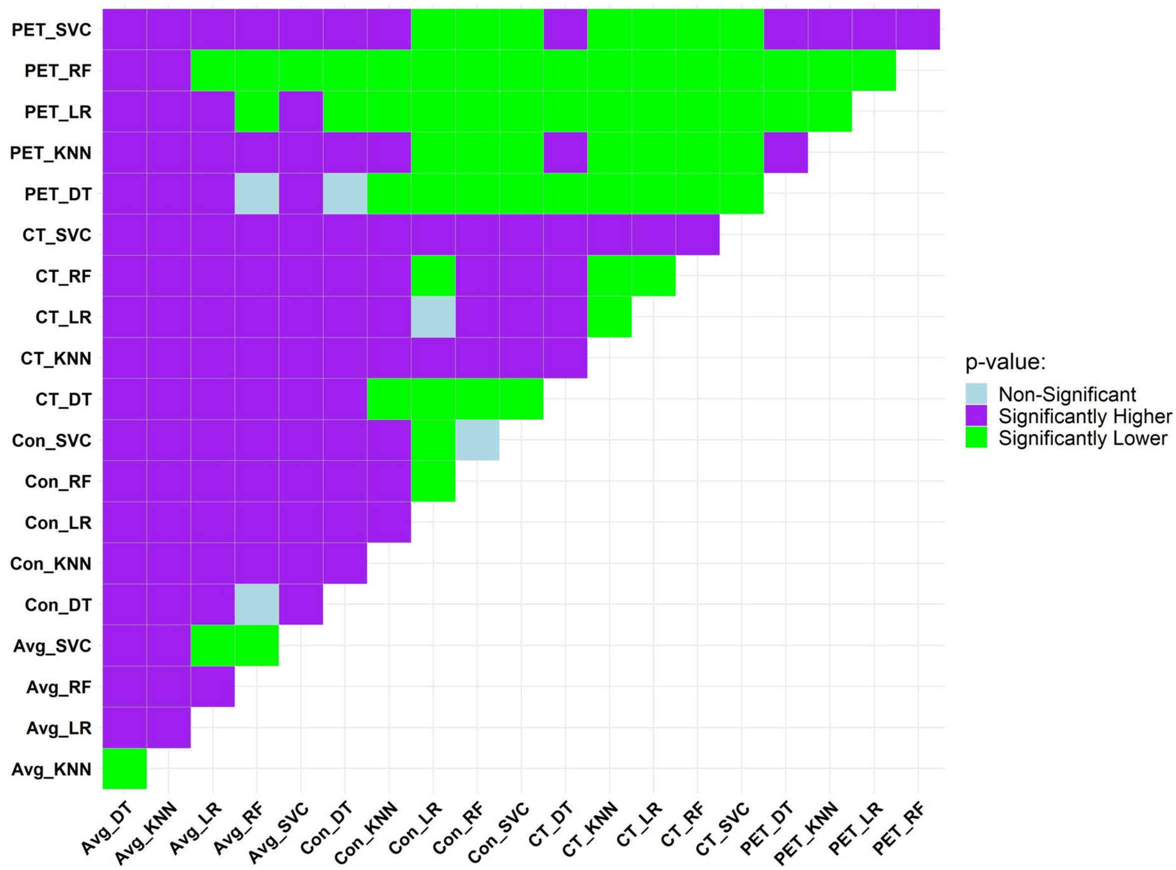


**Fig. 5** Wilcoxon test on predictive models of PR

PR, HER2, and Ki67 factors are demonstrated in Figs. 4–7, respectively. Conventional metrics including AUC, Accuracy, Sensitivity, Specificity, and ROC curve were reported for superior models.

#### 3.2.1 Statistical Analysis of ER

As shown in Fig. 4, the RF algorithm developed from Con features showed a significantly higher difference than all the compared models in the prediction of ER. This model was the best predictive model with an average AUC of $0.69 \pm 0.11$.

#### 3.2.2 Statistical Analysis of PR

To assess the relationship between features extracted from PET, CT, Avg, and Con with PR expression, the SVC model in CT features outperformed other algorithms with a mean AUC of $0.74 \pm 0.02$ (Fig. 5).

#### 3.2.3 Statistical Analysis of HER2

According to Fig. 6, two algorithms, namely SVC and RF models of CT features, were found to have a higher significance than the others and not significant to each other. Therefore, both algorithms could be adopted as the most predictive models for HER2. Average values of AUC of $0.72 \pm 0.04$ and $0.73 \pm 0.03$ were achieved for the SVC and the RF model, respectively.

#### 3.2.4 Statistical Analysis of Ki67

As is evident from Fig. 7, it was found that the most predictive models were RF and SVC algorithms from PET features with average values of $0.8 \pm 0.09$ and $0.83 \pm 0.03$ for AUC, respectively. These two algorithms had significantly higher difference compared to other predictive models without significant difference to each other. Both models could be accepted as superior ones for Ki67 prediction.

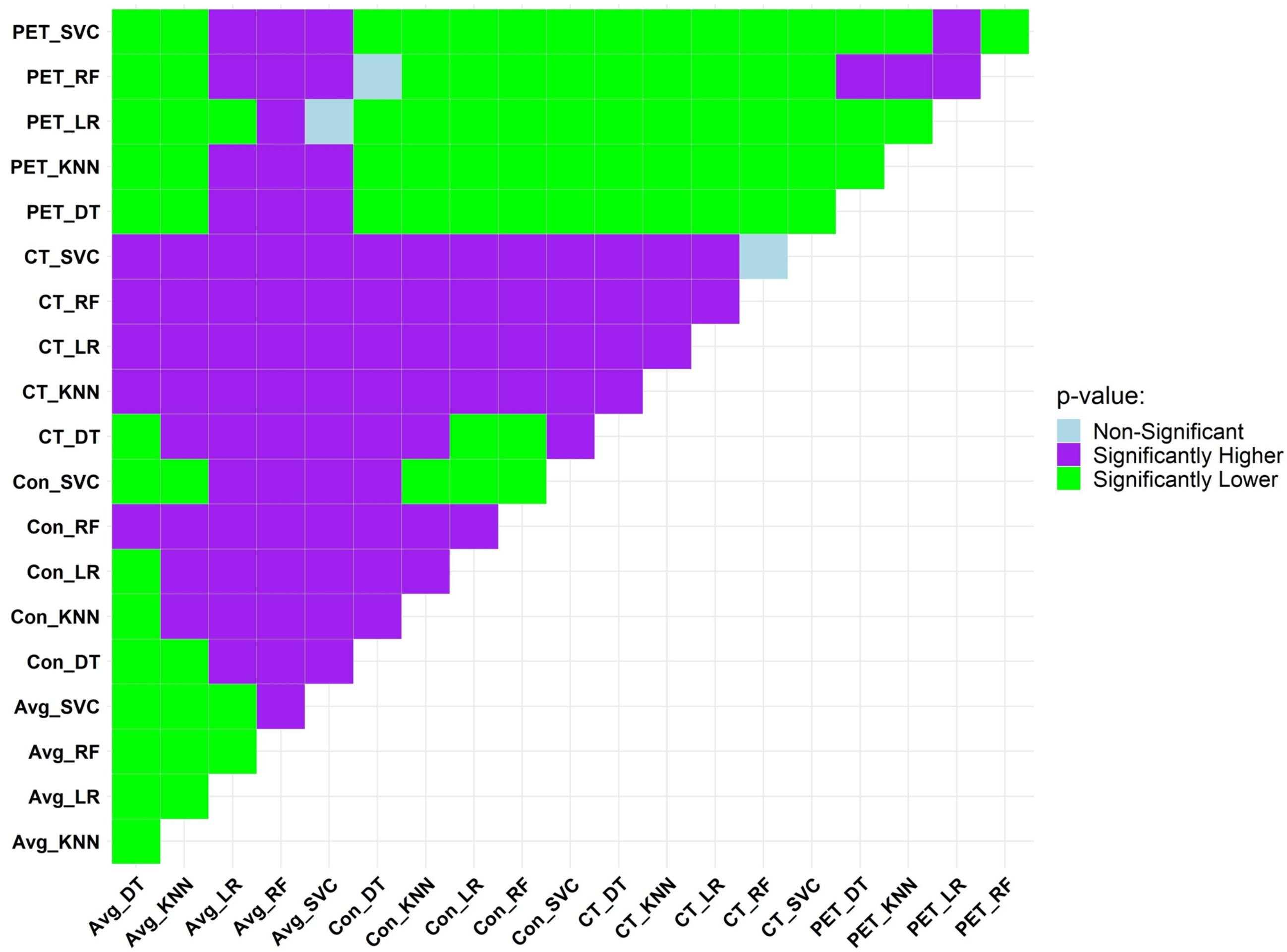


**Fig. 6** Wilcoxon test on predictive models of HER2

#### 3.2.5 Selected Features Associated with Factors

The most relevant selected features to train the best predictive models for ER, PR, HER2, and Ki67 factors in each fold of the outer loop are listed in Table 2.

#### 3.2.6 Conventional Metrics for the Best Models of Factors

The values of AUC, accuracy, sensitivity, and specificity from each fold in the outer loop of the superior prediction algorithms for ER, PR, HER2, and Ki67 are listed in Table 3. According to Fig. 8, the ROC curves of the predictive models for IHC factors were illustrated. The AUC of each fold in the outer loop and the average of three folds (Mean ± STD) are displayed.

## 4 Discussion

Various studies reported that the use of a radiomics framework based on PET/CT images significantly improves the efficiency of treatment options for patients with BC [40, 41]. In the present study, given the extremely important role of IHC factors in treatment selection and the importance of a comprehensive assessment of breast mass heterogeneity, machine learning algorithms were used to predict four key IHC factors (ER, PR, HER2, and Ki67). Five classification algorithms containing SVC, LR, DT, KNN, and RF models based on the four types of radiomic features, including CT, PET, Avg, and Con were created. In predicting PR and ER, CT and Con features performed best by building an SVC and RF, respectively. Two SVC and RF models were introduced as superior predictors for HER2 used in CT features and for Ki67 used in PET features.

A comprehensive radiomic analysis using CT, PET, and PET/CT fusion (Avg, Con) was conducted in this study. While prior studies examined the correlation of radiomic

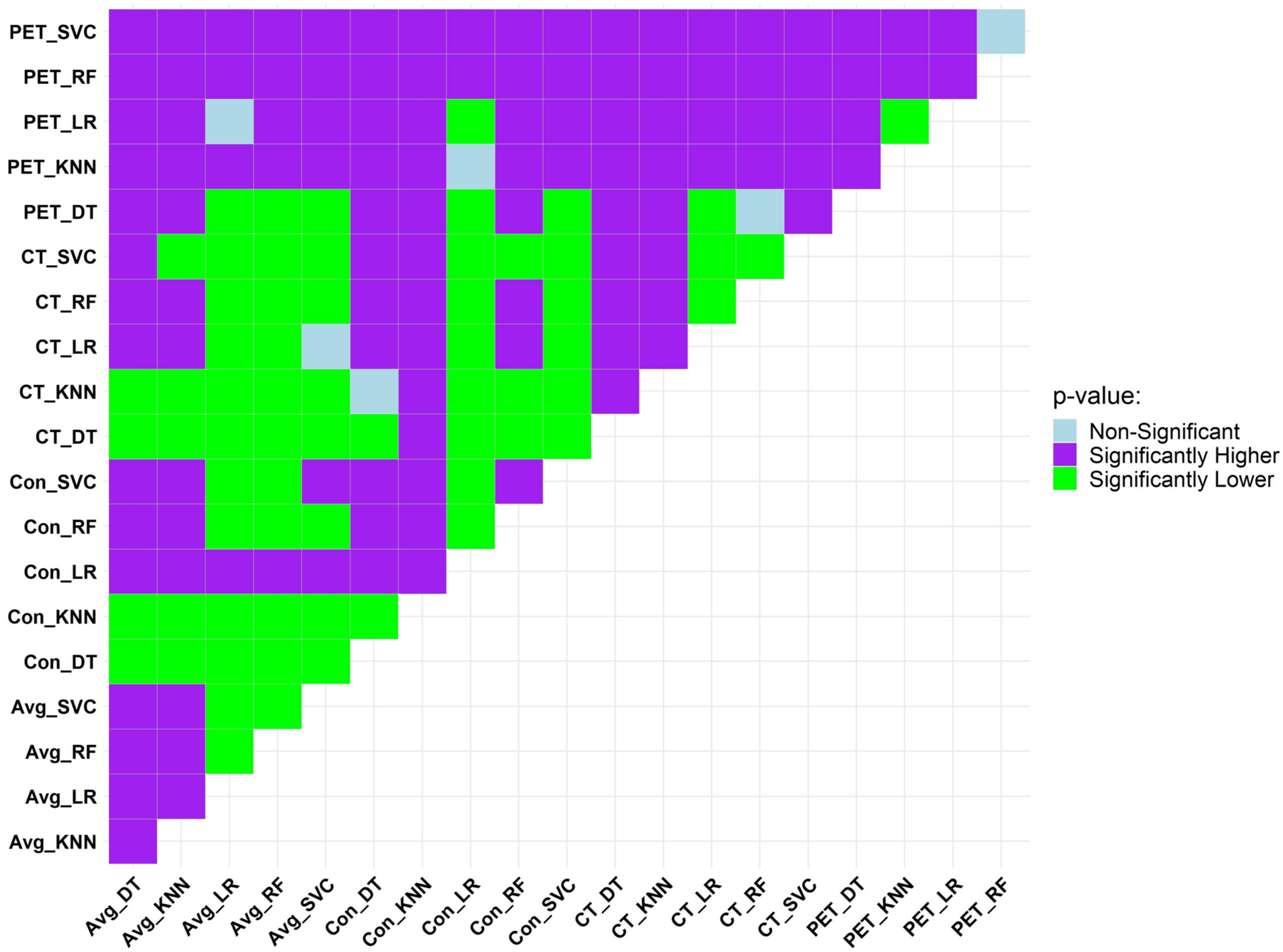


**Fig. 7** Wilcoxon test on predictive models of Ki67

features based on PET scans only and molecular subtypes of BC [10, 42–44]. In addition, Liu's study evaluated only CT and PET images separately [25]. Contrary to the results reported by Yang et al. [45], addressing only the distinction between ER + and ER–, our study included the analysis of ER + and ER–, PR + versus PR–, HER2 + and HER2–, and Ki67 ≥ 20% versus Ki67 < 20%.

For this study, we employed the MRMR and then the RFEcv algorithms to choose the optimal features. Ultimately, only features with the highest value and the least dependence on each other remained. To compare our study with other PET/CT studies, Chen et al. used five different methods [26]. These included the intraclass correlation coefficient (ICC), value variance, Mann–Whitney U test, decision tree, and variance inflation factor (VIF) for selecting radiomic features, while Liu et al. utilized a three-step process to identify reliable radiomic features [25]. In our study, the wavelet filter was found to yield a significantly higher percentage of selected features for each IHC factor compared to the Gaussian filter and original features (The percentages for wavelet, Gaussian, and original are 77%, 15%, and 8%, respectively). Since the computing time is fast in the wavelet domain and the independence of the coefficients is not considered, this filter can provide efficient results compared to other filters [46].

In the present study, in ER and PR, GLSZM features were the repetitive features included in all folds. The analysis showed that HER2 was associated with first-order features more frequently, whereas Chen's study [26] reported GLCM features as more dominant. Finally, in our study Ki67 was associated with NGTDM features more often.

One study using machine learning models was performed by Chen et al. to predict HER2 using PET/CT features. SVC and RF models achieved AUC of 0.6 and 0.61, respectively [26]. Our results of HER2 prediction showed SVC with mean AUC of 0.72 ± 0.04 and RF with mean AUC of 0.73 ± 0.03 based on CT features were the best models. In addition, in Chen's et al. study, XGboost model with AUC of 0.76 of Avg features and AUC of 0.72 of Con features significantly outperformed other models in predicting HER2. The results of Con features coincide well with

**Table 2** Top selected features to train the best models

| Feature (Factor) | Fold1 | Fold2 | Fold3 |
|---|---|---|---|
| Con ( ER) | CT.wavelet-HLH_glszm_HighGrayLevelZoneEmphasis<br>PET.wavelet-HHH_firstorder_Kurtosis<br>CT.wavelet-LHL_glszm_SizeZoneNonUniformityNormalized<br>CT.wavelet-HLH_glszm_LowGrayLevelZoneEmphasis<br>PET.wavelet-LHL_glcm_Correlation<br>CT.wavelet-HHL_glszm_SmallAreaLowGrayLevelEmphasis<br>PET.original_firstorder_Skewness<br>CT.wavelet-HLH_glszm_GrayLevelVariance<br>CT.wavelet-LHH_glszm_SizeZoneNonUniformityNormalized | CT.wavelet-HLH_glszm_HighGrayLevelZoneEmphasis<br>CT.wavelet-HLH_glszm_GrayLevelVariance<br>CT.wavelet-HLL_ngtdm_Contrast<br>CT.wavelet-HLH_glszm_LowGrayLevelZoneEmphasis<br>CT.log-sigma-2–0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis<br>CT.wavelet-HLH_glszm_GrayLevelNonUniformityNormalized<br>PET.wavelet-HHH_gldm_DependenceVariance<br>PET.wavelet-HLH_firstorder_Kurtosis<br>CT.wavelet-LHL_glrlm_ShortRunLowGrayLevelEmphasis | PET.wavelet-LHL_glcm_ClusterShade<br>CT.wavelet-LHL_glszm_SmallAreaEmphasis<br>CT.wavelet-LHH_firstorder_RobustMeanAbsoluteDeviation<br>PET.original_firstorder_Skewness<br>CT.original_shape_Flatness<br>PET.log-sigma-2–0-mm-3D_firstorder_Maximum<br>CT.wavelet-HLL_glszm_SmallAreaEmphasis<br>CT.wavelet-HHL_glcm_SumEntropy |
| CT (PR) | wavelet-LHL_glrlm_ShortRunLowGrayLevelEmphasis<br>log-sigma-2–0-mm-3D_glcm_ClusterShade<br>wavelet-HHL_glszm_LowGrayLevelZoneEmphasis<br>original_glszm_LowGrayLevelZoneEmphasis<br>wavelet-HLL_glszm_SmallAreaEmphasis<br>wavelet-HLL_glcm_SumSquares<br>wavelet-HLH_glszm_LowGrayLevelZoneEmphasis<br>wavelet-LHL_glrlm_LowGrayLevelRunEmphasis<br>wavelet-HLH_glszm_HighGrayLevelZoneEmphasis<br>original_glrlm_ShortRunLowGrayLevelEmphasis<br>wavelet-LHL_gldm_LowGrayLevelEmphasis<br>original_shape_Flatness | wavelet-HLH_glszm_LowGrayLevelZoneEmphasis<br>wavelet-LHL_glszm_GrayLevelNonUniformityNormalized<br>wavelet-LLH_gldm_DependenceEntropy<br>wavelet-HHL_glszm_LowGrayLevelZoneEmphasis<br>original_glrlm_ShortRunLowGrayLevelEmphasis<br>wavelet-HLL_glrlm_ShortRunLowGrayLevelEmphasis<br>wavelet-HLH_glszm_HighGrayLevelZoneEmphasis<br>log-sigma-2–5-mm-3D_firstorder_InterquartileRange<br>wavelet-LHL_glszm_SmallAreaEmphasis<br>wavelet-HHL_glszm_SizeZoneNonUniformityNormalized<br>wavelet-HHL_glszm_HighGrayLevelZoneEmphasis<br>wavelet-LHL_firstorder_Maximum | log-sigma-1–5-mm-3D_ngtdm_Contrast<br>wavelet-HLH_glszm_GrayLevelNonUniformityNormalized<br>wavelet-LHL_glrlm_ShortRunLowGrayLevelEmphasis<br>wavelet-HHH_glszm_SmallAreaLowGrayLevelEmphasis<br>wavelet-HHL_glszm_SizeZoneNonUniformityNormalized<br>log-sigma-2–0-mm-3D_glcm_Idmn<br>wavelet-HLH_glszm_LowGrayLevelZoneEmphasis<br>wavelet-HHL_glszm_ZoneEntropy<br>log-sigma-1–5-mm-3D_glcm_Idmn<br>wavelet-HLH_glszm_HighGrayLevelZoneEmphasis |
| CT (HER2) | log-sigma-1–5-mm-3D_glrlm_LongRunLowGrayLevelEmphasis<br>wavelet-LLL_firstorder_Minimum<br>wavelet-HLH_firstorder_RootMeanSquared<br>log-sigma-1–5-mm-3D_glrlm_LongRunEmphasis<br>log-sigma-2–5-mm-3D_glcm_Imc1<br>wavelet-LLH_glszm_GrayLevelNonUniformityNormalized<br>wavelet-LLH_glszm_GrayLevelVariance<br>wavelet-HHH_glcm_Autocorrelation<br>wavelet-HHL_firstorder_Mean<br>wavelet-LHH_glszm_HighGrayLevelZoneEmphasis | wavelet-LLH_glszm_GrayLevelNonUniformityNormalized<br>log-sigma-2–5-mm-3D_ngtdm_Coarseness<br>wavelet-HLL_glcm_ClusterProminence<br>wavelet-LLH_glszm_LowGrayLevelZoneEmphasis<br>wavelet-HLH_firstorder_Median<br>log-sigma-0–5-mm-3D_firstorder_InterquartileRange<br>wavelet-HHH_firstorder_Kurtosis<br>wavelet-HHL_firstorder_RootMeanSquared<br>log-sigma-1–5-mm-3D_glrlm_LongRunLowGrayLevelEmphasis<br>wavelet-LLH_gldm_LargeDependenceEmphasis<br>wavelet-HLL_glszm_LargeAreaHighGrayLevelEmphasis | wavelet-HLH_firstorder_RootMeanSquared<br>wavelet-HLL_glcm_ClusterProminence<br>wavelet-HHL_firstorder_Median<br>wavelet-LLL_firstorder_Minimum<br>wavelet-HLH_firstorder_Mean<br>wavelet-HHH_firstorder_Kurtosis<br>wavelet-HHH_glcm_Autocorrelation<br>wavelet-HHL_glszm_SmallAreaLowGrayLevelEmphasis<br>wavelet-LHH_glrlm_LongRunLowGrayLevelEmphasis<br>wavelet-LLH_glszm_GrayLevelNonUniformityNormalized |

**Table 2** (continued)

| Feature (Factor) | Fold1 | Fold2 | Fold3 |
|---|---|---|---|
| PET (Ki67) | wavelet-LHL_ngtdm_Coarseness<br>wavelet-HHH_firstorder_Skewness<br>log-sigma-0–5-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis<br>wavelet-LHH_glszm_SmallAreaLowGrayLevelEmphasis<br>wavelet-HLL_ngtdm_Coarseness<br>wavelet-HHL_ngtdm_Coarseness<br>wavelet-HLH_ngtdm_Coarseness<br>wavelet-HLL_firstorder_InterquartileRange<br>original_ngtdm_Busyness<br>wavelet-LHH_ngtdm_Coarseness | log-sigma-0–5-mm-3D_glszm_LowGrayLevelZoneEmphasis<br>wavelet-HLL_ngtdm_Coarseness<br>log-sigma-1–0-mm-3D_glszm_ZoneEntropy<br>wavelet-LHL_ngtdm_Coarseness | wavelet-HHH_glcm_MaximumProbability<br>wavelet-HLH_glcm_Idn<br>log-sigma-0–5-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis<br>log-sigma-0–5-mm-3D_glcm_MaximumProbability<br>wavelet-LHL_gldm_SmallDependenceLowGrayLevelEmphasis<br>wavelet-HHH_glszm_SmallAreaLowGrayLevelEmphasis<br>original_glszm_SizeZoneNonUniformityNormalized<br>wavelet-LLL_firstorder_Skewness<br>wavelet-LHH_glszm_SmallAreaLowGrayLevelEmphasis<br>wavelet-HHH_glcm_InverseVariance<br>wavelet-LHL_ngtdm_Coarseness |

**Table 3** Summary of the metrics calculated for each fold on the outer loop for the best predictive models of IHC factors

| Factor | Feature type | Algorithm | Mean_AUC | fold | AUC | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|---|---|---|
| ER | Con | RF | 0.69 ± 0.11 | 1 | 0.72 | 0.81 | 0.94 | 0.25 |
| | | | | 2 | 0.54 | 0.62 | 0.63 | 0.6 |
| | | | | 3 | 0.8 | 0.85 | 0.94 | 0.5 |
| PR | CT | SVC | 0.74 ± 0.02 | 1 | 0.72 | 0.71 | 0.87 | 0.33 |
| | | | | 2 | 0.77 | 0.76 | 0.8 | 0.67 |
| | | | | 3 | 0.74 | 0.65 | 0.71 | 0.5 |
| HER2 | CT | SVC | 0.72 ± 0.04 | 1 | 0.68 | 0.56 | 0.73 | 0.29 |
| | | | | 2 | 0.76 | 0.65 | 0.5 | 0.86 |
| | | | | 3 | 0.74 | 0.76 | 0.8 | 0.71 |
| | | RF | 0.73 ± 0.03 | 1 | 0.7 | 0.72 | 0.73 | 0.71 |
| | | | | 2 | 0.76 | 0.65 | 0.6 | 0.71 |
| | | | | 3 | 0.74 | 0.71 | 0.6 | 0.86 |
| Ki67 | PET | SVC | 0.8 ± 0.09 | 1 | 0.75 | 0.89 | 1 | 0.6 |
| | | | | 2 | 0.72 | 0.83 | 0.85 | 0.8 |
| | | | | 3 | 0.93 | 0.78 | 0.66 | 1 |
| | | RF | 0.83 ± 0.03 | 1 | 0.82 | 0.89 | 1 | 0.6 |
| | | | | 2 | 0.79 | 0.78 | 0.76 | 0.8 |
| | | | | 3 | 0.86 | 0.67 | 0.5 | 1 |

the results of the best models in our study. The modeling results of radiomic features in Liu et al. with one of the largest datasets on molecular subtyping PET/CT reported an average AUC value of 0.818 for HER2 + against HER2– subgroups [25]. Because the number of patients in our study is less than in Liu's study, the results of Liu's study are probably not comparable with our HER2 prediction results.

Similar investigations for the prediction of IHC factors have also been performed using digital mammography [47], and magnetic resonance imaging (MRI) [48, 49] radiomic features. The HER2 prediction results in the above study reported an AUC of 0.78 from the LR model [47] which had good agreement with our results (mean AUC values of 0.72 ± 0.04 and 0.73 ± 0.03 for SVC and RF models, respectively). Demircioglu et al. used RF, Naive Bayes (NB), LR, and NB models to achieve AUC values of 0.62, 0.69, 0.81, and 0.67 for predicting HER2, PR, Ki67, and ER, respectively [48]. Demircioglu's study had poor performance compared to the achieved results from the present study with mean AUC values of 0.73 ± 0.03 and 0.72 ± 0.04 for RF and SVC in HER2 prediction, and a mean AUC value of 0.74 ± 0.02 for SVC in PR prediction. According to Fig. 7, the results of ER and Ki67 prediction were in agreement with the results of Demircioglu's study [48]. Our research results on predicting Ki67 showed an AUC value of 0.8 ± 0.09 and 0.83 ± 0.03, which was higher than Liang's findings (AUC = 0.71) [49].

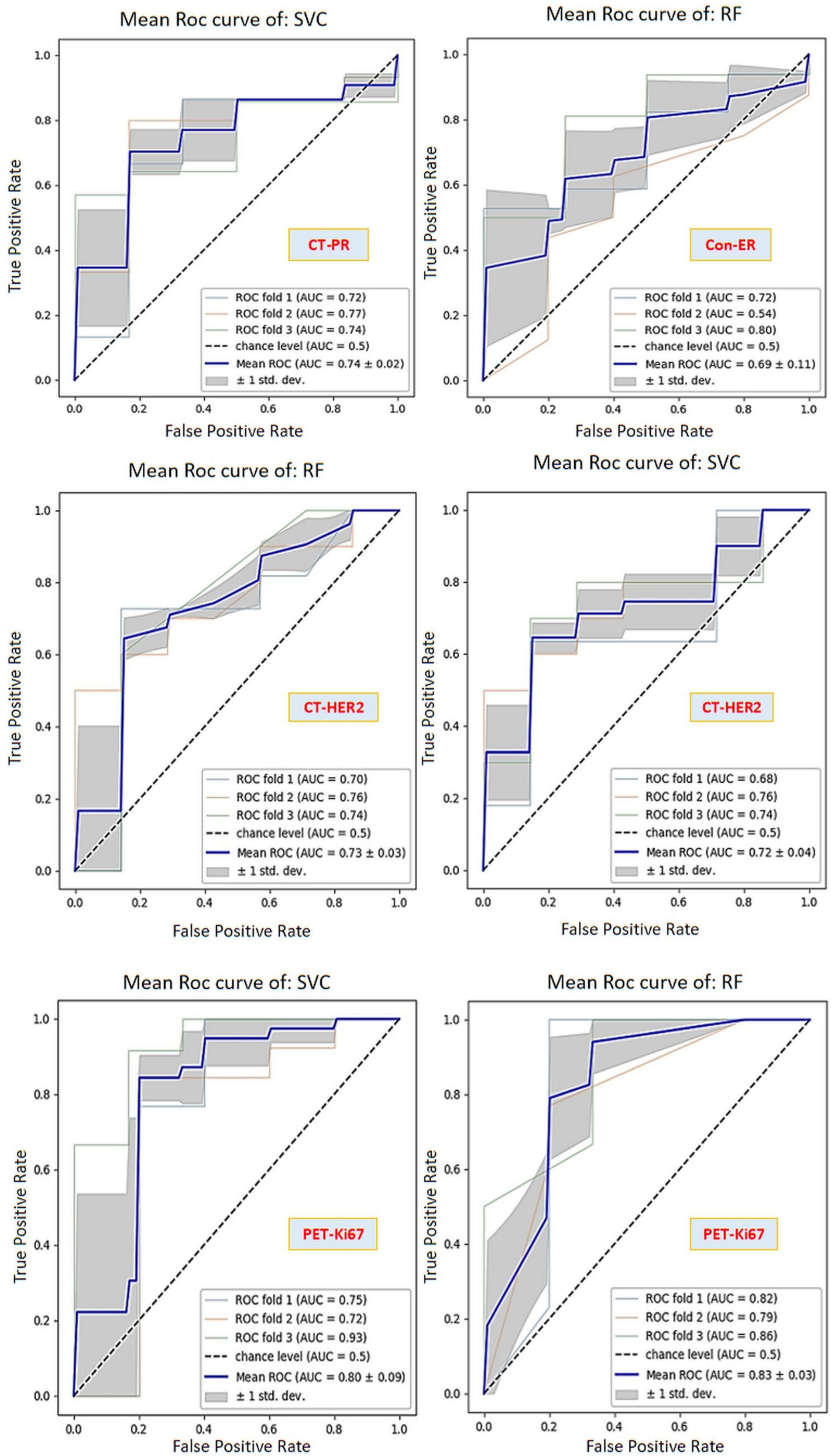


**Fig. 8** ROC curves of the best machine learning models for IHC factors

In general, combining PET/CT radiomics and machine learning could provide powerful biomarkers and models for predicting BC IHC factors. Meanwhile, these methods can be used repeatedly before and after therapy, contributing to individualized and precision medicine for BC patients. The study's main strength was the use of multi-centric radiomic features to introduce optimal models. This approach utilized harmonized feature sets despite the involvement of three independent institutions and different scanners, setting it apart from prior PET/CT studies conducted in a single center. However, one significant constraint of this study was the limited size of the datasets Thus, increasing the number of PET/CT centers to achieve a larger sample size will help confirming the study findings. In our study, some patients were excluded because the FISH test was not available to determine HER2 factor status. Therefore, future multi-center studies, combining various information to create a comprehensive model using a large dataset, could enhance the performance of models. The study only utilized binary classification for the prognostic models. Future multi-center studies could use regression models for Ki67 analyses and multiclass classification for HER2 (0, + 1, + 2, + 3) analyses. Another limitation of this study was that we did not simultaneously evaluate the impact of image acquisition or reconstruction parameters on radiomic features. Instead, we utilized the ComBat harmonization algorithm to mitigate multi-center parameter effects on radiomics features. The final limitation was that respiratory motion changed the tumor position in CT and PET images, and as such, segmentations were carried out independently for each modality.

## 5 Conclusions

The present research demonstrated that radiomic features extracted from $^{18}$F-FDG PET/CT images, combined with machine learning methods, hold promise in predicting IHC factors in BC. Our multicentric study revealed a good correlation between the expression of ER, PR, HER2, Ki67, and radiomic models. Nevertheless, it is important to note that future studies involving a larger cohort of patients could further enhance the development of more robust and powerful radiomic models for predicting IHC factors in breast cancer. Further research in this direction has the potential to advance our understanding and application of radiomics-based predictive models in clinical practice.

**Acknowledgements** We are grateful to the PET/CT centers for their cooperation; Nuclear Medicine Department of Razavi Hospital, Khatam PET/CT Center, and PET/CT Department of Mehraneh Charity Center.

**Author Contributions** Conception and design, Sahar Faraji, Farshad Emami, Ghasem Hajianfar, Habibeh Vosoughi and Parham Geramifar; Material preparation and data collection, Sahar Faraji, Farshad Emami, Rezvan Samimi, Habibeh Vosoughi and Parham Geramifar; data analyses, Sahar Faraji, Zahra Vosoughi, Ghasem Hajianfar and Shahrokh Naseri; funding acquisition, Shahrokh Naseri; writing, review and editing, Sahar Faraji, Ghasem Hajianfar, Habibeh Vosoughi, Parham Geramifar and Habib Zaidi. All authors read and approved the final manuscript.

**Funding** This study was part of a MSc thesis supported under grant number 4010663 by Mashhad University of Medical Sciences, Mashhad, Iran.

**Data Availability** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions.

## Declarations

**Conflicts of interest** The authors declare no conflict of interest.

**Informed Consent** Patient consent was waived due to the images of the patients were provided to us for analysis only, with their names and characteristics removed.

## References


1. Sollini, M., et al. (2021). PET/CT radiomics in breast cancer: Mind the step. *Methods, 188*, 122–132.
2. Koren, S., & Bentires-Alj, M. (2015). Breast tumor heterogeneity: Source of fitness, hurdle for therapy. *Molecular cell, 60*(4), 537–546.
3. Carrier-Vallières, M., *Radiomics: enabling factors towards precision medicine*2018: McGill University (Canada).
4. Burrell, R. A., et al. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature, 501*(7467), 338–345.
5. Grosselin, K., et al. (2019). High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nature genetics, 51*(6), 1060–1066.
6. Tselikas, L., et al. (2019). Role of image-guided biopsy and radiomics in the age of precision medicine. *Chinese Clinical Oncology, 8*(6), 02.
7. Agersborg, S., et al. (2018). Immunohistochemistry and alternative FISH testing in breast cancer with HER2 equivocal amplification. *Breast Cancer Research and Treatment, 170*, 321–328.
8. Tirada, N., et al. (2018). Breast cancer tissue markers, genomic profiling, and other prognostic factors: A primer for radiologists. *Radiographics, 38*(7), 1902–1920.
9. Fragomeni, S. M., Sciallis, A., & Jeruss, J. S. (2018). Molecular subtypes and local-regional control of breast cancer. *Surgical Oncology Clinics, 27*(1), 95–120.
10. Antunovic, L., et al. (2017). [18F] FDG PET/CT features for the molecular characterization of primary breast tumors. *European Journal of Nuclear Medicine and Molecular Imaging, 44*(12), 1945–1954.
11. Cain, E. H., et al. (2019). Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: A study using an independent validation set. *Breast cancer research and treatment, 173*, 455–463.
12. Jeong, J., et al., *Radiomics in cancer radiotherapy: a review.* arXiv preprint arXiv:1910.02102, 2019.
13. Urso, L., et al. (2022). PET-derived radiomics and artificial intelligence in breast cancer: A systematic review. *International Journal of Molecular Sciences, 23*(21), 13409.
14. Aerts, H. J., et al. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications, 5*(1), 1–9.

15. Hajianfar, G., et al. (2023). Prediction of ‘disease pathogenic variants using hybrid Machine learning systems and radiomic features. *Physica Medica, 113*, 102647.
16. Weber, W. A., Schwaiger, M., & Avril, N. (2000). Quantitative assessment of tumor metabolism using FDG-PET imaging. *Nuclear medicine and biology, 27*(7), 683–687.
17. Buvat, I., et al., *Quantitative nuclear medicine imaging: concepts, requirements and methods.* 2014.
18. Zhang, J., et al. (2013). The SUVmax for 18F-FDG correlates with molecular subtype and survival of previously untreated metastatic breast cancer. *Clinical nuclear medicine, 38*(4), 256–262.
19. Chicklore, S., et al. (2013). Quantifying tumour heterogeneity in 18 F-FDG PET/CT imaging by texture analysis. *European journal of nuclear medicine and molecular imaging, 40*, 133–140.
20. Uribe, C. F., et al. (2019). Machine learning in nuclear medicine: Part 1—introduction. *Journal of Nuclear Medicine, 60*(4), 451–458.
21. Park, E. K., et al. (2019). Machine learning approaches to radiogenomics of breast cancer using low-dose perfusion computed tomography: Predicting prognostic biomarkers and molecular subtypes. *Scientific reports, 9*(1), 17847.
22. Ha, R., et al. (2019). Predicting breast cancer molecular subtype with MRI dataset utilizing convolutional neural network algorithm. *Journal of digital imaging, 32*, 276–282.
23. Wu, T., et al. (2019). Machine learning for diagnostic ultrasound of triple-negative breast cancer. *Breast cancer research and treatment, 173*, 365–373.
24. Ma, W., et al. (2019). Breast cancer molecular subtype prediction by mammographic radiomic features. *Academic radiology, 26*(2), 196–201.
25. Liu, J., et al. (2021). Molecular subtype classification of breast cancer using established radiomic signature models based on 18 F-FDG PET/CT images. *Frontiers in Bioscience (Landmark Edition), 26*(9), 475–484.
26. Chen, Y., et al. (2022). Prediction of HER2 expression in breast cancer by combining PET/CT radiomic analysis and machine learning. *Annals of Nuclear Medicine, 36*(2), 172–182.
27. Van Griethuysen, J. J., et al. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer research, 77*(21), e104–e107.
28. Zwanenburg, A., et al. (2020). The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology, 295*(2), 328–338.
29. Whybra, P., et al. (2024). The image biomarker standardization initiative: standardized convolutional filters for reproducible radiomics and enhanced clinical insights. *Radiology, 310*(2), e231319.
30. Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics, 8*(1), 118–127.
31. Orlhac, F., et al. (2018). A postreconstruction harmonization method for multicenter radiomic studies in PET. *Journal of Nuclear Medicine, 59*(8), 1321–1328.
32. Mahon, R., et al. (2020). ComBat harmonization for radiomic features in independent phantom and lung cancer patient computed tomography datasets. *Physics in Medicine & Biology, 65*(1), 015010.
33. Zwanenburg, A. (2019). Radiomics in nuclear medicine: Robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *European journal of nuclear medicine and molecular imaging, 46*(13), 2638–2655.
34. Ishwaran, H., et al. (2011). Random survival forests for high-dimensional data. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 4*(1), 115–132.
35. Ishwaran, H., et al. (2010). High-dimensional variable selection for survival data. *Journal of the American Statistical Association, 105*(489), 205–217.
36. Vabalas, A., et al. (2019). Machine learning algorithm validation with a limited sample size. *PLoS ONE, 14*(11), e0224365.
37. Castiglioni, I., et al. (2021). AI applications to medical images: From machine learning to deep learning. *Physica Medica, 83*, 9–24.
38. Géron, A., *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*2022: " O’Reilly Media, Inc.".
39. Hatt, M., et al. (2023). Joint EANM/SNMMI guideline on radiomics in nuclear medicine : Jointly supported by the EANM physics committee and the SNMMI Physics, instrumentation and data sciences council. *European Journal of Nuclear Medicine and Molecular Imaging, 50*(2), 352–375.
40. Li, P., et al. (2020). 18 F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. *European Journal of Nuclear Medicine and Molecular Imaging, 47*, 1116–1126.
41. Tahmassebi, A., et al. (2019). Impact of machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotherapy and survival outcomes in breast cancer patients. *Investigative radiology, 54*(2), 110.
42. Önner, H., et al. (2022). Association of 18F-FDG PET/CT textural features with immunohistochemical characteristics in invasive ductal breast cancer. *Revista Española de Medicina Nuclear e Imagen Molecular (English Edition), 41*(1), 11–16.
43. Acar, E., et al. (2019). Comparison of the volumetric and radiomics findings of 18F-FDG PET/CT images with immunohistochemical prognostic factors in local/locally advanced breast cancer. *Nuclear Medicine Communications, 40*(7), 764–772.
44. Moscoso, A., et al. (2018). Texture analysis of high-resolution dedicated breast 18 F-FDG PET images correlates with immunohistochemical factors and subtype of breast cancer. *European journal of nuclear medicine and molecular imaging, 45*, 196–206.
45. Yang, Z., et al. (2017). The assessment of estrogen receptor status and its intratumoral heterogeneity in patients with breast cancer by using 18F-fluoroestradiol PET/CT. *Clinical nuclear medicine, 42*(6), 421–427.
46. Thakur, S.K. *Comparison of Filters used for Underwater Image Pre-Processing*. 2010.
47. Zhou, J., et al. (2019). Evaluating the HER-2 status of breast cancer using mammography radiomics features. *European Journal of Radiology, 121*, 108718.
48. Demircioglu, A., et al. (2020). A rapid volume of interest-based approach of radiomics analysis of breast MRI for tumor decoding and phenotyping of breast cancer. *PLoS ONE, 15*(6), e0234871.
49. Liang, C., et al. (2018). An MRI-based radiomics classifier for preoperative prediction of Ki-67 status in breast cancer. *Academic radiology, 25*(9), 1111–1117.