



Tumor Morphology for Prediction of Poor Responses Early in Neoadjuvant Chemotherapy for Breast Cancer: A Multicenter Retrospective Study

Wen Li^{1,*}, Nu N. Le¹, Rohan Nadkarni¹, Natsuko Onishi¹, Lisa J. Wilmes¹, Jessica E. Gibbs¹, Elissa R. Price¹, Bonnie N. Joe¹, Rita A. Mukhtar², Efstathios D. Gennatas³, John Kornak³, Mark Jesus M. Magbanua⁴, Laura J. van't Veer⁴, Barbara LeStage⁵, Laura J. Esserman² and Nola M. Hylton¹

- ¹ Department of Radiology and Biomedical Imaging, University of California, 550 16th Street, San Francisco, CA 94158, USA
- ² Department of Surgery, University of California, San Francisco, 550 16th Street, San Francisco, CA 94158, USA
- ³ Department of Epidemiology and Biostatistics, University of California, San Francisco, 550 16th Street, San Francisco, CA 94158, USA
- ⁴ Department of Laboratory Medicine, University of California, San Francisco, 2340 Sutter Street, San Francisco, CA 94115, USA
- ⁵ I-SPY 2 Advocacy Group, San Francisco, CA 94158, USA
- * Correspondence: wen.li@ucsf.edu; Tel.: +1-(415)-885-3764; Fax: +1-(415)-885-7876

Abstract: Background: This multicenter and retrospective study investigated the additive value of tumor morphologic features derived from the functional tumor volume (FTV) tumor mask at pre-treatment (T0) and the early treatment time point (T1) in the prediction of pathologic outcomes for breast cancer patients undergoing neoadjuvant chemotherapy. Methods: A total of 910 patients enrolled in the multicenter I-SPY 2 trial were included. FTV and tumor morphologic features were calculated from the dynamic contrast-enhanced (DCE) MRI. A poor response was defined as a residual cancer burden (RCB) class III (RCB-III) at surgical excision. The area under the receiver operating characteristic curve (AUC) was used to evaluate the predictive performance. The analysis was performed in the full cohort and in individual sub-cohorts stratified by hormone receptor (HR) and human epidermal growth factor receptor 2 (HER2) status. Results: In the full cohort, the AUCs for the use of the FTV ratio and clinicopathologic data were 0.64 \pm 0.03 (mean \pm SD [standard deviation]). With morphologic features, the AUC increased significantly to 0.76 ± 0.04 (p < 0.001). The ratio of the surface area to volume ratio between T0 and T1 was found to be the most contributing feature. All top contributing features were from T1. An improvement was also observed in the HR+/HER2and triple-negative sub-cohorts. The AUC increased significantly from 0.56 ± 0.05 to 0.70 ± 0.06 (p < 0.001) and from 0.65 \pm 0.06 to 0.73 \pm 0.06 (p < 0.001), respectively, when adding morphologic features. Conclusion: Tumor morphologic features can improve the prediction of RCB-III compared to using FTV only at the early treatment time point.

Keywords: magnetic resonance imaging; breast cancer; tumor morphology; neoadjuvant therapy; multicenter clinical trial; residual cancer burden

1. Introduction

For women with locally advanced breast cancer undergoing neoadjuvant chemotherapy (NAC), favorable recurrence-free survival (RFS) and overall survival (OS) are observed for those who achieve a pathologic complete response (pCR), defined as the absence of invasive disease after NAC [1,2]. For those who do not have pCR, the residual cancer burden (RCB) method is a standard measure of residual disease in the primary tumor and in the lymph nodes after neoadjuvant treatment [3–6]. The categorical RCB classes (0-III) quantify the extent of residual disease, with RCB-0 indicating no residual invasive disease (pCR) and RCB-III indicating a high burden of residual disease [7]. The RCB classes and



Citation: Li, W.; Le, N.N.; Nadkarni, R.; Onishi, N.; Wilmes, L.J.; Gibbs, J.E.; Price, E.R.; Joe, B.N.; Mukhtar, R.A.; Gennatas, E.D.; et al. Tumor Morphology for Prediction of Poor Responses Early in Neoadjuvant Chemotherapy for Breast Cancer: A Multicenter Retrospective Study. *Tomography* 2024, *10*, 1832–1845. https://doi.org/10.3390/ tomography10110134

Academic Editor: Sungheon Gene Kim

Received: 6 October 2024 Revised: 12 November 2024 Accepted: 19 November 2024 Published: 20 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). the RCB calculation on a continuous scale have been validated as a prognostic tool, with a higher RCB after NAC associated with a worse prognosis [8]. In patients undergoing NAC, the ability to identify poorly responding tumors at an early treatment time point can potentially facilitate appropriate early intervention, either by switching to a more effective therapy or by stopping an ineffective but potentially toxic one.

Magnetic resonance imaging (MRI) offers a non-invasive and real-time assessment of the breast cancer treatment response. The functional tumor volume (FTV), derived from dynamic contrast-enhanced (DCE) MRI, is an imaging biomarker that has been integrated into personalized decision-making workflows in the Investigation of Serial Studies to Predict your Therapeutic Response with Imaging and Molecular Analysis 2 (I-SPY 2 TRIAL) to identify and re-direct patients with tumors that respond well or poorly to a more suitable treatment by either de-escalating or escalating the treatment, respectively [9]. However, as a gross measurement of the tumor volume, the FTV does not reflect tumor morphological characteristics such as the shape or structure, which may also predict the response to treatment [10,11]. In BI-RADS, or the Breast Imaging Reporting and Data System, the tumor morphology is described based on observations by radiologists regarding the appearance, shape, and structure in breast imaging, but it is subject to inter-observer variability, especially for non-mass lesions [12]. An objective, quantitative assessment of the tumor morphology in breast DCE-MRI is needed. Previous studies have demonstrated that the tumor sphericity, a single shape feature measured from the same tumor segmentation used in FTV calculation, can predict the pCR for breast cancer treated by NAC [13].

In this study, we extracted a set of morphologic features from the existing FTV tumor segmentation and investigated the additive value of tumor morphologic features to FTV in the prediction of RCB-based pathologic outcomes using data from the multicenter I-SPY 2 study.

2. Materials and Methods

2.1. Patient Cohort

The study cohort comprised 990 patients enrolled in the I-SPY 2 clinical trial who had completed treatment by November 2016. Women who were 18 years or older and diagnosed with clinical stage II or III breast cancer with a tumor size of at least 2.5 cm by imaging or physical examination were eligible to enroll in the I-SPY 2 TRIAL. Patients who had a hormone receptor (HR)-positive/human epidermal growth factor receptor 2 (HER2)-negative status that were low risk according to MammaPrint (Agendia, Irvine, CA, USA) were excluded. Other patient exclusion criteria included not having a pathologic assessment at surgery or not having demographic data. All patients were treated with experimental drugs or standard chemotherapy for 12 weeks, followed by 4 cycles of doxorubicin and cyclophosphamide (AC). This was a Health Insurance Portability and Accountability Act (HIPAA)-compliant multicenter study. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Boards of the participating sites. All participating sites had received institutional review board approval, and all participating patients provided written informed consent for this research.

2.2. Image Acquisition

As per the I-SPY protocol, dynamic contrast-enhanced (DCE) MRI was acquired at multiple time points for each patient at the participating site. MRI examinations were performed on either 1.5T or 3T scanners, but the same scanner and configuration was always used for the same patient across time points. Three-dimensional fat-suppressed T1 images were acquired before and after contrast injection. Details of the DCE-MRI acquisition protocol can be found in Table S1. Postcontrast imaging started simultaneously with the injection. The gadolinium contrast agent was administrated intravenously at a dose of 0.1 mmol/kg body weight and at a rate of 2 mL/s, followed by a 20 mL saline flush. DCE-MRI acquired at pretreatment (T0) and 3 weeks after the start of NAC (T1) was analyzed in this study. The imaging data included in this study can be accessed via the

Cancer Imaging Archive (TCIA) [14]. Patients included in this study were required to have MRI at both T0 and T1, with acceptable MRI quality. MRI was excluded if the data were corrupted or the acquisition was not protocol-adherent (across visits) or had severe motion upon visual inspection.

2.3. Image Analysis

The MRI variables, including the FTV and tumor morphologic features, were calculated from the DCE-MRI. As described in prior publications [15], the FTV was calculated as the sum of the voxel volumes, in which the voxels were selected automatically using thresholds for the early percent enhancement (PE) and signal enhancement ratio (SER), after a 3D rectangular volume-of-interest box was delineated manually by a site radiologist or a trained imaging coordinator to cover the entire tumor area. The FTV was calculated at each treatment time point, T0 and T1, separately. The ratio of the FTV value at T1 relative to the FTV value at T0 (referred to as FTV Ratio (FTV_R)) was also calculated to capture the change between treatment time points.

To calculate the tumor morphologic features, three-dimensional shape features were extracted from the existing FTV tumor mask after simple preprocessing steps were applied to the tumor mask to remove noise and ensure that the radiomic features accurately reflected the tumor shape. First, all voxels were resampled to an isotropic voxel size of $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$ with nearest-neighbor interpolation. Second, a morphological closing operation with a kernel radius of 5 mm was applied to fill small holes. Third, antialiasing filtering was applied to the binary image to reduce the aliasing artifacts created by surface generation [16]. The number of iterations and maximum root-mean-square error were 50 and 0.01, respectively. Finally, objects with a size less than or equal to 100 connected voxels were removed.

FTV tumor masks were generated using in-house software developed in the Interactive Data Language (IDL Version 8.5, NV5 Geospatial Solutions, Broomfield, CO, USA). All preprocessing steps were implemented using the SimpleITK toolkit (Kitware, Inc., Clifton Park, New York, NY, USA) [17]. After preprocessing, radiomic shape features were extracted within the FTV tumor mask. A total number of 17 3D shape features were calculated for the tumor masks in each DCE-MRI exam using Pyradiomics [18]. See Figure 1 for an illustration of the shape feature extraction, and the list of features extracted can be found in Table S2. Missing data occurred when the number of tumor voxels was too small to calculate shape features. In these cases, a random forest-based imputation method was implemented [19]. A total of 51 shape feature variables were extracted, including 17 shape features extracted at T0, 17 shape features extracted at T1, and 17 ratios of shape features at T1 over shape features at T0.



FTV tumor mask

Tumor mask after preprocessing

Figure 1. Illustration of radiomic feature extraction from functional tumor volume (FTV) tumor mask in dynamic contrast-enhanced MRI. First, early percent enhancement (PE) and signal enhancement ratio (SER) thresholds were applied to generate the FTV tumor mask, from which the FTV was calculated. Second, multiple preprocessing steps were applied to the FTV tumor mask to fill small holes, smooth edges, and remove small clusters of connected voxels. Lastly, the Pyradiomics package was used to extract radiomic 3D shape features.

2.4. Pathologic Outcome

The pathologic outcome was assessed at the time of surgery, after the completion of NAC. The histopathologic analysis of the surgical specimens was performed at the participating sites by the institutional pathologist, following the I-SPY protocol [6]. RCB classes (RCB-0, RCB-I, RCB-II, RCB-III) were calculated based on pathology measurements [20], where RCB-0 was considered as a pCR. The patient cohort was dichotomized based on the RCB classes in three methods: (1) pCR (RCB-0) versus non-pCR (RCB-I, -II, -III) for the prediction of the best outcome; (2) RCB-0 and -I (RCB-0 and RCB-I versus RCB-III and RCB-III) for the prediction of minimal residual disease; (3) RCB-III versus non RCB-III (RCB-0, -I, -II) for the prediction of the worst outcome with extensive residual disease. This study focused on predicting RCB-III versus non RCB-III.

2.5. Statistical Analysis

The prediction of RCB outcomes was evaluated using four commonly used machine learning (ML) algorithms—elastic net [21], classification and regression tree (CART) [22], random forest [23], and gradient boosting machine (GBM) [24]. Each model was evaluated with respect to its performance using the FTV_R and clinicopathologic data, with and without 51 shape feature variables. The patient clinicopathologic data were the age, race, menopausal status, and HR/HER2 subtype defined by HR-positive/-negative and HER2postive/-negative status. Model evaluation was performed using automatic tuning and testing, in which nested resampling was used. The inner resampling was conduced for the hyperparameter tuning of the model, performed by a grid search using 5-fold cross validation. Lists of the hyperparameters tuned in each ML model can be found in Table S3. The outer resampling was conducted for model generalizability and was performed with 20 stratified subsamples. The area under the receiver operating characteristic curve (AUC) was used to evaluate the predictive performance. AUC estimations were given as the mean with standard deviation (SD) of the AUC values across all 20 testing sets of the outer resampling. An illustration of the nested cross-validation is shown in Figure S1. The model comparison was conducted by the pairwise Wilcoxon signed-rank test. Fisher's exact test was used to estimate *p*-values for the distributions of categorical clinicopathologic data between the two groups, and a two-sided Student *t* test was used to estimate the *p*-value for the age difference between the two groups. Numeric MRI variables were presented as medians with interquartile ranges, and a two-sided Wilcoxon signed-rank test was used to test the difference in the two groups. Correlation coefficients between the FTV and shape features were calculated by Spearman's method.

The additive value of the shape features was assessed by comparing the AUCs of the optimal models with and without shape features. The optimal model was selected by identifying the one with the highest mean AUC among the four ML algorithms listed above. Variable importance was used to rank the contributions of MRI and the clinicopathologic variables in the ML models. In the elastic net models, coefficients were used to generate the variable importance for numeric variables only (categorical variables were not included) after the coefficients were scaled by the multiplication of the standard deviation.

The statistical analysis was performed in R (version 4.2.2). The machine learning model evaluation was performed using the rtemis package [25]. The statistical significance level was nominally set to be 0.05.

3. Results

3.1. Patient Characteristics

A flowchart showing the patient inclusion and exclusion process is shown in Figure 2. Patients with missing pathologic outcomes (n = 45) or clinicopathologic data (n = 3), missing MRI data (n = 17), or poor MRI quality (n = 15) were excluded. Out of 15 excluded due to poor quality, six had non-usable MRI data; six had severe motion; and three had inconsistent sequence acquisition times between T0 and T1 or a non-protocol-adherent sequence acquisition time. Thus, a sub-cohort of 910 was included in the analysis. Table 1

shows the characteristics of the eligible (n = 990), analysis (n = 910), and excluded (n = 80) cohorts. The analysis cohort had the same mean age (48.8 years old) and standard deviation (10.5 years old) as the eligible cohort. The two cohorts also shared similar HR/HER2 and menopausal status, race, and RCB class distributions. However, the excluded cohort had different subtype (p = 0.024), race (p = 0.018), and RCB class (p < 0.001) distributions compared to the analysis cohort.



Figure 2. Data inclusion and exclusion. Patients were excluded from the analysis because of missing pathological outcomes, clinicopathologic data, or MRI or unusable imaging data.

Table 1.	Patient	characteristics.

Characteristic	Eligible Cohort n = 990	Analysis Cohort n = 910	Excluded Cohort n = 80	<i>p-</i> Value (Analysis versus Excluded)
Age (mean \pm standard deviation)	48.8 ± 10.5	48.8 ± 10.5	47.8 ± 11.0	0.41
HR/HER2 subtype (n, %)				
HR+/HER2-	382 (39%)	358 (39%)	24 (30%)	0.024
HR+/HER2+	156 (16%)	147 (16%)	9 (11%)	
HR-/HER2+	89 (9%)	75 (8%)	14 (18%)	
HR-/HER2- (triple-negative)	363 (37%)	330 (36%)	33 (41%)	
Menopausal status (n, %)				
Premenopausal	481 (49%)	438 (48%)	43 (54%)	0.63
Perimenopausal	35 (4%)	33 (4%)	2 (3%)	
Postmenopausal	301 (30%)	282 (31%)	19 (24%)	
Not applicable	135 (14%)	123 (14%)	12 (15%)	
Unknown	38 (4%)	34 (4%)	4 (5%)	
Race (n, %)				
White	783 (79%)	730 (80%)	53 (66%)	0.018
Black or African American	120 (12%)	104 (11%)	16 (20%)	
Asian	68 (7%)	61 (7%)	7 (9%)	
Mixed	7 (0.7%)	7 (0.8%)	0 (0%)	
Native Hawaiian or Pacific Islander	5 (0.5%)	5 (0.5%)	0 (0%)	
American Indian or Alaska Native	4 (0.4%)	3 (0.3%)	1 (1%)	
Unknown	1 (0.1%)	0 (0%)	1 (1%)	

Characteristic	Eligible Cohort n = 990	Analysis Cohort n = 910	Excluded Cohort n = 80	<i>p-</i> Value (Analysis versus Excluded)
Residual cancer burden (n, %)				
RCB-0 (pCR)	325 (33%)	315 (35%)	10 (13%)	
RCB-I	135 (14%)	127 (14%)	8 (10%)	
RCB-II	339 (34%)	327 (36%)	12 (15%)	<0.001
RCB-III	146 (15%)	141 (15%)	5 (6%)	
Unknown	45 (5%)	0 (0%)	45 (56%)	

Table 1. Cont.

HR: hormone receptor. HER2: human epidermal growth factor receptor 2. pCR: pathologic complete response. RCB: residual cancer burden. p-values were calculated by two-sided t test for age and Fisher's exact test for HR/HER2, menopausal, and pCR.

3.2. Additive Value of Shape Features

In the analysis cohort, 141 patients (15%) had RCB-III at the surgery and the remaining 769 patients (85%) were non RCB-IIIs (65%). The AUCs of the ML models estimated in the full cohort (n = 910) and in the sub-cohorts defined by HR/HER2 subtype are listed in Table S4. The optimal model for the prediction of RCB-III using the FTV_R and clinicopathologic data was based on GBM with an estimated AUC of 0.64 \pm 0.03. After the shape feature variables were added together with the FTV_R and clinicopathologic data, a higher AUC was achieved (0.76 \pm 0.04, *p* < 0.001) by random forest. Boxplots of the AUCs for the models with and without the shape are shown in Figure 3. The AUCs for the prediction of the pCR and RCB-0/-I using the ML models estimated in the full cohort (n = 910) and in the sub-cohorts defined by HR/HER2 subtype are listed in Tables S5 and S6.



Model 😑 Without shape 😑 With shape

Figure 3. Boxplots of area under the receiver operating characteristic curve (AUC) for prediction of residual disease with and without shape features. AUCs were evaluated by optimal machine learning models independently in 20 stratified subsamples of the analysis cohort (n = 910) for the prediction of RCB-III. Model—without shape: FTV_R and clinicopathologic data were used in the predictive model. Model—with shape: shape features were added to the predictive model together with FTV_R and clinicopathologic data.

The associations between the individual MRI variables (FTV_R and shape features) and RCB-III according to a univariate analysis can be found in Figure S3. The Spearman's correlation coefficients between the FTV and shape features calculated at T0 and T1 are

listed in Table S7. The variable importance for the prediction of RCB-III using shape features together with FTV_R and clinicopathologic data by random forest is shown in Figure 4, which shows that the most important variable was the ratio of the surface area to volume. The surface area to volume ratio was one of the radiomic shape features analyzed in this study (see Table S2 for the full list), and the ratio represented the change in this shape feature between T0 and T1. The beeswarm plots with overlaid boxplots of the surface area to volume ratio for the RCB-III and non RCB-III groups are shown in Figure 5, which also shows the plots for the FTV_R . The FTV_R was ranked fourth among the top ten contributing variables, and it was significantly higher in the RCB-III group (0.74 [0.51, 0.97]) than in the non-RCB-III group (0.49 [0.26, 0.76]), p < 0.001. The surface area to volume ratio was lower in the RCB-III group (1.10 [1.00, 1.26]) than in the non-RCB-III group (1.27 [1.09, 1.58]), p < 0.001.



Figure 4. Variable importance for prediction of RCB-III using shape features together with FTV ratio and demographic variables by random forest. Variable importance was ranked according to the mean decrease in accuracy (%) when a variable was excluded. A higher value means that a variable was more important. T0: pretreatment time point. T1: early treatment time point.



Figure 5. Beeswarm plots with overlaid boxplot of MRI features. (**a**) Functional tumor volume (FTV) ratio between pretreatment and early treatment. (**b**) Ratio of surface area to volume between pretreatment and early treatment. Number of patients in the RCB-III group was 141, with 769 patients in the nRCB-III group (RCB-0, -I, or -II).

Figure 6a shows an example of the RCB-0 cases. The FTV_R at T1 was 0.99, which indicated that the FTV did not change after three weeks of NAC. However, the surface area to volume ratio increased from 0.68 at T0 to 0.87 at T1, resulting in a ratio of 1.28. This example illustrates that, in good responders, the surface area to volume ratio increased even when the FTV remained the same. Figure 6b shows an example of the RCB-III cases. The FTV_R was 0.67 but the surface area to volume ratio remained the same (0.38 at T0 and 0.38 at T1). This suggests that, even if the tumor is reduced in size but retains its original shape after three weeks of NAC, it might still indicate a poor response.



Figure 6. Example cases. Representative slices of post-contrast dynamic contrast-enhanced MRI at pretreatment (T0) and early treatment (T1) time points are shown. Functional tumor volume (FTV) tumor masks were generated by voxels within the region-of-interest box in yellow that had an early percent enhancement (PE) above 70% and are shown superimposed on the representative slices. Colors within FTV tumor masks represent different levels of signal enhancement ratio (SER)—blue: 0 to 0.9; purple: 0.9 to 1.0; green: 1.0 to 1.3; red: 1.3 to 1.75; white: 1.75 and higher. Three-dimensional surface rendering of the preprocessed tumor mask is shown next to the representative slice. (**a**) An example of a patient with RCB-0. (**b**) An example of a patient with RCB-III.

The same four ML models were built separately in the HR/HER2 sub-cohorts, and boxplots of the AUCs for the models with and without the shape, evaluated in the individual sub-cohorts, are shown in Figure 7. In HR-/HER2+, the number of RCB-IIIs was too low (n = 4, 5%), so the results in this sub-cohort were omitted.



Figure 7. Boxplots of area under the receiver operating characteristic curve (AUC) for prediction of residual disease with and without shape features. AUCs were evaluated by optimal machine learning models independently in 20 stratified subsamples of the analysis cohort (n = 910) for the prediction of RCB-III in (**a**) HR+/HER2-, (**b**) triple-negative, and (**c**) HR+/HER2+. Model—without shape: FTV_R and clinicopathologic data were used in the predictive model. Model—with shape: shape features were added to the predictive model together with FTV_R and clinicopathologic data.

In the HR+/HER2– sub-cohort, the total number of patients was 358, and 81 (23%) of them were estimated as RCB-III at surgery. The optimal model for the prediction of RCB-III using the FTV_R and clinicopathologic data was based on the random forest algorithm, with an estimated AUC of 0.56 ± 0.05 . After the shape feature variables were added together with the FTV_R and clinicopathologic data, a higher AUC was achieved (0.70 ± 0.06) by the random forest, p < 0.001.

In the triple-negative sub-cohort, the total number of patients was 330, and 36 (11%) of them were estimated as RCB-III at surgery. The optimal model for the prediction of RCB-III using the FTV_R and clinicopathologic data was based on GBM, with an estimated AUC of 0.65 \pm 0.06. After the shape feature variables were added together with the FTV_R and clinicopathologic data, a higher AUC was achieved (0.73 \pm 0.06) by the GBM, *p* < 0.001.

In the HR+/HER2+ sub-cohort, the total number of patients was 147, and 20 (14%) of them were estimated as RCB-III at surgery. The optimal model for the prediction of RCB-III using the FTV_R and clinicopathologic data was based on the elastic net, with an estimated AUC of 0.84 ± 0.07 . After the shape feature variables were added together with the FTV_R and clinicopathologic data, a higher AUC was achieved (0.87 ± 0.06) by the random forest, p = 0.073.

4. Discussion

This retrospective study leveraged both the imaging and pathological data from the multicenter I-SPY trial and investigated the additive value of radiomic shape features for the FTV from DCE-MRI at pretreatment and after 3 weeks of NAC in the prediction of pathologic outcomes. Our results showed that, when adding radiomic shape features, the AUCs of the predictive models increased in predicting the worst outcome (RCB-III) substantially from 0.64 to 0.76. We also performed the same analysis by breast cancer subtype, defined by HR and HER2 status, and substantial improvements in predicting RCB-III after adding shape features were observed in HR+/HER2- (from 0.56 to 0.70) and triple negatives (from 0.65 to 0.73).

Established in the multicenter I-SPY clinical trial, the FTV is a quantitative imaging marker for the NAC treatment response in women with stage II or III breast cancer [9]. The FTV is currently used as one of the biomarkers in the I-SPY 2 randomization engine to adaptively assign patients to treatment drug arms and to redirect treatment [26,27]. Features that could assist the FTV in predicting pathologic outcomes at early treatment time points would be valuable to improve the performance of MRI-based models. Although the performance of the FTV in predicting the pCR has been reported multiple times over the years [9,28–30], the assessment of the FTV in predicting the RCB-III outcome has not been previously discussed. Recently, FTVs measured from 3-week and 6-week MRI have been used to identify inferior responses to NAC by predicting RCB-II/-III in the I-SPY 2 trial [26].

Radiomic analysis is a fast-evolving field of research for radiology and oncology, and radiomic feature extraction is often performed on intensity-based images to quantify the tumor size, shape, texture, and heterogeneity. A large number of image features, including shape features, can be extracted altogether using automated software packages. Associations between radiomic shape features and the treatment response in breast cancer and other types of cancer have been reported previously [31–35]. In a multicenter study of 320 tumors by Granzier et al. [31], a predictive model combining selected pretreatment MRI-based radiomics, including sphericity, and clinical data achieved an AUC of 0.73 in predicting the pCR. Similar shape features (tumor volume, sphericity, and compacity) were found to be predictive of the treatment response to neoadjuvant chemotherapy and tumor recurrence in patients with locally advanced rectal cancer, according to a study by Park et al. [35].

However, the present study was not a typical study that used radiomic features to predict pathologic outcomes. Instead, only radiomic shape features were extracted using existing tumor segmentations generated by FTV calculation, and they were used to quantify the tumor morphology.

This study observed a substantial improvement in predicting RCB-III after shape features were combined with the FTV and demographic data, suggesting that shape features may be associated with drug resistance, especially for HR+/HER2- and triple-negative tumors. Patients with HR+/HER2-, HR+/HER2+, or triple-negative tumors experienced significantly poorer survival when their residual disease was classified as RCB-III compared to those with other RCB classifications, according to a multicenter pooled study of 5161 patients [36]. Therefore, identifying RCB-III early on may aid in the customization of therapeutic strategies to improve patient outcomes. Our results regarding the variable importance demonstrated that most of the top contributing MRI variables were from the early treatment time point, not the pretreatment one. This finding suggests that the tumor morphology or a change in the tumor morphology in DCE-MRI can be more accurate in identifying tumors that are resistant to NAC shortly after treatment initiation compared to before treatment.

While the minimal reduction in tumor size indicated by the FTV could suggest resistance to treatment, our findings indicate that radiomic shape features could offer extra insights into this resistance. The surface area to volume ratio measured between T0 and T1 was ranked as the top contributor based on our variable importance analysis for the prediction of RCB-III, suggesting that non-responding tumors not only maintained their tumor sizes but also tended to maintain their shapes compared to other RCB classes after 3 weeks of NAC. This finding aligns with studies of tumor shrinkage patterns in DCE-MRI during NAC, in which tumor shrinkage patterns were categorized into groups based on concentric shrinkage or non-concentric shrinkage [37,38]. Fukada et al. found statistically significant differences in both the disease-free survival (DFS) and overall survival (OS) between the shrinkage patterns observed in MRI during NAC for low-grade luminal breast cancer [37]. Wang et al. found that the diffuse decrease pattern in HR+/HER2- and stable disease in HER2+ and triple-negative breast cancer could serve as indicators of the response to NAC in their stratified analysis by HR/HER2 subtype [38]. However, both studies were based on the visual assessment of the enhancement patterns on DCE-MRI, which could be subjective.

This study had a few limitations. The data in this study were from a multicenter clinical trial with variability in imaging protocol compliance. A recent study published from I-SPY 2 suggests that protocol compliance and longitudinal variations in FTV segmentation can affect the predictive performance of the FTV [30,39]. Since the tumor masks were generated based on voxels used for FTV calculation, these factors could have affected our analysis as well. Furthermore, our patient population was treated by drug combinations. This treatment heterogeneity could also cause the tumor response to be more heterogeneous, which would make the prediction process more challenging. The current study focused on the relative additive value of radiomic shape features, so we expected a minimal impact on our results. Our cohort was limited by the I-SPY 2 data availability and the targeted drugs used in the trial, so the pathologic outcome rates may be different from the population who receive standard NAC, especially for the best and worst outcomes. However, AUC evaluation is not affected by outcome prevalence, so this limitation should not affect the model's generalizability. Lastly, this study only included shape features, a small subset of the large number of radiomic features that can be extracted from DCE-MRI, which may limit the additive value of radiomic features. A future study to evaluate the additive value of a comprehensive set of radiomic features with the FTV in predicting the response to NAC is warranted.

5. Conclusions

In conclusion, this multicenter retrospective study showed that tumor morphologic features extracted from FTV tumor masks can add value to the FTV in the prediction of RCB-based outcomes early in NAC, particularly for the prediction of RCB-III, the worst residual disease. This improved performance could significantly impact patient care, as it could allow for the timely redirection of patients with poor responses to more effective treatments, thereby increasing their chances of achieving a pCR and improving their prognosis. Future investigation is warranted to assess the tradeoff between false positives and false negatives when the MRI-based model is used in the clinical decision-making for treatment redirection.

Supplementary Materials: The following supporting information can be downloaded at: https: //www.mdpi.com/article/10.3390/tomography10110134/s1, Table S1: I-SPY 2 DCE-MRI acquisition parameters; Table S2: List of 3D shape features extracted using Pyradiomics; Table S3: List of hyperparameters tuned for machine learning algorithms; Table S4: Model evaluation for predicting residual disease defined by RCB-III; Table S5: Model evaluation for predicting pCR; Table S6: Model evaluation for predicting residual disease defined by RCB-III and RCB-III; Table S7: Spearman's correlation coefficients between FTV and shape features; Figure S1: Illustration of nested crossvalidation for machine learning models; Figure S2: Boxplots of area under the receiver operating characteristic curve (AUC) of predicting residual disease with and without shape features; Figure S3: Plots of coefficients and p-values for predicting RCB-III.

Author Contributions: Conceptualization, W.L., L.J.E. and N.M.H.; Methodology, W.L., E.D.G., J.K. and N.O.; Software, E.D.G.; Validation, N.N.L. and W.L.; Formal Analysis, W.L., N.N.L. and R.N.; Investigation, W.L., N.N.L. and R.N.; Resources, N.M.H., L.J.v.V. and L.J.E.; Data Curation, L.J.W., J.E.G., E.R.P., B.N.J., L.J.E. and N.M.H.; Writing—Original Draft Preparation, W.L. and E.R.P.; Writing—Review and Editing, N.N.L., N.O., J.E.G., R.A.M., J.K., M.J.M.M. and B.L.; Visualization, N.N.L.; Supervision, N.M.H.; Project Administration, J.E.G.; Funding Acquisition, W.L., M.J.M.M., L.J.v.V., L.J.E. and N.M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially supported by NIH/NCI grants: R01 CA132870, U01 CA225427, P01 CA210961, R01 CA255442.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the University of California, San Francisco (Protocol Code: 161317; Approval Date: 12 April 2016).

the study.

Data Availability Statement: The data presented in this study are openly available in the Cancer Imaging Archive at 10.7937/TCIA.D8Z0-9T85.

Acknowledgments: The authors would like to acknowledge all I-SPY patients and their families for participating in the trial. The authors thank the members of the I-SPY 2 Trial Imaging Working Group, the I-SPY 2 Trial Investigators Network, and the study coordinators from all participating sites for their contributions to the project.

Conflicts of Interest: L.J.E. is on the Blue Cross Medical Advisory Panel, is an uncompensated board member of Quantum Leap Healthcare Collaborative, and is an Investigator who initiated trial for high-risk DCIS funded by Moderna for DCIS phase 1 study. L.J.v.V. is part-time employee and stocks Agendia NV, advisor and stock options Exai Inc. N.M.H. receives institutional research funding from NIH. B.N.J. received author royalties from UpToDate, received WorldClassCME honoraria for lectures, received Medicolegal consulting payment, serves as Board of Directors for Society of Breast Imaging, Board of Trustees for RSNA R&E Foundation, and Deputy Editor for Radiology: Imaging Cancer. All authors declare no conflicts of interest regarding this study.

References

- Spring, L.M.; Fell, G.; Arfe, A.; Sharma, C.; Greenup, R.; Reynolds, K.L.; Smith, B.L.; Alexander, B.; Moy, B.; Isakoff, S.J.; et al. Pathologic Complete Response after Neoadjuvant Chemotherapy and Impact on Breast Cancer Recurrence and Survival: A Comprehensive Meta-Analysis. *Clin. Cancer Res.* 2020, *26*, 2838–2848. [CrossRef] [PubMed]
- Yee, D.; DeMichele, A.M.; Yau, C.; Isaacs, C.; Symmans, W.F.; Albain, K.S.; Chen, Y.-Y.; Krings, G.; Wei, S.; Harada, S.; et al. Association of Event-Free and Distant Recurrence-Free Survival With Individual-Level Pathologic Complete Response in Neoadjuvant Treatment of Stages 2 and 3 Breast Cancer: Three-Year Follow-up Analysis for the I-SPY2 Adaptively Randomized Clinical Trial. *JAMA Oncol.* 2020, *6*, 1355–1362. [CrossRef] [PubMed]
- Symmans, W.F.; Peintinger, F.; Hatzis, C.; Rajan, R.; Kuerer, H.; Valero, V.; Assad, L.; Poniecka, A.; Hennessy, B.; Green, M.; et al. Measurement of Residual Breast Cancer Burden to Predict Survival after Neoadjuvant Chemotherapy. J. Clin. Oncol. 2007, 25, 4414–4422. [CrossRef] [PubMed]
- Provenzano, E.; Bossuyt, V.; Viale, G.; Cameron, D.; Badve, S.; Denkert, C.; MacGrogan, G.; Penault-Llorca, F.; Boughey, J.; Curigliano, G.; et al. Standardization of Pathologic Evaluation and Reporting of Postneoadjuvant Specimens in Clinical Trials of Breast Cancer: Recommendations from an International Working Group. *Mod. Pathol.* 2015, 28, 1185–1201. [CrossRef] [PubMed]
- Bossuyt, V.; Provenzano, E.; Symmans, W.F.; Boughey, J.C.; Coles, C.; Curigliano, G.; Dixon, J.M.; Esserman, L.J.; Fastner, G.; Kuehn, T.; et al. Recommendations for Standardized Pathological Characterization of Residual Disease for Neoadjuvant Clinical Trials of Breast Cancer by the BIG-NABCG Collaboration. *Ann. Oncol.* 2015, 26, 1280–1291. [CrossRef]
- Symmans, W.F.; Yau, C.; Chen, Y.-Y.; Balassanian, R.; Klein, M.E.; Pusztai, L.; Nanda, R.; Parker, B.A.; Datnow, B.; Krings, G.; et al. Assessment of Residual Cancer Burden and Event-Free Survival in Neoadjuvant Treatment for High-Risk Breast Cancer: An Analysis of Data From the I-SPY2 Randomized Clinical Trial. *JAMA Oncol.* 2021, 7, 1654–1663. [CrossRef]
- Symmans, W.F.; Yau, C.; Chen, Y.-Y.; Datnow, B.; Wei, S.; Feldman, M.D.; Ritter, J.; Duan, X.; Chen, B.; Tickman, R.; et al. Residual Cancer Burden (RCB) as Prognostic in the I-SPY 2 TRIAL. J. Clin. Oncol. 2018, 36 (Suppl. S15), 520. [CrossRef]
- Symmans, W.F.; Wei, C.; Gould, R.; Yu, X.; Zhang, Y.; Liu, M.; Walls, A.; Bousamra, A.; Ramineni, M.; Sinn, B.; et al. Long-Term Prognostic Risk After Neoadjuvant Chemotherapy Associated With Residual Cancer Burden and Breast Cancer Subtype. J. Clin. Oncol. 2017, 35, 1049–1060. [CrossRef]
- Hylton, N.M.; Blume, J.D.; Bernreuter, W.K.; Pisano, E.D.; Rosen, M.A.; Morris, E.A.; Weatherall, P.T.; Lehman, C.D.; Newstead, G.M.; Polin, S.; et al. Locally Advanced Breast Cancer: MR Imaging for Prediction of Response to Neoadjuvant Chemotherapy— Results from ACRIN 6657/I-SPY TRIAL. *Radiology* 2012, 263, 663–672. [CrossRef]
- 10. Reig, B.; Lewin, A.A.; Du, L.; Heacock, L.; Toth, H.K.; Heller, S.L.; Gao, Y.; Moy, L. Breast MRI for Evaluation of Response to Neoadjuvant Therapy. *Radiographics* **2021**, *41*, 665–679. [CrossRef]
- 11. Öztürk, V.S.; Polat, Y.D.; Soyder, A.; Tanyeri, A.; Karaman, C.Z.; Taşkın, F. The Relationship Between MRI Findings and Molecular Subtypes in Women With Breast Cancer. *Curr. Probl. Diagn. Radiol.* **2020**, *49*, 417–421. [CrossRef] [PubMed]
- Lunkiewicz, M.; Forte, S.; Freiwald, B.; Singer, G.; Leo, C.; Kubik-Huch, R.A. Interobserver Variability and Likelihood of Malignancy for Fifth Edition BI-RADS MRI Descriptors in Non-Mass Breast Lesions. *Eur. Radiol.* 2020, 30, 77–86. [CrossRef] [PubMed]
- Li, W.; Newitt, D.C.; La Yun, B.; Jones, E.F.; Arasu, V.; Wilmes, L.; Gibbs, J.; Nguyen, A.; Onishi, N.; Kornak, J.; et al. Tumor Sphericity Predicts Response in Neoadjuvant Chemotherapy for Invasive Breast Cancer. *Tomography* 2020, *6*, 216–222. [CrossRef] [PubMed]

- Li, W.; Newitt, D.C.; Gibbs, J.; Wilmes, L.J.; Jones, E.F.; Arasu, V.A.; Strand, F.; Onishi, N.; Nguyen, A.A.T.; Kornak, J.; et al. I-SPY 2 Breast Dynamic Contrast Enhanced MRI Trial (ISPY2) (Version 1). Available online: https://doi.org/10.7937/TCIA.D8Z0-9T85 (accessed on 1 October 2023).
- Newitt, D.C.; Aliu, S.O.; Witcomb, N.; Sela, G.; Kornak, J.; Esserman, L.; Hylton, N.M. Real-Time Measurement of Functional Tumor Volume by MRI to Assess Treatment Response in Breast Cancer Neoadjuvant Clinical Trials: Validation of the Aegis SER Software Platform. *Transl. Oncol.* 2014, 7, 94–100. [CrossRef]
- 16. Whitaker, R.T. Reducing Aliasing Artifacts in Iso-Surfaces of Binary Volumes. In Proceedings of the 2000 IEEE Symposium on Volume Visualization (VV 2000), Salt Lake City, UT, USA, 9–10 October 2000; pp. 23–32. [CrossRef]
- 17. Yaniv, Z.; Lowekamp, B.C.; Johnson, H.J.; Beare, R. SimpleITK Image-Analysis Notebooks: A Collaborative Environment for Education and Reproducible Research. J. Digit. Imaging **2018**, *31*, 290–303. [CrossRef]
- van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017, 77, e104–e107. [CrossRef]
- 19. Stekhoven, D.J.; Bühlmann, P. MissForest—Non-Parametric Missing Value Imputation for Mixed-Type Data. *Bioinformatics* **2012**, 28, 112–118. [CrossRef]
- MDAnderson. Residual Cancer Burden Calculator. Available online: https://www.mdanderson.org/breastcancer_RCB (accessed on 23 April 2024).
- Rydzewski, N.R.; Peterson, E.; Lang, J.M.; Yu, M.; Laura Chang, S.; Sjöström, M.; Bakhtiar, H.; Song, G.; Helzer, K.T.; Bootsma, M.L.; et al. Predicting Cancer Drug TARGETS -TreAtment Response Generalized Elastic-NeT Signatures. NPJ Genom. Med. 2021, 6, 76. [CrossRef]
- Brown, T.J.; Gimotty, P.A.; Mamtani, R.; Karasic, T.B.; Yang, Y.-X. Classification and Regression Trees to Predict for Survival for Patients With Hepatocellular Carcinoma Treated With Atezolizumab and Bevacizumab. *JCO Clin. Cancer Inform.* 2024, *8*, e2300220. [CrossRef]
- Wang, Z.; Li, X.; Zhang, H.; Duan, T.; Zhang, C.; Zhao, T. Deep Learning Radiomics Based on Two-Dimensional Ultrasound for Predicting the Efficacy of Neoadjuvant Chemotherapy in Breast Cancer. *Ultrason. Imaging* 2024, 46, 357–366. [CrossRef]
- 24. Rodríguez-Tomàs, E.; Arenas, M.; Baiges-Gaya, G.; Acosta, J.; Araguas, P.; Malave, B.; Castañé, H.; Jiménez-Franco, A.; Benavides-Villarreal, R.; Sabater, S.; et al. Gradient Boosting Machine Identified Predictive Variables for Breast Cancer Patients Pre- and Post-Radiotherapy: Preliminary Results of an 8-Year Follow-Up Study. *Antioxidants* **2022**, *11*, 2394. [CrossRef] [PubMed]
- Gennatas, E.D. Towards Precision Psychiatry: Gray Matter Development and Cognition in Adolescence. Ph.D. Thesis, University of Pennsylvania, Philadelphia, PA, USA, 2017.
- Onishi, N.; Gibbs, J.E.; Li, W.; Newitt, D.C.; Price, E.R.; LeStage, B.; Symmans, W.F.; DeMichele, A.M.; Yau, C.; The I-SPY 2 TRIAL Imaging Working Group; et al. Functional Tumor Volume at 3 and 6-Week MRI as an Indicator of Patients with Inferior Outcome after Neoadjuvant Chemotherapy. *Cancer Res.* 2022, 82 (Suppl. S4), P3-03-01. [CrossRef]
- Li, W.; Newitt, D.C.; Gibbs, J.; Wilmes, L.J.; Jones, E.F.; Onishi, N.; Joe, B.N.; Price, E.; Kornak, J.; Yau, C.; et al. Subtype-Specific MRI Models to Guide Selection of Candidates for de-Escalation of Neoadjuvant Therapy. *Cancer Res.* 2021, *81* (Suppl. S4), PD6-05. [CrossRef]
- Li, W.; Newitt, D.C.; Gibbs, J.; Wilmes, L.J.; Jones, E.F.; Arasu, V.A.; Strand, F.; Onishi, N.; Nguyen, A.A.T.; Kornak, J.; et al. Predicting Breast Cancer Response to Neoadjuvant Treatment Using Multi-Feature MRI: Results from the I-SPY 2 TRIAL. NPJ Breast Cancer 2020, 6, 63. [CrossRef] [PubMed]
- Musall, B.C.; Abdelhafez, A.H.; Adrada, B.E.; Candelaria, R.P.; Mohamed, R.M.M.; Boge, M.; Le-Petross, H.; Arribas, E.; Lane, D.L.; Spak, D.A.; et al. Functional Tumor Volume by Fast Dynamic Contrast-Enhanced MRI for Predicting Neoadjuvant Systemic Therapy Response in Triple-Negative Breast Cancer. J. Magn. Reson. Imaging 2021, 54, 251–260. [CrossRef]
- Onishi, N.; Bareng, T.J.; Gibbs, J.; Li, W.; Price, E.R.; Joe, B.N.; Kornak, J.; Esserman, L.J.; Newitt, D.C.; Hylton, N.M.; et al. Effect of Longitudinal Variation in Tumor Volume Estimation for MRI-Guided Personalization of Breast Cancer Neoadjuvant Treatment. *Radiol. Imaging Cancer* 2023, 5, e220126. [CrossRef]
- Granzier, R.W.Y.; Ibrahim, A.; Primakov, S.P.; Samiei, S.; van Nijnatten, T.J.A.; de Boer, M.; Heuts, E.M.; Hulsmans, F.-J.; Chatterjee, A.; Lambin, P.; et al. MRI-Based Radiomics Analysis for the Pretreatment Prediction of Pathologic Complete Tumor Response to Neoadjuvant Systemic Therapy in Breast Cancer Patients: A Multicenter Study. *Cancers* 2021, 13, 2447. [CrossRef]
- 32. Huynh, E.; Coroller, T.P.; Narayan, V.; Agrawal, V.; Romano, J.; Franco, I.; Parmar, C.; Hou, Y.; Mak, R.H.; Aerts, H.J.W.L. Associations of Radiomic Data Extracted from Static and Respiratory-Gated CT Scans with Disease Recurrence in Lung Cancer Patients Treated with SBRT. *PLoS ONE* **2017**, *12*, e0169172. [CrossRef]
- Rakshit, S.; Orooji, M.; Beig, N.; Alilou, M.; Pennell, N.A.; Stevenson, J.; Shapiro, M.A.; Madabhushi, A.; Velcheti, V. Evaluation of Radiomic Features on Baseline CT Scan to Predict Clinical Benefit for Pemetrexed Based Chemotherapy in Metastatic Lung Adenocarcinoma. J. Clin. Oncol. 2016, 34 (Suppl. S15), 11582. [CrossRef]
- Banerjee, I.; Malladi, S.; Lee, D.; Depeursinge, A.; Telli, M.; Lipson, J.; Golden, D.; Rubin, D.L. Assessing Treatment Response in Triple-Negative Breast Cancer from Quantitative Image Analysis in Perfusion Magnetic Resonance Imaging. *J. Med. Imaging* 2018, 5, 11008. [CrossRef]

- 35. Park, H.; Kim, K.A.; Jung, J.-H.; Rhie, J.; Choi, S.Y. MRI Features and Texture Analysis for the Early Prediction of Therapeutic Response to Neoadjuvant Chemoradiotherapy and Tumor Recurrence of Locally Advanced Rectal Cancer. *Eur. Radiol.* 2020, *30*, 4201–4211. [CrossRef] [PubMed]
- Yau, C.; Osdoit, M.; van der Noordaa, M.; Shad, S.; Wei, J.; de Croze, D.; Hamy, A.-S.; Laé, M.; Reyal, F.; Sonke, G.S.; et al. Residual Cancer Burden after Neoadjuvant Chemotherapy and Long-Term Survival Outcomes in Breast Cancer: A Multicentre Pooled Analysis of 5161 Patients. *Lancet Oncol.* 2022, 23, 149–160. [CrossRef] [PubMed]
- Fukada, I.; Araki, K.; Kobayashi, K.; Shibayama, T.; Takahashi, S.; Gomi, N.; Kokubu, Y.; Oikado, K.; Horii, R.; Akiyama, F.; et al. Pattern of Tumor Shrinkage during Neoadjuvant Chemotherapy Is Associated with Prognosis in Low-Grade Luminal Early Breast Cancer. *Radiology* 2017, 286, 49–57. [CrossRef] [PubMed]
- Wang, M.; Du, S.; Gao, S.; Zhao, R.; Liu, S.; Jiang, W.; Peng, C.; Chai, R.; Zhang, L. MRI-Based Tumor Shrinkage Patterns after Early Neoadjuvant Therapy in Breast Cancer: Correlation with Molecular Subtypes and Pathological Response after Therapy. Breast Cancer Res. 2024, 26, 26. [CrossRef]
- Onishi, N.; Li, W.; Gibbs, J.; Wilmes, L.J.; Nguyen, A.; Jones, E.F.; Arasu, V.; Kornak, J.; Joe, B.N.; Esserman, L.J.; et al. Impact of MRI Protocol Adherence on Prediction of Pathological Complete Response in the I-SPY 2 Neoadjuvant Breast Cancer Trial. *Tomography* 2020, *6*, 77–85. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.