

Breast Cancer Molecular Subtype Prediction by Mammographic Radiomic Features

Wenjuan Ma, PhD, Yumei Zhao, BS, Yu Ji, MS, Xinpeng Guo, BE, Xiqi Jian, PhD, Peifang Liu, PhD, Shandong Wu, PhD

Rationale and Objectives: This study aimed to investigate whether quantitative radiomic features extracted from digital mammogram images are associated with molecular subtypes of breast cancer.

Materials and Methods: In this institutional review board–approved retrospective study, we collected 331 Chinese women who were diagnosed with invasive breast cancer in 2015. This cohort included 29 triple-negative, 45 human epidermal growth factor receptor 2 (HER2)-enriched, 36 luminal A, and 221 luminal B lesions. A set of 39 quantitative radiomic features, including morphologic, grayscale statistic, and texture features, were extracted from the segmented lesion area. Three binary classifications of the subtypes were performed: triple-negative vs non-triple-negative, HER2-enriched vs non-HER2-enriched, and luminal (A + B) vs nonluminal. The Naive Bayes machine learning scheme was employed for the classification, and the least absolute shrink age and selection operator method was used to select the most predictive features for the classifiers. Classification performance was evaluated by the area under receiver operating characteristic curve and accuracy.

Results: The model that used the combination of both the craniocaudal and the mediolateral oblique view images achieved the overall best performance than using either of the two views alone, yielding an area under receiver operating characteristic curve (or accuracy) of 0.865 (0.796) for triple-negative vs non-triple-negative, 0.784 (0.748) for HER2-enriched vs non-HER2-enriched, and 0.752 (0.788) for luminal vs nonluminal subtypes. Twelve most predictive features were selected by the least absolute shrink age and selection operator method and four of them (ie, roundness, concavity, gray mean, and correlation) showed a statistical significance ($P < .05$) in the subtype classification.

Conclusions: Our study showed that quantitative radiomic imaging features of breast tumor extracted from digital mammograms are associated with breast cancer subtypes. Future larger studies are needed to further evaluate the findings.

Key Words: Molecular subtypes; mammogram; radiomics; breast cancer.

© 2018 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.

INTRODUCTION

Breast cancer subtyping has important therapeutic implications on clinical management of the disease. The major breast cancer molecular subtypes include luminal A, luminal B, human epidermal growth factor receptor 2 (HER2)-enriched, and triple-negative types (1). In general,

luminal tumors represent a majority (70%) of invasive breast cancers and respond well to endocrine therapy. HER2-enriched tumors are better candidates for targeted antibody therapy (2). Triple-negative cancers can be more aggressive and difficult to treat, but some of these tumors may respond well to chemotherapy (2,3). Breast cancer subtypes can be categorized by genetic array testing or approximated in typical clinical practice using immunohistochemistry markers. Immunohistochemistry requires tissue specimens typically obtained by a needle biopsy. Because of the relatively small tissue sample size and tumor heterogeneity, the subtyping assessment performed on a needle biopsy sample may not be representative of the tumor entirety. Clinical imaging is a noninvasive approach that has the ability to capture a broader range of tumor heterogeneity than from a single tissue sample (4). Recently, breast cancer subtype characterization has made progress on using radiological images. For example, certain qualitative and visual information of the imaging characteristics assessed on breast magnetic resonance imaging (MRI), mammography, or ultrasound has been shown to be related to the molecular subtypes of breast cancer (5–8).

Acad Radiol 2018; ■:■■–■■■

From the Department of Breast Imaging, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, 24 Huanhuxi Rd, Hexi District, Tianjin 300060 (W.M., Y.Z., Y.J., X.G., P.L.); Key Laboratory of Breast Cancer Prevention and Therapy, 24 Huanhuxi Rd, Hexi District, Tianjin 300060 (W.M., Y.Z., Y.J., X.G., P.L.); Tianjin's Clinical Research Center for Cancer, 24 Huanhuxi Rd, Hexi District, Tianjin 300060, China (W.M., Y.Z., Y.J., X.G., P.L.); Department of Biomedical and Engineering, Tianjin Medical University, Tianjin (W.M., X.J.); Departments of Radiology, Biomedical Informatics, and Bioengineering, University of Pittsburgh, 3362 Fifth Ave, Pittsburgh, PA 15213 (S.W.). Received December 19, 2017; revised January 23, 2018; accepted January 25, 2018. **Address correspondence to:** S.W. e-mail: wus3@upmc.edu; P.L. e-mail: liupf2017@126.com

© 2018 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.
<https://doi.org/10.1016/j.acra.2018.01.023>

Radiomics refers to the extraction and analysis of a large amount of quantitative imaging features from medical images, which has been extensively studied in recent literature, demonstrating predictive or prognostic associations between quantitative imaging features and medical outcome (9,10). A few recent studies reported that there are differences in quantitative radiomic features extracted from breast dynamic contrast-enhanced (DCE) MRI with respect to the four breast cancer subtypes (11–13). However, breast MRI is an expensive examination and currently not widely available especially in less-developed countries. Digital mammography is less expensive and widely used for breast cancer screening and diagnosis; but unlike DCE-MRI, mammography lacks the capability of characterizing certain biological or physiological properties of the breast tissue. Yet, several studies have demonstrated that quantitative imaging features can distinguish malignant and benign lesions in digital mammogram images (14–16). In this study, we aimed to employ machine learning techniques to investigate whether quantitative radiomic imaging features extracted from digital mammography are associated with breast cancer subtypes on a Chinese women population.

MATERIALS AND METHODS

Study Cohort and Imaging Dataset

We conducted a retrospective study and it was approved by our institutional review board, and the informed consent requirement was waived. We collected a study cohort of 331 Chinese women who were diagnosed with invasive breast cancer (confirmed by pathology) during August 2015 to October 2015. All of the 331 cancers were detected on mammography and 253 (76.4%) were palpable. This cohort included 29 triple-negative, 45 HER2-enriched, 36 luminal A, and 221 luminal B lesions. All mammogram images were acquired with the Hologic Lorad Selenia full-field digital mammography systems (HOLOGIC Gen-Probe, San Diego, USA). The full-field digital mammography images were at 14-bit quantization with a pixel size of $70 \times 70 \mu\text{m}$. Each patient or breast has both the craniocaudal (CC) view and the mediolateral oblique (MLO) view images. A total of 662 mammogram images were analyzed.

Radiomic Feature Extraction

Radiomic features were extracted from the lesion area on each image. To do this, the first step was to outline the lesion area from each image. An experienced breast imaging-specialized radiologist (25 years' experience) manually outlined the contours of the diagnosed breast tumor in each image of each individual patient. The largest lesion was selected if there are multiple lesions in a breast. These contours circumscribe the breast tumor region, and over this lesion region, we applied existing automated computer programs to calculate a set of 39 quantitative radiomic features from the mammographic imaging data. The 39 features include (1) morphologic features,

such as shape, size, area, perimeter, roundness, concavity, and Fourier coefficient descriptors; (2) grayscale statistic features calculated from the histogram of tumor voxel intensities, such as mean, standard deviation, skewness, and kurtosis (17); and (3) Haralick texture features for quantifying intratumor heterogeneity calculated using the gray level co-occurrence matrix, such as contrast, correlation, energy, entropy, homogeneity, inertia, and inverse different moments (18). All the 39 radiomic features were computed based on standard mathematical algorithms or formulas, as reported in previous literature. These features provide a quantitative way to capture important phenotypic information of the segmented lesions. Note that all these features were normalized to a standard range before used in a machine learning model for breast cancer subtype classification.

Molecular Subtype Classification

We performed three binary classification tasks for subtype prediction: (A) triple-negative vs non-triple-negative, (B) HER2-enriched vs non-HER2-enriched, and (C) luminal (A and B) vs nonluminal. The Naive Bayes machine learning scheme was employed for the classification (19). Considering that some of the 39 radiomic features may be correlated, we used the least absolute shrinkage and selection operator (LASSO) feature selection method (20) to preidentify those top-ranked or most predictive features before the classification experiments. LASSO is a regression analysis process to improve the prediction accuracy and interpretability of statistical models by altering the model fitting process to select only a subset of the provided variables for use in the final model rather than using all of them. LASSO utilizes both variable selection and regularization to select the subset of variables that minimizes predicting error of the outcome. Statistical significance of a feature was assessed by the Kruskal-Wallis test. In addition, because the sample numbers of the subtypes are unbalanced for certain subtypes (eg, 29 triple-negative vs 153 non-triple-negative images), we adopted the synthetic minority oversampling technique to balance the sample numbers. The synthetic minority oversampling technique has been used in many previous works to address a similar data imbalance problem for image classification (21–23), and here in our study, we have followed the routine procedures of this technique. We utilized a 10-fold cross validation and repeated it 10 times to calculate an average classification performance. Classification performance was evaluated by the area under receiver operating characteristic curve (AUC) and accuracy. In addition, to evaluate the respective effects of the MLO and CC view images for the subtype classification, we conducted and compared the classification experiments on the MLO view images alone, CC view alone, and their combination.

RESULTS

Table 1 shows the characteristics of our study cohort and lesions. There was no statistically significant difference in age (all

TABLE 1. Characteristics of the Study Cohort and Lesions

Characteristics	Triple-negative	HER2-enriched	Luminal (A + B)
No. of patients	29	45	108
Age (y), mean (range)	52 (30–82)	51 (31–72)	53 (29–78)
Masses			
With microcalcifications	1 (3.4%)	2 (4.4%)	6 (5.5%)
Without microcalcifications	24 (82.8%)	13 (28.9%)	44 (40.7%)
Nonmass (including architectural distortion and asymmetrical density)			
With microcalcifications	0 (0%)	19 (42.2%)	20 (18.5%)
Without microcalcifications	4 (14%)	10 (22.2%)	37 (34.3%)
Microcalcifications	0 (0%)	1 (2.2%)	1 (1%)

HER2, human epidermal growth factor receptor 2.

TABLE 2. Overall Classification Performance (Format: AUC/Accuracy) of the Three Binary Classifications of Breast Cancer Subtypes

Data	Triple-negative vs Non-Triple-negative	HER2-enriched vs Non-HER2-enriched	Luminal vs Nonluminal
CC view	0.695/0.634	0.741/0.725	0.755/0.785
MLO view	0.853/0.772	0.756/0.702	0.791/0.782
CC and MLO	0.865/0.796	0.784/0.748	0.752/0.788

AUC, area under receiver operating characteristic curve; CC, craniocaudal; HER2, human epidermal growth factor receptor 2; MLO, mediolateral oblique.

$P > .05$) among the three subtype groups. The majority of the breast cancers are mass (49.5%) and nonmass (49.5%, including architectural distortion and asymmetrical density) with only two microcalcifications. Of the 331 cancer cases, 305 (92.1%) were diagnosed as invasive ductal carcinomas of no special type and 26 (7.9%) were diagnosed as invasive carcinomas of special type (including 10 mucous carcinomas, 9 lobular carcinomas, and 7 papillary carcinomas).

Table 2 shows the comparisons of the overall classification performance. As shown, AUC ranged from 0.695 to 0.865, whereas accuracy ranged from 0.634 to 0.796 in all the experiments. In terms of AUC, MLO view shows a higher AUC than the CC view across the three binary classifications, and when the CC and MLO view were combined, AUC increased except for the luminal vs nonluminal classification. In terms of accuracy, there were no consistent trends between the CC and the MLO view data across the three subtype classifications; however, accuracy of the combination of the CC and MLO view outperformed either of them alone. In general, it is clear that the combination of CC and MLO view yielded the overall best classification performance.

In terms of the most significant features selected by the LASSO method, there were 11, 10, and 12 top-ranked features for the CC view, MLO view, and their combination, respectively. Figure 1 shows the boxplots of the 12 selected features from CC or MLO view images when they were combined for classification, including perimeter (CC), concavity (CC), correlation (CC), inverse different moment (CC), roundness (MLO), concavity (MLO), 10th Fourier coefficients (MLO), 24th Fourier coefficients (MLO), gray mean (MLO),

correlation (MLO), energy (MLO), and inverse different moment (MLO). Note that the three features, concavity, correlation, and inverse different moment, in both CC and MLO views were selected. Of the 12 features, 4 showed a statistically significant ($P < .05$) difference: concavity ($P = .027$), correlation ($P = .0015$), roundness ($P = .00016$), and gray mean ($P = .026$). More specifically, it can be seen that the concavity values of triple-negative samples tend to be smaller than those of HER2-enhanced and luminal samples (Fig 1f and b), the correlation values of luminal lesions tend to be larger than those of the other lesion types (Fig 1c), the roundness values of triple-negative samples were significantly larger than those of HER2-enhanced and luminal lesions, and Figure 1h shows that the gray mean values of triple-negative samples tend to be larger than those of the other subtypes.

DISCUSSION

In this study, we employed a radiomic approach to investigate the potential association between breast cancer molecular subtypes and quantitative imaging features extracted from digital mammogram images. Our results on the three binary classifications of subtypes (ie, triple-negative vs other types, HER2-enhanced vs other types, and luminal vs other types) showed that a set of such quantitative radiomic features is predictive of the molecular subtypes of breast cancer.

Although several previous studies have shown that breast MRI-derived radiomic features are associated with the subtypes (11,13,24–26), our study complements answering a question of whether imaging features in digital mammography

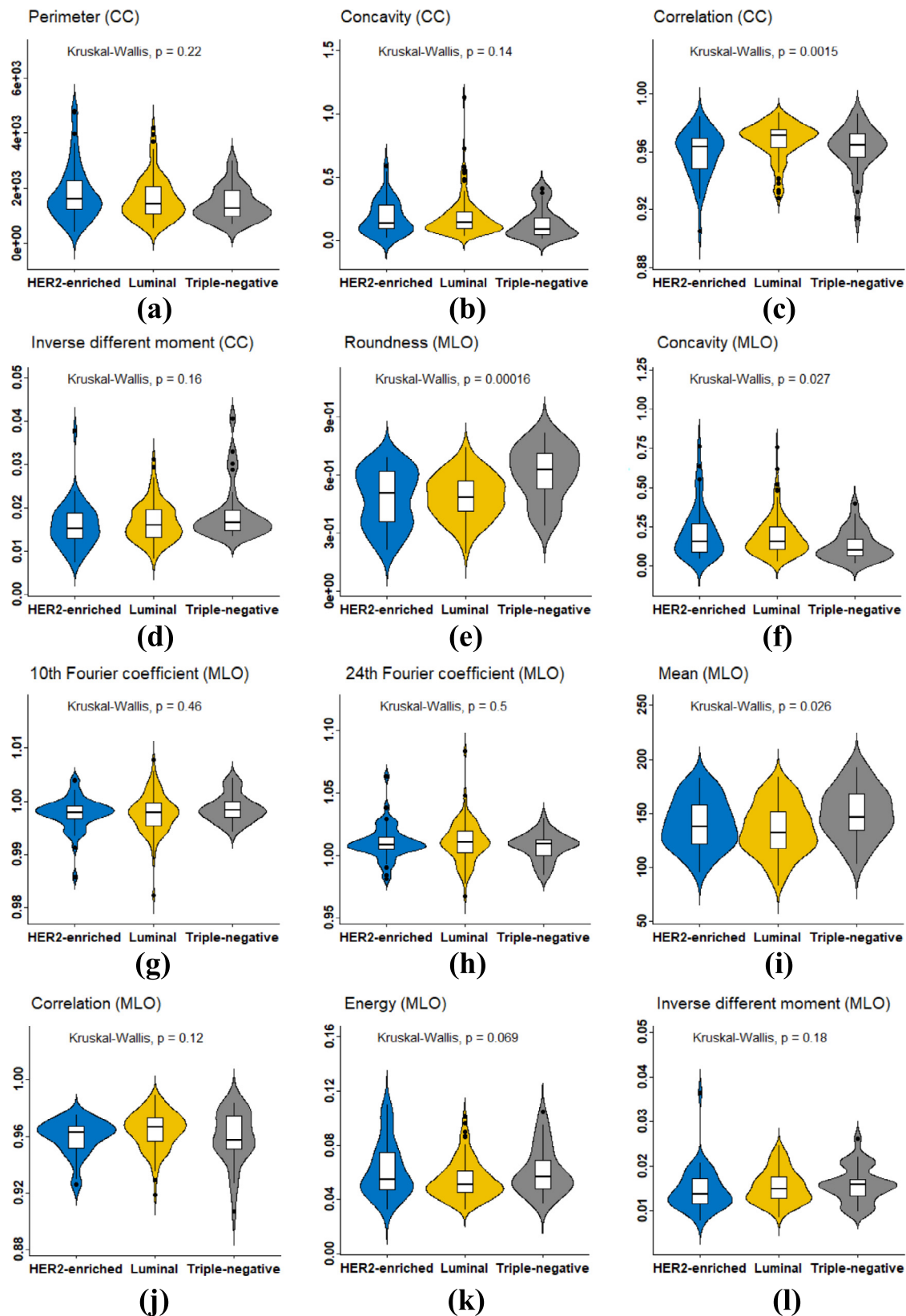


Figure 1. The top 12 ranked radiomic imaging features selected in the craniocaudal (CC) or mediolateral oblique (MLO) view images. (a) perimeter (CC), (b) concavity (CC), (c) correlation (CC), (d) inverse different moment (CC), (e) roundness (MLO), (f) concavity (MLO), (g) 10th Fourier coefficient (MLO), (h) 24th Fourier coefficient (MLO), (i) gray mean (MLO), (j) correlation (MLO), (k) energy (MLO), and (l) inverse different moment (MLO). Four of them, that is, (c) correlation (CC), (e) roundness (MLO), (f) concavity (MLO), and (i) gray mean (MLO), have statistically significant difference ($p < 0.05$) among the subtypes.

would have a similar association effect or not. Although mammography is different from MRI in terms of potential imaging traits that they may be able to capture, our study demonstrated the value of mammogram images that capture only morphologic or anatomic properties of the breast in helping assess breast cancer subtypes from the image-derived features.

Several previous studies have shown some correlation between breast cancer subtypes and certain qualitative mammographic characteristics or reading experience. For example, triple-negative subtypes may be more likely to manifest as an ill-defined mass, whereas non-triple-negative subtypes more likely to present as a spiculated mass; HER2-enriched subtypes are often characterized by structural distortion, and luminal lesions are more often to present like oval on mammogram images. Our finding is in line with these previous studies and we showed a further relationship between the molecular subtypes and the quantitative radiomic imaging features automatically extracted from digital mammograms.

In our study, four features were identified as most significant predictors of the subtypes: roundness, concavity, gray mean, and correlation. Roundness depicts a lesion's shape, whereas concavity reflects the irregularity of a lesion boundary. Gray mean simply represents the brightness of a lesion. Correlation represents the smoothing gradient of patterns, and a larger value indicates a smoother pattern. Based on these interpretations, our results indicate that triple-negative breast cancers may be more round, regular, and brighter than other subtypes. Similarly, our results imply that luminal lesions may be more smoothing than HER2-enriched and triple-negative breast cancers. These findings, however, will need further evaluation by larger future studies. Meanwhile, we noted that some of our identified features are different from those reported in previous breast MRI-based studies (11). This may be interpreted by the intrinsic difference in the nature of the two-dimensional mammogram images and three-dimensional DCE-MRI. This merits a further analysis and comparison of the two imaging modalities in future work. Because in our study cohort only a very small portion of the patients had breast MRI available, this prevented us from performing a comparative study between the two modalities for radiomic-based prediction of breast cancer subtypes.

We compared the effects of using MLO view, CC view, and their combination in this specific subtype classification task. Our finding that the combination of the two views achieved a higher performance may be due to the fact that the two views can provide more information than either of them alone.

Our study has several limitations. First, this was a retrospective analysis of a single-vendor image acquired at a single institution. It will be critical to evaluate whether our findings will generalize on other vendor images and external data. A future multicenter study may help address this question. Second, we used clinical immunohistochemical surrogate markers to categorize breast cancer subtypes. This is a routine approach in most such studies since in standard clinical practice genetic assay is not available to define the subtypes. Third,

we have mentioned that we were not able to compare the performance of mammogram images and DCE-MRI in this cohort due to the lack of MRI data. However, we would like to point out that this will be an important study to follow up. In addition, note that we did not separate luminal A and luminal B subtypes in our classification, mainly because they are similar and there is only a small difference between them in clinical management. Finally, we believe it will be worthwhile to test and compare to other machine learning techniques for the subtype classification work and we plan to do so in next steps.

Mammogram images are the most commonly available examination for breast cancer screening and diagnosis. If the automated radiomic features like we identified in this study are validated to be predictive of the molecular subtypes, it can provide further information from the images to aid radiologists in mammographic reading and to better inform clinical diagnosis and decision-making. This would have important additional value too for patients who do not have a breast MRI scan available.

In summary, this pilot radiomics study showed that quantitative imaging features extracted from digital mammograms are associated with breast cancer subtypes. Future larger studies are needed to further evaluate the findings and examine the relationship with breast MRI-identified features.

ACKNOWLEDGMENTS

This work was supported by the Key Project of Tianjin Science and Technology Committee Foundation grant (12ZCDZSY16000) and the Tianjin Municipal Government of China (15JCQNJC14500). This work was also partially supported by a National Institutes of Health (NIH)/National Cancer Institute (NCI) R01 grant (No. 1R01CA193603) and an R01 Supplement grant (No. 3R01CA193603-03S1).

REFERENCES

1. Goldhirsch A, Wood WC, Coates AS, et al. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* 2013; 18:1133–1144.
2. Lam SW, Jimenez CR, Boven E. Breast cancer classification by proteomic technologies: current state of knowledge. *Cancer Treat Rev* 2014; 40:129–138.
3. Huber KE, Carey LA, Wazer DE. Breast cancer molecular subtypes in patients with locally advanced disease: impact on prognosis, patterns of recurrence, and response to therapy. *Semin Radiat Oncol* 2009; 19:204–210.
4. Lambin P, Riosvelazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; 48:441–446.
5. Çelebi F, Pilancı KN, Ordu Ç, et al. The role of ultrasonographic findings to predict molecular subtype, histologic grade, and hormone receptor status of breast cancer. *Diagn Interv Radiol* 2015; 21:448–453.
6. Uematsu T, Kasami M, Yuen S. Triple-negative breast cancer: correlation between MR imaging and pathologic findings. *Radiology* 2009; 250:638–647.
7. Wu M, Jie M. Association between imaging characteristics and different molecular subtypes of breast cancer. *Acad Radiol* 2016; 24:426–434.

8. Luck AA, Evans AJ, James JJ, et al. Breast carcinoma with basal phenotype: mammographic findings. *AJR Am J Roentgenol* 2008; 191:346–351.
9. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging* 2012; 30:1234–1248.
10. Limkin EJ, Sun R, Dercle L, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* 2017; 28:1191–1206.
11. Li H, Zhu Y, Burnside ES, et al. Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. *NPJ Breast Cancer* 2016; 2:16012.
12. Wu J, Sun X, Wang J, et al. Identifying relations between imaging phenotypes and molecular subtypes of breast cancer: model discovery and external validation. *J Magn Reson Imaging* 2017; 46:1017–1027.
13. Sutton EJ, Dashevsky BZ, Oh JH, et al. Breast cancer molecular subtype classifier that incorporates MRI features. *J Magn Reson Imaging* 2016; 44:122–129.
14. Tang J, Rangayyan RM, Xu J, et al. Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *IEEE Trans Inf Technol Biomed* 2009; 13:236–251.
15. Baltzer PAT, Dietzel M, Kaiser WA. A simple and robust classification tree for differentiation between benign and malignant lesions in MR-mammography. *Eur Radiol* 2013; 23:2051–2060.
16. Mu T, Nandi AK, Rangayyan RM. Classification of breast masses using selected shape, edge-sharpness, and texture features with linear and kernel-based classifiers. *J Digit Imaging* 2008; 21:153–169.
17. Karsten R, Ewert B. A feature set for cytometry on digitized microscopic images. *Anal Cell Pathol* 2003; 25:1–36.
18. Haralick R, Shanmugam K, Dinstein I. Texture parameters for image classification. *IEEE Trans SMC* 1973; 3:610–621.
19. Peter H. *Machine learning in action*. Shelter Island, NY: Manning Publications, 2012.
20. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997; 16:385–395.
21. Chawla N, V, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002; 16:321–357.
22. Emaminejad N, Wang Y, Qian W, et al. Applying a radiomics approach to predict prognosis of lung cancer patients. In: *Medical imaging 2016: computer-aided diagnosis*. 2016; 97851E.
23. Maciejewski T, Stefanowski J. Local neighbourhood extension of SMOTE for mining imbalanced data. In: *Proceeding of the IEEE symposium on computational intelligence and data mining*. Paris, France: IEEE, 2011; 104–111.
24. Mazurowski MA, Zhang J, Grimm LJ, et al. Radiogenomic analysis of breast cancer: luminal B molecular subtype is associated with enhancement dynamics at MR imaging. *Radiology* 2014; 273:365–372.
25. Chang RF, Chen HH, Chang YC, et al. Quantification of breast tumor heterogeneity for ER status, HER2 status, and TN molecular subtype evaluation on DCE-MRI. *Magn Reson Imaging* 2016; 34:809–819.
26. Fan M, Li H, Wang S, et al. Radiomic analysis reveals DCE-MRI features for prediction of molecular subtypes of breast cancer. *PLoS ONE* 2017; 12:e0171683.