Published in final edited form as:

Med Phys. 2020 December ; 47(12): 6334–6342. doi:10.1002/mp.14538.

Machine learning prediction of axillary lymph node metastasis in breast cancer: 2D versus 3D radiomic features

Dooman Arefan,

Department of Radiology, University of Pittsburgh, School of Medicine, 4200 Fifth Ave, Pittsburgh, PA 15260, USA

Ruimei Chai,

Department of Radiology, First Hospital of China Medical University, Shenyang, Liaoning Province, China

Min Sun,

UPMC Hillman Cancer Center at St. Margaret, University of Pittsburgh School of Medicine, Pittsburgh, PA 15215, USA

Margarita L. Zuley,

Department of Radiology, University of Pittsburgh, School of Medicine, 4200 Fifth Ave, Pittsburgh, PA 15260, USA; Magee-Womens Hospital of University of Pittsburgh Medical Center, 300 Halket St, Pittsburgh, PA 15213, USA

Shandong Wu^{a)}

Departments of Radiology of Biomedical Informatics of Bioengineering, University of Pittsburgh, 4200 Fifth Ave, Pittsburgh, PA 15260, USA

Abstract

Purpose: The purpose of this study was to distinguish axillary lymph node (ALN) status using preoperative breast DCE-MRI radiomics and compare the effects of two-dimensional (2D) and three-dimensional (3D) analysis.

Methods: A retrospective study including 154 breast cancer patients all confirmed by pathology; 80 with ALN metastasis and 74 without. All MRI scans were achieved at a 3.0 Tesla scanner with 7 post-contrast MR phases sequentially acquired with a temporal resolution of 60 s. MRI radiomic features were extracted separately from a 2D single slice (i.e., the representative slice) and the 3D tumor volume. Several machine learning classifiers were built and compared using 2D or 3D analysis to distinguish positive vs negative ALN status. We performed independent test and 10-fold cross validation with multiple repetitions, and used bootstrap test, least absolute shrinkage selection operator, and receiver operating characteristic (ROC) curve analysis as statistical tests.

^{a)} Author to whom correspondence should be addressed. wus3@upmc.edu; Telephone: (412) 641-2567; Fax: (412) 641-2582.

CONFLICT OF INTEREST

The authors have no relevant conflicts of interest to disclose.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Results: The highest area under the ROC curve (AUC) was 0.81 (95% confidence intervals [CI]: 0.80–0.83) and 0.82 (95% CI: 0.81–0.82) for 2D and 3D analysis, respectively; the corresponding accuracy was 79% and 80%. The linear discriminant analysis (LDA) classifier achieved the highest classification performance. None of the AUC differences between 2D and 3D analysis was statistically significant for the several tested machine learning classifiers (all $P > 0.05$).

Conclusions: Radiomic features from segmented tumor region in breast MRI were associated with ALN status. The separate radiomic analysis on 3D tumor volume showed a similar effect to the 2D analysis on the single representative slice in the tested machine learning classifiers.

Keywords

2D/3D analysis; axillary lymph node (ALN) metastasis; breast cancer; MRI radiomics

1. INTRODUCTION

Axillary lymph node (ALN) is an important anatomic information for breast cancer staging. They can reflect the risk of distant recurrence and death subsequent to local therapy. The preoperative identification of lymph node metastasis will allow to apply more precise treatments and to better inform patient outcome. Tumor size has been shown to be positively correlated with the lymph node status, but not always; a tumor may begin to spread and progress through the lymphatic system even if it is in small size, or still stay local even if it develops to a larger size.¹ This observation may have to do with the heterogeneous biological characteristics of breast cancer.²

Radiomics is an approach to extract quantitative imaging features for building classification/prediction models^{3–5} and has been shown promising effect in many studies. Within the realm of breast cancer, dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) radiomics has been investigated in predicting breast cancer biology, including receptor status,⁶ subtype,⁷ and genomics.⁸ More specifically, MRI-derived radiomic features of breast tumor have been shown to be associated with certain gene expression profiles of breast cancer,⁹ responsiveness of neoadjuvant chemotherapy,¹⁰ recurrence likelihood and recurrence-free survival,¹¹ and measuring potentials of tumor metastasis.¹²

Radiomic features are usually extracted from a region of interest (ROI), for example, the segmented tumor region. Tumor segmentation is thus a critical step for many radiomics-related studies. While there are various semi-automated or fully automated segmentation methods published, either two-dimensional (2D) or three-dimensional (3D), most of them lack enough validation on reproducibility, robustness, and generalizability. As a result, many radiomic studies heavily rely on manual segmentation of ROIs from domain experts. Obviously, manual segmentation is time-consuming, tedious, and prone-to-error. Due to this reason, tumor segmentation has been done only on a single center/representative slice, or very few selected slices. Volumetric 3D segmentation on the full slices that include complete tumor region is not common, because it requires tremendous time to perform manual segmentation especially for large study cohorts or datasets. In principle, compared to using a single tumor slice (2D), 3D analysis covers the full tumor volume and thus can better depict spatial heterogeneity,^{13,14} an important aspect for tumor characterization. However,

the actual effects of using radiomic features extracted from a 2D slice or the 3D volume vary. It was shown that in a colorectal cancer study¹³ the overall heterogeneity of tumor volume was more closely correlated with the survival rates than the largest cross-sectional area of tumor. The study by Lubner et al.¹⁵ shows that there is no significant difference between the 2D and 3D texture analysis of hepatic metastatic colorectal cancer where the 2D analysis is enough to reflect tumor heterogeneity. Some studies¹⁶ also indicates that the heterogeneity of the tumor on 2D analysis in CT images is more accurate than the 3D volumetric characteristics of the tumor in predicting therapeutic response for patients with liver metastasis from colorectal cancer. These studies show that the effects of 2D or 3D radiomic analysis are not consistent, remaining to be further studied by different patient cohorts and tasks.

It has been showed that breast MRI has superior performance than some other imaging techniques such as ultrasound or positron emission tomography/computed tomography (PET-CT) to visually diagnose ALN status.¹⁷ While breast MRI is useful to evaluate of ALN status, it suffers from some limitations such as: the positive ALNs may locate at the border of the coil or may even be out of the imaging FOV due to being centered of the surface coil on the breast region,¹⁸ and the air–tissue interfaces usually produce large spatial variations in the local magnetic field that may affect the visibility of ALN.¹⁹

In this study, we performed an investigation on comparing the effects of 2D and 3D tumor analysis using preoperative breast DCE-MRI radiomics for distinguishing the ALN status.

2. MATERIALS AND METHODS

2.A. Study cohort and imaging protocol

This retrospective study was approved by the Institutional Review Board of our institution and informed consent was waived. We identified the invasive breast cancer patients newly diagnosed between April 2013 and September 2017 and excluded those if they had neoadjuvant chemotherapy history, postoperative recurrence, inflammatory breast cancer, obvious magnetic susceptible artifacts on DCE-MRI, lacking a pre-operative DCE-MRI scan. All breast cancer diagnoses were confirmed by pathological examination. A total of 154 breast cancer patients were enrolled in this study.

The following clinical factors for each patient were collected: patient age, tumor size, ALN metastasis status. The ALN metastasis was categorized as positive and negative. The positive ALN metastasis was confirmed by either lymph node dissection or ultrasound-guided fine-needle aspiration, and the negative status was confirmed by lymph node dissection (16).

Breast DCE-MRI examinations were acquired for each patient using the same 3.0 T MR scanner (Siemens Magnetom Verio, Erlangen, Germany) with an 8-channel breast-specific phased-array coil. The DCE imaging was performed using axial 3D volumetric interpolated breath-hold examination (VIBE) technique of Siemens (TR/TE = 4.67 ms/1.66 ms, field of view = 360 × 360 mm, matrix = 384 × 296, flip angle = 10°, section thickness = 1.2 mm, gap = 0 mm). After the pre-contrast acquisition, Magnevist (gadopentetate dimeglumine, Bayer Healthcare, Berlin, German) was injected in a 0.2 ml/kg (or 0.1 mmol/kg body

weight) through the vein at a rate of 3.0 ml/s, followed by a flush of equal volume of saline solution, and then 7 post-contrast phases were sequentially acquired with a temporal resolution of 60 s.

2.B. Tumor segmentation

Breast tumor was segmented semi-automatically by an experienced radiologist (R.C., with 11 yr of experience in breast MRI) blinded to the pathological outcome information. Tumor segmentation was performed on the second post-contrast sequence (denoted by CE2). The reason of using CE2 was because this sequence exhibits the best lesion-background contrast according to the assessment of our expert radiologist. Tumor was segmented in two steps: (1) the radiologist first selected a center/representative slice out of all tumor-affected slices and drawn an initial rectangular ROI to cover the tumor region; the initial ROI was automatically propagated to other slices, and it was redrawn if needed on the center/representative slice, to ensure the entire tumor volume across different slices was not missed; and (2) an in-house developed interactive segmentation software was applied to the ROIs to automatically segment the tumor on the selected slice first and then all the other tumor-affected slices. The segmentation algorithm consisted of three steps: preprocessing, segmentation, and postprocessing. In preprocessing, we reduced noise by applying a median filter with a kernel size of 3-by-3 pixels on each slice, followed by image contrast enhancement using histogram equalization.²⁰ Tumor segmentation was initiated by the energy minimization approach in the bounding box drawn by radiologists, followed by the extended seeded region growing (SRG) technique²¹ to implement segmentation. At postprocessing, we used the 3D connected component algorithm to reduce noise and small speckles.

Note that all segmentation results from the software were reviewed, adjusted, and finally confirmedly accepted by the radiologist in a slice-by-slice basis for each MRI scan. We denoted “2D segmentation” as the segmented tumor on the center/representative slice only and “3D segmentation” as the tumor volume generated by stacking the segmentations on all tumor-affected slices. Figure 1 illustrates the segmentation processes.

2.C. Radiomic features and definition of analysis

We defined 2D analysis and 3D analysis as follows: 2D analysis was based on radiomic features extracted from the 2D segmentation while 3D analysis was based on radiomic features extracted from the 3D segmentation. We studied three types of features (Fig. 2): kinetic (45 features), morphological (17 and 10 shape-based features in 3D and 2D analysis, respectively), and texture (94 features) in both 2D and 3D analysis. The calculation of kinetic and texture features followed the same expression in 2D and 3D but different operation at pixel or voxel basis. The majority of our radiomic features extracted by Pyradiomics were in compliance with the Imaging Biomarker Standardization Initiative (IBSI) recommendation.

2.C.1. Kinetic features (45 features)—The pre-contrast and all seven time-point post-contrast sequences were used to calculate three categories of kinetic features: aggregate, heterogeneity, and textural kinetic features, using an in-house developed algorithm.

Aggregate kinetic features (6 features): They were calculated directly from the characteristic kinetic curve of each voxel over the tumor region. These features included: wash-in slope (WIS), wash-out slope (WOS), peak enhancement (PE), and time to peak (TTP). They were computed using the average values for each pixel/voxel's kinetic parameters over the tumor region. In addition, we also calculated the most enhanced region/spot, that is, the maximum PE value and hot spot (the highest connected pixel/voxels average²²) signal enhancement ratio over the tumor region, as two additional features.

Heterogeneity kinetic features (21 features): The segmented tumor was first partitioned into three sub-clusters²³ according to their TTP values (TTP ranges from 1 to 7, corresponding the 7 post-contrast sequences). The three subclusters (denoted as TTP characteristic maps) were formed by categorizing each pixel to represent quick, intermediate, and slow arrivals to the peak point²³ as following: a pixel with TTP value of 1 or 2, 3 or 4, and greater than 4 was categorized as quick, intermediate, and slow, respectively. The heterogeneity kinetic features were calculated on the TTP characteristic maps, using the average and standard deviation (SD) of kinetic parameters (i.e., WIS, WOS, and PE) and the proportion of pixels in each of the three subclusters.

Textural kinetic features (18 features): We also computed typical textural features on the above-created TTP characteristic maps, including inverse difference moment (IDM), absolute value of differences (AVD), contrast, skewness, kurtosis, angular second moment (ASM), energy, SD, etc. a total of 18 features.

2.C.2. Morphological and texture features—Typical morphological and texture features were calculated using Pyradiomics open-source python package²⁴ on the original intensity space of the segmented tumor. These included 17 and 10 shape-related/morphological features in 3D and 2D analysis, respectively (e.g., elongation, flatness, mesh volume, surface area, sphericity, and major axis length, etc.), 19 first-order features (e.g., energy, entropy, minimum, 10th percentile, mean, kurtosis, and skewness, etc.), and 75 higher-order statistics features based on gray-level size zone matrix (GLSZM), gray-level dependence matrix (GLDM), gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), and neighboring gray tone difference matrix (NGTDM). All images were rescaled to 8 bits/voxel and gray level quantization was performed using a fixed bin width (binWidth = 25). More details on the texture feature calculation were specified in the supplementary material (Radiomics_SuppInfo.docx). In contrast to GLCM and GLRLM, the GLDM, NGTDM, and GLSZM are rotation independent.

2.D. Modeling and statistical analysis

Binary machine learning (ML) classifiers were built to distinguish positive vs negative ALN status. Here, five classifiers were used and compared, including linear discriminant analysis (LDA), random forest (RF), naive Bayes (NB), K-nearest neighbor (KNN), and support vector machine (SVM).

To compare the effects of radiomic analysis on 2D vs 3D tumor segmentation, different ML models were built and compared using 2D and 3D features. We implemented an independent test approach to evaluate the models. Out of the 154 patients, a group of 103 (67%) patients

were randomly selected as the training set, and the remaining unseen 51 (33%) cases were held out as an independent testing set. Furthermore, in view of the overall scale of our study cohort, we performed a secondary evaluation by 10-fold cross-validation on the full 154 patients as an additional means to examine the model's robustness. We repeated the 10-fold cross-validation 50 times, and in each iteration, the whole dataset was split randomly to reduce potential bias induced by the split.

The least absolute shrinkage and selection operator (LASSO) technique was employed to select the best-performing subset features according to mean square error curves.²⁵ Note that the feature selection was conducted on the training set only for both the independent test and 10-fold cross-validation. The top LASSO-ranked features were identified and compared for their distributions in 2D vs 3D analysis. Model performance was evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The bootstrap test method was used to compare the difference between two AUCs ($P < 0.05$ was considered statistically significant).

3. RESULTS

3.A. Patient characteristics

Table I summarizes some key patient characteristics. The mean age of the 154 patients was 46.1 yr old and the average tumor size was 3.1 cm. Both age and tumor size were not significantly different between the training and independent test groups ($P > 0.05$). The proportions of positive and negative cases were similar in the training and independent test sets.

3.B. Classification performance comparison

Table II shows results of the independent test and 10-fold cross-validation under 2D and 3D radiomic analysis by different classifiers. As can be seen, for the independent test, the 3D analysis showed a similar performance to the 2D analysis, where the differences among the classifiers were small; similar results were observed for the other performance metrics, i.e., accuracy, sensitivity, specificity, PPV, and NPV. The AUC difference between 3D and 2D analysis was not statistically significant for all the tested classifiers (all $P > 0.05$). Figure 3 shows the ROC curves for the LDA classifier.

In terms of the additional evaluation by 10-fold cross-validation with 50-time repeats, similar observations were found that 3D analysis showed a similar performance to 2D analysis. Specifically, the highest average AUC and accuracy were achieved in the LDA classifier in both 2D and 3D analysis. The AUC and accuracy were 0.81 [95% confidence interval (CI): 0.80–0.83] and 79% (95% CI: 0.75–0.83), respectively in 2D analysis. In comparison, the 3D analysis achieved an average AUC of 0.82 (95% CI: 0.81–0.82) and accuracy of 80% (95% CI: 0.78–0.83).

3.C. Comparison of selected features under independent test

To further evaluate the difference of the radiomic models, we used the independent test to measure the statistics and distribution of the features selected in both 2D and 3D analyses.

To better understand the features that lead to the different classification performance in both 2D and 3D analyses, we ranked all extracted features. Table III lists the top-five LASSO-ranked features on the training set for both 2D and 3D analysis. Figure 4 shows the distributions of each feature type among the top-ten LASSO-ranked features for each model. There are several observations on Fig. 4. First, as shown in (a) and (b), more kinetic features drove the models than morphological and texture features, regardless in 2D or 3D analysis. Second, as shown in (c), textural kinetic features played the major role among all kinetic features; furthermore, more textural kinetic features were selected in 3D analysis compared to 2D analysis. Third, among the morphological and texture features, higher-order and first-order statistics features played the main role for both 3D and 2D analyses, as indicated in (d). In addition, for first-order features, more 2D features were selected than 3D features, as shown in (d).

4. DISCUSSION

In this study, we evaluated ALN status on breast cancer patients using the tumor-derived radiomic features from breast MRI. Specifically, we investigated the respective effects of radiomic features extracted from a 2D representative slice and the 3D tumor volume in distinguishing the ALN status. Our results on the five classifiers showed that the 2D and 3D analyses had a similar effect in expressing the association with the ALN status in breast cancer. Previous studies achieved 83–100% sensitivity and 90–56% specificity, based on human experts in preoperative diagnosis of ALN metastasis status on DCE-MRI images.^{26,27} The radiomics models showed a similar performance to human experts, and thus it may provide an automatic assessment to potentially assist clinical diagnosis of ALN metastasis in breast cancer patients.

Most prior studies have employed either a single slice (e.g., a clinician-selected representative slice, the slice with the largest tumor diameter, or simply the center slice)²⁸ or whole 3D tumor volume (generated from multiple slices) to extract features for radiomics analysis.⁷ When automated and reliable segmentation algorithms are not widely available, manual segmentation by expert physicians are heavily relied. Obviously, manual segmentation has to be used, 3D tumor segmentation needs more efforts than 2D because more slices need to be processed. This is part of the reasons that many studies focused only on using a single slice in order to reduce the time and cost associated with 3D segmentation.

While that, the critical point here is what would be the extra gain to be obtained from by using 3D tumor volume over a single slice, and how likely the extra gain may or may not result in a different model performance in different classification tasks.

In the specific classification task of this study, both the independent test and 10-fold cross-validation results from the five classifiers showed that there was no significant difference between the AUCs achieved by the 3D vs 2D analysis. This study contributes more insights

to the community on the effects of radiomic analysis in 2D vs 3D space for a breast cancer-related classification task.

While many tumor-affected slices were not even included in the 2D radiomic analysis, the actual effects resulted from a single representative slice were in general comparable to the whole tumor volume. This observation was in line with some of previous reported studies.¹⁵ This may support some of the promising model performance reported in the literature where only a single slice was used.²⁸ The reasons behind this observation may be multi-fold, where one may be that a “representative” slice may indeed be capable of capturing most of the tumor characteristics conveyed in the 3D tumor volume. Although, 2D analysis may be susceptible to the choices of the single representative slice chosen by human experts, while 3D analysis is not subject to this variation.

Another consideration factor for 2D or 3D analysis is the availability of mature segmentation methods. If validated automated computer methods are widely available for both 2D and 3D segmentation, using 2D or 3D analysis will be less annoying because automated segmentation is fast, reliable, and reproducible. Unfortunately, most computerized segmentation methods remain to be fully validated by large-scale evaluation. Thus, when it comes to use such methods to process a large scale of images for segmenting tumors, researchers often are not confident enough to simply employ these partially-evaluated methods to process large imaging datasets without manual inspection. Because if the segmentation is not robust/accurate, it is very likely to lead to biased or even wrong radiomic analysis results. From this perspective, sufficient validation of automated segmentation methods is always important and demanded for the radiomics study community.

In terms of the feature distribution breakdown, the majority of top-ranked features were kinetic features in both 2D and 3D analyses. This shows the stronger utility of kinetics-related tumor enhancement information in breast DCE-MRI for the targeted classification tasks. In particular, the textural kinetic features, that is, texture features extracted from TTP characteristic maps, had the dominant role in both 2D and 3D models. This result is consistent with previous studies that showed tumor enhancement texture features are prevalent across many differentiation tasks for predicting triple-negative cancers.²⁹ This also indicates that these texture features are more capable of capturing effective information than directly calculation from raw intensity data. Among the morphological and texture features, higher-order texture features (i.e., GLSZM, GLDM, GLCM, GLRLM, and NGTDM) contributed more in 3D than 2D analysis. This indicates that more texture details of tumor areas captured from the entire tumor volume (i.e., 3D higher-order texture features) were more informative than those captured from the 2D single slice. In contrast, the first-order texture features which capture general texture information of tumor areas were more dominant in 2D than 3D. On the other side, the features captured from the shape of tumor in 3D volume (i.e., 3D shape-based features) were more predictive than those captured from a single slice. This is not surprising as shape-based radiomic features are highly correlated to the tumor contours, and 2D segmentation is less robust than 3D due to the variation in picking the representative slice by human experts.

In view of the scale of our study cohort, we implemented two evaluation schemes, i.e., independent test and cross-validation with multiple repeats, which provided a more extensive evaluation on the model analysis and result comparison. The cross-validation exhibited a comparable effect to the independent test in all the tested classifiers.

Our study has few limitations. The tumor segmentation was semi-automated with necessary user interactions, which may have introduced variations and is not fully reproducible. However, we feel for our specific study purpose, it is important to have such semi-automated and user-guided segmentation in order to ensure that all segmentation is precise or at least acceptable to experienced expert radiologists. As mentioned earlier, it will be ideal to use mature and validated fully automated segmentation methods as they are available. The potential influence of different experts or algorithms on tumor segmentation and thus radiomic features needs further evaluation in order to fully investigate the model performance comparison between 2D and 3D analysis. In addition, this was a retrospective study with a single-institutional dataset which may have limited our analyses. Large studies with more diverse patient cohorts and imaging datasets are warranted for further study. Finally, the radiomics features used in our study may not be comprehensive but we intended to keep the feature set compact and relevant considering our sample size of study cohort.

In conclusion, we investigated and compared the effects of 2D and 3D radiomic analysis on diagnosis of ALN metastasis status. The main motivation behind was the concern on information extracted from a single MR slice vs. the entire tumor volume and how that may lead to a difference in radiomics model performance. This specific study showed a statistically insignificant difference on the association effect between 3D tumor volume and 2D single representative slice, but further evaluation on large datasets is warranted.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health (NIH)/National Cancer Institute (NCI) R01 grants (#1R01CA193603, #3R01CA193603-03S1, and #1R01CA218405), a Radiological Society of North America (RSNA) Research Scholar Grant (#RSCH1530), an Amazon Machine Learning Research Award, and a University of Pittsburgh Physicians (UPP) Academic Foundation Award. We thank the computational support by the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation (NSF) grant number ACI-1548562. Specifically, it used the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center.

REFERENCES

1. Jardines L, Goyal S, Royce M, Jaiyesimi I, Goldfarb SB. Stages III and IV breast cancer. 2009. <https://www.cancernetwork.com/view/stages-iii-and-iv-breast-cancer>
2. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012;366:883–892. [PubMed: 22397650]
3. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2015;278:563–577. [PubMed: 26579733]
4. Arefan D, Talebpour A, Ahmadinejad N, Asl AK. Automatic breast density classification using neural network. *J Instrum*. 2015;10:T12002.

5. Talebpour A, Arefan D, Mohamdlou H. Automated abnormal mass detection in the mammogram images using chebyshev moments. *Res J Appl Sci Eng Technol*. 2013;5:513.
6. Guo W, Li H, Zhu Y, et al. Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data. *J Med Imaging*. 2015;2:41007.
7. 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. [PubMed: 27853751]
8. Zhu Y, Li H, Guo W, et al. Deciphering genomic underpinnings of quantitative MRI-based radiomic phenotypes of invasive breast carcinoma. *Sci Rep*. 2015;5:17787. [PubMed: 26639025]
9. Ashraf AB, Daye D, Gavenonis S, et al. Identification of intrinsic imaging phenotypes for breast cancer tumors: preliminary associations with gene expression profiles. *Radiology*. 2014;272:374–384. [PubMed: 24702725]
10. Fan M, Wu G, Cheng H, Zhang J, Shao G, Li L. Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients. *Eur J Radiol*. 2017;94:140–147. [PubMed: 28712700]
11. Li H, Zhu Y, Burnside ES, et al. MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays. *Radiology*. 2016;281:382–391. [PubMed: 27144536]
12. Yamamoto S, Han W, Kim Y, et al. Breast cancer: radiogenomic biomarker reveals associations among dynamic contrast-enhanced MR imaging, long noncoding RNA, and metastasis. *Radiology*. 2015;275:384–392. [PubMed: 25734557]
13. Ng F, Kozarski R, Ganeshan B, Goh V. Assessment of tumor heterogeneity by CT texture analysis: can the largest cross-sectional area be used as an alternative to whole tumor analysis? *Eur J Radiol*. 2013;82:342–348. [PubMed: 23194641]
14. Skogen K, Ganeshan B, Good T, Critchley G, Miles KA. Imaging heterogeneity in gliomas using texture analysis. *Cancer Imaging*. 2011;11:S113.
15. Lubner MG, Stabo N, Lubner SJ, et al. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. *Abdom Imaging*. 2015;40:2331–2337. [PubMed: 25968046]
16. Ahn SJ, Kim JH, Park SJ, Han JK. Prediction of the therapeutic response after FOLFOX and FOLFIRI treatment for patients with liver metastasis from colorectal cancer using computerized CT texture analysis. *Eur J Radiol*. 2016;85:1867–1874. [PubMed: 27666629]
17. Valente SA, Levine GM, Silverstein MJ, et al. Accuracy of predicting axillary lymph node positivity by physical examination, mammography, ultrasonography, and magnetic resonance imaging. *Ann Surg Oncol*. 2012;19:1825–1830. [PubMed: 22227922]
18. Yun SJ, Sohn Y-M, Seo M. Differentiation of benign and metastatic axillary lymph nodes in breast cancer: additive value of MRI computer-aided evaluation. *Clin Radiol*. 2016;71:403–e1. [PubMed: 26875621]
19. Huang SY, Seethamraju RT, Patel P, Hahn PF, Kirsch JE, Guimaraes AR. Body MR imaging: artifacts, k-Space, and solutions. *Radiographics*. 2015;35:1439–1460. [PubMed: 26207581]
20. Zhou Y, Shi C, Lai B, Jimenez G. Contrast enhancement of medical images using a new version of the World Cup optimization algorithm. *Quant Imaging Med Surg*. 2019;9:1528. [PubMed: 31667139]
21. Adams R, Bischof L. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell*. 1994;16:641–647.
22. Almendro V, Fuster G. Heterogeneity of breast cancer: etiology and clinical relevance. *Clin Transl Oncol*. 2011;13:767–773. [PubMed: 22082639]
23. Chai R, Ma H, Xu M, et al. Differentiating axillary lymph node metastasis in invasive breast cancer patients: a comparison of radiomic signatures from multiparametric breast MR sequences. *J Magn Reson Imaging*. 2019;50:1125–1132. [PubMed: 30848041]
24. Van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77:e104–e107. [PubMed: 29092951]
25. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B*. 1996;58:267–288.

26. Murray AD, Staff RT, Redpath TW, et al. Dynamic contrast enhanced MRI of the axilla in women with breast cancer: comparison with pathology of excised nodes. *Br J Radiol.* 2002;75:220–228. [PubMed: 11932214]
27. Kvistad KA, Rydland J, Smethurst H-B, Lundgren S, Fjøsne HE, Haraldseth O. Axillary lymph node metastases in breast cancer: preoperative detection with dynamic contrast-enhanced MRI. *Eur Radiol.* 2000;10:1464–1471. [PubMed: 10997438]
28. 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. [PubMed: 26756416]
29. Wang J, Kato F, Oyama-Manabe N, et al. Identifying triple-negative breast cancer using background parenchymal enhancement heterogeneity on dynamic contrast-enhanced MRI: a pilot radiomics study. *PLoS One.* 2015;10:e0143308. [PubMed: 26600392]

Fig. 1.

Tumor segmentation illustration on breast dynamic contrast-enhanced magnetic resonance imaging for a 57-yr-old woman diagnosed with invasive ductal carcinoma in the right breast. (a) The second post-contrast sequence (i.e., CE2) was used for segmentation. (b) The selected center/representative slice of all tumor-affected slices with a manually drawn rectangular ROI (blue line) by the radiologist. (c) Two-dimensional tumor segmentation (red contour). (d) Three-dimensional tumor segmentation (red contour).

Fig. 2.
Schematic illustration of the radiomic features extracted in our study.

Fig. 3.

ROC curves for classifying ALN metastasis status using the linear discriminant analysis classifier, comparing the effects of two-dimensional (2D) and three-dimensional (3D) analysis. The area under curve (AUC) difference between 2D and 3D analysis was not statistically significant ($P > 0.05$).

Fig. 4.

The distribution/contribution of top 10 radiomic features in two-dimensional (2D) and three-dimensional (3D) analysis. (a and b) The percentage of kinetic vs morphological and texture features. (c) The distribution of the three categories of kinetic features. (d) The distribution of the three categories of morphological and texture features.

Table I.

Patient characteristics

	Full cohort (n = 154)	Training set (n = 103)	Independent test set (n = 51)	P- value
Age (mean $\pm$ SD ^a), years old	46.1 $\pm$ 10.3	46.6 $\pm$ 10.6	44.5 $\pm$ 9.6	0.33
Tumor size (mean $\pm$ SD), cm	3.1 $\pm$ 1.9	3.0 $\pm$ 1.9	3.4 $\pm$ 1.9	0.32
ALN: Positive/Negative	80/74	53/50	27/24	-

^aStandard deviation

Table II.

The model performance for classifying axillary lymph node (ALN) status under the independent test and 10-fold cross validation evaluation in two-dimensional (2D) and three-dimensional (3D) radiomic analysis for different machine learning (ML) classifiers

Model	Analysis	AUC	P	Accuracy ^a %	Sensitivity%	Specificity%	NPV ^a %	PPV ^b %
Independent test								
LDA	2D	0.75	0.35	74	60	87	66	84
	3D	0.79		78	63	92	69	89
RF ^c	2D	0.74	0.32	73	60	86	65	82
	3D	0.78		77	64	90	67	88
NB ^d	2D	0.78	0.46	74	81	67	76	73
	3D	0.78		74	85	62	79	72
KNN ^e	2D	0.64	0.91	64	74	54	65	64
	3D	0.63		63	67	58	61	64
SVM ^f	2D	0.77	0.40	71	70	71	68	73
	3D	0.72		66	81	50	71	65
10-fold cross validation [mean (95% CI)]								
LDA	2D	0.81 (0.80-0.83)	0.63	79 (75-83)	82 (79-85)	76 (71-81)	82 (80-84)	82 (79-84)
	3D	0.82 (0.81-0.82)		80 (78-83)	82 (80-85)	78 (76-80)	82 (79-84)	82 (81-83)
RF	2D	0.78 (0.76-0.80)	0.53	79 (74-83)	89 (85-93)	68 (63-73)	89 (85-92)	79 (76-82)
	3D	0.80 (0.78-0.82)		80 (74-86)	89 (85-93)	70 (62-78)	88 (85-91)	79 (74-84)
NB	2D	0.78 (0.77-0.79)	0.49	78 (73-83)	73 (70-77)	82 (76-88)	77 (74-79)	84 (79-89)
	3D	0.75 (0.72-0.78)		73 (65-81)	69 (60-78)	76 (69-83)	71 (60-83)	82 (79-85)
KNN	2D	0.77 (0.75-0.79)	0.55	77 (74-81)	79 (76-82)	75 (71-79)	78 (75-81)	79 (76-81)
	3D	0.75 (0.73-0.76)		75 (71-80)	75 (71-79)	75 (70-80)	74 (72-77)	77 (74-80)
SVM	2D	0.80 (0.79-0.82)	0.28	77 (74-80)	65 (62-69)	88 (86-91)	72 (70-74)	87 (85-90)
	3D	0.76 (0.75-0.78)		76 (73-80)	73 (71-76)	79 (75-83)	75 (74-76)	81 (78-85)

^aNPV: Negative predictive value.

^bPPV: Positive predictive value (Precision).

^cRF: Random Forest.

* Note that we calculated the balanced accuracy for each observer.

f_{SVM} : Support Vector Machine

e_{KNN} : *K-nearest neighbor*

p_{NB} : Naive Bayes.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table III.
The top 5 least absolute shrinkage and selection operator (LASSO)-ranked features selected on the training set

Rank	Features in 2D analysis	Type	Features in 3D analysis	Type
1	Energy	Textural kinetic	Maximum PE ^a /GLSZM (GLN) ^b	Aggregate kinetic/Higher-order
2	Maximum diameter	Shape-based	AVD ^c	Textural kinetic
3	GLCM (CS) ^d	Higher-order	GLSZM (SZN) ^e	Higher-order
4	GLDM (SDLGLE) ^f	Higher-order	Least Axis Length	Shape-based
5	Skewness	Textural kinetic	Maximum diameter/Skewness	First-order/Textural kinetic

^aPE: Peak enhancement.
^bGLSZM (GLN): Gary level size zone matrix (**Gray Level Non-Uniformity**).
^cAVD: Absolute Value of Differences.
^dGLCM (CS): Gray level Co-occurrence matrix (Cluster Shade).
^eGLSZM (SZN): Gary level size zone matrix (**Size-Zone Non-Uniformity**).
^fGLDM: Gray level dependence matrix (**Small Dependence Low Gray Level Emphasis**).