10.1148/radiol.2018180608

Appendix E1

Methods: Three Compartment Breast Imaging and Feature Extraction for Quantitative 3CB Image Analysis

Three compartment breast (3CB) imaging is a dual-energy imaging technique that is easily integrated into existing full-field digital mammography (or digital breast tomosynthesis) units with inexpensive modifications. 3CB imaging produces quantitative and reproducible information of the water, lipid, and protein content, the three 'compartments', throughout the imaged breast (24). Hologic Selenia full-field digital mammography systems (Hologic, Inc.) were used to image women with 3CB at two clinical sites. This system configuration has a molybdenum x-ray anode and two internal x-ray filters of either molybdenum or rhodium. Two mammographic images were acquired on each woman's affected breast using a single compression. The first exposure was made to mimic the clinical screening or diagnostic mammogram conditions such that Selenia's internal software chooses the voltage and current settings based on breast thickness (usually below 30 kVp). The second mammographic image was acquired at a fixed voltage of 39 kVp (the highest attainable voltage on the Selenia unit) and fixed current for all participants. This high-energy exposure was made using an additional 3-mm plate of aluminum in the beam to raise the average energy of the high-energy image. We limited the total dose of this procedure to be approximately 110% of the mean-glandular dose of an average screening mammogram.

Derivation of thickness maps of the 3 breast compartments of water, lipid, and protein is possible through 1) the use of dual-energy mammography and 2) the use of phantoms modeling each compartment and different compartment thickness combinations. In the phantoms, water is modeled by plastic water, lipid by wax, and protein by Delrin. 3CB imaging uses two phantoms: a calibration phantom of 51 compartment thickness combinations imaged prior to each study participant, and a smaller phantom (the "SXA phantom") attached to the mammography paddle imaged concurrently with the study participant (Fig E1). The SXA phantom contains 9 modeled thickness combinations and 9 metal beads. The position of this within-image phantom was used for the breast thickness estimation since the compressed thickness given by the mammography unit was not accurate enough due to the paddle tilt angle (33). It was also used to solve the attenuation equations at the two x-ray energies. For monochromatic x-rays, the attenuation equations at two energies can be solved analytically to yield the three compartments, ie, the water-lipid-protein thicknesses, for a known breast thickness. For polychromatic x-rays (as in mammography) this is no longer the case and a Taylor expansion up to second order was used to derive the compartment thicknesses. It is important to note that 3CB imaging yields quantitative and reproducible compartment thickness maps throughout the imaged breast. The calibration standards and 3CB algorithms are described in full elsewhere (24). It takes only a few seconds to generate the 3CB maps from the dual-energy mammography acquisitions.

In quantitative 3CB (q3CB) image analysis, ie, in the analysis of the quantitative 3CB thickness maps, 'simple' features were extracted from the water, lipid, and protein thickness maps: the mean, median, standard deviation, and skewness within a lesion (based on the computer segmentation, see Appendix E2) and within a 2 mm band surrounding the lesion. Moreover, the difference, as well as the ratio, in values between a lesion and its surrounding were calculated, resulting in 16 features for each compartment (16 for water, 16 for lipid, and 16 for protein thickness maps, respectively) (Table E1). However, in our current work we used a predefined q3CB feature signature derived from reanalysis of

pilot study data (28): median water thickness within a mass ($f_{q,1}$), the median water thickness within the surrounding parenchyma ($f_{q,2}$), the ratio of the median water thicknesses of a mass and surrounding parenchyma ($f_{q,3}$), and the skewness of the lipid thickness within a mass ($f_{q,4}$, quantifying the asymmetry of the lipid thickness distribution or, in other words, the extent to which the lipid thickness distribution differs from a normal distribution).

Appendix E2

Methods: Lesion Segmentation and Feature Extraction for Mammography Radiomics

The mammographic masses were segmented on the low-energy mammograms (equivalent to conventional diagnostic digital mammograms) with a dual-step method that uses as input the images and the approximate lesion centers (in this case the centers of the radiologist mass delineations) (26). The first step was to obtain an initial estimate of the mass boundary and the second step was to further refine that initial estimate. The initial estimate for the boundary was obtained through a method based on the radial gradient index which is a measure combining boundary irregularity and mass size (34). The radial gradient index ranges between negative one (spherical dark mass in a bright background, an 'ideal' mass in breast ultrasound) and one (spherical bright mass in a dark background, an 'ideal' mass in mammography). In this initial segmentation step, a mammogram was first multiplied by a Gaussian function that was centered at the manually indicated approximate mass center and had a fixed width (15 mm in our application). Candidate mammographic mass boundaries were then obtained though iterative gray value thresholding that image and for each candidate boundary the radial gradient index was calculated (34). The initial estimate for the mass margin was chosen as the candidate boundary with the maximum radial gradient index. The final computer mass delineation was obtained using the initial boundary estimate and original mammogram as input to an active contour method (26).

For mammography radiomics, 32 features pertaining to mass size, shape, morphology (including margin), and texture were calculated. In this work, however, we used a predefined feature signature of 5 features that was derived in our previous work on a different dataset (27) (Table E2).

Appendix E3

Methods: Deriving Models for q3CB, Mammography Radiomics, and Combined Mammography Radiomics Plus q3CB Analyses, and Assessing the Potential for Additive Benefit

The computer extracted tumor features (as described above) served as input to a linear discriminant (LDA) classifier, which was trained and tested within 10-fold cross-validation (Fig E2). For each of the 10 training folds of the cross-validation, a linear model was obtained that merged for each mass the multiple input features into a single output of probability of malignancy, ie, the classifier weights of the LDA were determined using the 'ground truth' for the training fold. The model obtained for a given training fold was used to predict the probabilities of malignancy in the corresponding (independent) test fold and the values for the model coefficients were stored. After completion of the 10-fold cross-validation, the estimated probabilities of malignancy for all the test folds were aggregated (since each mass was used in the testing capacity exactly one time) and used in combination with the 'ground truth' to assess performance.

Since 10-fold cross-validation results in 10 classification models, one for each of the 10 training folds, we assessed the stability (trustworthiness) of the estimated values for the coefficients of the linear classification models (those for the q3CB features, mammography radiomics features, and combined q3CB plus mammography radiomics features), we repeated 10-fold cross-validation 10 times. In each of the 10 cross-validations we randomly repartitioned the data into 10 nonoverlapping training and testing folds. Thus, we obtained 100 models for each analysis. For each model coefficient, the median value, 25th and 75th percentiles were calculated and box plots were constructed. The narrower a 'box' in these box plots, the more stable, ie, reliable, the estimate for a model coefficient and the more likely that the values for the performance metrics will be reproducible. A 'final' model then, can be constructed by using the median values for each coefficient of the linear model.

Our current study sought to 1) confirm that the mass signatures (feature combinations) obtained in previous work (through stepwise multilinear regression) (27,28) translated to the current dataset and 2) to assess whether combining mammography radiomics and q3CB has potential for improving classification performance. The Pearson correlation coefficient (29) between probabilities of malignancy estimated from mammography radiomics and from the q3CB analysis was calculated to gain insight in the potential of mammography radiomics and q3CB analysis to complement each other. In other words, a high correlation between probabilities of malignancy estimated from mammography radiomics and q3CB would imply little potential for synergy, while a low correlation would imply that both approaches yield unique information. Bland-Altman analysis (35,36) was performed to assess whether the combined mammography radiomics plus q3CB analysis estimated overall higher probabilities of malignancy for invasive cancers and lower probabilities for the benign lesions than mammography radiomics.

Appendix E4

Results: The q3CB, Mammography Radiomics, and the Combined q3CB Plus Mammography Radiomics Models and Additive Benefit

The coefficients for the linear models for q3CB, mammography radiomics, and combined q3CB plus mammography radiomics, proved to be quite stable (Fig E3). Interestingly, the weaker the predictive power of an individual feature, the wider the distribution for the corresponding model coefficient. The median water thickness within a mass and that within the surrounding parenchyma ($f_{q,1}$ and $f_{q,2}$) are slightly predictive of malignancy on their own with an area under the ROC curve (AUC) just better than random guessing (AUC = 0.5) and display a much larger spread in the values for their model coefficient than for the ratio of median water thicknesses ($f_{q,3}$), which is the strongest single q3CB predictor of malignancy with an AUC of 0.69 (standard error 0.05). Similarly, for mammography radiomics, the distribution for the 'weaker' features diameter and average gray value ($f_{m,1}$ and $f_{m,2}$, respectively, both with AUC = 0.62 (0.05)) demonstrate a wider distribution in model coefficients than the 'stronger' features, with the full-width at half maximum of the region of interest radial gradient histogram, $f_{m,4}$, being the strongest mammography radiomics predictor of malignancy with an AUC of 0.74 (0.04).

Correlation between the probabilities of malignancy estimated by mammography radiomics and q3CB was fair (37), with a correlation coefficient of 0.38, but statistically significant ($P < .001$). Bland-Altman analysis demonstrated additive benefit of q3CB and mammography radiomics, in that the combined mammography and q3CB approach yielded estimates for the probability of malignancy that were overall higher for the malignant lesions (by 0.03) and overall lower for the benign lesions (by-0.07) than estimated through mammography radiomics alone (Fig E4).

References

33. Malkov S, Wang J, Kerlikowske K, Cummings SR, Shepherd JA. Single x-ray absorptiometry method for the quantitative mammographic measure of fibroglandular tissue volume. Med Phys 2009;36(12):5525–5536 .

34. Kupinski MA, Giger ML. Automated seeded lesion segmentation on digital mammograms. IEEE Trans Med Imaging 1998;17(4):510–517 .

35. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307–310 .

36. Bland JM, Altman DG. Difference versus mean plots. Ann Clin Biochem 1997;34(Pt 5):570–571 .

37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159–174 .

Table E1: q3CB features extracted from the lesion (L) as segmented automatically (5), its surrounding parenchyma (P, a 2 mm thick band around the segmented lesion), the difference in values for lesion and parenchyma (L-P), and the ratio of values for lesion and parenchyma (L/P) where √ indicates a feature was used in the q3CB signature, and – indicates a feature was calculated but not used in this work

q3CB features derived from 3CB thickness maps:	Water					Lipid				Protein		
	L	P	L-P	L/P	L	P	L-P	L/P	L	P	L-P	L/P
Mean (cm)	—	—	—	—	—	—	—	—	—	—	—	—
Median (cm)	√	√	—	√	—	—	—	—	—	—	—	—
Standard deviation (cm)	—	—	—	—	—	—	—	—	—	—	—	—
Skewness	—	—	—	—	√	—	—	—	—	—	—	—

Table E2: Mammography Radiomics Features as Extracted Based on Computer-Segmented (26) Lesions

Mammography radiomics features (27)	
$f_{m,1}$	lesion maximum diameter
$f_{m,2}$	average gray value within lesion
$f_{m,3}$	contrast
$f_{m,4}$	full-width-at-half-maximum of the histogram of radial gradients of the lesion margin
$f_{m,5}$	full-width-at-half-maximum of the histogram of radial gradients within region of interest encompassing the lesion