Automated Percentage of Breast Density Measurements for Full-field Digital Mammography Applications

Erin E. E. Fowler, BS, Celine M. Vachon, PhD, Christopher G. Scott, MS, Thomas A. Sellers, PhD, John J. Heine, PhD

Rationale and Objectives: Increased mammographic breast density is a significant risk factor for breast cancer. A reproducible, accurate, and automated breast density measurement is required for full-field digital mammography (FFDM) to support clinical applications. We evaluated a novel automated percentage of breast density measure (PD_a) and made comparisons with the standard operator-assisted measure (PD) using FFDM data.

Methods: We used a nested breast cancer case–control study matched on age, year of mammogram and diagnosis with images acquired from a specific direct x-ray conversion FFDM technology. PD_a was applied to the raw and clinical display (or processed) representation images. We evaluated the transformation (pixel mapping) of the raw image, giving a third representation (raw-transformed), to improve the PD_a performance using differential evolution optimization. We applied PD to the raw and clinical display images as a standard for measurement comparison. Conditional logistic regression was used to estimate the odd ratios (ORs) for breast cancer with 95% confidence intervals (CI) for all measurements; analyses were adjusted for body mass index. PD_a operates by evaluating signal-dependent noise (SDN), captured as local signal variation. Therefore, we characterized the SDN relationship to understand the PD_a performance as a function of data representation and investigated a variation analysis of the transformation.

Results: The associations of the quartiles of operator-assisted PD with breast cancer were similar for the raw (OR: 1.00 [ref.]; 1.59 [95% CI, 0.93–2.70]; 1.70 [95% CI, 0.95–3.04]; 2.04 [95% CI, 1.13–3.67]) and clinical display (OR: 1.00 [ref.]; 1.31 [95% CI, 0.79–2.18]; 1.14 [95% CI, 0.65–1.98]; 1.95 [95% CI, 1.09–3.47]) images. PD_a could not be assessed on the raw images without preprocessing. However, PD_a had similar associations with breast cancer when assessed on 1) raw-transformed (OR: 1.00 [ref.]; 1.27 [95% CI, 0.74–2.19]; 1.86 [95% CI, 1.05–3.28]; 3.00 [95% CI, 1.67–5.38]) and 2) clinical display (OR: 1.00 [ref.]; 1.79 [95% CI, 1.04–3.11]; 1.61 [95% CI, 0.90–2.88]; 2.94 [95% CI, 1.66–5.19]) images. The SDN analysis showed that a nonlinear relationship between the mammographic signal and its variation (ie, the biomarker for the breast density) is required for PD_a. Although variability in the transform influenced the respective PD_a distribution, it did not affect the measurement's association with breast cancer.

Conclusions: PD_a assessed on either raw-transformed or clinical display images is a valid automated breast density measurement for a specific FFDM technology and compares well against PD. Further work is required for measurement generalization.

Key Words: Breast density; automated measure; breast cancer risk; full field digital mammography; differential evolution.

Mammographic breast density is a significant breast cancer risk factor (1–3). Many of the breast density research studies to date have been based on an operator-assisted measure (PD) to estimate the percentage of breast density within a mammogram. There are various methods under development to automate the estimation of breast density (4–21). Developing a fully automated and standardized breast density measurement has proven somewhat difficult, but at least two commercial standardized measures are available for raw full-field digital mammography (FFDM) images: Volpara and Quantra (19,21–23). However, these have not been shown to be associated with breast cancer risk to date.

Although there are various FFDM manufacturers, the two predominant FFDM technologies used today consist of direct and indirect x-ray conversion systems (24–26) that produce images with different characteristics. The data representation produced by FFDM systems may vary because of the x-ray detection technology, x-ray generation, or postacquisition processing. FFDM systems produce both raw and clinical display (ie, processed) representation mammograms. A given clinical display, or processed image, is derived from its respective raw image with methods developed by the unit's

Acad Radiol 2014; 21:958–970

From the Department of Cancer Epidemiology, Division of Population Sciences, H. Lee Moffitt Cancer Center, Tampa, Florida (E.E.E.F., T.A.S.); Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, Minnesota (C.M.V., C.G.S.); and Department of Cancer Imaging and Metabolism, H. Lee Moffitt Cancer Center, 12901 Magnolia Drive, Tampa, FL 33612 (J.J.H.). Received December 20, 2013; accepted April 24, 2014. Conflicts of Interest: The authors have a pending patent application for this work. **Address correspondence to:** J.J.H. e-mail: john.heine@moffitt.org

http://dx.doi.org/10.1016/j.acra.2014.04.006

manufacturer. The raw images are normally not considered in the clinical evaluation. When applying automated methods, it is not clear if both representations result in similar breast density measurements, if there is a preferred representation, or what impact the technology plays.

Because of the long-standing merit of PD, we are developing an automated measure for FFDM applications referred to as PD_a that provides the same metric as PD. Our automated measurement evolved from earlier work in modeling the Fourier power spectra of digitized-film mammograms. In our prior work, we estimated the spectral form of a given mammogram and removed it with a deconvolution process, resulting in a *noise field* (ie, the filtered image). The degree of local variation in the filtered image (ie, noise) corresponded to the degree of mammographic density (ie, the signal) in the raw image at the same location (27), which is indicative of a signal-dependent noise (SDN) relationship. We developed a statistical method for detecting these areas of increased variation in the filtered image forming the basis of the PD_a technique (28). In subsequent work, the deconvolution process was replaced (approximated) with a high-pass wavelet filter, increasing the algorithm speed, and PD_a was validated using digitized-film mammograms with breast cancer status as the end point (29). As of yet, PD_a has not been evaluated in depth with FFDM images.

In this report, we are generalizing the PD_a algorithm for FFDM applications and developing metrics to evaluate the algorithm's performance relative to the data representation. Because our study focused on developing an automated density measure, we controlled for factors known to be related to breast density and breast cancer. We applied PD_a to a nested breast cancer case–control data set for patients with images (raw and processed image representations) acquired from a specific direct x-ray conversion FFDM technology. We applied an empirically determined data transform to the raw images as a preprocessing step to improve the PD_a raw image processing (ie, to improve the agreement with PD). This transform produces a third data format, defined as the raw-transformed representation. We used an evolutionary optimization strategy to determine the parameters of this transform. We applied PD_a to the three FFDM image representations and compared the respective associations with breast cancer. We compared these associations to those provided by PD (from the Cumulus program described in the following), considered as the standard for comparison. We also characterized the SDN as a function of the data representation using methods developed previously (30) to understand the impact of each representation on the automated PD_a processing.

METHODS

Study Population and Mammography

The patients for this study were derived from the Mayo Mammography Health Study (MMHS) cohort, Rochester, MN and described previously (31,32). Briefly, the MMHS is a prospective cohort study of women living in Minnesota, Wisconsin, or Iowa, aged >35 years, who had a film screening mammography at the Mayo Clinic between 2003 and 2006, and no personal history of breast cancer at study entry. Participants completed a questionnaire and provided written informed consent to use their mammograms, medical records, and blood samples and to link their data to state cancer registries. The 19,924 subjects who participated (51% of the 38,883 subjects who were eligible) were monitored for incident cancer events through the tristate cancer and Mayo Clinic tumor registries. Through December 31, 2010, a total of 492 incident and histologically confirmed primary breast cancers were identified. The analysis was restricted only to cases who had an FFDM examination at least 6 months before diagnosis, limiting our analysis to 228 breast cancer cases and 456 age- and interval-matched controls (two per case), which formed our nested case–control study. All patient mammograms were acquired from Hologic Selenia FFDM units. This FFDM unit has 70-μm spatial resolution (pixel pitch) and a 24 cm × 29 cm field of view (FOV). Screening mammograms are most often acquired with two image sizes depending on the compression paddle choice, inducing an FOV change: 2560 × 3328 pixels (18 cm × 24 cm) and 3328 × 4098 pixels (24 cm × 29 cm). The raw and processed representation images (ie, clinical display images) have 14 bit and 12 bit per pixel dynamic range, respectively. For cases, we used the noncancerous breast, and for controls, the same side used for the matched case was analyzed. We used the craniocaudal (CC) views as the study images. This study was approved by the Mayo Clinic Institutional Review Board.

Statistical Analysis

Patient characteristics and breast measures were summarized with the distribution mean and standard deviation (SD), and differences between case and control groups were tested using conditional logistic regression. Quartiles and the SD of each breast density measure were defined based on the distribution of that density measure among the control subjects. Conditional logistic regression (33,34) was used in the primary analysis to examine the association between quartiles or SD of PD with breast cancer status. As the primary metric, the magnitudes of the associations were summarized by odds ratios (ORs) with 95% confidence intervals (CIs). Models were adjusted for body mass index (BMI) measured in kilogram per square meter. Missing BMI values for cases and controls were imputed using the mean BMI of the respective distribution. Additionally as a secondary means to summarize the strength of association, the area under the receiver operating characteristic curve (Az) was computed as a summary of the ability of each model to discriminate between cases and controls. To match the study design, Az was calculated only within matched case–control pairs. A 95% CI was calculated for each Az based on 1000 bootstrap samples, and these samples were also used to compare Az.