Collection Names and Descriptions:
Collection 1 Name: An Automated System for Breast Cancer Biomarker Analysis
Instrument: General Electric (GE) Senographe 2000D full field digital mammography (FFDM), two-dimensional (2D) units
Data Description: 180 case-control pairs; 2D mammograms in both *for presentation* and *for processing* image data representations; demographical-clinical; and pathological data from women, 18 years and older.
Grant Number: R01CA114491 breast density and breast cancer risk study (2006-2011).

Collection 2 Name: Automated Quantitative Measures of Breast Density
Instrument: Hologic Selenia 2D FFDM units and Dimensions digital breast tomosynthesis (DBT) units operating in the 2D mode
Data Description**:** 319 case-control pairs; 2D mammographic images in *for presentation* and *for processing* image data representations; demographical-clinical; and pathological data, from women 18 years and older.
Grant Number: R01CA166269 breast density and breast cancer study (2013-2017).

Collection 3 Name: Moffitt Imaging Biomarker Validation Center
Instrument: Hologic Dimensions DBT mammography units
Description: 348 case-control pairs; mammographic images: (2D) *for presentation* and *for processing* image data representations; DBT volumetric (2D slices) and C-View synthetic 2D mammograms; demographical-clinical; and pathological data from women 18 years and older.
Grant Number: U01CA200464 breast density and breast cancer risk study (2016-2022).

Results and Conclusions Summary: select findings from studies using these collections are briefly described, as the investigations spanned roughly 19 years. Concise findings and detailed descriptions of the algorithms and methods can be found from the respective citations.

From Collection 1, a calibration system was established to adjust *for processing* 2D mammograms acquired with FFDM to a common normalized effective x-ray attenuation coefficient scale [1]. This system included a serial quality assurance (QA) monitoring system based on the cumulative sum (CUSUM) technique. This QA approach was used to check the serial accuracy of baseline calibration data so it could be updated if calibration accuracy moved beyond a given tolerance [2]. This technique was able to detect signs of impending x-ray tube failure long before the actual failure. Calibration also required an accurate estimate of the compressed breast thickness during the acquisition. Compression paddle force on the breast caused deviations in the compressed breast thickness spatial distribution; a method was developed to compensate for the compression paddle contortion [3]. Using calibrated data, the mean, variation, BI-RADS ordinal breast composition measures, and a PD-type measure produced significant odds ratios (ORs) [4-6]. The variation from both *for presentation* and *for processing* mammograms produced significant ORs, and that BI-RADS descriptors could be developed without calibration.

From Collection 2, the same calibration system was extended from the Collection 1 studies [7], and the QA monitoring approach was advanced with 2D mammograms acquired with a different FFDM technology. Here we showed how to monitor a given acquisition's baseline calibration data with a 50/50 percent adipose/glandular phantom and correct a given curve when it drifted out of a present calibration tolerance with the same 50/50 phantom measurement [8]; this system also required a compressed breast thickness correction technique to compensate for the paddle deflection due to the compression force [9]. The calibrated variation measure produced significant ORs [10]. An automated PD measure developed earlier [11, 12], modified to operate on *for presentation* and *for processing* images (i.e., non-calibrated data), also produced significant ORs [13] when applied to this collection, unrelated to the calibration studies.

From collections 1 and 2, Fourier ring measurements (texture analyzed in the Fourier domain) produced significant findings across the power spectrum with calibrated, *for processing,* and *for presentation* images [14]; illustrations were provided showing how these measures translate to image texture and spatial scales. In particular,

a low frequency measurement produced significant ORs that was in agreement with earlier work with digitized film mammograms [15] and subsequently validated with two large disparate populations with FFDM images [16]. Summarized local spatial correlation measures produced significant ORs also related to specific spatial scales [17].

From Collection 3, several PD type measures produced significant ORs [18]: PD as a volumetric quantity, PD averaged over the DBT slices (area quantity), and PD determined from 2D synthetic images. ORs were about the same across these measures. Theoretical derivations predicted the volumetric and average slice measures are the same, agreeing with the respective OR findings. PD was modeled as a function compressed breast thickness (CBT). Maximum PD location was approximately 0.41×CBT and similar across case-controls. PD determined from the slice where it was a maximum value produced significant ORs very similar to the other PD methods. Both the average pixel values from the DBT volume and from the 2D synthetic images produced significant ORs. Unlike the analysis of 2D FFDM images, variation measured in multiple ways did not produce significant findings.

Method Details

Approach: all collections were developed to make *breast density* measurements with a matched (1:1) case-control design. Mammograms in the craniocaudal view were used in all studies. Breast density is a generalized term used in our context to imply an arbitrary measurement from a mammogram. For the most part, citations were reserved for the results section unless prior work was required for replicating the algorithms or to put investigations in context. Image definitions and descriptions are provided in the Data collection protocol Section. Collection 1 used two-dimensional (2D) mammograms acquired with full field digital mammography (FFDM). This collection's main study purpose was to standardize (*for processing*) mammograms using a breast tissue equivalent calibration phantom imaging approach to account for acquisition technique differences (i.e., target-filter combination, mAs, and compressed breast thickness). This approach was founded on making effective x-ray coefficient measurements from breast tissue equivalent phantoms [19] and developing a serial quality control method using the cumulative sum approach [20]. Images were mapped to a standardized scale (0 – 100) representing total adipose to total glandular tissue respectively, prior to making image measurements (i.e., mean, and standard deviation). The mapping required developing baseline calibration curves at time-zero (about 160 phantom acquisitions). Additionally, differential evolution optimization [21] was used to develop a four-state ordinal measurement system that paralleled the BI-RADS [22] breast composition descriptors using both calibrated and non-calibrated images. The implementation of this DE approach is also described in detail in this work [23]. For risk modeling in all related studies, conditional logistic regression was used to estimate odds ratios (ORs) for given image measurement in a standard deviation increment with 95% confidence intervals (CIs). ORs were the primary risk metric for the matched design with our intent to isolate a given image measurement. For most modeling outputs, ROC curve areas were provided as well with CIs. Both body mass index and ethnicity were controlled for in the modeling with a given image measurement. In all studies discussed here, breast density measurements were made from the non-affected breast for a case and same-side breast for the matched control. Calibrated measures were compared against breast density measured with the user-assisted Cumulus percentage of breast density (PD) method [24-26]. For a given image, this is binary labeling technique based on thresholding (dense / non-dense area), where the dense area normalized by the total breast area multiplied by 100% produces the PD measure. This PD measure is often used as the standard for comparison. Automated measures were based on eroding the breast area first by 25% unless noted. Here, we assumed the breast was an approximate half semi-circle, and the breast area was eroded radially inward by 25%. Erosion was used to approximate the breast area that was in contact with the compression paddle surface during the acquisition. Collection 2 was designed with the same intent as Collection 1 to evaluate if the same calibration approach was applicable to a different type of FFDM technology. In both collections, various measures were also made from both *for processing* and *for presentation* images without calibration within the eroded area. Measures included the mean, standard deviation, Fourier based on dividing the power spectrum into concentric rings and summing the power within a given ring

[27] (texture analysis in the Fourier domain), local correlation, and automated PD (without erosion). The same image processing techniques and risk modeling used for Collection 1 were applied to Collection 2. When making Fourier measures, the largest rectangle that would fit within the breast area was used as the measurement-region. This *largest-rectangle* algorithm is described in detail in this work [14]. Collection 3 was designed to make breast density measurements from DBT data and the related 2D synthetic images. Here, an automated PD type measurement was applied to the volume slices producing (1) a volumetric PD measure, and (2) an average PD measurement taken over the DBT slices. The same PD approach was applied to the synthetic 2D images. The same modeling design was used as with the other collections, where a given breast density measurement was modeled with conditional logistic regression; analyzing DBT data did not require breast area erosion.

Data collection protocol: all datasets were developed with the same IRB-approved protocol. Cases (unilateral breast disease) were either: (1) women attending the breast clinics at Moffitt Cancer Center (MCC) diagnosed with breast cancer (type-1) or (2) attendees of surrounding area clinics sent to MCC for breast cancer treatment or diagnostic purposes and found to have breast cancer (type-2). Cases have pathology verified unilateral (first time) breast cancer. Controls were attendees of MCC with no history of breast cancer. Controls were individually matched to cases on age (± 2 years), hormone replacement therapy (HRT) usage and current duration, screening history, and mammography unit. The HRT match was based on status of current users or non-users. Nonusers included women that have not taken HRT for at least two years. If a case was a current HRT user, the control was matched on this duration (± 2 years). Controls were matched by screening history using a three-category classification. Group 1 included women with prior screening history by any means; the duration between the last screening and the study image date must be no more than 30 months. Group 2 included women with a screening history that does not fit within in Group 1 or Group 3. Group 3 included women with no screening history. We used mammograms in craniocaudal (CC) views as study images. The unaffected breast was used as the study image for cases (image acquired before treatment) and the matching lateral breast for controls. Women that had breast implants were excluded from this study. Cases were selected retrospectively (type-1) via electronic medical records search or recruited (type-2). Controls were selected retrospectively via electronic medical records search. Multiple suitable controls were matched to a given case and one control was selected randomly for the study.

The collections represent mammograms from different technology designs and manufactures. For conventional 2D images acquired with full field digital mammography (FFDM), there are two sets of related images available at the acquisition time termed *for processing* and *for presentation*; *for processing* mammograms can be considered as raw data and are not used clinically. Manufacturer specific algorithms are applied to these images to produce enhanced *for presentation* images that are used clinically. Both types of images can be used for experimental measurement investigations. FFDM images from Collection 1 have 100μm pitch and mammograms (FFDM) from collection 2 have 70μm pitch; these images were acquired with two different spatial sizes depending on breast size (i.e., different size compression paddles). Digital breast tomosynthesis (DBT) data includes *volumetric* images that are 2D slices of the breast about 1mm thick (or about 10 slices per cm of compressed breast thickness) and 2D synthetic images referred to as C-View images for the technology associated with Collection 3. We will refer to these as 2D synthetic images. Pixel spacing in DBT acquired images is about 100μm but varies across women from roughly 80μm to 110μm in these datasets (but is the same for a given woman's DBT dataset). DBT units can take both 2D FFDM and DBT acquisitions in tandem without patient repositioning; due the way the images from Collection 2 and 3 were acquired, a larger DBT dataset can be constructed by combining elements from both collections (see citation in Collection 3 results); likewise, a larger 2D FFDM dataset can be constructed from these collections is a similar fashion. There is an accompanying data dictionary with these collections. The collections do not contain image annotations; only the cancerous breast is known. Of importance for automated processing, an intricate image file naming convention was developed to inform the user of study and image characteristics such as study number, case-control status, mammographic view, study image eligibility, and image type. The file name string convention can be searched automatically to find a given image type relatively easily and to ensure cases and matching controls can be assembled.

Limitations: sampling of cases and controls was not population-based, but rather a mixture of cases ascertained at an NCI-designated comprehensive cancer center inclusive of referrals from the community. There is no evidence from our studies that the cases are not representative, but findings should be replicated in population-based studies. Data fields allow for selection of the population-based cases (discounting the referrals) but will reduce the case-control numbers. Image data from the General Electric Senographe 2000D FFDM units do not include women with large breasts due to the x-ray detector design limitations. Images may contain artifacts (documented in the data fields) such as nipple markers, mole markers, biopsy clips, and scar markers. These artifacts are documented in the data fields. All images were visually inspected and approved for automated processing. Here, a judgement was made to exclude a sample with too many markers or to include the sample because artifacts were deemed negligent.

References

[1] J. J. Heine and J. A. Thomas, "Effective x-ray attenuation coefficient measurements from two full field digital mammography systems for data calibration applications," (in Eng), *Biomedical engineering online,* vol. 7, no. 1, p. 13, Mar 28 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18373863

[2] J. J. Heine, K. Cao, and C. Beam, "Cumulative sum quality control for calibrated breast density measurements," *Med Phys,* vol. 36, no. 12, pp. 5380-90, Dec 2009, doi: 10.1118/1.3250842.

[3] J. J. Heine, K. Cao, and J. A. Thomas, "Effective radiation attenuation calibration for breast density: compression thickness influences and correction " *BioMedical Engineering OnLine* vol. 9, p. 73, 2010.

[4] E. E. Fowler, T. A. Sellers, B. Lu, and J. J. Heine, "Breast Imaging Reporting and Data System (BI-RADS) breast composition descriptors: automated measurement development for full field digital mammography," *Med Phys,* vol. 40, no. 11, p. 113502, Nov 2013, doi: 10.1118/1.4824319.

[5] J. J. Heine, K. Cao, D. E. Rollison, G. Tiffenberg, and J. A. Thomas, "A Quantitative Description of the Percentage of Breast Density Measurement Using Full-field Digital Mammography," (in eng), *Acad Radiol,* vol. 18, no. 5, pp. 556-64, May 2011, doi: 10.1016/j.acra.2010.12.015.

[6] J. J. Heine, E. E. E. Fowler, and C. I. Flowers, "Full field digital mammography and breast density: comparison of calibrated and noncalibrated measurements," (in English), *Acad Radiol,* vol. 18, no. 11, pp. 1430-6, 2011, doi: 10.1186/1475-925X-12-114.

[7] E. E. Fowler, B. Lu, and J. J. Heine, "A comparison of calibration data from full field digital mammography units for breast density measurements," *Biomedical engineering online,* vol. 12, p. 114, 2013, doi: 10.1186/1475-925X-12-114.

[8] B. Lu, A. M. Smallwood, T. A. Sellers, J. S. Drukteinis, J. J. Heine, and E. E. Fowler, "Calibrated breast density methods for full field digital mammography: a system for serial quality control and inter-system generalization," *Med Phys,* vol. 42, no. 2, pp. 623-36, Feb 2015, doi: 10.1118/1.4903299.

[9] E. E. E. Fowler, A. M. Smallwood, N. Z. Khan, K. Kilpatrick, T. A. Sellers, and J. Heine, "Technical challenges in generalizing calibration techniques for breast density measurements," *Med Phys,* vol. 46, no. 2, pp. 679-688, Feb 2019, doi: 10.1002/mp.13325.

[10] E. E. Fowler *et al.*, "Calibrated Breast Density Measurements," *Acad Radiol,* Dec 10 2018, doi: 10.1016/j.acra.2018.10.009.

[11] J. J. Heine *et al.*, "An automated approach for estimation of breast density," (in eng), *Cancer Epidemiol Biomarkers Prev,* vol. 17, no. 11, pp. 3090-7, Nov 2008. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18990749

[12] J. J. Heine and R. P. Velthuizen, "A statistical methodology for mammographic density detection," *Med Phys,* vol. 27, no. 12, pp. 2644-51, Dec 2000. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11190946.

[13] E. E. Fowler, C. M. Vachon, C. G. Scott, T. A. Sellers, and J. J. Heine, "Automated Percentage of Breast Density Measurements for Full-field Digital Mammography Applications," *Acad Radiol,* vol. 21, no. 8, pp. 958-70, Aug 2014, doi: 10.1016/j.acra.2014.04.006.

[14] E. E. E. Fowler, A. Smallwood, C. Miltich, J. Drukteinis, T. A. Sellers, and J. Heine, "Generalized breast density metrics," (in English), *Physics in Medicine and Biology,* vol. 64, no. 1, Jan 2019, doi: 10.1088/1361-6560/aaf307.

[15] A. Manduca *et al.*, "Texture features from mammographic images and risk of breast cancer," *Cancer Epidemiol Biomarkers Prev,* vol. 18, no. 3, pp. 837-45, Mar 2009, doi: 10.1158/1055-9965.EPI-08-0631.

[16] J. Heine *et al.*, "Mammographic Variation Measures, Breast Density, and Breast Cancer Risk," *AJR. American journal of roentgenology,* vol. 217, no. 2, pp. 326-335, Aug 2021, doi: 10.2214/AJR.20.22794.

[17] E. E. E. Fowler, C. Hathaway, F. Tillman, R. Weinfurtner, T. A. Sellers, and J. Heine, "Spatial correlation and breast cancer risk," *Biomedical Physics & Engineering Express,* vol. 5, no. 4, p. 045007, 2019/05/22 2019, doi: 10.1088/2057-1976/ab1dad.

[18] J. Heine, E. E. E. Fowler, R. J. Weinfurtner, E. Hume, and S. S. Tworoger, "Breast density analysis of digital breast tomosynthesis," *Sci Rep,* vol. 13, no. 1, p. 18760, Oct 31 2023, doi: 10.1038/s41598-023-45402-x.

[19] J. J. Heine and M. Behera, "Effective x-ray attenuation measurements with full field digital mammography," *Med Phys,* vol. 33, no. 11, pp. 4350-66, Nov 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17153414.

[20] D. M. Hawkins and D. H. Olwell, *Cumulative Sum Charts and Charting for Quality Improvement* (Statistics for Engineering and Physical Science). New York: Springer-verlag 1997.

[21] K. V. Price, R. M. Storn, and J. A. Lampinen, *Differential evolution : a practical approach to global optimization* (Natural computing series). Berlin ; New York: Springer, 2005, pp. xix, 538 p.

[22] C. D'Orsi, C. J, and e. al., *ACR BI-RADS Atlas: Breast Imaging Reporting and Data System; Mammography, Ultrasound, Magnetic Resonance Imaging, Follow-up and Outcome Monitoring, Data Dictionary*. ACR, American College of Radiology, 2013.

[23] E. E. Fowler, A. Berglund, T. A. Sellers, S. Eschrich, and J. Heine, "Empirically-Derived Synthetic Populations to mitigate Small Sample Sizes " *Journal of Biomedical Informatics,* vol. 105, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046420300368.

[24] J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe, "Automated analysis of mammographic densities," *Phys Med Biol,* vol. 41, no. 5, pp. 909-23, May 1996, doi: 10.1088/0031-9155/41/5/007.

[25] J. W. Byng, N. F. Boyd, E. Fishell, R. A. Jong, and M. J. Yaffe, "The quantitative analysis of mammographic densities," *Phys Med Biol,* vol. 39, no. 10, pp. 1629-38, Oct 1994, doi: 10.1088/0031-9155/39/10/008.

[26] N. F. Boyd *et al.*, "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study," *J Natl Cancer Inst,* vol. 87, no. 9, pp. 670-5, May 3 1995, doi: 10.1093/jnci/87.9.670.

[27] J. J. Heine and R. P. Velthuizen, "Spectral analysis of full field digital mammography data," *Med Phys,* vol. 29, no. 5, pp. 647-61, May 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12033559.