Advancements in an Automated Breast Density Detection Technique for Breast Cancer Risk Prediction: a Synthetic Signal-dependent Noise Construct

John Heine*, Erin Fowler, Matthew Schabath, and Kathleen M. Egan

Moffitt Cancer Center & Research Center
Cancer Epidemiology Department
12902 Magnolia Drive
Tampa, Florida 33612

*Corresponding author: john.heine@moffitt.org

Keywords: percentage of breast density, signal-dependent noise, noise shaping, breast cancer risk, ensemble averaging, full field digital mammography, and digital breast tomosynthesis

Abstract ~ Breast density is a significant breast cancer risk factor estimated from mammograms and useful for breast cancer risk prediction. We describe enhancements to a previously developed automated percentage of breast density detection method. The approach relies on signal-dependent noise (SDN), where the expected variance and mean (i.e., the signal) are functionally related. SDN variations in different image data representations caused degradation in the algorithm's performance; this work addressed this problem. Image data used in the analysis were from three breast cancer case-control studies employing different mammographic technologies i.e., full field digital mammography (including raw and clinical images) and digital breast tomosynthesis. Advancements demonstrated that the algorithm produced significant odds ratios across all image data representations by adding these capabilities: (1) the ability to transform noise to a specific optimized quadric SDN form given an arbitrary image data representation; and (2) the ability to perform ensemble averaging over a given image, boosting the signal. We also demonstrated methods to combine measurements from different technologies using a probability density transformation approach. The technique can be applied to images from different technologies with minimal adjustments thereby making it suitable for both research and clinical applications.

Background

Breast density is observable in mammograms with or without automation and is a widely accepted significant breast cancer risk factor [1, 2]. The aim of the present work is to refine procedures for the automation of a breast density measurement from varied digital mammograms. Inclusion of this image measure in models that predict absolute risk over time can be useful for personalized breast screening strategies [3-5] and may facilitate epidemiologic and clinical research on this important breast cancer risk factor. Our goal is to refine a measurement approach for an automated breast density detection technique described previously [6-8]. Modifications to this approach were derived from a model used in image-sensor research, described below.

In mammograms used for clinical purposes, breast density is loosely defined as the *bright* tissue in the image. The breast is normally treated as a *two-component* system comprised of adipose and the remaining bright radiographically more opaque regions of dense breast tissue. For the x-ray energies used in mammography, this two-component system is a good approximation. Nevertheless, dense tissue is a combination of both epithelial (ductal and lobular) and stromal (supporting or connective) tissues [9, 10], which are likely inseparable in x-ray mammograms due to their similar x-ray attenuation coefficient values.

The user-assisted measure of breast density, referred to as Cumulus (University of Toronto, Toronto Canada) [11-13], has been considered the benchmark for comparison with new measures [14]. This technique computes the percentage of dense tissue within the breast area. In Cumulus, the operator performs the following tasks: (1) segments the breast area from the background; and

(2) sets a threshold, labeling everything above as dense and below as adipose tissue. This results in a binary labeled image within the breast area and yields an estimate of the proportion of the tissue equivalent to the percentage of breast density (PBD). In this user-assisted measurement, both segmentation and labeling thresholds are subject to individual user definition, and their assignment can be prone to random intra-user variation. In contrast, automation of breast density measurements should remove subjectivity, and provide a more reproducible measurement, as well as more timely and less labor-intensive ascertainment of breast density in clinical and research applications.

The history and merits of developing quantitative breast density measurements are extensive [15-17]. Image-analytics have evolved to include automated features that may not be precisely observable in mammograms. For example, these measures could be generated with filtering or other techniques that summarize spatial relationships [18, 19] as surrogates for breast density or possibly standalone independent predictors of risk [20]. Such measures are often collectively referred to as texture features. More recently, texture-type risk prediction features have been generated with more advanced machine learning and artificial intelligence (AI) techniques [4, 21-25]. In some AI applications, texture is captured with convolutional neural networks (CNNs). Texture extracted by CNNs can be considered as resulting from a learned composite filter derived through multiple processing steps [26, 27]. Although these applications show much potential and have improved risk prediction capabilities, the image features are problematic because their description is not readily available, hindering clinical adoption [28]. As a remedy, ongoing research shows promise in interpreting features captured by AI architectures that use CNNs [27].

One obstacle in developing a universally applicable automated method for breast density estimation when incorporating both current and historical data is the changing landscape of digital mammography over time beginning with digitizing screen-film images, advancing to full field digital mammography (FFDM), and then to the current digital breast tomosynthesis (DBT). Each technology is associated with unique image data representations and accompanied by an array of distinguishing technical features (e.g. different film-digitizers, enhancements for clinical purposes, 2D images from different FFDM detector designs, adding a third dimension, or varied synthetic 2D images) that can differ both within and across manufacturers. When building a universally applicable model that includes breast density, these potential variants must be accommodated. A prime research example of the evolution in technology and ensuing challenges comes from the Nurses' Health Study (NHS). The NHS originated in the mid-1970s [29, 30] and spans the entire era of mammographic screening and adoption of new mammography advancements over 40-plus years [31, 32]. In research based on the NHS, inclusion of mammography data from the early years of the cohort required the physical collection, and digitization of film images for thousands of mammograms prior to applying image analytics [33].

The Breast Imaging and Reporting Data System (BI-RADS) ordinal breast density assessment method (i.e., the breast composition descriptors described in the American College of Radiology Atlas 5th Edition, 2013) [34] has also been used for risk prediction (also described in earlier atlas editions). This mammography reporting technique trades measurement resolution (i.e., a continuous scale), and in theory enhanced risk prediction potential, for contrast (i.e., a four-state ordinal scale) and ease of use in the clinical workflow. Although initially not developed for risk prediction purposes, it is commonly used for this application [35-37]. As an alternative, a viable

automated measurement approach should be equally useful in research and clinical setting when compared to the BI-RADS measure. Transition to the clinic from research applications should require minimal adjustments and accommodate a range of image formats in current and near-future use.

The algorithm advancements investigated in this report are intended to produce a more stable measure and assist in producing an automated PBD measure with the potential for universal applications. This measurement approach operates by analyzing signal-dependent noise. In past analyses, this method's performance varied depending on the image data representation (i.e., image format). In radiographic imaging systems, such as mammography, quantum noise (i.e., Poisson noise) is one source of variation [38, 39] that is signal-dependent. This type of noise produces a grainy appearance especially noticeable in underexposed image areas. At small spatial scales, noise variation is approximately signal-dependent (i.e., quantum noise) in FFDM units [6, 40]. We have noted that the noise variance form affected the method's performance [6]. For example, in the ideal situation, a larger noise variance locally is the signature of dense tissue and lower adipose tissue. In practice, this relationship was not in the *optimal* form (shape) in some image formats. In these situations, ad-hoc adjustments were implemented with no attempt to arrive at an optimal algorithm correction [6, 7]. The modified approach described here no longer relies on a given mammogram's intrinsic noise characteristics. Instead, the detection's input is a synthesized image, derived from the respective mammogram, that has a signal-dependent noise form shaped to enhance the algorithm's performance, regardless of a given mammogram's intrinsic, or natural, noise characteristics. This synthesis follows from a signal-dependent noise model that raises the uncorrupted image to a positive power, with restrictions, and modulating it with random noise (i.e.,

random noise multiplication). This modification added two additional free parameters to the algorithm, discussed below. In theory, these modifications should produce a more stable PBD estimation regardless of image format differences.

In the context of the above-described PBD algorithm modifications, we evaluated the technique's risk prediction capabilities over varied image formats. The research uses three breast cancer matched case-control studies, each study accompanied by images from different mammographic technologies in use at this center between 2007-2022. The primary objectives were to: (1) evaluate how to determine the operating characteristics of the detection process (i.e., set the four free detection parameters) over a range of image formats; (2) investigate how to process PBD measures that yield consistent results across studies; and (3) develop a standardized approach to facilitate aggregation of data derived from different mammography technologies, a frequent goal in epidemiologic research.

Methods

Automated breast density detection technique

The automated approach was described previously [6-8] and is provided here for context with current advancements. The method produces an output like Cumulus and is referred to as PD_a (i.e., automated PBD); we use this abbreviation for both the technique and the measure it produces. It operates by analyzing signal-dependent noise (SDN). In images formed by x-ray fluence, the signal captured by the ideal linear detector would be represented by a Poisson process at the pixel level (i.e., neglecting other noise type contributions). In this idealized situation, the expected mean

equals the expected variance in an ensemble sense. Ensemble expectation quantities can be derived at the pixel level when it is possible to average multiple realizations of the same image acquired at the same time; this implies acquiring multiple mammograms of the same participant without repositioning, changing the x-ray tube orientation, or changing the breast compression, which is normally not possible. This ideal Poisson construct approximately holds locally in practice with FFDM images in some formats but not universally. To a good approximation, there is a linear relationship between the noise variance and expected signal mean in both indirect and direct x-ray detection FFDM units when analyzing raw images [6, 40]. As an approximation, a high pass filter was applied to a given mammogram to separate it into two image components: chatter (i.e., SDN) and the lower resolution image structure (the signal). The type of high pass filter has changed since the technique's inception for algorithm processing time reductions, while serving the same purpose. Because there is only one mammogram per participant (i.e., per breast considering its laterality and orientation) for a given time point normally, ensemble expectations were approximated with local spatial averages taken over relatively small image regions. This approximation is valid because mammograms have long range positive spatial correlation (discussed below). The high pass filtered image is scanned with n×n pixel search window (n~2-4 pixels). At the pixel level, using these pixels in an expectation calculation approximates acquiring n^2 consecutive mammograms of a given participant without repositioning and taking the expectation of a given quantity using n^2 realizations. In each window, the local variance is calculated and compared with the global adipose reference variance using a chi-square test (two-sided) with a preset significance level (defining detection parameter 1). The reference variance is first estimated by using all pixels in the high pass filtered image within the breast area (see below for an important qualification) in the calculation. To capture the rarer very dense breast event, the

detection is repeated after the first stage is completed by setting a second significance level (defining the second detection parameter) and updating the adipose reference variance. The updated reference variance is calculated in the high pass filtered image using pixels detected as non-dense in the first detection stage. This second detection stage produces the PBD binary output image. The second detection stage is required especially when an upper-density image is under analysis. In this situation, the first adipose reference estimation will be less than optimal. In this modified approach, the window is *centered* on a given pixel, and the decision is made at approximately at the central pixel location. In either stage, when the local variance deviates significantly, the central pixel is labeled as dense or otherwise it is labeled as adipose. The window is shifted in pixel steps across the filtered mammogram's breast area. When operating on images in a different format, detection parameters were adjusted with training data, where case-control data was preferred but not necessary.

For context, PD_a is defined as detection technique for these reasons. In classical (Neyman–Pearson) detection theory, a true detection event (dense tissue) is based on a fixed probability of a false-alarm (i.e. here, labeling adipose tissue as dense.. In our variant of this approach, a mammogram is transformed into a form suitable for applying a standard statistical test using the ratio of two-variances; this is a practical surrogate for the Neyman-Pearson likelihood ratio test because we do not know the statistical specifics for dense tissue [41]. The detection approach is contingent upon estimating the adipose reference variance, used in the ratio test. Supporting work for the two-stage reference variance determination comes from a study that showed the 50% dense breast by volume is rare [42]. In a similar sense, recent work demonstrated how the volumetric breast density structure from DBT projects onto a plane to form the PBD measure [43]; these two

studies taken together show why the first adipose reference variance calculation is an acceptable estimation for many images that requires fine tuning is some instances.

Over the course of our research with this PBD approach, a simple adjustment was overlooked that eliminates two algorithm *restrictions*: (1) the requirement for the presence of SDN in the appropriate form in each image; and (2) estimating the SDN component with high pass filtering. Because the SDN form was not always conducive to the algorithm's performance, image transformations were explored on a somewhat ad hoc basis [6, 7]. Yet, the optimal SDN form was not understood, and the algorithm could not produce significant findings with some image data representations. In this modification, SDN is shaped to a predictable form. Filtering was eliminated because the SDN is now synthesized rather than approximated. In pursuit of our goals, the methodology was never evaluated over the full range of detection parameters in the context provided in this current study. The main innovation here is the capability of transforming a given mammogram into a form that the appropriate standard statistical test can be applied.

Modified detection input image

The main breakthrough follows from changing the detection input image. An image corrupted by SDN can be modeled as follows

$$i_o(x,y) = n(x,y) \times i(x,y)^{\alpha} + i(x,y), \qquad \text{Eq. [1]}$$

where $i_o(x, y)$ is the observed noise-corrupted image, $i(x,y)$ is the ideal noiseless image, $n(x, y)$ is random noise, $0 < \alpha \leq 1$, and (x, y) are spatial pixel coordinates. Here, additive random noise has been neglected. Variants of this model have been used in a wide field of investigations as reviewed

elsewhere [44]: (1) images formed with charge-coupled device and complementary metal-oxide semiconductor arrays; (2) ultrasound images; and (3) images constructed from synthetic aperture radar. Taking the expected ensemble averages from Eq.[1] gives the mean and variance, respectively: using $E[i_0(x, y)] = i(x, y)$, it follows $E[i_0(x, y)^2] - E[i_0(x, y)]^2 = i(x, y)^{2\alpha}$, where $E[z]$ is the expectation operator applied to the argument z. For n(x, y), we used a standardized normal distribution (zero mean, unit variance). A Poisson process can be modeled by setting $\alpha = ½$, which gives i(x, y) as the expected ensemble mean and variance. When $\alpha = 1$, the mean and noise variance bear a quadratic relationship. The key here is to recognize that the SDN image component is defined in Eq. [1] and could be synthesized if i(x, y) was known.

The original approach relied on applying a high pass filter to a given mammogram (i.e., $i_o(x, y)$ in Eq. [1]) to approximate the SDN image component [i.e., an approximation for $n(x, y) \times i(x, y)^{\alpha}$]. This estimation was valid because the power spectrum of the SDN component is statistically uniform (analogous to white noise), and this specific high pass filter (i.e., a wavelet filter) captures ¾ of the noise variance when applied to white noise (i.e., noise that has a uniform power spectrum). This filter-variance relationship applies to both traditional 2D white noise images and to this SDN image component expressed in Eq. [1] [44]. Because spatial averaging is still used and in furtherance of the approximation, the power spectrum of mammograms follows an inverse power law [8, 45-47]. This characteristic gives rise to long range positive spatial correlation (i.e., mammograms vary slowly spatially), and the high frequency image component is often relatively parasitic.

To generalize the algorithm, the detection input image is synthesized using the SDN component from Eq. [1], in isolation, expressed specifically as

$$s_r(x, y) = n(x, y) \times i(x, y)^{\alpha}, \qquad \text{Eq. [2]}$$

where $\alpha > 0$ (eliminating the restriction from above), although only two values of α (½ and 1) were investigated here. The subscript r was added to indicate random noise image. Signal-dependent noise is synthesized using Eq. [2] with a suitable estimation of i(x, y). The exponent, α, is the third free parameter. The detection is applied to an image derived by squaring Eq. [2] and ensemble averaging expressed as

$$s_m(x, y) = [s_r(x, y)]_m^2, \qquad \text{Eq. [3]}$$

where the integer subscript $m > 0$ specifies the ensemble size used in the averaging. A given image in this ensemble (specific to one patient) is generated with a different realization of n(x, y) for the same i(x, y) in Eq. [2] and then squaring Eq. [2]. Averaging m ensemble realizations gives the image in Eq. [3]. Because $E[s_r(x, y)] = 0$, this produces a scaled chi-square variate with m degrees of freedom at each pixel, collectively expressed in Eq. [3]. Here, m is the fourth degree of freedom. When scanning $s_m(x, y)$ with the search window, the detection is now based on the average of n×n pixels about position (x, y). This produces approximately a chi-square variate with $(n^2-1) \times m$ degrees of freedom at (x, y), where the degree of freedom reduction is because variance is calculated within each window in the detection stages.

In theory, starting from Eq. [2] with Eq. [3] will ensure there is a predictable SDN relationship for a given image format. To approximate i(x, y), a 7×7 boxcar averaging filter was applied to the respective mammogram to destroy high frequency chatter. In this format, the scaled chi-square

estimate becomes almost exact when applying ensemble averaging followed by spatial averaging because the noise signal at each pixel in $s_r(x, y)$ lacks linear spatial correlation [44]. This synthetic construction becomes crucial if the input image is highly processed or where the traditional Poisson or SDN relationship would not be expected to hold. This could be due to the heavy processing used in the reconstruction process for DBT images, images with a compressed density-adipose data scale, or other images used clinically that have been enhanced. We refer to an image formed by averaging (both ensemble and spatial) of the squared noise component as a chi-square image.

Modified signal-dependent noise characteristics

In the modified approach, ensemble expectation quantities can be calculated; the respective mean (μ) and standard deviation (σ) from applying Eq. [3] are discussed here. In Eq. [2], we make the substitution $d = i(x,y)^{\alpha}$ for a fixed pixel location and use this substitution Eq. [3] with $m = k$ (i.e., averaging k identically disturbed chi-square variates with one degree of freedom). This produces a variate distributed as chi-square distribution with k degrees of freedom with $\mu = d^2$ and $\sigma = \sqrt{\frac{2}{k}}\, d^2$. This shows, in principle, the uncertainty (i.e., σ) in the estimated noise variance at a pixel can be driven arbitrarily small by increasing the number of realizations in the averaging. We can estimate the noise variance of a given mammogram as a function of the expected pixel average without ensemble averaging by applying a spatial average approximation [40]. Here, clinical mammograms were used as examples, where the intrinsic SDN relationships were compared with those derived from the respective synthetic chi-square images using a suitable choice for α.

Additional preprocessing for raw mammograms

Prior experiments indicated raw Hologic-FFDM images required a non-linear transformation to facilitate the density detection [6]. Prior to this report, we explored alternative transformations to raw mammograms from different technologies. Raw images from FFDM units used at this center are/were in monochrome 2 format (dense tissue is dark). First, after the boxcar filter application, the intensity scale was reversed by subtracting the (pixel bit-rate -1) from a given image. The resultant image was then divided by its maximum pixel value giving t(x, y); the noiseless image approximation is then given by $i_0(x, y) = \exp [c \times t(x, y)]$. The constant, c, is found by setting the largest value in each $i_0(x, y)$ image equal to 16,000, a choice based on heuristics. The same transformation was used for both Hologic and GE raw FFDM images. The transformation was derived through experimentation in the past and may not be globally optimal.

Algorithm qualifiers

For algorithm replication purposes, the breast area was eroded radially inward [48] (treating the breast area as an approximate half hemisphere) by 2% before estimating reference values and performing the detection. This is to mitigate edge effects in each image. Edge effects can skew both global and local variance calculations. The full breast area was included in the normalization for PD_a.

Spectral attributes of the chi-square images

The $s_m(x, y)$ image expressed in Eq. [3] appears as *mammogram* with sufficient m, but its spectral properties differ from the respective mammogram. Although this artifact will not be investigated here in detail, we show why this assertion holds. To illustrate the spectral differences, we take the Fourier transform (FT) of Eq. [2] without considering its power spectrum. Briefly, using the convolution relationship gives

$$\mathbf{FT}[n(x, y) \times i(x, y)^{\alpha}]^2 = [N(f) ** N(f)] * [G(f) ** G(f)], \qquad \text{Eq. [4]}$$

where **FT** is the 2D-FT operator, $g(x, y) = i(x, y)^{\alpha}$, $**$ is the 2D convolution operator, capitals indicate a given factor's FT, and f is a 2D frequency coordinate. At this point, it is apparent that in general Eq. [4] $\neq G(f)$ without further manipulation. The Fourier spectral properties will be studied in the future.

Measurement standardization

For making breast density measurements, we assume if two or more populations are statistically similar, their breast density measures could differ only to image format differences (i.e. due technological variations only). Our definition of technological differences includes both format differences and methods used to assess breast density (for the same metric). The same argument applies to the same population when mammograms are in different formats. We are going to take a general approach to standardization. However, this study is limited in what can be discerned from these modeling experiments directly. Conditional logistic regression is the appropriate approach for evaluating the matched case-control design (described below). This model does not permit the input of site-specific indicator variables. Here, we can evaluate scale differences but not translational differences from modeling alone. Standardization could take the form of transforming

all like measurements to a reference distribution using a pdf mapping approach, where the reference could be from one of the populations or an external pdf. The transformation approach is more general than a linear z-score transformation. However, the pdf mapping approach will resolve to a linear transformation (in this situation to a scaling transformation) if it is the solution. In either event, the approach should accommodate small datasets in the limiting case. We used the transformation approach as it may be useful in future studies as well as for comparisons in this investigation. The pdf transformation technique will be compared with the z-transform, logarithmic transformation, and without applying a transformation, where the transformation applies to PD_a.

Transformations and comparisons are described in detail. The first approach used the target pdf of a specific measurement that showed an association with breast cancer risk such as PBD, labeled as u. Specifically, u (the target) will be measured from synthetic 2D mammograms: clinical images from current technology. When the measurement is applied to images of a different format, it is labeled as w. The task is to make the measurement distribution in w appear as that of u. More formally given $w = f(u)$, the probability, $P(\mu_u - \sigma_u < u < \mu_u + \sigma_u)$, is generally not equivalent to analogous transformed probability, $P(\mu_w - \sigma_w < w = f(u) < \mu_w + \sigma_w)$, where, u_u and u_w are the respective intrinsic and transformed distribution means, σ_u and σ_w are the respective intrinsic and transformed distribution standard deviations, and f(u) is a monotonic transformation. Centering one or both distributions will not generally give equivalent probabilities. For other than linear, the transformation can distort the distribution shape and shift the center of mass. Both, the non-equivalence pdf property and regression coefficient rescaling can make it difficult to compare odds ratios (ORs) across studies without standardization. Of note in mammography studies, ORs from

image measurements are often cited as per-standard deviation shift. This approach is often used because a shift of this degree is likely to induce a significant change in risk. However, for comparisons, both the image measurement's standard deviation and respective logistic regression coefficient should be provided. Otherwise, different scales and image formats make comparisons across studies tenuous. The second transform used a reference pdf not related to u or w; a standardized normal pdf was used, taking the place of w. In the third scenario, we applied a z-score transformation. These were compared with ORs generated from the intrinsic pdfs (no-transformation) as well. For all formats, ORs and regression coefficients will be cited in both unit and standard deviation shifts when applicable.

We elaborate on the pdf transformation approach to address the small sample size problem. First, both the u and w pdfs were mapped to a standardized normal distribution (zero mean, unit variance) constructed with an *infinite* number of normal samples as an intermediate step. This step is to accommodate small datasets and put the measurements on a more universal scale. If the measurement scale is irrelevant, mapping to the normal pdf is likely sufficient, particularly for analytics. However clinically, the possibility of negative breast density reading may appear odd. A univariate kernel density estimation (KDE) technique was used to complement the number of samples (i.e., generate synthetic samples) in both u and w to construct the transformations. The KDE width parameter for a given format was optimized such that the samples added to the intrinsic pdfs or mapped u or w pdf were not significantly different (or target pdf) at the 5% significance level. An *infinite* number of synthetic samples was used to construct the finalized maps. This step ensures the maps are smooth, *continuous,* and that output values extend beyond the dynamic range of the respective starting distributions; integer valued data can be mapped to real numbers as well.

The inverse map to u is the final product, where the pdf of PD_a derived from the C-View images defined the target. The same process applies to multiple datasets that each take the position of w. A given w was mapped to a standardized normal and then mapped to u. Mapping to a normal is not required as the same process applies when mapping w to the C-View's pdf directly. Uncertainties caused by this type of manipulation would tend to push the breast cancer risk prediction relationship towards the null. Specifics for the univariate KDE (non-parametric) pdf generation technique based on differential evolution optimization were described earlier [49]. For large population datasets, the KDE application becomes less important.

Study Population

The current automated approach was applied to three successive case-control (CC) studies conducted over the course of the investigations, each defined by a given technology used to acquire its mammograms. Table 1 shows the mammographic technology for each study and image format details. Specifically, Study 1 encompassed 2D mammograms from a GE unit; Study 2 encompassed 2D mammograms from Hologic units; and Study 3 encompassed DBT datasets (volume and 2D synthetic images) from Hologic units. For brevity, we refer to these 3 studies as GE/Hologic raw-FFDM, GE/Hologic clinical-FFDM, or C-View (DBT), respectively. Although the Hologic and GE systems produce FFDM images, their detector technologies are different. Hologic units use direct x-ray conversion detector technology, whereas the GE unit used indirect x-ray conversion (i.e., their data representations are different).

The participant selection protocol was the same for each of the three studies, described previously [50, 51]. These studies were designed specifically to investigate image measures that could be used for risk prediction modeling. To isolate effects of image measurements, we selected controls that were matched to cases on key variables known to strongly influence breast physiology i.e., age within ± 2 years, hormone replacement (HR) usage (ever or never) and HR duration intervals (± 2 years). Controls were matched to cases on mammography units and screening history as well. All cases had unilateral breast cancer. The unaffected breast was analyzed with PD_a for cases and the corresponding latterly matched breast of the control was analyzed to form a matched case-control pair. Only craniocaudal views were considered to avoid chest wall interference.

When this institution's mammography centers were transitioning to Hologic-DBT from Hologic-FFDM, the center used the HD Combo Mode in the DBT units that acquired both 3D tomosynthesis data (volumetric and synthetic 2D images) and a separate 2D FFDM acquisition. In this mode, DBT and 2D (FFDM) acquisitions are taken in sequence without repositioning the patient or changing the breast compression. A total of 78 CC pairs underwent mammography during this transition period straddling studies 2 and 3 and had both DBT acquisitions and 2D FFDM images for analysis. For purposes of the present study, to optimize numbers of independent observations in each study and weighting toward the newer DBT technology, DBT data (only) from these, 78 CC pairs were included in the present analysis with Study 3; hence, data were analyzed for a total of 180, 241 and 426 CC pairs in studies 1, 2 and 3, respectively.

Risk prediction modeling

To *optimize* PD_a, OR tables were constructed for each image format as a function of the number of ensemble averages (m), detection search window size (n) in pixels, and search range of the two chi-square detection parameters. Odds ratios and related regression coefficients were determined with conditional logistic regression. These are relatively large tables, where model parameters were generated after taking the natural logarithm of PD_a giving another transformation comparison. For large batch automated processing tasks, this is our standard approach when investigating image measurements because intrinsic distributions may be positive-skewed (i.e., normally no chance of a negative valued finding). This step often gives more accurate estimates but can compress the intrinsic pdf to a point where a unit shift can span most of the distribution (spanning left to right tails), artificially inflating both the ORs and regression coefficients. For models using a logarithmic transform, parameters were generated in per standard deviation shift only of the log-transformed pdf. In all models, we controlled for body mass index (BMI) and ethnicity; age and hormonal attributes were matching factors in case-control selections.

To assist in making parameter choices, historical PBD measurements from the Cumulus technique were used as reference. These are tabulated in Table 2 and were distilled from these studies: (1 and 2) are from [6]; (3) is from [48]; (4) is from [7]; (5 and 6) are from [52]; and (7 and 8) are from [53]. This list includes measures derived from GE-FFDM, Hologic-FFDM, and digitized film mammograms from three different types of digitizers.

Frequently used abbreviations and conventions

The following abbreviations and conventions apply throughout this report: artificial intelligence (AI); case-control (CC); digital breast tomosynthesis (DBT); Fourier transform (FT); full field digital mammography (FFDM); General Electric (GE); odds ratio (OR); signal-dependent noise (SDN); three dimensional (3D); and two-dimensional (2D). A normalized histogram is referred to as an empirical probability density function (pdf). When a pdf is made from a given measurement without a transformation, it is referred to as the intrinsic pdf. Intrinsic is used in the same sense for other quantities as well. The inter parameter n is used to define both the number of CC pairs and the detention search window size (context is clear). Transformations are mappings or maps. A standardized normal distribution has zero-mean with unit variance. C-View images are 2D, constructed from DBT acquisitions, and are used for clinical purposes. DBT provides a limited angle quasi-3D reconstructed breast volume. In the DICOM header fields, images labeled *for presentation* are used for clinical purposes and those labeled *for processing* are loosely raw images. Clinical 2D FFDM images are created from the respective raw images with vendor proprietary image processing algorithms. Many facilities destroy the raw images. The various images have different data representations, which are referred to as format differences.

Results

Image format differences and breast density detection examples

Figure 1, top-row shows mammograms acquired *simultaneously* for two women in different formats: (1) panes 1-3 show three images from woman-1, and (2) panes 4-5 show two images from woman-2. These are used to exemplify the approach and illustrate natural variations throughout this work. Pane-1 shows a Hologic raw-FFDM image with inverted intensities, pane-2 shows the

corresponding clinical FFDM-image, pane-3 shows a C-View image from Hologic DBT (panes 1-3 show images from woman-1 at the same time point), pane-4 shows a GE raw-FFDM image with inverted intensities, and pane-5 shows the associated GE clinical FFDM-image (both GE images are from woman-2 at the same time point). These examples were selected at random from the respective studies. For a given woman, the multiple images are all valid image formats, where the bright tissue is breast density. The intra-variation within each woman's images is clear visually. The second row shows the respective chi-square images derived with varying degrees of ensemble averaging: from left to right m = (25, 5, 50, 25, 100). The corresponding PD_a images are shown in the bottom row: PD_a = (1) 13.2%; (2) 23.8%; (3) 25%; (4) 36.3%; and (5) 52.0% respectively. Each of these percentages were derived from a specific set of detection parameters discussed below. These demonstrate the scale variation problem; within each format, the CC distributions for PD_a produced significant ORs, although PD_a clearly varies across formats for the same woman due to the aforementioned differences. We are left with the obvious quandary of choosing the best representation, assuming one exists.

Signal-dependent noise shaping

In this section we show why gaining the ability to shape SDN to a specific functional form is the major innovation responsible for producing significant ORs given an arbitrary image format. With $\alpha = 1$, we expect that the noise variance to be a quadradic function of the signal when the variance is derived from the synthesized image in Eq. [3]. The intrinsic SDN functional forms for the three types of clinical mammograms shown in the top row of Figure 1 (panes 2, 3, and 5 corresponding with Hologic-clinical, Hologic C-view, and GE-clinical respectively) are shown in the top row of Figure 2, respectively. These show the expected noise variance (y-axis) as a function of the local

average (x-axis) in each mammogram. Potential problems for the automated detection can be noted in each example. Both the Hologic-clinical and C-View plots are concave-down with first derivatives that have maximums over the signal-average range. Here, two average signal values correspond with the same noise variance. When the detection threshold is located between these horizontal line SDN function intersects, it will cause algorithm confusion. The GE-clinical plot has approximately two increasing *linear* sections connected with an elbow, which is also less than optimal. The corresponding plots from Eq. [3] with $\alpha = 1$ are shown in the bottom row of Figure 2. These plots were derived with both ensemble and spatial averaging applied to the respective chi-square image. In each pane, function values from the data are shown in black and red shows fitted curves using a quadratic polynomial model, noting all fitted curves show close agreement with the data (similar agreement was noted across all images and formats, which are not shown). Such predictable relationships derived from different formats, both within and across, should stabilize the density detection algorithm. We only compared OR Tables with $\alpha = ½$ and 1. Tables with $\alpha = 1$ produced more cells with significant ORs (tables with $\alpha = ½$ are not shown). Thus, we cannot conclude that α has been globally optimized but the quadratic form provides enough contrast between the two tissue types to produce a predictable output with significant ORs. This also shows why the algorithm's performance was not optimal when the mammogram's local noise characteristic follows that of a Poisson process (i.e., when $\alpha = ½$).

Detection parameter operating points

The methodology for setting the operating points for PD_a is discussed in this section. Ideally, a CC sample from a given image format would be used to set these parameters. Historical distribution parameters were used to help select the operating points in conjunction with significant OR

findings. Because these CC studies were designed for image measurement investigations, one guiding premise is that significant ORs provide an objective truth given breast density is a significant risk factor. Initial experimentation indicated that setting $\alpha = 1$ provided superior results, which are discussed here. The degree of ensemble averaging (m) for each format was determined through initial experimentation (not shown here). The spatial averaging extent was set a priori by considering the intrinsic pixel spacing (or pitch) with $n = 4$ for the higher resolution images (i.e., Hologic FFDM images) and $n = 2$ for the other image formats. The respective OR findings are shown in tables 3-7. Odds ratios are provided in each cell with confidence intervals as a function of the two detection parameters. Although these were generated by first log-transforming the intrinsic distributions, the cited moments are from the intrinsic data. The blue shaded cells in each table show operating parameters that provided significant ORs that have intrinsic distribution moments (means and standard deviations) approximately within the range of the historical values cited in Table 2; these were used as a loose objective truth because other cells with relative outlier moments with significant ORs could also be valid. The objective of the optimization procedure was to adjust the four parameters (two detection parameters, m, and $\alpha = ½$ or 1) to optimize the number of possible acceptable operating points for a given format. We assume these parameters are the most stable choices because perturbing them slightly produced similar findings. Of note for the Hologic-raw-FFDM, and C-View formats, all parameters (all cells) produced significant ORs as did the GE clinical-FFDM cells; most Hologic clinical-FFDM cells produced significant ORs, as well. In each table, the cell with bold fonts and asterisk marks the respective operating points selected for this investigation. Odds ratios differed by image format. In summary (in standard deviation shift) images used for clinical purposes ORs ~ 1.31-1.60 and from raw-FFDM

images ORs ~ 1.20 – 1.31. Note, the blue shaded cells also provide examples of the scaling variations in the PD_a distributions both within and across formats.

General transformation analyses

The general pdf transformation findings are discussed first followed by specific examples. Table 8 shows the transform analyses restricted to clinical mammograms. Headings in the top (left to right) show the number of CC pairs used in the respective analyses. The total n in the last column is found by combining unique pairs from each format: n = 180 + 241 + 426 = 847 unique pairs. For PD_a, the regression parameter (β) for each type of transformation is cited in the top-row. Here (1 and 2) are best compared, and (3 and 4) are best compared. The largest impact is found within the GE clinical-FFDM findings. We posit the GE findings are due to small sample size uncertainty and pdf differences (discussed below). There were small changes in these coefficients when considering the z-score transformation (comparing 3 with 4). The middle row shows the respective ORs in unit shift [again, (1 and 2) are best compared, and (3 with 4) are best compared]. Within these comparison groups, the respective ORs in unit shift were similar within and across formats. In the bottom row, ORs in standard deviation shift were similar within formats (comparing 1-5); excluding the GE format, the other ORs were similar within and across formats within the comparison groups. Similarities in the last (total) column findings may be driven by the relatively large n in the C-View format, which is investigated further below with a specific example. The similarity in the respective z-score and pdf transformations suggest the transformations are approximately linear: determined by comparing 3 and 4 within formats. Across formats, ORs in unit shift varied from 1.05 - 1.07, which are relatively consistent considering the image format differences. The bottom row shows the OR findings in per standard deviation shift. Within format,

the ORs varied slightly over the transformation type. Across formats, the OR from the GE format was the largest, again probably due to sampling uncertainty. For the other formats, the findings were similar when considering the CIs. The sampling uncertainly is exemplified when comparing the total Hologic column (n = 667 pairs) with the corresponding smaller sample column (n = 241 pairs). The combined column indicates PD_a can be combined to form a larger case-control study with n = 847 pairs. The similarities between the probability transformation and z-score findings are an indication that the respective transformations are approximately linear.

Table 9 shows the raw FFDM formats compared with C-View format findings (repeated from above for ease of comparison). For PD_a, the regression parameters (βs) determined from the various transformations are cited in the top row, noting (1 and 2) are best compared and (3 and 4) best compared. Here, none of the regression parameters were significantly different from zero for the GE-format, in contrast with βs from the Hologic and C-View formats (all significantly different from zero). The other βs showed variation within and across formats. The middle row shows the ORs in unit shift. When comparing (1 with 2) and (3 with 4), ORs for GE were not significant. The other ORs were both similar and significant within and across formats, restricted to comparison groups. Odds ratios were similar when comparing 1-4 in the bottom row particularly within formats and the OR for the GE logarithmic transformation (5) was significant. Similarities between the respective pdf and z-transformation findings indicate the pdf transformations were also approximately linear, as above.

Additional transformation analyses

To assess whether selecting a different cell for the GE raw-FFDM findings would change the transformation metric outcomes, detection parameters from a cell representing outlier operating points relative to the historical PBD findings were selected from Table 5: specifically, (column 7, row 7) corresponding to the cell with $\mu = 8.48$ and $\sigma = 8.14$. Different outcomes were produced relative to those shown in Table 9, cited in Table 10. The βs from the standard normal (sn) and C-View pdf transformations (cv) were significant as were the three ORs cited in the middle (cv, sn, and z) and in the bottom row (cv, sn, and lg). Besides gaining significance, comparing the respective ORs across the findings in tables 9 and 10, there were no significant differences but those in Table 10 were consistently larger in all comparisons. This experiment demonstrates: (1) ORs are a function of the measurement distribution; and (2) the distribution is a function of the algorithm's operating parameters. These represent valid density representations when considering significant ORs provide an objective truth that the measurements are predictive of risk; this cautiously suggests historical distributions may not be the most relevant. This observation agrees with related work that showed higher density thresholds provided a measure that is more predictive of risk than the Cumulus application's thresholds [54].

Another alternative experiment was used to illustrate potential uncertainties due to distribution variations in PD_a when merging data from two formats to create a larger dataset. Here, the GE clinical-FFDM and C-View images were evaluated. The relevant intrinsic distributions for PD_a are shown in Figure 3. Figure 4 shows the three relevant maps: GE clinical-FFDM to normal (left); C-View to normal (middle); and GE-clinical to C-View (right). Figure 5 provides the pdf comparisons. The left pain shows the C-View pdf (black) compared with the corresponding mapped GE clinical-FFDM distribution (red). The right pane shows the normal pdfs for both

formats: C-view (black) and clinical-FFDM (red). The intrinsic distribution moments are cited in the respective OR table (bold fonts and asterisks). For comparisons purposes, the relevant ORs and CIs in standard deviation shifts are repeated here: (1) GE clinical-FFDM, OR =1.62 (1.25, 2.10); (2) C-View = 1.38 (1.14, 1.67); and (3) mapped GE clinical-FDMM to C-View with OR = 1.59 (1.23, 2.06). Because the number of CC pairs in the C-View study is considerably larger than the GE study, we randomly sampled 180 CC pairs from the C-View study for this investigation. For the reduced C-View pairs analysis: OR = 1.32 (0.99, 1.76), noting significance was lost. Combining these two populations (n = 360 pairs) without mapping gave OR = 1.80 (1.34, 2.41). Then, combining the populations after mapping the GE clinical-FFDM gave OR = 1.47 (1.21, 1.77), which appears to be a more feasible finding. Mapping both to normal pdfs and combining gave the *same* finding with OR = 1.47 (1.21, 1.78). The normal C-View pdfs were likely similar because the C-View pdf is approximately symmetric. The notable finding is from comparing the combined non-mapped and mapped ORs. The OR from the combined dataset without mapping is outside of the upper CI of the combined sample when the mapping was applied. Although further investigations are required, this illustration shows that it is possible to combine PD_a measurements from different formats (or different studies) to form a larger dataset when applying conditional logistic regression modeling. The maps shown in Figure 4 all appear to have a strong partial linear characteristic over most of the data ranges on the x-axes, indicating they are behaving as quasi-linear transformations.

Discussion and Conclusion

The work demonstrated how to generalize PD_a by adding two free parameters to the algorithm by incorporating a specific model that permitted: (1) shaping SDN to a predictable form; and (2)

ensemble averaging. Although not apparent, the four parameters in isolation affect the detection in these general ways for a given format: smaller significance levels reduce the detected area in either stage; increasing α, increases the detected area, and increasing the number of realizations (m) in the ensemble averaging increases the detected area. Likewise, changing a given parameter in the opposite direction reverses the respective effect. In combination, increasing m beyond a practical point will drive the significance parameters to parasitic levels, making it difficult to set detection thresholds. Ensemble averaging over a given mammogram and the α parameter were permitted by using the image defined in equations [2 and 3] instead of the mammogram as the detection input image. Equation [3] is a SDN transformation that eliminates (makes irrelevant) the requirement for a traditional SDN image as the detection input and can accommodate mammograms that were heavily processed. As demonstrated, the SDN relationship can be forced to the same quadratic form across different formats (the same results hold for the raw FFDM images but are not shown). The shaped and predictive SDN construct follows from the synthesis prescription. It is also clear that the ability to shape the SDN to quadratic form is format independent. Additionally, the intrinsic pixel spacing of a given mammogram is irrelevant. We also showed how to select the operating parameters using the OR table approach in conjunction with historical measurements. For fixed m and α, a given table specifies the problem within these two detection constraints. Strengths of the approach are its simplicity, the ability to generate a predictable SDN form, and it does not require ground truth (in the traditional sense) for training but just a sample from a foreign format (given an existing valid format), preferably derived from a CC sample to find operating parameters the produce first and second moments within the ranges shown here. The other intrinsic algorithm innovation is the capability of transforming a given mammogram into a form such that a standard statistical test is the canonical operation for detecting dense tissue. These are positive attributes

when compared to more complicated approaches with many free parameters. These attributes may be particularly useful for DBT applications. In the DBT, the format differences and image appearances are far more varied across manufactures than the corresponding FFDM images due to the respective technical differences in the image acquisitions [55], as well as the inclusion of AI enhanced synthetic 2D images (e.g., FDA # K162505**:** Hologic Intelligent 2D™ Image Reconstruction Software).

There are several qualifications that require mention. Hospital based populations were studied in the analysis. A small sample from the population of possible image formats was used in this study. However, the evidence shows our approach can adapt, especially for formats other than raw images (i.e., images that do not have the non-compressed breast thickness artifact). Likewise, the number of ensemble averages for a given format was not settled nor was the value for the exponent α. In our study, only two values for α were considered, both within the prior constraints. Precluding $\alpha < 1$ for further experimental consideration probably can be safely deduced from our findings. However, letting α exceed the upper $\alpha = 1$ bound will be explored in the future. Although the findings indicate the quadratic SDN form provided sufficient separation between adipose and dense tissue, we have not explicitly shown why this is the case. Planned simulation work using adipose and glandular x-ray attenuation coefficients from FFDM units[56, 57] with a commonly encountered range of compressed breast thicknesses may help to illustrate the gained separation. We have used these datasets in other breast density analyses so there is always a chance of sampling variation bias and overfitting. However, these concerns are mitigated for several reasons: (1) significant findings were found in all formats investigated over a wide range of operating

parameters; (2) the detection process operated on a *tailored* input image, and (3) PD_a is limited in the number of free operating parameters.

When considering absolute risk prediction over time, there are unresolved issues that include image measurement and modeling techniques. A few salient points are addressed here. Current clinical-models key in on different measures (with or without density) [5, 58] and do not consider molecular subtypes across breast lesions [59]. There are many automated approaches for estimating breast density. Whether or not a given approach is superior, or preferable for clinical workflow, requires a performance metric for comparison that includes the totality of complexities and requirements for a given approach. For example, a study that compared breast density metrics produced by commercial products with AI that computes PBD [14] illustrate the choice problem when considering the nuances and complexities associated with applying a given technique (e.g., considering free parameters and number of images for training and evaluation). There is limited work in evaluating PBD from DBT data. One AI approach showed ORs were larger from DBT than those estimated from 2D FFDM [60]; this study used the 2D projection images from the DBT acquisition as a starting point to produce the outputted PBD (for clarity, 2D projection images are similar to substantially lower dose FFDM images each acquired at a different projection angle that are then used to construct the DBT volume slices). Another technique provided similar findings from measures derived from the DBT volume slices and synthetic 2D images [43]. As the work in this report illustrated, measurement distributions have an impact on ORs making cross-study comparisons difficult, adding to the measurement choice problem.

Models used to predict absolute risk may require further investigations to exact how image measurements are included. A given breast density measure is only one component in this process, whereas CNN based absolute risk prediction can incorporate something akin to combining multiple image measurements simultaneously (all four views), whether representing a given component in a wider model or as the singular predictor. There is subtlety with CNN based methods that use all four mammographic views. These techniques may be finding early signs of disease if bilateral asymmetry is included in the learning architecture. For example, an AI approach was used to derive breast density and features from benign findings from DBT in a model that predicts short-term absolute risk, with promising findings for screening recommendations [61]. In contrast, the singular breast density measurement approach is most likely insensitive to bilateral differences related to disease. That is, the singular breast density metric is more closely aligned with pure risk prediction when used as a model input rather than risk prediction in combination with early detection. Whether there is an important distinction between risk and risk combined with early detection is not currently evident. It may be for optimized care, different models with varied endpoints could be used in concert for a given patient. Importantly, the appearance of the breast can change over time, more likely shifting to a less mammographically dense state [62] with significant differences between cases and controls [63]. To the best of our knowledge, evaluating the multivariate distribution inclusive of breast density, age and other time changing covariates (e.g., mass or BMI) has not been considered in the absolute risk prediction context to any extent. Current models, accepted for clinical applications, use breast density from a given time point [58]. As mentioned above, to the best of our knowledge, an objective performance comparison metric is lacking that accounts for the complexities required for a given approach to produce a stable prediction over longer-time frames with changing image formats.

Many open-ended measurement/modeling problems will likely be addressed using some form of DBT data. Of special note, including the breast volume slices from DBT to discern an image derived risk factor is a massive data-analytic task, especially when considering four views and age-related distribution changes. We posit side-by-side comparisons of models that use 2D synthetic images versus those the use the respective volume slices are required to document gains (or gained precision) when considering all of complexities between the two processing tasks. There are multiple models in use now for clinical applications without CNN based AI methods. When considering AI models with the virtually *unlimited* possible variants, developing a standard across institutions and regions fuels an open-ended and required field of inquiry.

Disclosures

Author contributions

JH, primary and corresponding author, designed the detection approach, developed code, and directed the experiments. EF is a co-author, assisted in algorithm and investigation designs, performed most of the coding tasks, applied the statistical tests, and developed the graphics in accord with JH MS and KE. MS and KE are co-authors, both of whom provided the required epidemiological expertise for the investigation.

Internal Review Board (IRB)

This work was performed under an approved IRB at the University of South Floridia titled, Automated Quantitative Measures of Breast Density: IRB # 104715_CR000002.

Use of Artificial Intelligence

Artificial intelligence (AI) was not used for writing this manuscript, nor was it used for grammatical manuscript proofing. AI was used in part to search for relevant research articles that were then inspected by the authors for relevance.

Data Sharing

Image and clinical data used for this work were made public by JH and MS. Data can be found at this site [64]. Code for this work is not public. The algorithm can be replicated with the details provided in this manuscript, referenced work, and by contacting JH for details.

Funding

The work was funded in part by these NIH (NCI) grants: R01CA11491, R01CA166269, and U01 CA200464.

Financial and other related disclosures

The authors have no financial or other conflicts to disclose.

References

[1] M. J. Yaffe, "Mammographic breast density: A parameter with strong implications for breast cancer risk," *Radiologic Clinics of North America,* vol. 53, no. 4, pp. 601-616, 2015, doi: 10.1016/j.rcl.2015.03.002.

[2] N. F. Boyd, L. J. Martin, M. Bronskill, M. J. Yaffe, N. Duric, and S. Minkin, "Breast tissue composition and susceptibility to breast cancer," *Journal of the National Cancer Institute,* vol. 102, no. 16, pp. 1224-1237, 2010.

[3] G. Kim and M. Bahl, "Assessing risk of breast cancer: a review of risk prediction models," *Journal of breast imaging,* vol. 3, no. 2, pp. 144-155, 2021.

[4] S. Hussain *et al.*, "Breast cancer risk prediction using machine learning: a systematic review," *Frontiers in oncology,* vol. 14, p. 1343627, 2024.

[5] J. A. Cintolo-Gonzalez *et al.*, "Breast cancer risk models: a comprehensive overview of existing models, validation, and clinical applications," *Breast cancer research and treatment,* vol. 164, no. 2, pp. 263-284, 2017.

[6] 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," *Academic radiology,* vol. 21, no. 8, pp. 958-970, 2014.

[7] J. J. Heine *et al.*, "An automated approach for estimation of breast density," *Cancer Epidemiology Biomarkers & Prevention,* vol. 17, no. 11, pp. 3090-3097, 2008.

[8] J. J. Heine and R. P. Velthuizen, "A statistical methodology for mammographic density detection," *Medical physics,* vol. 27, no. 12, pp. 2644-2651, 2000.

[9] T. Li *et al.*, "The association of measured breast tissue characteristics with mammographic density and other risk factors for breast cancer," *Cancer Epidemiology Biomarkers & Prevention,* vol. 14, no. 2, pp. 343-349, 2005.

[10] K. Ghosh *et al.*, "Tissue composition of mammographically dense and non-dense breast tissue," *Breast cancer research and treatment,* vol. 131, no. 1, pp. 267-275, 2012.

[11] N. F. Boyd *et al.*, "Mammographic density and the risk and detection of breast cancer," *New England journal of medicine,* vol. 356, no. 3, pp. 227-236, 2007.

[12] A. Pettersson *et al.*, "Mammographic density phenotypes and risk of breast cancer: a meta-analysis," *Journal of the National Cancer Institute,* vol. 106, no. 5, p. dju078, 2014.

[13] N. F. Boyd *et al.*, "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study," *JNCI: Journal of the National Cancer Institute,* vol. 87, no. 9, pp. 670-675, 1995.

[14] O. H. Maghsoudi *et al.*, "Deep-LIBRA: An artificial-intelligence method for robust quantification of breast density with independent validation in breast cancer risk assessment," *Medical image analysis,* vol. 73, p. 102138, 2021.

[15] C. E. Edmonds, S. R. O'Brien, and E. F. Conant, "Mammographic breast density: current assessment methods, clinical implications, and future directions," in *Seminars in Ultrasound, CT and MRI*, 2023, vol. 44, no. 1: Elsevier, pp. 35-45.

[16] W. He, A. Juette, E. R. Denton, A. Oliver, R. Martí, and R. Zwiggelaar, "A review on automatic mammographic density and parenchymal segmentation," *International journal of breast cancer,* vol. 2015, no. 1, p. 276217, 2015.

[17] S. D. Pawar, P. T. Joshi, V. A. Savalkar, K. K. Sharma, and S. G. Sapate, "Past, present and future of automated mammographic density measurement for breast cancer risk

prediction," in *Journal of Physics: Conference Series*, 2022, vol. 2327, no. 1: IOP Publishing, p. 012076.
[18] A. Gastounioti, E. F. Conant, and D. Kontos, "Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment," *The Breast,* vol. 31, pp. 261-272, 2016, doi: 10.1016/j.breast.2016.06.008.
[19] 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.
[20] D. Kontos *et al.*, "Radiomic phenotypes of mammographic parenchymal complexity: toward augmenting breast density in breast cancer risk assessment," *Radiology,* vol. 290, no. 1, pp. 41-49, 2019.
[21] K. Dembrower, E. Wåhlin, and Y. Liu, "Time to stratify mammographic screening using artificial intelligence: a review of current evidence," *The Breast,* vol. 50, pp. 19-25, 2020, doi: 10.1016/j.breast.2020.05.004.
[22] T. Li, H. Liu, and Z. Liu, "Artificial intelligence in breast imaging: where are we and what's next?," *European Radiology,* vol. 29, no. 11, pp. 5664-5675, 2019, doi: 10.1007/s00330-018-5744-3.
[23] A. Yala, T. Schuster, and R. Barzilay, "Breast cancer risk assessment with deep learning based on mammograms and health records: A review," *npj Breast Cancer,* vol. 7, no. 1, pp. 1-9, 2021, doi: 10.1038/s41523-021-00249-0.
[24] A. Gastounioti, S. Desai, V. S. Ahluwalia, E. F. Conant, and D. Kontos, "Artificial intelligence in mammographic phenotyping of breast cancer risk: a narrative review," *Breast Cancer Research,* vol. 24, no. 1, p. 14, 2022.
[25] C. M. Schopf *et al.*, "Artificial intelligence-driven mammography-based future breast cancer risk prediction: a systematic review," *Journal of the American College of Radiology,* vol. 21, no. 2, pp. 319-328, 2024.
[26] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning* (no. 2). MIT press Cambridge, 2016.
[27] Z. Teng *et al.*, "A literature review of artificial intelligence (AI) for medical image segmentation: from AI and explainable AI to trustworthy AI," *Quantitative imaging in medicine and surgery,* vol. 14, no. 12, p. 9620, 2024.
[28] D. T. Huff, A. J. Weisman, and R. Jeraj, "Interpretation and visualization techniques for deep learning models in medical imaging," *Phys Med Biol,* vol. 66, no. 4, p. 04TR01, Feb 2 2021, doi: 10.1088/1361-6560/abcd17.
[29] Y. Bao *et al.*, "Origin, methods, and evolution of the three Nurses' Health Studies," *American journal of public health,* vol. 106, no. 9, pp. 1573-1581, 2016.
[30] G. A. Colditz and S. E. Hankinson, "The Nurses' Health Study: lifestyle and health among women," *Nature Reviews Cancer,* vol. 5, no. 5, pp. 388-396, 2005.
[31] L. Nicosia *et al.*, "History of mammography: analysis of breast imaging diagnostic achievements over the last century," in *Healthcare*, 2023, vol. 11, no. 11: MDPI, p. 1596.
[32] J. Heggie *et al.*, "Position paper: recommendations for a digital mammography quality assurance program V4. 0," *Australasian Physical & Engineering Sciences in Medicine,* vol. 40, no. 3, pp. 491-543, 2017.
[33] H. Oh *et al.*, "Early-life and adult anthropometrics in relation to mammographic image intensity variation in the Nurses' Health Studies," *Cancer Epidemiology, Biomarkers & Prevention,* vol. 29, no. 2, pp. 343-351, 2020.

[34] "BI-RADS® Atlas: Breast Imaging Reporting and Data System ", 5th Edition ed. Reston, VA: American College of Radiology, 2013.
[35] C. M. Vachon *et al.*, "Mammographic density, breast cancer risk and risk prediction," *Breast Cancer Research,* vol. 9, no. 6, p. 217, 2007.
[36] V. A. McCormack and I. dos Santos Silva, "Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis," *Cancer Epidemiology Biomarkers & Prevention,* vol. 15, no. 6, pp. 1159-1169, 2006.
[37] F. Bodewes, A. Van Asselt, M. Dorrius, M. Greuter, and G. De Bock, "Mammographic breast density and the risk of breast cancer: a systematic review and meta-analysis," *The Breast,* vol. 66, pp. 62-68, 2022.
[38] M. B. Williams, P. A. Mangiafico, and P. U. Simoni, "Noise power spectra of images from digital mammography detectors," *Medical physics,* vol. 26, no. 7, pp. 1279-1293, 1999.
[39] P. Monnin, H. Bosmans, F. R. Verdun, and N. Marshall, "A comprehensive model for quantum noise characterization in digital mammography," *Physics in Medicine & Biology,* vol. 61, no. 5, p. 2083, 2016.
[40] J. J. Heine and M. Behera, "Aspects of signal-dependent noise characterization," *J Opt Soc Am A Opt Image Sci Vis,* vol. 23, no. 4, pp. 806-15, Apr 2006, doi: 10.1364/josaa.23.000806.
[41] J. D. Perezgonzalez, "Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing," *Frontiers in psychology,* vol. 6, p. 223, 2015.
[42] M. Yaffe *et al.*, "The myth of the 50-50 breast," *Medical physics,* vol. 36, no. 12, pp. 5437-5443, 2009.
[43] 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.
[44] J. Heine, E. Fowler, and M. B. Schabath, "Fourier analysis of signal dependent noise images," *Sci Rep,* vol. 14, no. 1, p. 30686, Dec 28 2024, doi: 10.1038/s41598-024-78299-1.
[45] J. J. Heine, S. R. Deans, R. P. Velthuizen, and L. P. Clarke, "On the statistical nature of mammograms," *Medical physics,* vol. 26, no. 11, pp. 2254-2265, 1999.
[46] J. J. Heine and R. P. Velthuizen, "Spectral analysis of full field digital mammography data," *Medical physics,* vol. 29, no. 5, pp. 647-661, 2002.
[47] A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," *Medical physics,* vol. 28, no. 4, pp. 419-437, 2001.
[48] J. J. Heine, K. Cao, and D. E. Rollison, "Calibrated measures for breast density estimation," *Academic radiology,* vol. 18, no. 5, pp. 547-555, 2011.
[49] J. Heine, E. E. E. Fowler, A. Berglund, M. J. Schell, and S. Eschrich, "Techniques to produce and evaluate realistic multivariate synthetic data," *Sci Rep,* vol. 13, no. 1, p. 12266, Jul 28 2023, doi: 10.1038/s41598-023-38832-0.
[50] E. E. E. Fowler, A. Smallwood, C. Miltich, J. Drukteinis, T. A. Sellers, and J. Heine, "Generalized breast density metrics," *Phys Med Biol,* vol. 64, no. 1, p. 015006, Dec 19 2018, doi: 10.1088/1361-6560/aaf307.

[51] 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.
[52] C. M. Vachon *et al.*, "Comparison of percent density from raw and processed full-field digital mammography data," *Breast Cancer Research,* vol. 15, no. 1, p. R1, 2013.
[53] E. T. Warner *et al.*, "Automated percent mammographic density, mammographic texture variation, and risk of breast cancer: a nested case-control study," *NPJ breast cancer,* vol. 7, no. 1, p. 68, 2021.
[54] T. L. Nguyen *et al.*, "Mammographic density defined by higher than conventional brightness threshold better predicts breast cancer risk for full-field digital mammograms," *Breast Cancer Research,* vol. 17, no. 1, p. 142, 2015.
[55] K. J. Geras, R. M. Mann, and L. Moy, "Artificial intelligence for mammography and digital breast tomosynthesis: current concepts and future perspectives," *Radiology,* vol. 293, no. 2, pp. 246-259, 2019.
[56] J. J. Heine and J. A. Thomas, "Effective x-ray attenuation coefficient measurements from two full field digital mammography systems for data calibration applications," *Biomedical engineering online,* vol. 7, no. 1, p. 13, 2008.
[57] 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, no. 1, p. 114, 2013.
[58] R. J. Acciavatti *et al.*, "Beyond breast density: risk measures for breast cancer in multiple imaging modalities," *Radiology,* vol. 306, no. 3, p. e222575, 2023.
[59] A. M. McCarthy *et al.*, "Performance of breast cancer risk-assessment models in a large mammography cohort," *JNCI: Journal of the National Cancer Institute,* vol. 112, no. 5, pp. 489-497, 2020.
[60] A. Gastounioti *et al.*, "Fully automated volumetric breast density estimation from digital breast tomosynthesis," *Radiology,* vol. 301, no. 3, pp. 561-568, 2021.
[61] M. Eriksson *et al.*, "A risk model for digital breast tomosynthesis to predict breast cancer and guide clinical care," *Science Translational Medicine,* vol. 14, no. 644, p. eabn3971, 2022.
[62] O. Ginsburg, L. Martin, and N. Boyd, "Mammographic density, lobular involution, and risk of breast cancer," *British journal of cancer,* vol. 99, no. 9, pp. 1369-1374, 2008.
[63] S. Jiang, D. L. Bennett, B. A. Rosner, and G. A. Colditz, "Longitudinal analysis of change in mammographic density in each breast and its association with breast cancer risk," *JAMA oncology,* vol. 9, no. 6, pp. 808-814, 2023.
[64] Hackathon Data Collections. Partnerships, Public-Private Markers, CLIA-Approved

Tables

Image Format	FFDM *for processing* (raw)	FFDM *for presentation* (clinical)	DBT C-View (clinical)
General Electric	14-bit n = 180	12-bit n = 180	n/a n/a
Hologic	14-bit n = 667	12-bit n = 667	10-bit n = 426

Table 1. Image format descriptions. Shown are each mammography unit's manufacturer, pixel dynamic range (bit), number of case-control pairs (n), and image type/usage (top row). Full field digital mammography (FFDM) formats involve case-control pairs with data available both for processing (raw) and for presentation. Hologic includes images acquired with Selenia FFDM and Dimensions digital breast tomosynthesis (DBT) units. The intended use for each type of format is provided parenthetically in the top row: *for processing* defines raw mammograms not used for clinical purposes and *for presentation* defines images used for clinical purposes when considering FFDM.

Unit / detector type	n	Image type	μ	σ
1.Senographe 2000D	360	raw-FFDM	19.52	15.22
2.Senographe 2000D	360	clinical-FFDM	21.12	13.72
3.Selenia & Dimensions	638	clinical-FFDM	25.3	11.6
4.Kodak Lumiscan 75	712	digitized film	26.5	14.6
5.Lumysis 85 and VIDAR CAD PRO Advantage	1900	digitized film	37.6	19.8
6.Lumysis 85 and VIDAR CAD PRO Advantage	3921	digitized film	32.1	19.6
7.Hologic Selenia	550	raw-FFDM	14.1	12.4
8.Hologic Selenia	550	clinical-FFDM	17.4	10.1

Table 2. Historical percentage of breast density distribution moments: a brief survey of prior percentage of breast density (PBD) distribution moments are cited: mean (μ) and standard deviation (σ). The unit or detector specifics and participant numbers (n) are also cited. PBD measurements were made with the user-assisted Cumulus method. Study citations are identified in a subsection within the Methods Section titled *Risk prediction modeling* and 4-6 cite film digitizers.

Detection Parameters	1^{-1}	1^{-2}	1^{-3}	1^{-4}	1^{-5}	1^{-6}	1^{-7}	1^{-8}	
1^{-1}	49.49	41.77	36.11	31.72	28.18	25.24	22.69	3.56	μ
	6.25	10.29	12.32	13.26	13.64	13.72	13.64	6.3	σ
	1.04 (0.83, 1.30)	1.12 (0.89, 1.40)	1.15 (0.92, 1.44)	1.18 (0.94, 1.48)	1.20 (0.96, 1.51)	1.22 (0.97, 1.53)	1.23 (0.98, 1.55)	n/a	OR
1^{-2}	43.48	34.75	28.86	24.59	21.32	18.73	16.56	2.2	μ
	7.27	10.81	12.19	12.64	12.67	12.47	12.15	4.77	σ
	1.12 (0.90, 1.40)	1.19 (0.95, 1.49)	1.22 (0.97, 1.54)	1.25 (0.99, 1.58)	1.27 (1.01, 1.61)	1.29 (1.02, 1.63)	1.20 (1.03, 1.66)	n/a	OR
1^{-3}	39.77	30.65	24.83	**20.77***	17.76	15.43	13.5	1.57	μ
	7.35	10.53	11.62	**11.85**	11.71	11.4	11.01	3.85	σ
	1.16 (0.93, 1.45)	1.22 (0.97, 1.54)	1.26 (1.00, 1.59)	**1.29 (1.02, 1.63)**	1.31 (1.03, 1.66)	1.32 (1.04, 1.68)	1.34 (1.05, 1.70)	n/a	OR
1^{-4}	37.24	27.97	22.25	18.36	15.54	13.38	11.62	1.18	μ
	7.1	10.05	10.98	11.1	10.89	10.53	10.1	3.16	σ
	1.17 (0.94, 1.46)	1.24 (0.98, 1.56)	1.28 (1.01, 1.62)	1.30 (1.03, 1.65)	1.32 (1.04, 1.68)	1.34 (1.05, 1.71)	1.35 (1.06, 1.72)	n/a	OR
1^{-5}	35.42	26.1	20.46	16.7	14.01	11.97	10.32	0.92	μ
	6.76	9.54	10.37	10.43	10.18	9.79	9.35	2.64	σ
	1.17 (0.94, 1.46)	1.24 (0.99, 1.57)	1.29 (1.02, 1.63)	1.31 (1.04, 1.67)	1.34 (1.05, 1.70)	1.35 (1.06, 1.72)	1.37 (1.07, 1.74)	n/a	OR
1^{-6}	34.06	24.71	19.15	15.48	12.87	10.92	9.35	0.73	μ
	6.41	9.03	9.8	9.83	9.55	9.15	8.71	2.23	σ
	1.17 (0.93, 1.46)	1.25 (0.99, 1.57)	1.29 (1.02, 1.64)	1.32 (1.04, 1.68)	1.34 (1.06, 1.71)	1.36 (1.07, 1.73)	1.37 (1.08, 1.75)	n/a	OR
1^{-7}	32.98	23.63	18.12	14.52	11.98	10.09	8.58	0.59	μ
	6.09	8.57	9.28	9.28	8.99	8.59	8.14	1.92	σ
	1.16 (0.93, 1.45)	1.25 (0.99, 1.57)	1.29 (1.02, 1.64)	1.33 (1.04, 1.69)	1.35 (1.06, 1.72)	1.37 (1.07, 1.74)	1.38 (1.08, 1.76)	n/a	OR
1^{-8}	28.37	19.16	13.87	10.48	8.15	6.45	5.15	0.08	μ
	4.89	5.94	6.15	5.89	5.43	4.92	4.4	0.53	σ
	1.01 (0.81, 1.26)	1.20 (0.95, 1.50)	1.28 (1.01, 1.62)	1.33 (1.04, 1.69)	1.36 (1.07, 1.73)	1.38 (1.08, 1.76)	1.40 (1.10, 1.79)	n/a	OR

Table 3. General Electric raw-FFDM odds ratio table: this gives the odd ratios (ORs) for PD_a with m =25 (number of ensemble averages) and n= 2 (detection search window size in pixels) over the range of detection parameters; the top row gives detection parameter 1 and the first column detection parameter 2. There are three metrics cited in each cell: (1) distribution mean (μ, top); (2) distribution standard deviation (σ, middle); and (3) an OR with confidence intervals parenthetically. Blue shading indicates possible operating parameters for PD_a, The cell with bold fonts and asterisk gives the detection parameters used throughout the report's examples.

Detection Parameters	1^{-1}	1^{-2}	1^{-3}	1^{-4}	1^{-5}	1^{-6}	1^{-7}	1^{-8}	
1^{-1}	51.70	33.91	**24.00***	17.91	13.86	10.98	8.80	0.49	μ
	9.67	11.94	**12.11**	11.51	10.63	9.70	8.78	1.33	σ
	1.62 (1.26, 2.08)	1.63 (1.26, 2.11)	**1.62 (1.25, 2.10)**	1.61 (1.24, 2.09)	1.61 (1.24, 2.09)	1.61 (1.24, 2.08)	1.60 (1.24, 2.07)	n/a	OR
1^{-2}	42.29	26.36	18.28	13.49	10.35	8.13	6.45	0.31	μ
	9.42	11.23	10.98	10.12	9.14	8.18	7.29	0.86	σ
	1.65 (1.28, 2.13)	1.62 (1.25, 2.10)	1.60 (1.24, 2.08)	1.60 (1.23, 2.07)	1.59 (1.22, 2.06)	1.58 (1.22, 2.04)	1.56 (1.21, 2.02)	n/a	OR
1^{-3}	37.85	23.01	15.75	11.50	8.74	6.79	5.33	0.23	μ
	8.76	10.37	9.98	9.07	8.09	7.17	6.31	0.62	σ
	1.65 (1.28, 2.14)	1.61 (1.24, 2.09)	1.59 (1.23, 2.07)	1.58 (1.22, 2.05)	1.57 (1.21, 2.04)	1.56 (1.20, 2.02)	1.54 (1.19, 1.99)	n/a	OR
1^{-4}	35.27	21.08	14.26	10.31	7.75	5.97	4.64	0.18	μ
	8.05	9.57	9.15	8.24	7.29	6.40	5.58	0.48	σ
	1.65 (1.28, 2.14)	1.61 (1.24, 2.08)	1.58 (1.22, 2.05)	1.57 (1.21, 2.04)	1.56 (1.21, 2.02)	1.54 (1.19, 1.99)	1.52 (1.18, 1.97)	n/a	OR
1^{-5}	33.6	19.83	13.28	9.51	7.08	5.40	4.16	0.15	μ
	7.43	8.9	8.48	7.58	6.65	5.79	5.00	0.40	σ
	1.65 (1.28, 2.14)	1.60 (1.24, 2.07)	1.58 (1.22, 2.05)	1.57 (1.21, 2.03)	1.55 (1.20, 2.01)	1.54 (1.19, 1.98)	1.51 (1.17, 1.95)	n/a	OR
1^{-6}	32.45	18.96	12.59	8.94	6.60	4.99	3.81	0.14	μ
	6.91	8.35	7.93	7.05	6.14	5.29	4.53	0.34	σ
	1.65 (1.28, 2.14)	1.60 (1.23, 2.07)	1.58 (1.22, 2.04)	1.56 (1.21, 2.02)	1.55 (1.20, 2.00)	1.53 (1.19, 1.98)	1.51 (1.17, 1.95)	n/a	OR
1^{-7}	31.61	18.32	12.07	8.51	6.24	4.68	3.55	0.12	μ
	6.49	7.90	7.48	6.61	5.71	4.88	4.14	0.31	σ
	1.65 (1.28, 2.14)	1.60 (1.23, 2.07)	1.58 (1.22, 2.04)	1.56 (1.21, 2.02)	1.55 (1.20, 2.00)	1.53 (1.19, 1.97)	1.51 (1.17, 1.94)	n/a	OR
1^{-8}	29.09	16.40	10.5	7.18	5.10	3.70	2.71	0.09	μ
	5.16	6.30	5.83	4.96	4.09	3.32	2.66	0.19	σ
	1.62 (1.26, 2.10)	1.60 (1.23, 2.08)	1.58 (1.22, 2.05)	1.57 (1.21, 2.03)	1.56 (1.20, 2.01)	1.54 (1.19, 1.98)	1.51 (1.17, 1.95)	n/a	OR

Table 4. General Electric clinical-FFDM odds ratio table: this gives the odd ratios (ORs) for PD_a with m =100 (number of ensemble averages) and n = 2 (detection search window size in pixels) over the range of detection parameters; the top row gives detection parameter 1 and the first column detection parameter 2. There are three metrics cited in each cell: (1) distribution mean (μ, top); (2) distribution standard deviation (σ, middle); and (3) an OR with confidence intervals parenthetically. Blue shading indicates possible operating parameters for PD_a, The cell with bold fonts asterisk give the detection parameters used throughout this report's examples.

Detection Parameters	1^{-1}	1^{-2}	1^{-3}	1^{-4}	1^{-5}	1^{-6}	1^{-7}	1^{-8}	
1^{-1}	49.32	38.93	31.98	27.02	23.30	20.4	18.02	4.61	μ
	10.79	15.82	17.74	18.35	18.36	18.08	17.63	9.71	σ
	1.21 (1.07, 1.37)	1.23 (1.08, 1.39)	1.24 (1.09, 1.41)	1.25 (1.10, 1.42)	1.25 (1.10, 1.43)	1.26 (1.11, 1.43)	1.26 (1.11, 1.44)	n/a	OR
1^{-2}	42.40	31.55	**24.98***	20.59	17.46	15.08	13.17	3.09	μ
	12.03	15.95	**16.95**	16.91	16.49	15.91	15.25	7.56	σ
	1.25 (1.10, 1.42)	1.26 (1.11, 1.43)	**1.26 (1.11, 1.44)**	1.27 (1.12, 1.44)	1.27 (1.12, 1.45)	1.28 (1.12, 1.45)	1.28 (1.12, 1.45)	n/a	OR
1^{-3}	38.59	27.66	21.37	17.31	14.48	12.38	10.71	2.38	μ
	11.92	15.26	15.84	15.54	14.94	14.25	13.53	6.33	σ
	1.27 (1.11, 1.44)	1.27 (1.12, 1.45)	1.28 (1.12, 1.46)	1.28 (1.13, 1.46)	1.29 (1.13, 1.46)	1.29 (1.13, 1.46)	1.28 (1.13, 1.46)	n/a	OR
1^{-4}	36.24	25.26	19.13	15.28	12.63	10.69	9.17	1.95	μ
	11.44	14.44	14.81	14.38	13.71	12.99	12.27	5.55	σ
	1.28 (1.12, 1.45)	1.28 (1.13, 1.46)	1.29 (1.13, 1.47)	1.29 (1.13, 1.47)	1.29 (1.13, 1.47)	1.29 (1.13, 1.47)	1.28 (1.13, 1.46)	n/a	OR
1^{-5}	34.66	23.66	17.63	13.89	11.37	9.54	8.13	1.66	μ
	10.91	13.67	13.91	13.42	12.73	12.01	11.3	4.99	σ
	1.28 (1.13, 1.46)	1.29 (1.13, 1.47)	1.29 (1.14, 1.47)	1.29 (1.14, 1.47)	1.30 (1.14, 1.47)	1.29 (1.14, 1.47)	1.28 (1.13, 1.46)	n/a	OR
1^{-6}	33.55	22.52	16.55	12.89	10.45	8.71	7.37	1.45	μ
	10.4	12.99	13.15	12.63	11.93	11.22	10.53	4.55	σ
	1.29 (1.13, 1.46)	1.29 (1.14, 1.47)	1.30 (1.14, 1.48)	1.30 (1.14, 1.48)	1.30 (1.14, 1.48)	1.29 (1.14, 1.47)	1.28 (1.13, 1.46)	n/a	OR
1^{-7}	32.70	21.66	15.73	12.13	9.75	8.06	6.78	1.28	μ
	9.94	12.38	12.48	11.94	11.24	10.55	9.87	4.19	σ
	1.29 (1.13, 1.46)	1.30 (1.14, 1.47)	1.30 (1.14, 1.48)	1.30 (1.14, 1.48)	1.30 (1.14, 1.48)	1.29 (1.14, 1.47)	1.28 (1.13, 1.46)	n/a	OR
1^{-8}	29.53	18.52	12.73	9.29	7.06	5.53	4.40	0.46	μ
	7.69	9.27	8.98	8.23	7.45	6.73	6.09	2.02	σ
	1.24 (1.09, 1.40)	1.28 (1.12, 1.46)	1.29 (1.13, 1.47)	1.29 (1.14, 1.47)	1.29 (1.14, 1.47)	1.29 (1.13, 1.47)	1.28 (1.12, 1.45)	n/a	OR

Table 5. Hologic raw-FFDM odds ratio table: this gives the odd ratios (ORs) for PD_a with m = 25 (number of ensemble averages) and n= 4 (detection search window size in pixels) over the range of detection parameters; the top row gives detection parameter 1 and the first column detection parameter 2. There are three metrics cited in each cell: (1) distribution mean (μ, top); (2) distribution standard deviation (σ, middle); and (3) an OR with confidence intervals parenthetically. Blue shading indicates possible operating parameters for PD_a, The cell with bold fonts and asterisk gives the detection parameters used throughout this report's examples.

Detection Parameters	1^{-1}	1^{-2}	1^{-3}	1^{-4}	1^{-5}	1^{-6}	1^{-7}	1^{-8}	
1^{-1}	43.03	37.28	33.41	30.48	28.11	26.13	24.39	19.88	μ
	5.14	6.07	6.73	7.19	7.51	7.73	7.87	7.99	σ
	1.03 (0.90, 1.17)	1.12 (0.98, 1.28)	1.17 (1.02, 1.35)	1.21 (1.05, 1.39)	1.23 (1.07, 1.42)	1.25 (1.08, 1.44)	1.26 (1.10, 1.46)	1.29 (1.12, 1.49)	OR
1^{-2}	40.39	34.69	30.91	28.07	25.79	23.9	22.25	18.00	μ
	4.87	5.9	6.57	7.01	7.3	7.48	7.59	7.61	σ
	1.07 (0.94, 1.23)	1.17 (1.02, 1.34)	1.22 (1.06, 1.40)	1.25 (1.08, 1.43)	1.26 (1.10 , 1.46)	1.28 (1.11, 1.47)	1.29 (1.12, 1.48)	1.31 (1.14, 1.50)	OR
1^{-3}	38.62	32.97	29.26	26.49	24.28	22.45	20.85	16.78	μ
	4.7	5.75	6.42	6.84	7.1	7.26	7.35	7.3	σ
	1.11 (0.97, 1.27)	1.20 (1.04, 1.38)	1.24 (1.08, 1.43)	1.27 (1.10, 1.46)	1.28 (1.11, 1.48)	1.29 (1.12, 1.49)	1.30 (1.13, 1.49)	1.31 (1.14, 1.51)	OR
1^{-4}	37.27	31.68	28.02	25.3	23.15	21.37	19.82	15.87	μ
	4.55	5.61	6.26	6.67	6.91	7.05	7.12	7.03	σ
	1.13 (0.98, 1.29)	1.22 (1.06, 1.40)	1.26 (1.09, 1.45)	1.28 (1.11, 1.47)	1.29 (1.12, 1.49)	1.30 (1.13, 1.49)	1.30 (1.13, 1.50)	1.31 (1.14, 1.51)	OR
1^{-5}	36.18	30.63	27.03	**24.36***	22.24	20.5	18.99	15.14	μ
	4.42	5.48	6.11	**6.50**	6.73	6.85	6.91	6.79	σ
	1.14 (0.99, 1.31)	1.23 (1.07, 1.42)	1.27 (1.10, 1.46)	**1.29 (1.12, 1.48)**	1.30 (1.13, 1.49)	1.30 (1.13, 1.50)	1.31 (1.14, 1.50)	1.32 (1.15, 1.51)	OR
1^{-6}	35.27	29.76	26.2	23.57	21.49	19.78	18.3	14.54	μ
	4.29	5.35	5.97	6.34	6.55	6.67	6.72	6.57	σ
	1.15 (1.00, 1.32)	1.24 (1.08, 1.43)	1.27 (1.10, 1.47)	1.29 (1.12, 1.48)	1.30 (1.13, 1.50)	1.31 (1.14, 1.50)	1.31 (1.14, 1.51)	1.32 (1.15, 1.51)	OR
1^{-7}	34.46	29	25.47	22.88	20.83	19.15	17.69	14.01	μ
	4.18	5.21	5.82	6.18	6.38	6.49	6.53	6.37	σ
	1.16 (1.01, 1.33)	1.25 (1.08, 1.44)	1.28 (1.11, 1.47)	1.29 (1.12, 1.49)	1.30 (1.13, 1.50)	1.31 (1.14, 1.50)	1.31 (1.14, 1.51)	1.32 (1.15, 1.51)	OR
1^{-8}	32.36	27.01	23.59	21.08	19.12	17.51	16.12	12.64	μ
	3.83	4.80	5.36	5.68	5.86	5.94	5.96	5.76	σ
	1.16 (1.01, 1.33)	1.25 (1.09, 1.44)	1.28 (1.11, 1.48)	1.30 (1.13, 1.49)	1.31 (1.14, 1.50)	1.31 (1.14, 1.51)	1.31 (1.14, 1.51)	1.31 (1.15, 1.51)	OR

Table 6. Hologic clinical-FFDM odds ratio table: this gives the odd ratios (ORs) for PD_a with m = 5 (number of ensemble averages) and n= 4 (detection search window size in pixels) over the range of detection parameters; the top row gives detection parameter 1 and the first column detection parameter 2. There are three metrics cited in each cell: (1) distribution mean (μ, top); (2) distribution standard deviation (σ, middle); and (3) an OR with confidence intervals parenthetically. Blue shading indicates possible operating parameters for PD_a, The cell with bold fonts and asterisk gives the detection parameters used throughout this report's examples.

Detection Parameters	1^{-1}	1^{-2}	1^{-3}	1^{-4}	1^{-5}	1^{-6}	1^{-7}	1^{-8}	
1^{-1}	44.73	34.49	28.32	**23.97***	20.67	18.04	15.81	8.13	μ
	5.79	6.76	6.87	**6.68**	6.37	6.02	5.64	3.66	σ
	1.36 (1.14, 1.63)	1.38 (1.15, 1.67)	1.38 (1.15, 1.67)	**1.38 (1.14, 1.67)**	1.37 (1.13, 1.67)	1.37 (1.13, 1.66)	1.36 (1.12, 1.65)	1.33 (1.10, 1.62)	OR
1^{-2}	40.58	30.95	25.22	21.21	18.18	15.77	13.74	6.84	μ
	5.49	6.24	6.23	5.98	5.65	5.28	4.91	3.05	σ
	1.39 (1.15, 1.67)	1.39 (1.15, 1.68)	1.38 (1.14, 1.67)	1.37 (1.13, 1.66)	1.36 (1.12, 1.65)	1.35 (1.11, 1.64)	1.34 (1.11, 1.63)	1.30 (1.08, 1.58)	OR
1^{-3}	38.04	28.81	23.35	19.55	16.68	14.41	12.5	6.08	μ
	5.21	5.83	5.76	5.49	5.15	4.79	4.42	2.68	σ
	1.39 (1.16, 1.68)	1.39 (1.15, 1.68)	1.38 (1.13, 1.67)	1.36 (1.12, 1.65)	1.35 (1.11, 1.64)	1.34 (1.10, 1.62)	1.33 (1.09, 1.61)	1.28 (1.06, 1.55)	OR
1^{-4}	36.23	27.29	22.03	18.37	15.62	13.45	11.63	5.56	μ
	4.96	5.49	5.39	5.1	4.76	4.41	4.06	2.41	σ
	1.40 (1.16, 1.69)	1.39 (1.15, 1.68)	1.37 (1.13, 1.66)	1.36 (1.12, 1.64)	1.34 (1.10, 1.62)	1.33 (1.09, 1.61)	1.31 (1.09, 1.59)	1.27 (1.05, 1.53)	OR
1^{-5}	34.83	26.13	21.01	17.47	14.81	12.72	10.97	5.16	μ
	4.75	5.2	5.07	4.79	4.45	4.11	3.77	2.21	σ
	1.40 (1.16, 1.69)	1.39 (1.15, 1.68)	1.37 (1.13, 1.66)	1.35 (1.11, 1.64)	1.33 (1.10, 1.62)	1.32 (1.09, 1.60)	1.31 (1.08, 1.58)	1.26 (1.05, 1.52)	OR
1^{-6}	33.7	25.19	20.2	16.75	14.17	12.13	10.44	4.85	μ
	4.56	4.95	4.81	4.52	4.19	3.86	3.53	2.04	σ
	1.40 (1.16, 1.70)	1.39 (1.14, 1.68)	1.37 (1.13, 1.66)	1.35 (1.11, 1.63)	1.33 (1.10, 1.61)	1.31 (1.08, 1.59)	1.30 (1.07, 1.57)	1.25 (1.04, 1.50)	OR
1^{-7}	32.75	24.4	19.51	16.14	13.62	11.64	9.99	4.59	μ
	4.39	4.74	4.58	4.29	3.97	3.64	3.32	1.9	σ
	1.41 (1.16, 1.70)	1.39 (1.14, 1.69)	1.37 (1.13, 1.66)	1.34 (1.11, 1.63)	1.32 (1.09, 1.60)	1.31 (1.08, 1.58)	1.29 (1.07, 1.56)	1.24 (1.04, 1.49)	OR
1^{-8}	29.45	21.68	17.17	14.07	11.77	9.97	8.49	3.73	μ
	3.78	3.94	3.74	3.46	3.16	2.87	2.6	1.43	σ
	1.40 (1.16, 1.69)	1.39 (1.15, 1.69)	1.36 (1.12, 1.66)	1.34 (1.10, 1.62)	1.32 (1.09, 1.59)	1.30 (1.07, 1.57)	1.28 (1.06, 1.54)	1.22 (1.02, 1.46)	OR

Table 7. Hologic C-View odds ratio table: this gives the odd ratios (ORs) for PD_a with m = 50 (number of ensemble averages) and n= 2 (detection search window size in pixels) over the range of detection parameters; the top row gives detection parameter 1 and the first column detection parameter 2. There are three metrics cited in each cell: (1) distribution mean (μ, top); (2) distribution standard deviation (σ, middle); and (3) OR with confidence intervals parenthetically. Blue shading indicates possible operating parameters for PD_a. The cell with bold fonts and asterisk gives the detection parameters used throughout this report's examples.

Metrics	GE clinical-FFDM n = 180	Hologic clinical-FFDM n = 667	Hologic clinical-FFDM n = 241	C-View n = 426	Combined n = 847
1. β (in)	0.0384 (0.0170, 0.0599)	0.0411 (0.0191, 0.0631)	0.0657 (0.0293, 0.1021)	0.0488 (0.0202, 0.0775)	0.0441 (0.0286, 0.0595)
2. β (cv)	0.0704 (0.0311, 0.1097)	0.0397 (0.0182, 0.0612)	0.0576 (0.0237, 0.0914)	0.0488 (0.0202, 0.0775)	0.0545 (0.0357, 0.0733)
3. β (sn)	0.4657 (0.2054, 0.7259)	0.2672 (0.1241, 0.4103)	0.4271 (0.1905, 0.6637)	0.3262 (0.1348, 0.5176)	0.3729 (0.2459, 0.4999)
4. β (zt)	0.5074 (0.2325, 0.7824)	0.2724 (0.1249, 0.4198)	0.4298 (0.1880, 0.6716)	0.3418 (0.1452, 0.5384)	0.3909 (0.2599, 0.5220)
1. OR (in)	1.04 (1.02, 1.06)	1.04 (1.02, 1.07)	1.07 (1.03, 1.11)	1.05 (1.02, 1.08)	1.05 (1.03, 1.06)
2. OR (cv)	1.07 (1.03, 1.12)	1.04 (1.02, 1.06)	1.06 (1.02, 1.10)	1.05 (1.02, 1.08)	1.06 (1.04, 1.08)
3. OR (sn)	1.59 (1.23, 2.07)	1.31 (1.13, 1.51)	1.53 (1.21, 1.94)	1.39 (1.14, 1.68)	1.45 (1.28, 1.65)
4. OR (zt)	1.66 (1.26, 2.19)	1.31 (1.13, 1.52)	1.54 (1.21, 1.96)	1.41 (1.16, 1.71)	1.48 (1.30, 1.69)
1. OR SD (in)	1.59 (1.23, 2.07)	1.31 (1.13, 1.51)	1.48 (1.19, 1.83)	1.39 (1.14, 1.68)	1.42 (1.26, 1.62)
2. OR SD (cv)	1.59 (1.23, 2.06)	1.30 (1.13, 1.50)	1.44 (1.16, 1.78)	1.39 (1.14, 1.68)	1.44 (1.27, 1.63)
3. OR SD (sn)	1.59 (1.23, 2.07)	1.31 (1.13, 1.51)	1.48 (1.19, 1.83)	1.39 (1.14, 1.68)	1.45 (1.28, 1.64)
4. OR SD (zt)	1.62 (1.25, 2.10)	1.30 (1.13, 1.50)	1.47 (1.18, 1.83)	1.40 (1.15, 1.69)	1.45 (1.28, 1.65)
5. OR SD (lg)	1.61 (1.25, 2.10)	1.29 (1.12, 1.48)	1.52 (1.22, 1.89)	1.38 (1.14, 1.67)	1.44 (1.26, 163)

Table 8. Transformation findings from PD_a restricted to clinical images: each larger row contains multiple findings for a given metric. The top row provides β coefficients (unit shift) from conditional logistic regression with 95% confidence intervals (CIs), for each type of transformation, referenced parenthetically: (1) intrinsic pdf (in); (2) transformation to the C-View pdf (cv); (3) transformation to a standardized normal pdf (sn); and (4) z-score transformation (zt). The middle row provides odds ratio (OR) for each transformation in per unit shift with 95% CIs with the same labeling as the top row. The bottom row gives the OR for each transformation in standard deviation shift with 95% CIs with the same labeling; a fifth row was added to include the logarithmic transformation (lg) corresponding with the respective OR table. The number of case-control pairs (n) used for the respective calculations is given in the top row. For the combined column n = 180+241+426=847 derived from the respective number of unique pairs in each format.

Metrics	GE raw-FFDM n = 180	Hologic raw-FFDM n = 667	Hologic raw-FFDM n = 241	C-View n = 426	Combined n = 847
1. β (in)	0.0157 (-0.0039, 0.0353)	0.0153 (0.0077, 0.0229)	0.0133 (0.0007, 0.0259)	0.0488 (0.0202, 0.0775)	0.0174 (0.0074, 0.0275)
2. β (cv)	0.0270 (-0.0084, 0.0625)	0.0391 (0.0199, 0.0584)	0.0329 (0.0001, 0.0658)	0.0488 (0.0202, 0.0775)	0.0318 (0.0139, 0.0497)
3. β (sn)	0.1862 (-0.0464, 0.4188)	0.2593 (0.1309, 0.3876)	0.2260 (0.0125, 0.4395)	0.3262 (0.1348, 0.5176)	0.2213 (0.1031, 0.3395)
4. β (zt)	0.2098 (-0.0371, 0.4567)	0.2675 (0.1355, 0.3995)	0.2296 (-0.0027, 0.4620)	0.3418 (0.1452, 0.5384)	0.2238 (0.1000, 0.3475)
1. OR (in)	1.02 (1.00, 1.04)	1.02 (1.01, 1.02)	1.01 (1.00, 1.03)	1.05 (1.02, 1.08)	1.02 (1.01, 1.03)
2. OR (cv)	1.03 (0.99, 1.06)	1.04 (1.02, 1.06)	1.03 (1.00, 1.07)	1.05 (1.02, 1.08)	1.03 (1.01, 1.05)
3. OR (sn)	1.21 (0.96, 1.52)	1.30 (1.14, 1.47)	1.25 (1.01, 1.55)	1.39 (1.14, 1.68)	1.25 (1.11, 1.40)
4. OR (zt)	1.23 (0.96, 1.58)	1.31 (1.15, 1.49)	1.26 (1.00, 1.59)	1.41 (1.16, 1.71)	1.25 (1.11, 1.42)
1. OR SD (in)	1.21 (0.96, 1.52)	1.30 (1.14, 1.47)	1.23 (1.01, 1.51)	1.39 (1.14, 1.68)	1.22 (1.09, 1.36)
2. OR SD (cv)	1.19 (0.95, 1.51)	1.30 (1.13, 1.48)	1.22 (1.00, 1.49)	1.39 (1.14, 1.68)	1.23 (1.10, 1.38)
3. OR SD (sn)	1.22 (0.97, 1.54)	1.30 (1.14, 1.47)	1.23 (1.01, 1.51)	1.39 (1.14, 1.68)	1.24 (1.11, 1.40)
4. OR SD (zt)	1.21 (0.96, 1.52)	1.30 (1.14, 1.47)	1.22 (1.00, 1.49)	1.40 (1.15, 1.69)	1.23 (1.10, 1.39)
5. OR SD (lg)	1.29 (1.02, 1.63)	1.26 (1.11, 1.44)	1.26 (1.03, 1.53)	1.28 (1.14, 1.67)	1.22 (1.09, 1.37)

Table 9. Transformation findings from PD_a restricted to raw FFDM and C-view images: each larger row contains multiple findings for a given metric. The top row provides β coefficients (unit shift) from conditional logistic regression with 95% confidence intervals (CIs), for each type of transformation, referenced parenthetically: (1) intrinsic pdf (in); (2) transformation to the C-View pdf (cv); (3) transformation to a standardized normal pdf (sn); and (4) z-score transformation (zt). The middle row provides the OR for each transformation in per unit shift with 95% CIs with the same labeling. The bottom row gives the OR for each transformation in standard deviation shift with 95% CIs with the same labeling; a fifth row was added to include the logarithmic transformation (lg) corresponding with the respective OR table. The number of case-control pairs (n) used for the respective calculations is given in the top row. For the combined column n = 180+241+426=847, derived from the respective number of unique pairs each format C-view findings are the same as in Table 8 and repeated here for ease of companions.

Metrics	GE raw-FFDM n = 180
1. β (in)	0.0274 (-0.0011, 0.0559)
2. β (cv)	0.0427 (0.0071, 0.0783)
3. β (sn)	0.3126 (0.0642, 0.5610)
4. β (zt)	0.2229 (-0.0093, 0.4551)
1. OR (in)	1.03 (1.00, 1.06)
2. OR (cv)	1.04 (1.01, 1.08)
3. OR (sn)	1.37 (1.07, 1.75)
4. OR (zt)	1.25 (0.99, 1.58)
1. OR SD (in)	1.25 (0.99, 1.58)
2. OR SD (cv)	1.33 (1.05, 1.69)
3. OR SD (sn)	1.35 (1.06, 1.72)
4. OR SD (zt)	1.25 (0.99, 1.58)
5. OR SD (lg)	1.38 (1.08, 1.76)

Table 10. Outlier transformation findings for PD_a determined from GE raw-FFDM images: here, we selected outlier operating points relative to the historical findings from Table 5, (column 7, row 7), or the cell with $\mu = 8.48$ and $\sigma = 8.14$. Each larger row contains multiple findings for a given metric. The top row provides β coefficients (unit shift) from conditional logistic regression with 95% confidence intervals (CIs), for each type of transformation, referenced parenthetically: (1) intrinsic pdf (in); (2) transformation to the C-View pdf (cv); (3) transformation to a standardized normal pdf (sn); and (4) z-score transformation (zt). The middle row provides the odds ratio (OR) for each transformation in per unit shift with 95% CIs with the same labeling. The bottom row gives the OR for each transformation in standard deviation shift with 95% CIs using the same labeling; a fifth row was added to include the logarithmic transformation (lg) corresponding with the respective OR table. The number of case-control pairs (n) is given in the top row.

Figures

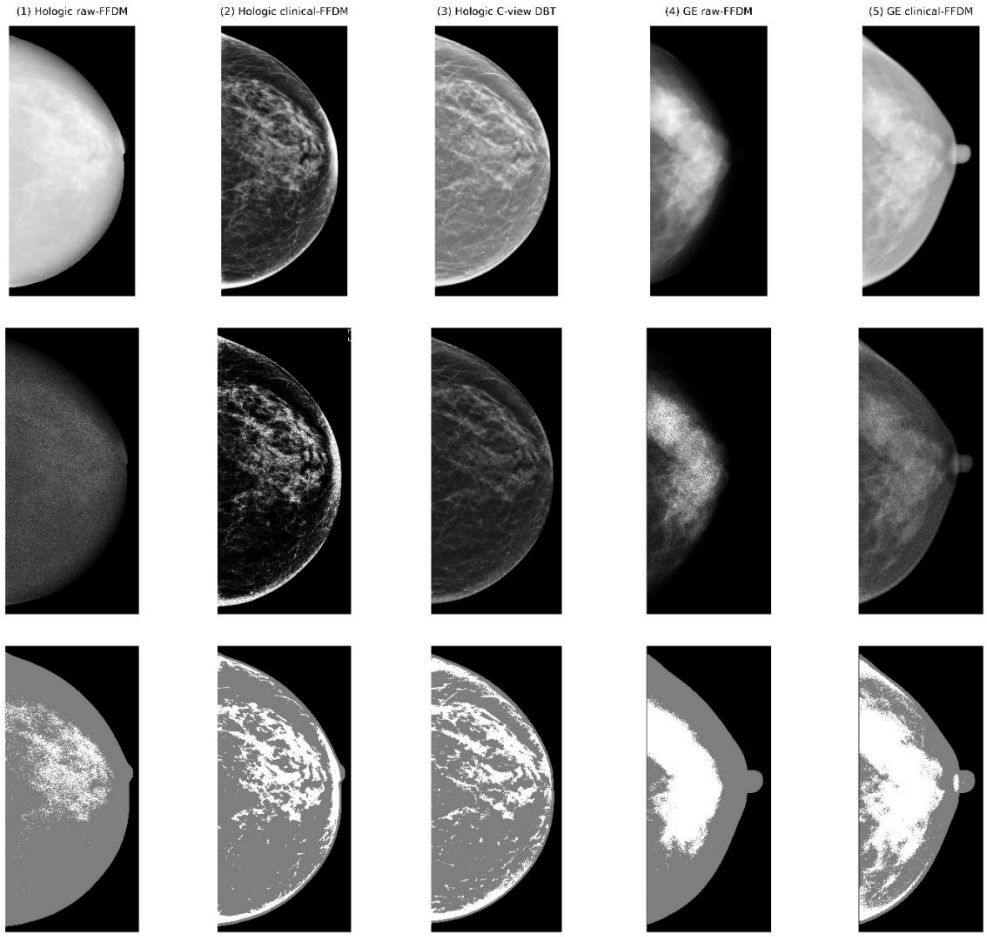

Figure 1. Image format differences and corresponding PD_a examples: in the top row (left to right) panes 1-3 show mammograms from woman-1, acquired *simultaneously* in three different formats, and panes 3-4 show mammograms from women-2 acquired *simultaneously* in two different formats: (1) woman-1, Hologic raw FFDM; (2) woman-1, Hologic clinical FFDM; (3) woman-1, Hologic C-View, (4-5) woman-2, GE raw FFDM and GE clinical FFDM images, respectively. The middle row shows the corresponding $s_m(x, y)$ images from Eq. [3] obtained with m ensemble averages followed by an n×n spatial average: (1) m = 25 and n = 4; (2) m = 5 and n = 4; (3) m = 50 and n = 2; (4) m = 25 and n = 2; and (5) m = 100 and n = 2. The bottom row shows the adipose and dense areas detected by PD_a: (1) 13.2%; (2) 23.8%; (3) 25%; (4) 36.3%; and (5) 52.0%. Images shown in the middle row were adjusted for viewing purposes and all have the same window level and width.

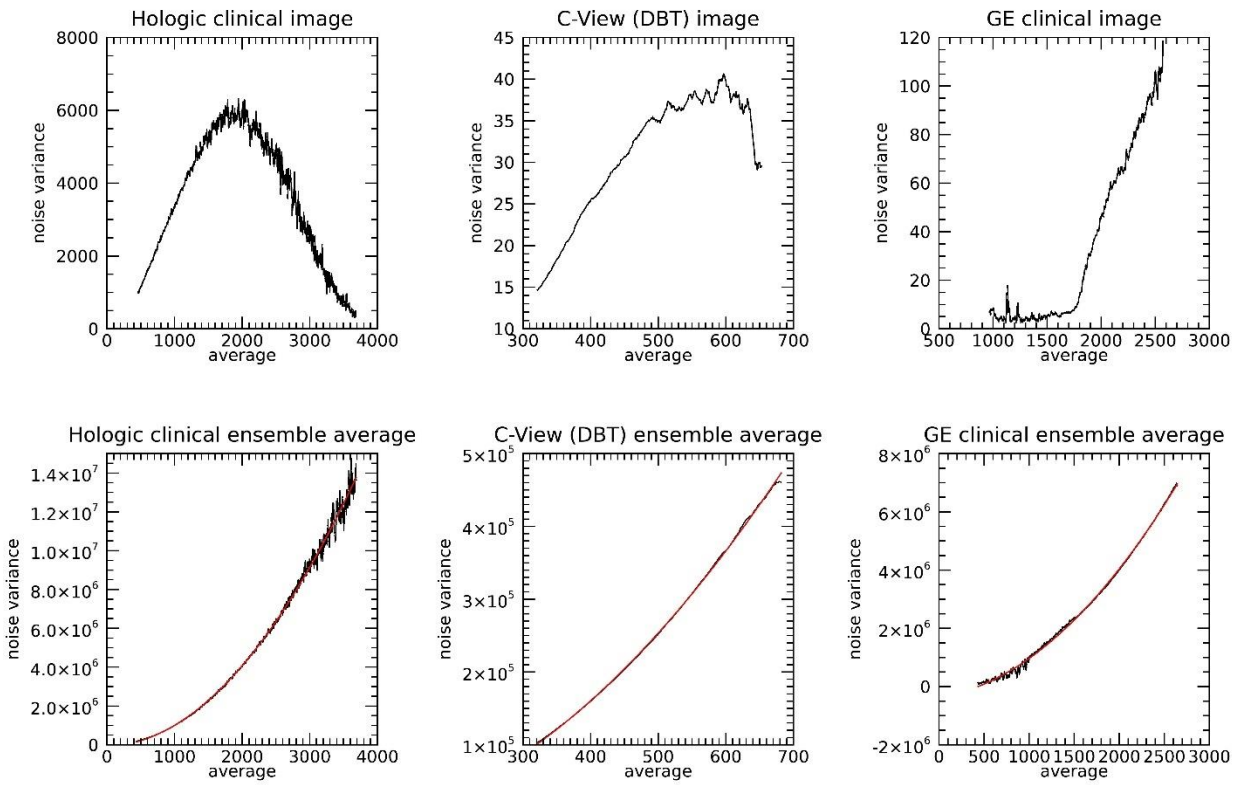

Figure 2. Signal-dependent noise variance plot examples: these plots show expected pixel values (average) on the x-axes and their expected variances on the y-axes determined by single chi-square mammograms. The top row shows intrinsic plots derived from the images used for clinical purposes shown in Figure 1: Hologic-clinical FFDM; Hologic C-View from DBT; and GE-clinical FFDM. The bottom row shows the respective plots derived from the chi-square images, also shown in Figure 1: black curves show data and red shows the fitted quadratic-model curves derived from regression analyses.

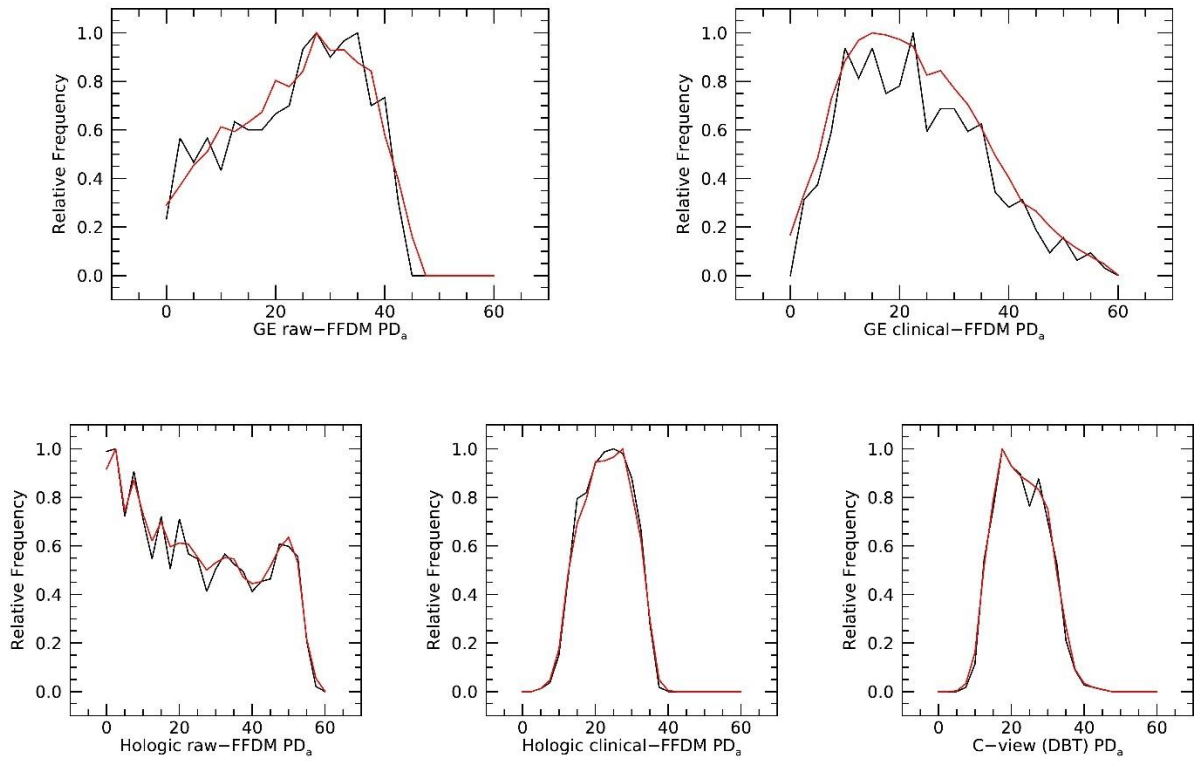

Figure 3. Empirical probability density function (pdf) comparisons with kernel density estimation pdfs: pdfs for PD_a distributions are provided for each image format using the detection parameters identified in the odds ratio tables. Solid black lines show the data and red lines pdfs derived from the kernel density estimations, respectively.

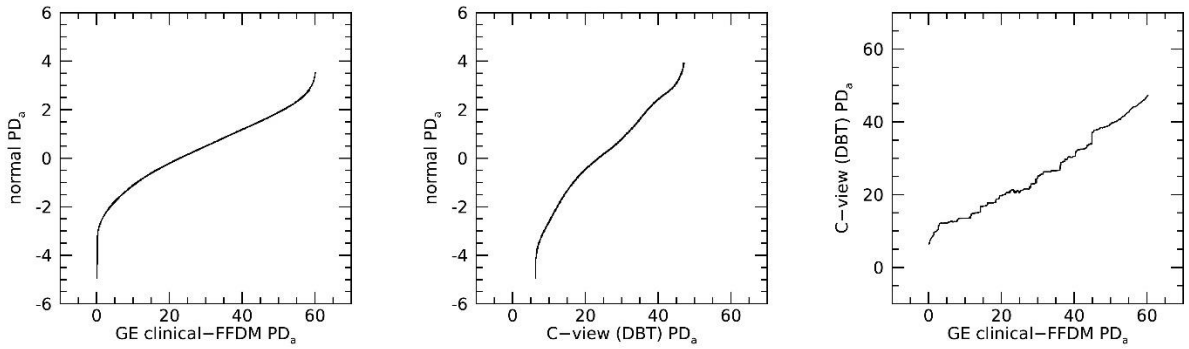

Figure 4. PD_a empirical probability density function (pdf) transformation examples for the GE-clinical pdf to C-View pdf: (1) GE intrinsic to normal pdf (left); (2) C-View intrinsic to normal pdf (middle); and GE pdf to C-View pdf (right).

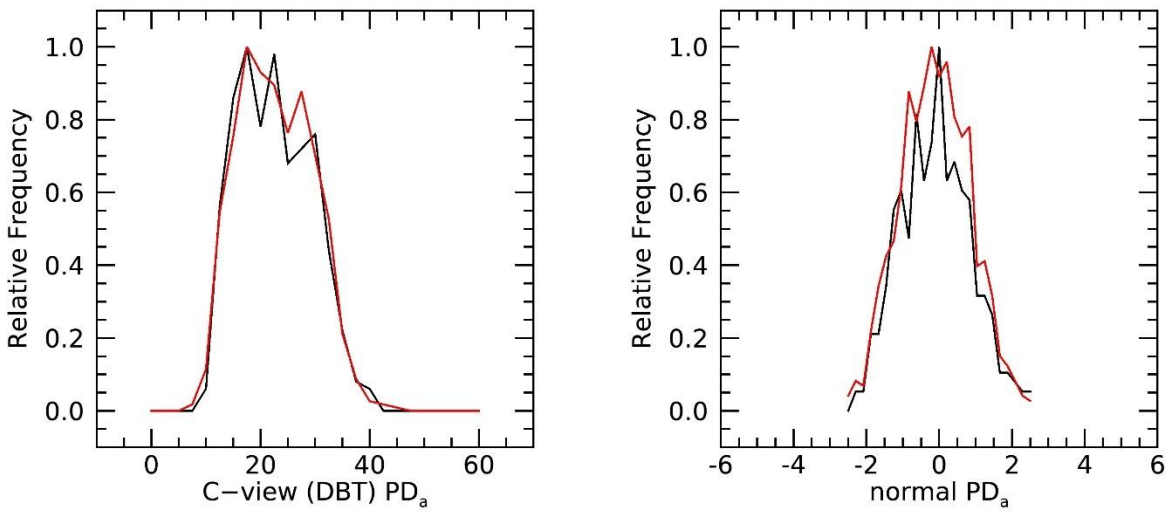

Figure 5. PD_a transformation empirical probability density function (pdf) example comparisons: the right pane shows the mapped GE-clinical pdf (red) compared with the C-View pdf (black). The right pane shows both distributions mapped to a standardized normal: GE-clinical pdf (red) and C-view pdf (black).