HOKKAIDO UNIVERSITY

Title	Imaging biomarkers of breast cancer derived from breast parenchyma
Author(s)	WANG, Jeffrey
Degree Grantor	北海道大学
Degree Name	博士(保健科学)
Dissertation Number	乙第7028号
Issue Date	2017-06-30
DOI	https://doi.org/10.14943/doctoral.r7028
Doc URL	https://hdl.handle.net/2115/66578
Type	doctoral thesis
File Information	Jeffrey_WANG.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学 位 論 文

Imaging biomarkers of breast cancer derived from breast parenchyma
（乳腺から派生した乳がんのイメージバイオマーカー）

Jeffrey Wang

北海道大学大学院保健科学院

２０１７年度

Contents

List of Tables 4
List of Figures 5
List of Abbreviations 10
Chapter 1 Introduction 12
1.1 Background 12
1.1.1 Breast cancer 12
1.1.2 Tumor microenvironment 13
1.1.3 Radiomics and machine learning 14
1.2 Purpose 15
Chapter 2 Agreement of Mammographic Measures of Volumetric Breast Density to MRI 17
2.1 Introduction 17
2.2 Methods 18
2.2.1 Subjects 19
2.2.2 MR imaging and breast density analysis 19
2.2.3 Mammographic imaging 20
2.2.4 SXA breast density analysis 20
2.2.5 Quantra and Volpara volumetric assessment 21
2.2.6 Statistical analysis 22
2.3 Results 23
2.4 Discussion 28
2.5 Conclusions 33
Chapter 3 Automatic Estimation of Volumetric Breast Density Using Artificial Neural Network-Based Calibration of Full-Field Digital Mammography: Feasibility on Japanese Women with and without Breast Cancer 34
3.1 Introduction 34
3.2 Materials and Methods 37
3.2.1 Study population 37
3.2.2 Breast density modeling 38
3.2.3 Breast density analysis 41
3.2.4 Statistical analysis 44
3.3 Results 46
3.3.1 Phantom validation 46
3.3.2 Patient validation 46
3.4 Discussion 51
3.5 Conclusions 55
Chapter 4 Identifying Triple-Negative Breast Cancer Using Background Parenchymal Enhancement Heterogeneity on Dynamic Contrast-Enhanced MRI: A Pilot Radiomics Study 56

4.1 Introduction 56
4.2 Materials and Methods 58
4.2.1 Study Population 58
4.2.2 Pathological Subtyping 58
4.2.3 Image Acquisition 59
4.2.4 Image Segmentation and Feature Extraction 60
4.2.5 Predictive Modeling for Differentiation 62
4.2.6 Cluster Analysis 63
4.2.7 Statistical Analysis 63
4.3 Results 64
4.3.1 Differentiation of Molecular Subtypes 64
4.3.2 Optimal Imaging Features 66
4.3.3 Clustering 69
4.4 Discussion 70
4.5 Conclusion 74
Chapter 5 Summary 75
References 77
Selected List of Relevant Works 89
Acknowledgements 91
Appendix 92

List of Tables

Table 2.1. Description of population characteristics 24

Table 2.2. Density measures for study participants by method (n=99) 25

Table 2.3. Quartile density ranges by method (n=99) 25

Table 3.1: Study population health demographics 38

Table 3.2: Imaging and breast parameter determinants of signal intensity on FFDM considered for ANN-VBD calibrations 40

Table 3.3: Comparisons of ANN-estimated measures of VBD between cancer and non-cancer patients 51

Table 4.1: Patient and subtype demographics 59

Table 4.2: Performance results of predictive modeling 65

Table 4.3: Optimal imaging features selected 67

Supplemental Table Appendix.S1: Optimal imaging feature distribution and prognostic significance 92

List of Figures

Figure 2.1: Best linear regression fit line with 95% confidence interval bands for percentage fibroglandular density (top), log fibroglandular volume (middle), and total breast volume (bottom) for MRI versus either SXA (left), Quantra (center), or Volpara (right) measures. Solid points correspond to example images in Figure 2.4..... 26

Figure 2.2: Comparison of quartiles classification for percent fibroglandular density (top) and log fibroglandular volume (middle) for MRI versus SXA (left), Quantra (center), and Volpara (right). The bottom row of plots show quartiles comparisons between mammographic density measures. Legend at right defines categories of agreement, where either the two compared method's agree completely (black) or are off by one or two quartiles up or down in comparison with the other method..... 27

Figure 2.3: Validation of SXA model using breast biology and adipose volume estimates from MRI. The amount of water volume in the MRI adipose volume was estimated to be 15% of the volume, which is consistent with previous work estimating it to be between 8% (64) and 20% (65). The MRI model does not include adipose density in the fibroglandular volume while SXA does. Subtracting out the adipose water volume from the SXA fibroglandular volume improved the agreement between SXA and MRI from R^2=0.78 to 0.83 and removed most of the bias between the measures. 28

Figure 2.4: Six comparisons of the LCC mammograms to their respective left central breast axial-slice MR images on five different women (c and d are the same woman). The white line connecting points in the MR images define the total breast volume. The MRI fibroglandular volume is shown delineated with white lines without points. Solid data points 2.4a-2.4f in Figure 2.1 correspond to the image labels a-f. Compared to the mammographically-derived SXA values, a) MRI percent density is higher, b) MRI percent density is lower, c) MRI breast volume is higher, d) of the same woman as c (this mammogram not part of analyses, only here and measures plotted in Figure 2.1 to illustrate one reason for discrepancy between methods' results), MRI breast volume is better

segmented due to the breast being extended more into the mammographic image field, e) MRI breast volume is lower, f) all MRI measures of density and volume were in substantial agreement. 30

Figure 3.1: GEN III phantom image. Example FFDM image of GEN III phantom, with 9 breast tissue-equivalent regions used in the study as references annotated with their percent volumetric breast density (VBD, also at top) and thickness combinations (also at right). 38

Figure 3.2: An example of one GEN III phantom image's 9 tissue composition regions plotted with its fitted function for Volumetric Breast Density (z-axis), given tissue thickness (x-axis) and signal intensity (y-axis). Each function was used to derive two separate sets of virtual phantom data using random combinations of thickness and intensity - the first for training the Artificial Neural Network calibrations and the second for its validation. 40

Figure 3.3: Breast thickness modeling. Example breast tissue thickness map, modeled using compression paddle thickness and breast edge as described in methods. Breast tissue thickness (in z-direction, per color bar) is projected atop the bottom detector platform (0cm, represented as black). The sum of thicknesses at all pixels make up total breast volume (TBV). 42

Figure 3.4: Summary of mammographic breast density modeling and analysis performed in this study. Repeated imaging of a GEN III Volumetric Breast Density (VBD) phantom was performed on one FFDM system over a broad range of imaging parameters, determined by a survey of screening exams acquired as part of standard clinical practice (A). Physical phantom images were used to derive a set of virtual phantom data for the purposes of calibrating the FFDM system (B). This set of virtual calibration phantom data was used to train an Artificial Neural Network (ANN), taking imaging parameters and data (anode and filter materials used, tube voltage, current exposure time, image exposure sensitivity, background detector signal, tissue thickness, and signal intensity) as input and outputting VBD (C). Physical phantom images were used to derive a second set of virtual phantom data for the purposes of validating the ANN calibrations of the FFDM system for calculating VBD (D). Furthermore, the original set of physical phantom images were used directly for validating the ANN calibrations (E). ANN VBD was calculated for a set of cancer and non-

cancer patient's FFDM images (F). Statistical analysis was performed five-fold: by reviewing performance on two sets of validation phantom data, intra-patient, against BI-RADS breast composition scoring, against MRI VBD, and inter-patient (G)..... 43

Figure 3.5: Phantom validation of ANN-VBD. Linear regression results comparing ANN-estimated VBD against actual VBD (derived with virtual phantom in A, known with physical phantom in B). In considering random combinations of imaging parameters within expected ranges seen clinically, virtual phantom data derivation produced VBD values beyond 0 and 100% VBD (A). Comparison against physical phantom data shown as a boxplot since values quantized at 0, 50, and 100% VBD (B). Plots are inlayed with squared Pearson's correlation coefficients and root-mean-square errors. 46

Figure 3.6: Clinical images and ANN-VBD maps. Example images of low- (A), mid- (B), and high-VBD (C) breasts. In each panel, the vendor post-processed image used by radiologists for diagnostic reading is shown at left (not used in study, but shown here to illustrate images used clinically and commonly used to estimate mammographic density) and the ANN-calculated VBD map is shown at right. 48

Figure 3.7: Intra-patient validation of ANN-VBD. Linear regression plots comparing ANN-estimated VBD measures of the left (y-axis) and right (x-axis) breast of non-cancer patients. Plots are inlayed with squared Pearson's correlation coefficients, root-mean-square errors, and fit equations for comparisons of VBD (A), DV (B), and TBV (C). 48

Figure 3.8: Comparison of BI-RADS breast tissue composition scoring and ANN VBD. Tukey's multiple comparisons test reveal VBD in all categories as significantly different from each other, except between 'a' and 'b' as well as 'a' and 'c' (circles not overlapping at right, p-values detailed in text)..... 49

Figure 3.9: Linear regression plots comparing ANN-estimated measures of VBD (y-axis) with those measured on MRI (x-axis); total breast volume (A), dense volume (B), and volumetric breast

density (C). Plots are inlayed with squared Pearson's correlation coefficients, root-mean-square errors, and fit equations for comparisons. 49

Figure 3.10: Inter-patient validation of ANN-VBD. Comparison plots of ANN-estimated measures of VBD with significant difference against classical risk factors of breast cancer; age (A), BMI (B, D, E), and menopause status (C). Though not reaching statistical significance, VBD compared with parity status (F) is also shown. Linear regression plots are inlayed with squared Pearson's correlation coefficients and fit equations. Boxplots are inlayed with X^2 test statistics. 50

Figure 4.1: Example of tissue segmentation performed of all cancer patients' affected breast images. At top left (a), a dynamic contrast-enhanced MRI exam at t3 is seen in the axial plane, illustrating one slice of the view used for contouring the breast and tumor. At top right (b), the result of breast segmentation is shown. At bottom left (c), the segmented tumor is highlighted in blue. Finally at bottom right (d), the parenchyma segmented at t1 is highlighted in pink. Breast subcompartment segmentation was performed in 3-dimensions. 61

Figure 4.2: Summary of radiomic analysis performed in this study. Clinical features were evaluated by a radiologist according to Breast Imaging Reporting and Data System directly from dynamic contrast-enhanced MRI (a). 3-Dimensional tumor (red) and parenchyma (light blue) compartments were segmented (b), from which volumetric breast density was immediately estimated (c). Enhancement maps were then generated (d), from which textural features of tissue compartments were extracted and defined as enhancement heterogeneity (e). Subsequently, two analyses were conducted using extracted features: supervised learning of breast cancer subtype was performed with a support vector machine classifier (f) and unsupervised learning of background parenchymal enhancement feature expression pattern was performed with *k*-means clustering (g). 64

Figure 4.3: Box plots illustrating differences in distributions (quartiles as red boxes, grand mean indicated as spanning line) of the three most predictive quantitative features found in differentiation tasks: the lesion's 'mass size' feature (a), parenchyma's 'skewness of Signal Enhancement Ratio' feature (b), and parenchyma's 'standard deviation of rate in' feature (c) compared between the

triple-negative (TN) and non-TN groups. p-values were calculated by Wilcoxon Mann-Whitney tests. 68

Figure 4.4: Examples of 'parenchyma rate in' parameter (also Figure 4.3b) maps from a non-triple-negative (non-TN) patient (left) and a TN patient (right) illustrating the difference of a statistical texture feature between members of the two groups in image form. Slices of the 'parenchyma rate in' parameter map void of tumor tissue are presented in the sagittal plane. It is evident the variation of this background parenchymal enhancement texture feature's value is greater in TN cancers, where standard deviation is markedly higher at 352.9 as opposed to 133.8 in the non-TN patient. 69

Figure 4.5: Unsupervised *k*-means clustering of breast cancer patients (n=88) on the x-axis and quantitative background parenchymal enhancement (BPE) feature expression (n=39) on y-axis (as z-scores, with scale at bottom left. std=standard deviation). Correspondence of patient groups with similar radiomic expression patterns can be seen where the majority of triple-negative (TN) breast cancers have grouped together in the left cluster (9 of 11 TN in partition highlighted orange at top left) due to association of the BPE heterogeneity feature signatures. 1st order statistical texture features are highlighted as purple and similarly 2nd order statistical texture features are green at right indicating correspondence of feature groups with clustered expression patterns. 70

List of Abbreviations

%FGV	Percent fibroglandular volume (same as VBD)
2D	2-dimensional
3D	3-dimensional
ACR	American College of Radiology
ANN	Artificial neural networks
AUC	Area under the receiver operating characteristic curve
BI-RADS	Breast Imaging Reporting and Data System
BMI	Body Mass Index
BPE	Background parenchymal enhancement
CAD	Computer-Aided Diagnosis
CT	Computated Tomography
DCE-MRI	Dynamic Contrast-Enhanced MRI
DV	Absolute dense volume (same as FGV)
ER	Estrogen receptor
FCM	Fuzzy C-means
FDA	United States Food and Drug Administration
FFDM	Full-Field Digital Mammography
FGV	Absolute fibroglandular volume (same as DV)
GE	General Electric Medical Systems
HER2	Human epidermal growth factor 2 receptor
IQR	Inter-quartile range
kV	Kilovoltage
LCC	Left Craniocaudal
LumA	Luminal A
LumB	Luminal B

mAs	Milliampere-second
MRI	Magnetic Resonance Imaging
PE	Percent enhancement
PR	Progesterone receptor
R^2	Square of the Pearson correlation coefficient
RMSE	Root-mean-square errors
SAS	Statistical Analysis System Institute
SER	Signal enhancement ratio
std	Standard deviation
SVM	Support vector machine
SXA	Single-energy X-ray Absorptiometry
TBV	Total breast volume
TN	Triple-negative
UCSF	University of California, San Francisco
VBD	Volumetric breast density (same as %FGV)
WEKA	Waikato Environment for Knowledge Analysis

Chapter 1 Introduction

1.1 Background

As surrogate endpoints for clinical tests, biomarkers have proved to be immensely valuable, making for their widespread use in medicine. The field of imaging biomarkers specifically, has exploded in parallel with technological advances of computers and imaging devices, and their software, all of which continues to see greater overlap with the field of bioinformatics. Once specific imaging biomarkers are shown to have value, they can easily be integrated into the digital environment of the radiologist, communicated to health care providers as well as patients, and their impact on health can be quickly realized. Such efforts are in the process of bringing radiology from the days of largely qualitative description and reporting into the quantitative era, leveraging informatics tools to extract information from imaging alone, or together with other medical record data.

1.1.1 Breast cancer

Currently, breast cancer is the most prevalent form of invasive cancer in women worldwide (1). While it is the first leading cause of cancer death with women in developing nations, it has become the second in developed nations, due to earlier detection using screening by imaging, increased awareness, and better treatments with more personalized approaches. Incidence and mortality rates also vary greatly between countries, though there appears to be a common trend of increasing cases worldwide in both less and more developed nations (2). Screening X-ray mammography remains the most common and effective tool for early detection of breast cancer (3–7). MR imaging has also been realized as a tremendously sensitive technology for the diagnosis of breast cancers (8). While mammography is able to capture morphological and textural properties of breast tissues, MRI not only captures anatomical properties and heterogeneity of whole tissues being

imaged, but is also able to extract functional properties of tissues, such as angiogenesis, expression patterns, as well as metabolism. Over recent years, the ability to stratify patients based on genetic, hormonal, and lifestyle markers has proven to be of great value toward personalized medicine (9,10). Incorporating quantitative biomarkers from routine radiographic imaging, though a relatively recent trend, is expected to enhance healthcare even further.

1.1.2 Tumor microenvironment

Recent research has shown that beyond a tumor itself, a tumor's microenvironment also contains information about the nature of the present cancer, potentially defining and regulating cancer progression. Specifically with the breast, it is being gradually recognized as the key contributor to cancer progression and aggressiveness (11–15). The quantity of fibrous and glandular (fibroglandular) parenchyma tissue and temporal variations of parenchyma enhancement in dynamic imaging of the breast itself has been shown to be a strong risk factor for the developing of cancer (16–21). Furthermore, the various breast cancer subtypes appear to exhibit different growth and angiogenesis patterns in parenchyma tissue (22). Despite considerable progress in medicine having been made with the use of imaging biomarkers, topics such as heterogeneity of tumors and their microenvironments remain a diagnostic and prognostic challenge. Effectively detecting and quantifying heterogeneity in breast cancer is important for treatment. Characterization of the heterogeneity of breast cancer using genetic, pathologic, and image analysis methods provides insight during the process of selecting an effective therapeutic plan and is thought to be able to reveal significant improvements. Therefore, to correlate imaging parameters of the parenchyma tissue, in addition to the tumor itself, with outcomes of breast cancer such as risk and molecular subtype, is being investigated heavily.

1.1.3 Radiomics and machine learning

Computer-Aided Diagnosis (CAD) systems have been in development for many years, making major contributions to the improvements of healthcare. Early successes of CAD were realized in breast cancer imaging, where systems aimed to detect the presence of a lesion or cancer (17,23). As digital medical imaging technologies continue to develop, high-throughput image processing and bioinformatics tools are able now to extract a multitude of quantitative features from them, producing high-dimensional data mineable for information which reflects underlying pathophysiological relationships. This natural extension of CAD now gaining much momentum, known as radiomics, is poised to offer a nearly limitless number of imaging biomarkers that could potentially aid not only cancer detection and diagnosis, but also further classification, prognosis of response to treatment or other disease outcomes, and monitoring disease status (24–26). The ability to mine features from images that may not be visually discernible by clinicians offers the opportunity for better quantitative modeling of disease appearance and hence possibly improve predictions of disease outcome. Given the complexity and heterogeneity of cancer, as well as high-quality radiomics datasets of imaging features extracted using an ever growing amount of new methods, exploratory methods able to mine and validate useful biomarkers from the sea of available features are required. This process can take advantage of new statistical and machine learning approaches (27–31). Though many of the ideas have been around for decades, recent progress in this area has been driven by data availability and computational scale. Algorithms such as clustering, neural networks, and support vector machines prove tremendously powerful and efficient in elucidating the relationships in data, aiding the process of its interpretation and understanding.

To understand key characteristics of breast parenchyma tissue as potential cancer biomarkers has become clinically meaningful. Imaging biomarkers that are accurate and reproducible are becoming essential, and so to investigate and facilitate new approaches

to their discovery and inclusion for closing the gap between clinical practice and imaging research becomes more and more important.

1.2 Purpose

The aim of the studies integrated in this thesis is to address two areas concerning precision medicine: the identification and validation of imaging biomarkers useful in predicting outcomes of breast cancer, specifically those derived from the parenchyma breast tissue microenvironment, and in parallel investigate radiomics and machine learning methods in the discovery and measurement of imaging surrogates of breast cancer outcomes.

In Chapter 2, the agreement between three novel automated Full-Field Digital Mammography (FFDM) techniques for measuring breast parenchyma tissue volume quantitatively as breast density with one derived from MRI are investigated. The current standard of manually scoring mammographic breast tissue density is highly subjective and expensive. Automated methods to quantify the metric objectively exist, but whether these techniques accurately measure the quantity of breast parenchyma tissue is not well known. The accuracy of several of these measures is elucidated against MRI as a standard of volumetric anatomical imaging.

In Chapter 3, a fully automated machine learning algorithm is investigated for its potential of calibrating FFDM for quantitative analysis of breast parenchyma tissue. Artificial Neural Network modeling is used to estimate measures of volumetric breast density (VBD), which are validated against phantom data, compared intra-patient, with a qualitative scoring standard, with MRI VBD measures, and inter-patient with classical risk factors of breast cancer as well as cancer status.

In Chapter 4, the added discriminative value of characterizing breast parenchyma enhancement heterogeneity on Dynamic Contrast-Enhanced MRI (DCE-MRI), in addition

to the tumor itself, for identifying “triple-negative” breast cancers is investigated using radiomics and machine learning approaches. Heterogeneity of background parenchymal enhancement, measured quantitatively by textural imaging features, is used in parallel with tumor characteristics to create differentiation models to classify breast cancer subtype.

Chapter 5 summarizes the conclusions made in chapters 2-4 and the specific points found in the study.

Chapter 2 Agreement of Mammographic Measures of Volumetric Breast Density to MRI

2.1 Introduction

Other than age and specific genetic mutations, dense breast tissue is one of the strongest risk factors to predict who will develop breast cancer (32,33). Like body mass index, it is one of the few modifiable breast cancer risk factors. Breast density also impacts the sensitivity of mammography. Sensitivity decreases from over 90% in women with low breast density to less than 60% for women with high breast density (34,35). In addition, monitoring for a density reduction in women taking risk reduction therapy can in some cases be an effective method to monitor whether a woman is responding to therapy (36–41). Several states in the United States, including California, Texas, and Connecticut, have recognized the influence of breast density on cancer detection and require the reporting of breast density to women with dense breasts as part of their mammography examination. However, how to report breast density is under debate.

Mammographic breast density in clinical practice is assessed using a 4-category score defined in the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) (42). Several risk models have been developed using BI-RADS density including the Breast Cancer Surveillance Consortium's 1-year and 5-year models, developed using over 1 million women (43,44). BI-RADS density categories, however, have limitations. Agreement between radiologists is only moderate, with poor agreement between radiologists in the middle two density categories (45). Second, BI-RADS categories are too coarse to monitor breast density changes in individual women on prevention therapy: approximately 6-9% over 2 years for Tamoxifen (36,41,46).

Continuous and objective measures of breast density, such as percentage areal (47) and volumetric (19) mammographic density, were developed to improve on the semi-

quantitative and subjective nature of scoring. Percent areal density has been shown to have a slightly stronger risk association with breast cancer than categorical scores (48), but does not represent the true volume of dense tissue, could have errors associated with its two-dimensional projection, and requires a trained reader (49).

In vivo tissue volume and mass measures can be made from projection x-ray images. One example measures fat and lean mass using dual-energy x-ray absorptiometry (50,51). Several mammographic measures of absolute volumetric breast density have also been reported (52–55), but agreement between different methods of quantifying breast density is unclear. Three-dimensional imaging methods, such as MRI or CT, can be used to examine agreement with volumetric measures derived from projection x-ray systems. In 3D MRI and CT images, adipose and fibroglandular tissues do not overlap because the organ is imaged from several angles and reconstructed in 3-dimensions. Such images can therefore be easily segmented to quantify breast tissue compartments volumetrically (56,57).

In this study, we compared the agreement of three measures of automatic volumetric mammographic breast density to MRI breast density to determine the accuracy of volumetric breast density from mammography.

2.2 Methods

The study design was a retrospective analysis to compare breast density measured from screening MRI exams to that from screening mammograms on a population of women referred for both. Four methods were used to assess volumetric breast density: a fuzzy-clustering segmentation method on MRI (58), the Single-energy X-ray Absorptiometry (SXA) method (59), the Quantra method (Hologic, Inc., Bedford, MA, USA), and the Volpara method (Mātakina, Wellington, New Zealand) on Full-Field Digital Mammography (FFDM) images (60). The study was fully HIPAA compliant and approved by the University

of California, San Francisco (UCSF) Institutional Review Board for passive consenting processes or a waiver of consent to enroll participants, link data, and perform analyses for research purposes. A Federal Certificate of Confidentiality also protects the identities of research subjects.

2.2.1 Subjects

Women aged 18 years or older undergoing screening mammography and screening MRI between 2007 and 2010 at UCSF were included. Sample size was not estimated before beginning the study. For a woman to have been included in the study, she must have had a set of screening digital mammograms and a screening MRI exam acquired within 1 year of each other, have completed a breast health questionnaire, had no previous history of breast cancer or breast surgery, and have had a BI-RADS assessment of either 1 or 2 (negative or benign finding, respectively). Only images of the left laterality were used for the entire study, of craniocaudal (LCC) views for mammography. Each subject contributed one LCC mammogram matched to a left breast MRI. When multiple mammography examinations were available within the study period, that closest in date to the MRI exam was used.

2.2.2 MR imaging and breast density analysis

T1-weighted non-contrast fat-saturated images were acquired on either a 1.5 or 3 Tesla GE system (General Electric Medical Systems, Milwaukee, WI) using a bilateral phased-array breast coil (Medical Devices, Madison, WI) with women lying in a prone position. Slice thicknesses across subject were consistently 2mm, though in-plane spatial resolutions varied with breast size, averaging approximately 0.7 mm x 0.7 mm. The images were analyzed using a quantitative fuzzy C-means (FCM) technique previously described (58). Percent fibroglandular volume (%FGV) was calculated as the ratio of

clusters of high intensity voxels determined to be fibroglandular volume (FGV) to total breast volume (TBV). The reader reviewed all slices to insure accuracy of the segmentation of fibroglandular tissue.

2.2.3 Mammographic imaging

All mammograms were acquired on one of six Hologic Selenia FFDM systems at UCSF. These systems used a molybdenum anode x-ray tube and have a pixel spatial resolution of 70 µm x 70 µm. The raw ("For Processing") format images were archived, from which all analyses for mammographic techniques were obtained. The images available to the study were collected as part of a large cohort study and previously downsized by 50% in both dimensions before analysis to conserve server storage space, creating 140 µm x 140 µm averaged pixels. These downsized images were used directly for SXA analysis and upsized to original dimensions for Quantra and Volpara analyses. A test sample of images were compared using all three mammographic density techniques before and after downsizing and no significant differences were found in breast volume, FGV, and %FGV measures.

2.2.4 SXA breast density analysis

SXA is an established method for measuring breast density, which previously has been described (19,59). In brief, the technique compares the breast image pixel grayscale values to that of a reference phantom that is imaged with the breast. Additional weekly quality control scans were acquired using a phantom named GEN III. GEN III was imaged in the location normally occupied by a breast and was constructed with three tissue-equivalent density materials at three thicknesses as well as other features to test the mammography system's geometric accuracy. Differences between the SXA calibration and the GEN III measures were used to update the SXA calibration continually. This study

reports results using version 7.1 of the SXA analysis software package (UCSF, San Francisco, CA).

2.2.5 Quantra and Volpara volumetric assessment

Quantra and Volpara are FDA approved, commercially available, and fully-automated software for estimating volumetric breast density. Quantra has been previously described by Harman et al (60) and Volpara by Aitken et al (61). Version 3.2 of the Quantra Algorithm (Cenova 1.3) and version 1.4.3 of the Volpara Algorithm (Imaging Software 1.5.7) were used in this study. In both, FGV is found by referencing each pixel's attenuation to the attenuation of pixels that are labeled as exclusively adipose (i.e. the lowest attenuation pixels). The estimated FGV is then divided by the TBV to calculate the %FGV of the breast.

There are three primary differences between the two commercial mammographic methods (Quantra and Volpara), FCM technique for MRI, and SXA algorithms regarding how they define FGV and %FGV:

1. Reference Definition: Quantra, Volpara, and the FCM use relative references of adipose tissue attenuation (Quantra, Volpara) or signal intensity (FCM), defining the area with the lowest value in each woman's image as pure adipose, while SXA uses an in-image phantom with fixed fibroglandular density and fat references. In addition, SXA defines the total lack of dense breast tissue (i.e. 0 %FGV) as pure fat (versus adipose, containing both fat and water). If the fractions of fat and water in adipose tissue were the same in all women, then SXA would still differ from the FCM and the commercial systems' calibrations due to the water volume in adipose. The SXA algorithm does this because phantom references for breast adipose have been found to be too dense (62,63) and because a fixed reference is always available even in dense breasts without many available pixels of pure adipose tissue.

2. Tissue Compartment Model: Quantra and Volpara use a 3-compartment model of skin, fibroglandular, and adipose tissue (60), while SXA and FCM utilize 2-compartment models; fat and fibroglandular for SXA and fat and water for FCM.
3. Pixel/Voxel Subdivision: SXA, Quantra, and Volpara make no assumptions with regards to labeling pixels as either all adipose or fibroglandular, but ultimately subdivide each pixel into some fraction of fibroglandular and a second compartment (fat for SXA and adipose+skin for Quantra and Volpara). In contrast FCM groups each voxel to either a fat or fibroglandular tissue cluster.

A modified SXA model, SXA without adipose water volume, was derived to assess compatibility of the MRI and SXA models further by approximating the hydration of adipose to be 15% water. Adipose volume was defined as TBV less the FGV. Previous work using water saturated MRI has estimated the adipose water volume as 8% (64) to 20% (65) of the total adipose volume. After normalizing for any differences in measured TBV between the MRI and SXA, the adipose water volume was subtracted from the SXA FGV.

2.2.6 Statistical analysis

Descriptive statistics of the subjects' age, weight, and BMI were used to summarize the population. The frequencies of first degree family history of breast cancer and BI-RADS density categorization were also calculated. Descriptive statistics of the mammographic and MRI density measures were used to summarize the breast density data. Medians and inter-quartile ranges were used instead of means and standard deviations with the density measures, as their distributions were not normal. Wilcoxon signed-rank tests were used to determine whether mammographic measures' were significantly different than those of MRI.

Linear regressions were performed to model the relationships between mammographic and MRI breast density measures of FGV, log FGV, and %FGV, using

difference in time between exams as a possible covariate. The natural logarithm of FGV was used to normalize its distribution and to be comparable to our previous reporting of these measures, where the log transformation was found amongst a variety of scaling factors to best normalize and improve breast cancer risk classification ability of the FGV measure (19). Squared Pearson's correlation coefficients were calculated from the relationships of density measures between methods. Differences between regression equation fit parameters were tested with paired Student's t-tests for significance. Root-mean-square errors (RMSE) were also calculated to aggregate the magnitude of individual differences between mammographically and MRI derived measures.

Quartile groupings based on distributions of %FGV and log FGV measures were compared using weighted kappa statistics to determine clinical agreement between measures. We interpreted the kappa statistics with the following categories (66): slight agreement (less than or equal to 0.20), fair agreement (0.21 to 0.40), moderate agreement (0.41 to 0.60), substantial agreement (0.61 to 0.80), and almost perfect agreement (0.81 to 1.00).

All statistical analyses were performed using SAS Version 9.3 software (SAS Institute, Cary, NC) and p-values of less than 0.05 interpreted as significant.

2.3 Results

Ninety-nine women met selection criteria. A high proportion of women had a family history of breast cancer and most had normal BMI (Table 2.1). The median breast density and quartile ranges are shown for each density measure in **Table 2.**2 and **Table 2.**3 respectively. Linear regression plots with regression coefficients comparing the mammography techniques to MRI for %FGV, log FGV, and TBV are shown in Figure 2.. All mammographic measures of TBV were highly correlated to MRI TBV (R^2=0.91 for all three). Volpara showed a higher correlation to MRI for log FGV while SXA showed the

highest correlation to MRI for %FGV. All best fit regression lines had significant intercepts (mammographic value not 0 when MRI value is 0), except that of Quantra vs MRI TBV. In general, the mammographically-derived values tracked MRI values in a similar way for log FGV and TBV (SXA to MRI slopes of 0.65, and 0.95, Quantra to MRI slopes of 0.65 and 0.92, and Volpara to MRI slopes of 0.68 and 1.06 respectively). However, the slopes were substantially different for %FGV to MRI (0.91, 0.33 and 0.35 for SXA, Quantra, and Volpara respectively). The time difference between the mammogram and MRI acquisition was not significantly associated with any measures except %FGV for SXA and FGV for Quantra. Including time between MRI and mammography measures explained less than 3% of the variance (not shown). RMSE results are also shown in Figure 2.1, ranging from 4.4% to 10% for %FGV, 0.37 to 0.42 for FGV, and 108 ml to 121 ml for TBV.

Table 2.1. Description of population characteristics

Demographic Variables	**Mean (std)**	**Min**	**Max**
Age (years)	47.2 (12.1)	26	79
Weight (pounds)	139.1 (24.1)	85	260
BMI (kg/m^2)	23.1 (3.9)	14.6	46.1
Frequency Variables	**Number**	**Percent of population**	
Family history (1st degree)			
No	21	21.2	
Yes	78	77.7	
BI-RADS density category			
1	12	12.1	
2	35	35.4	
3	30	30.3	
4	22	22.2	
Time difference between exams (mammography MRI)			
- 1 to +1 years	99	100	

BMI=Body Mass Index, std=standard deviation.

Table 2.2. Density measures for study participants by method (n=99)

Density Measures	Median (IQR)	Min	Max
Percent Fibroglandular Volume (%)			
SXA	36.7* (38.8)	6.1	85.8
Quantra	22.0* (14.0)	9.0	49.0
Volpara	13.3* (12.6)	2.6	37.3
MRI	24.0 (36.0)	2.0	85.0
Absolute Fibroglandular Volume (ml)			
SXA	192.6* (160.5)	31.1	533.8
Quantra	101.0* (102.0)	9.0	425.0
Volpara	64.8* (61.6)	13.0	278.2
MRI	102.3 (120.4)	14.3	338.1
Total Breast Volume (ml)			
SXA	514.0* (383.3)	69.4	1882.8
Quantra	441.0* (379.0)	51.0	1870.0
Volpara	558.3* (452.3)	76.2	2178.3
MRI	460.4 (412.1)	49.9	1828.4

Table 2.3. Quartile density ranges by method (n=99)

Percent Fibroglandular Density Quartiles				
	Q1	**Q2**	**Q3**	**Q4**
SXA percent fibroglandular volume (%)	6.1-21.4	21.4-36.7	36.7-60.2	60.2-85.8
Quantra percent fibroglandular volume (%)	9.0-15.0	15.0-22.0	22.0-29.0	29.0-49.0
Volpara percent figroglandular volume (%)	2.6-7.9	7.9-13.3	13.3-20.5	20.5-37.3
MRI percent fibroglandular volume (%)	2.0-13.0	13.0-24.0	24.0-49.0	49.0-85.0

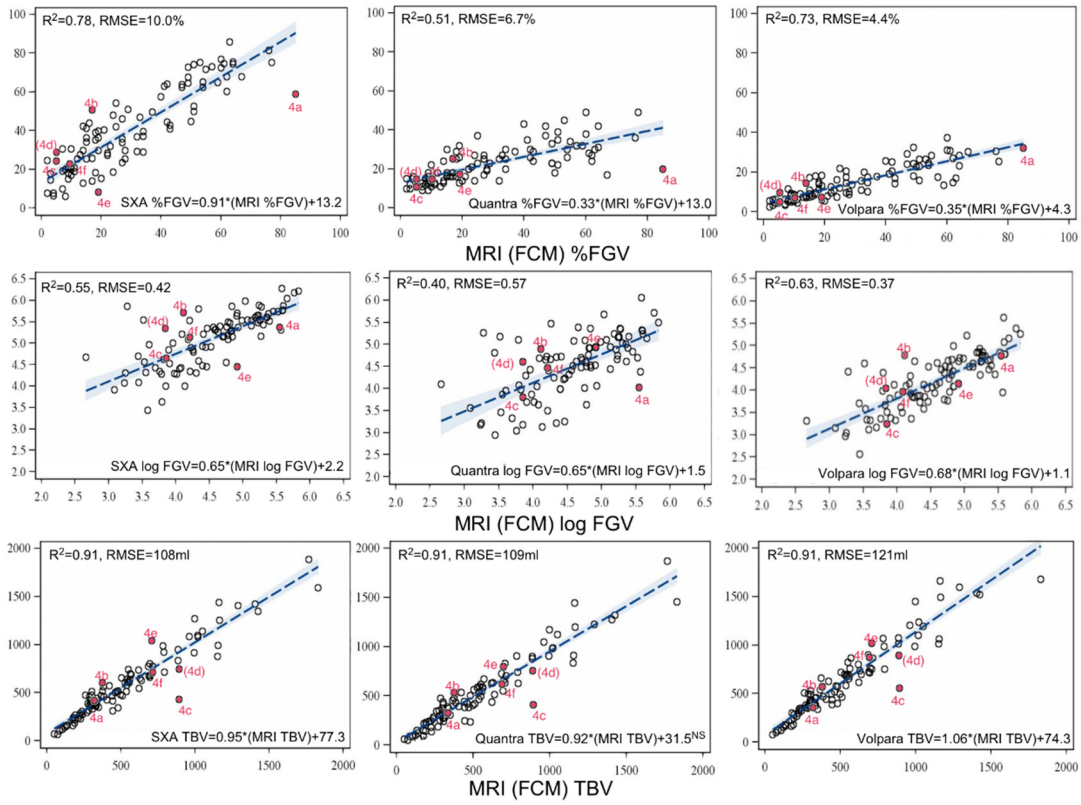

Figure 2.1: Best linear regression fit line with 95% confidence interval bands for percentage fibroglandular density (top), log fibroglandular volume (middle), and total breast volume (bottom) for MRI versus either SXA (left), Quantra (center), or Volpara (right) measures. Solid points correspond to example images in Figure 2.4.

Distribution quartile groups were compared for MRI, SXA, Quantra, and Volpara (Figure 2.2). Overall, the percentage agreements within each %FGV quartile were similar for each mammographic measure to MRI. SXA %FGV had the highest percentage of agreement to MRI. The mammographic methods placed 60% (Quantra) to 85% (SXA) of the same women in the highest density quartile as MRI. A lower agreement was seen for the middle two quartiles of %FGV, with agreement ranging from 46% to 55%. Similar trends were seen for log FGV. Disagreement of 2 quartiles for density was less common but did exist for Quantra %FGV measures and log FGV for all three mammographic methods. The weighted kappa statistics are also shown in Figure 2.2 for each relationship in the respective plot. Substantial agreement was found between all density comparisons with the highest kappa coefficient being k=0.72 for MRI versus %FGV. Moderate

agreement was found between all log FGV measures, where weighted kappa scores ranged from k=0.47 (Quantra vs. MRI) to 0.64 (Volpara vs. MRI). The kappa coefficients for %FGV between mammographic measures was also moderate from 0.67 to 0.74.

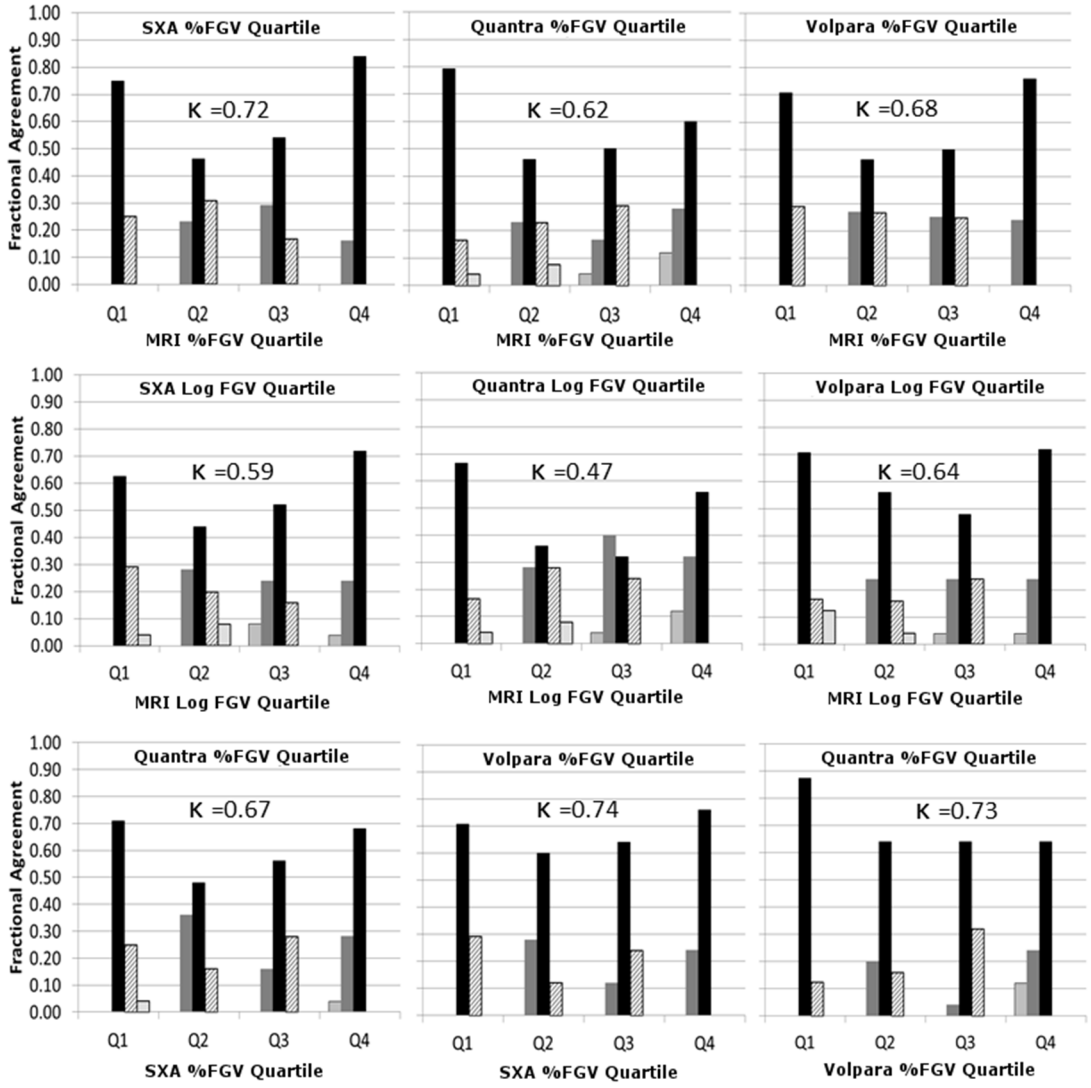

Figure 2.2: Comparison of quartiles classification for percent fibroglandular density (top) and log fibroglandular volume (middle) for MRI versus SXA (left), Quantra (center), and Volpara (right). The bottom row of plots show quartiles comparisons between mammographic density measures. Legend at right defines categories of agreement, where either the two compared method's agree completely (black) or are off by one or two quartiles up or down in comparison with the other method.

In the modified SXA model, where adipose water volume was subtracted from the measures (Figure 2.3), the best fit slope for log FGV improved from 0.65 to 1.012 and the intercept term decreased from 2.2 to 0.27 (not significant). The best fit slope for SXA %FGV changed from 0.91 to 1.015, and the intercept term decreased from 13.2 to -2.2 (not significant). Agreement between SXA and MRI %FGV improved slightly from $\kappa = 0.72$ to 0.74 (not shown).

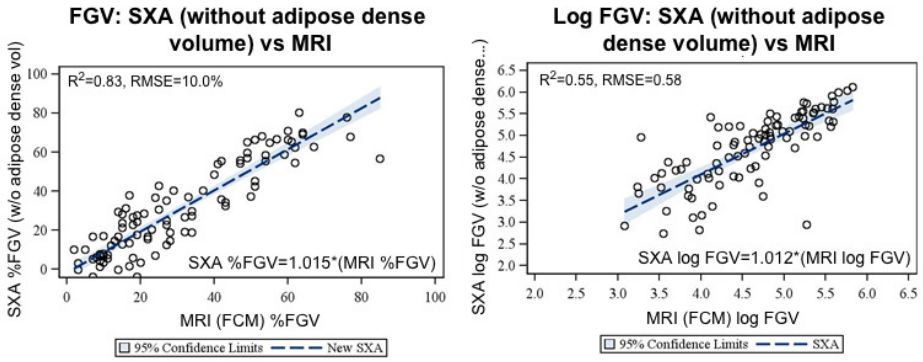

Figure 2.3: Validation of SXA model using breast biology and adipose volume estimates from MRI. The amount of water volume in the MRI adipose volume was estimated to be 15% of the volume, which is consistent with previous work estimating it to be between 8% (64) and 20% (65). The MRI model does not include adipose density in the fibroglandular volume while SXA does. Subtracting out the adipose water volume from the SXA fibroglandular volume improved the agreement between SXA and MRI from R^2=0.78 to 0.83 and removed most of the bias between the measures.

2.4 Discussion

We compared three mammographic techniques to an MRI technique for quantifying volumetric breast density. TBV, as measured by MRI and mammography techniques were well correlated with regression slopes ranging from 0.92 to 1.06 times that of MRI. One may expect that the MRI TBV to be higher because MRI has access to delineate around the pectorals muscle. However, we found that average TBV was higher for SXA and Volpara but lower for Quantra when compared to MRI. We found most cases of

disagreement were driven by differences in TBV measured for MRI versus mammography. Figure 2.4 shows comparative images selected from the results in Figure 2.1. From Figure 2.1, the RMSE between SXA and MRI TBV was 108 ml but 65 ml for FGV (not shown), indicating that the observed lack of agreement between mammographic density and MRI density is most likely driven by differences in the total volume. Figure 2.4a is an example where the difference in TBV caused a large difference in density due to glandular density that extended to the chest wall (i.e. a lack of retroglandular adipose). However, differences seen in Figure 2.4b seem to be related to the breast having higher attenuation throughout, including the adipose that impacted the mammographic measures. The mammographic images in Figures 2.4c and 2.4d were acquired from the same woman on the same day (left, MR images on right are the same) and are examples of how measured TBV can vary substantially due to breast positioning. The mammogram in Figure 2.4c was a reimage of that in 4d to ensure a good nipple profile in the image. Only that in Figure 2.4c was used in our study results since we chose the last mammogram acquired for each visit. Figure 2.4d and its calculated measures plotted in Figure 2.1 were not actually part of the study's quantitative analyses, but only included to illustrate one reason for discrepancy.

Comparing the MR images of Figures 2.4c to 2.4e, there were substantial differences in how the TBV was delineated from truncal subcutaneous in the retroglandular region. In Figure 2.4e the TBV is substantially less than the mammographically-defined measure.

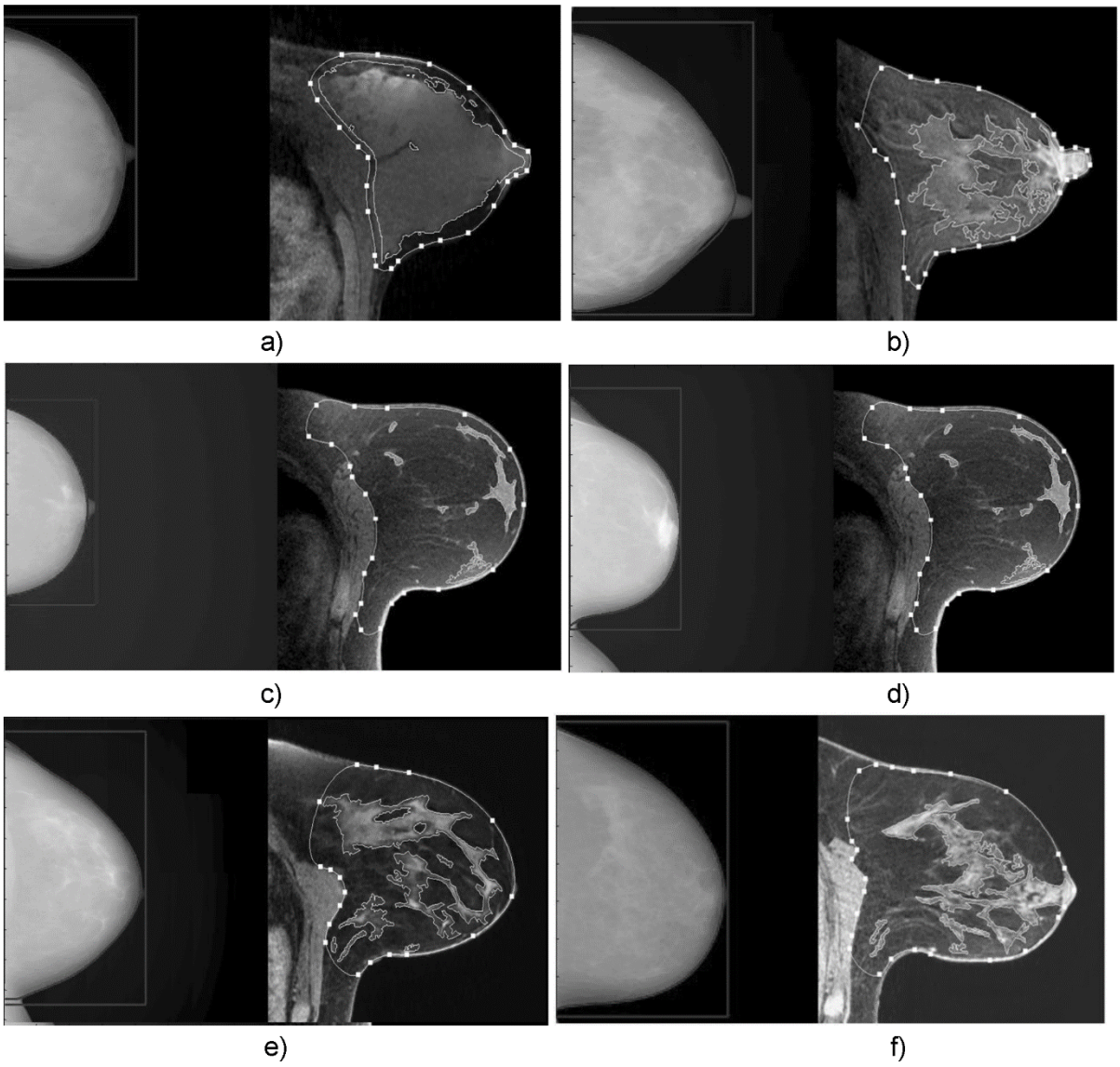

Figure 2.4: Six comparisons of the LCC mammograms to their respective left central breast axial-slice MR images on five different women (c and d are the same woman). The white line connecting points in the MR images define the total breast volume. The MRI fibroglandular volume is shown delineated with white lines without points. Solid data points 2.4a-2.4f in Figure 2.1 correspond to the image labels a-f. Compared to the mammographically-derived SXA values, a) MRI percent density is higher, b) MRI percent density is lower, c) MRI breast volume is higher, d) of the same woman as c (this mammogram not part of analyses, only here and measures plotted in Figure 2.1 to illustrate one reason for discrepancy between methods' results), MRI breast volume is better segmented due to the breast being extended more into the mammographic image field, e) MRI breast volume is lower, f) all MRI measures of density and volume were in substantial agreement.

As evidenced in Figure 2.4, the differences in TBV were mainly due the difficulties in delineating the breast from truncal subcutaneous adipose (for MRI) and variations in breast positioning (for mammography). TBV errors, either by incomplete breast imaging in craniocaudal views or ambiguity in delineating between breast and truncal adipose, seem to be the limiting factor on both accuracy and precision for volumetric breast density. It is unclear if screening mammography mediolateral oblique views would be any better in this regard as this was not tested.

The primary differences between the three mammography techniques and MRI were in the type of references used for defining fibroglandular and adipose tissue. The mammographic techniques do not segregate fibroglandular from adipose tissue while MRI does ultimately label each voxel as one or the other. There were differences in the best fit RMSE values to MRI especially for %FGV and FGV. It was not possible to directly test for explanations. Kallenberg et al. (67) found that paddle tilt correction improved the agreement between both percent and absolute FGV of their mammography measures to MRI. The SXA method also attempts to accurately assess breast thickness variations due to compression paddle tilt and warp, and uses a mammogram-specific phantom for a tissue density reference. The Volpara and Quantra methods are proprietary with respect to corrections they may make regarding paddle tilt. Because the SXA model references to fat and fibroglandular tissue, we expected SXA to have higher %FGV and FGV than MRI, Volpara, or Quantra since the water volume of adipose is included in the SXA FGV compartment. We found that eliminating the adipose water volume from the FGV slightly improved the MRI and SXA agreement.

There are several previous comparisons of MRI to mammographic measures of volumetric breast density. The most methodologically similar study to the present study was smaller (n=32) and compared Quantra to MRI density, Kontos et al. (68). The MRI were analyzed using a similar fuzzy C-means segmentation algorithm. Like the present

study, TBV and density were found to be highly associated (R^2=0.71 and 0.80 respectively), with the average MRI TBV being higher and density being lower than Quantra. However, Kontos et al. found lower overall association of FGV than in our study (R^2=0.15). These finding are similar but not identical to our findings and we attribute the differences to the small size of their study and only three subjects with a density greater than 30%. In our study, 48 subjects had breast density higher than 30%. Van Engeland et al. (55) compared a proprietary mammographic volumetric breast density to MRI on 22 women. MRI breast density was measured using a manual segmentation technique and the authors reported a high correlation of R^2=0.94. In 26 young women, Highnam et al. (52) compared mammographic volumetric density using Volpara version 1.2.1 and found a correlation of R=0.94 but no further analysis was offered on how the MRI density was measured or on further statistical description. Thus, including our own study, there are at least four studies with four different mammographic volumetric breast density measures that show a high correlation to volumetric breast density by MRI.

It appears that not all volumetric measures of breast density, either by MRI or by mammography, are equivalent. This lack of equivalency may or may not impact their association with breast cancer risk. For example, in fully-adjusted models of 275 breast cancer cases and 825 controls, the SXA method has been shown to have a greater association to breast cancer risk than percentage mammographic breast density (19) where the fifth to first quintile odds ratios were 4.1 for SXA breast density and 2.5 for two-dimensional breast density. To date, there have not been reports of breast cancer associations for MRI, Volpara, or Quantra measures of volumetric breast density techniques.

Our study had the following limitations: First, most MR and mammography images were not acquired at the same visit. Images acquired on the same visit could have potentially eliminated some of the observed differences. Second, we did not compare the

associations to breast cancer risk across techniques. This will be done in a larger dataset now being collected.

2.5 Conclusions

We conclude that volumetric breast density measures of total breast volume, fibroglandular volume, and percent fibroglandular volume from screening digital mammograms calculated from the techniques used in this study are in moderate to substantial agreement with the volume measures derived from MRI. The SXA measure of density showed a higher association to MRI than Volpara or Quantra density measures. However, classification of women by volumetric density by any of the three mammographic techniques is comparable to classifications by MRI density.

Chapter 3 Automatic Estimation of Volumetric Breast Density Using Artificial Neural Network-Based Calibration of Full-Field Digital Mammography: Feasibility on Japanese Women with and without Breast Cancer

3.1 Introduction

Breast cancer is the most frequently diagnosed invasive cancer and the most common cause of cancer death among women globally (1). Though incidence and mortality rates vary greatly around the world, an increase of cases worldwide has been observed in both less and more developed countries (2).

In Japan, an increase of breast cancer incidence in recent years has been found to be the predominant contributor to the overall rise of cancer incidence in women. Despite a decrease in cancer mortality overall, breast cancer mortality in Japan has only recently seen a plateau after a gradual increase since the 1960s. A closer look with respect to age reveals that while this rate has been decreasing among women aged 30-54, this is not the case among older groups (69).

Recent improvements in breast cancer prognosis can be largely attributed to the ability to detect cancers earlier while tumors are small in size and have not spread to lymph nodes or further, aided by advances in technology and efforts toward personalization of medicine. Screening X-ray mammography remains the most common and effective method for early detection of breast cancer and several randomized and controlled trials worldwide (3–5) and meta-analyses/reviews of trials (6,7) have shown evidence of its benefits. Additionally, the ability to stratify women based on risk of developing cancer with genetic, hormonal, and lifestyle markers has proved to be of great value (9,10). Though breast cancer risk assessment models using multidisciplinary approaches have been in development for many years and are promising, there still appears to be much room for improvement in terms of positive predictive values of those

available thus far (70). Extracting and incorporating quantitative biomarkers from routine radiographic imaging, such as mammographic breast density, is a relatively recent trend seen likely to enhance existing risk models (43,44,71).

Breast density, as measured by mammography, is the amount and appearance of glandular and fibrous (more generally, fibroglandular) tissue in the breast. High breast density is a strong risk factor for developing breast cancer (18–20,48,72). It also has significant influence over radiologist screening sensitivity (73) as well as computer-aided diagnosis (CAD) sensitivity and specificity (74,75) due to dense tissue and cancer having similar appearances on mammograms. On the other hand, having low breast density has recently been shown as related to worse prognosis irrespective of patient age, BMI or menopausal status (76). There is also some evidence indicating that density may be a correlate of the rate of breast tissue aging (77,78). Since these associations have been shown in both pre- and post-menopausal women (48) and the measure itself is one of the few modifiable risk factors for breast cancer (79), it is currently a major focus for breast cancer imaging and many are undertaking its research.

Standard in the industry are breast density metrics typically estimated using categorical methods (42) or measuring the area occupied by dense tissue on a mammogram using threshold segmentation (80). Such methods tend to be subjective and can be cumbersome as well as costly since they usually need manual operation by experts with special training. As a product of this, these measurements may suffer limitations in distinguishing all dense parts of the breast, thereby not sufficiently reflecting tissue composition and their associations with cancer risk may be weakened. More recently, with the emergence and growth of digital mammography technology in parallel with spread of computerized algorithms for image interpretation and diagnosis (23), interest in measuring breast density in a fully automatic, quantitative, as well as volumetric manner has grown (19,52,81–83). Noticeably, due to the rise in information processing power and growing

amounts of data being created, machine learning algorithms have also seen an increased presence in such computer-aided analyses in breast imaging and radiology as a whole (17,27,83–85).

Artificial neural networks (ANN) are a family of statistical learning algorithms that simulate the structure of biological neural networks for approximating mathematical functions and are well suited for medical applications (86). In brief, they are made up of interlinked synthetic neurons that process information using a connectionist approach of computation and allow for machine learning due to their adaptive ability to change structure based on information processed (87). As a product of this, ANN are capable of learning complicated patterns from data and are able to provide accurate predictions despite vague or even missing data (88). The multilayer feedforward perceptron specifically, a popular ANN architecture, has been shown as capable of approximating any measurable function to any desired degree of accuracy (87).

As of this writing, 26 of the 50 states in the United States have enacted bills that mandate informing patients of their breast density. Several other states have proposed legislation of the same nature. Furthermore, federal legislation has been introduced that would require national reporting of whether or not the patient has dense tissue. Such efforts have also stirred up interest in other countries, including Japan (89–92). Despite existing work reporting on mammographic breast density worldwide however, it does not appear that an automatic method of measuring quantitative volumetric breast density (VBD) on Full-Field Digital Mammography (FFDM) has yet been accepted into routine practice. Moreover, to the best of our knowledge, works using ANN to model VBD do not exist so far.

It appears a breast cancer risk predictor with high discriminatory power that would enable an increase in efficiency, efficacy, and cost-effectiveness of breast screening programs is a global objective and the effort may be facilitated by considering quantitative

imaging biomarkers. The purpose of this study was to develop an innovative method to automatically quantify VBD from FFDM using a statistical machine learning approach. Specifically we modeled VBD phantom calibration data with a feed-forward multilayer ANN given its universal approximation ability and evaluated our method on phantom data, against a qualitative standard of scoring breast composition, against quantitative VBD estimated from MRI, and on a population of Japanese women with and without breast cancer.

3.2 Materials and Methods

3.2.1 Study population

This retrospective study was approved by the Institutional Review Board of the recruitment hospital and informed consent was waived. Mammograms acquired between February 2012 and June 2013 on one Amulet f FFDM system (Fujifilm Medical, Tokyo, Japan) in Hokkaido University Hospital (Sapporo, Japan) of 46 women, subsequently diagnosed with unilateral invasive carcinoma, were included in the study and their cranial-caudal (CC) images contralateral to the tumor were used in the analysis. MRI studies of these 46 women were also collected from the same period (3D T1-weighted images, Achieva 3.0T system, Philips Healthcare, Best, Netherlands). Additionally, screening CC mammograms acquired on the same system in June 2013 of 53 women with negative findings were also included in the analysis. Patient demographics are summarized in Table 3.1. In addition, age at menarche (12.8±1.5 years), post-menopausal status (n=33), and status as nulliparous (n=12) were available for women with cancer.

Table 3.1: Study population health demographics

Parameter	Cancer	Non-cancer	p-value, by *t or* X^2 Test
Patients, n	46	53	
BMI, kg/m²*	23.6 (3.9)	22.5 (4.1)	0.189
Age, years*	58.8 (11.0)	60.3 (11.9)	0.516
Age > 50, n	30 (65%)	40 (75%)	0.264

*Parameters reported as mean (standard deviation)

3.2.2 Breast density modeling

One "GEN III" VBD quality control (QC) phantom (University of California, San Francisco, CA, USA), consisting of radiographically breast tissue-equivalent materials at several fibroglandular densities and thicknesses (Computerized Imaging Reference Systems, Inc., Norfolk, VA, USA), served as the VBD reference for the ANN modeling performed in our study. In brief, the GEN III phantom was specially designed with features for calibrating parameters of FFDM systems to quantify VBD accurately and precisely (93). It includes 9 regions covering fat-equivalent 0% VBD, 50/50 water/fat-equivalent 50% VBD, and water-equivalent 100% VBD compositions at 2, 4, and 6 cm thicknesses (Figure 3.1), useful for our study's calibration purposes.

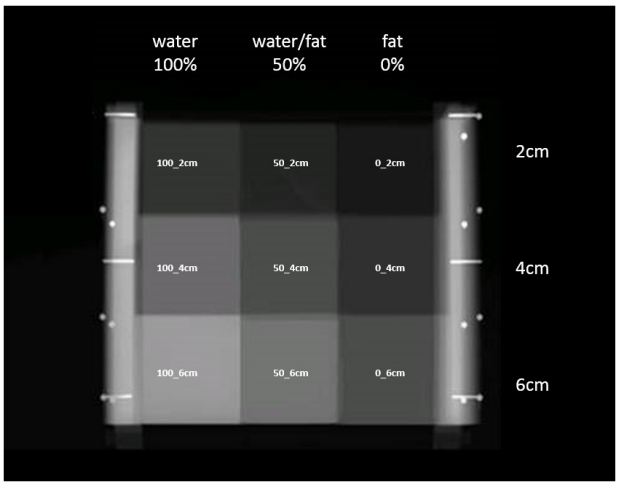

Figure 3.1: GEN III phantom image. Example FFDM image of GEN III phantom, with 9 breast tissue-equivalent regions used in the study as references annotated with their percent volumetric breast density (VBD, also at top) and thickness combinations (also at right).

A survey of all standard screening exams acquired on the FFDM system used in the study during June 2013 was performed to estimate expected ranges of clinical X-ray imaging parameters that affect appearance of mammograms. Table 3.2 briefly describes parameters included and their observed ranges: X-ray anode target and filter materials, tube voltage, current exposure time, relative image exposure sensitivity, and background detector signal. These five parameters, along with those two presented from the breast being imaged - tissue thickness and fibroglandular density, were considered primary determinants of signal intensity on FFDM for the purposes of our modeling. Next, repeated imaging of the GEN III phantom was performed across the ranges of imaging parameters seen in the survey of clinical images. In total, 300 images of the GEN III phantom were acquired. From this set of images, it would be possible to learn the relationship between the 7 imaging and breast parameters with signal intensity. Conversely, fibroglandular density (VBD) could be modeled from signal intensity and 6 other parameters (Table 3.2). Before going forward with the latter however, virtual phantom data was generated from the set of physical phantom data in order to train the ANN model with a more robust calibration set than just that of the GEN III's 9 tissue composition regions. The 9 VBD regions' values of each calibration image acquired were fitted with empirical functions given thicknesses and observed signal intensity. An example of this fitting can be seen in Figure 3.2 for one GEN III image. Each function was then used to compute sets of 100 interpolated points using random combinations of thickness and intensity values within expected limits given the particular image parameters used (300 images * 100 points = 30,000 virtual phantom data points total). Physical phantom data was ultimately not used directly to train the ANN model here, but instead used as one of two phantom datasets for validation of the ANN method as discussed below.

Table 3.2: Imaging and breast parameter determinants of signal intensity on FFDM considered for ANN-VBD calibrations

Parameter source	Parameter	Description	Ranges seen in survey of screening exams	
FFDM system	Anode target and filter materials	Combination of X-ray anode target and filter materials used (Molybdenum and Rhodium)	Mo/Mo	Mo/Rh
	Tube voltage	Voltage of X-ray tube (kV)	26-28	28-30
	Current exposure time	Time of exposure at X-ray tube current (mAs)	26-84	40-78
	Relative image exposure sensitivity	Image exposure sensitivity optimization factor, based on median value of image histogram	33-71	31-53
	Background detector signal	Background signal intensity at location without breast tissue	680-1023	470-1023
Breast imaged	Tissue thickness	Thickness of breast at pixel location (cm)		
	Fibroglandular density*	Volumetric density of breast at pixel location (%)		

*ANN calibration of FFDM was ultimately performed so as to output fibroglandular density (VBD), taking the remaining 7 parameters as input (signal intensity, in addition to the 6 other parameters listed)

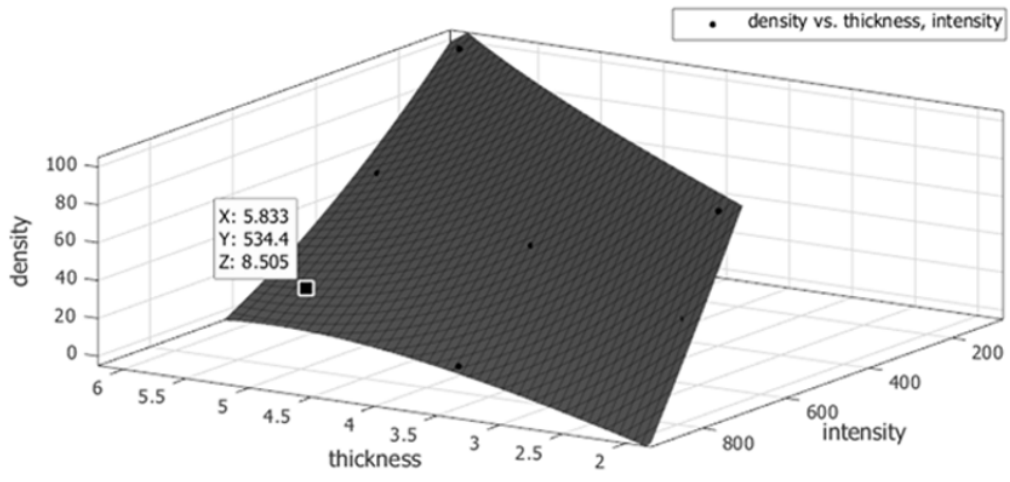

Figure 3.2: An example of one GEN III phantom image's 9 tissue composition regions plotted with its fitted function for Volumetric Breast Density (z-axis), given tissue thickness (x-axis) and signal intensity (y-axis). Each function was used to derive two separate sets of virtual phantom data using random combinations of thickness and intensity - the first for training the Artificial Neural Network calibrations and the second for its validation.

One 3-layer (1 input, 1 hidden, 1 output) ANN model was trained using all virtual calibration phantom data generated. The input layer consisted of 7 nodes (1 for each imaging and breast parameter considered, Table 3.2), the hidden layer consisted of 3 sigmoid nodes (approximately half the number of parameters) introducing non-linearity, and the output layer consisted of 1 un-thresholded linear node for numeric output (VBD). Training time was set to 500 epochs, with backpropagation weight updates set to a learning rate of 0.3 and a momentum of 0.2. ANN modeling and VBD estimation was performed using the WEKA 3.6.12 framework (University of Waikato, Hamilton, New Zealand).

3.2.3 Breast density analysis

First, any protrusion of nipple at the breast image's edge on the mammogram was removed by replacement with a smooth polynomial function fit to the surrounding edge. As only the small variant 18 x 24 cm compression paddle of the extremely rigid and non-tilting variety was in use at the study institution, the first approximation of patient breast thickness making contact with both top and bottom compression surfaces was obtained directly from the FFDM system's digital compression thickness readout (67,94). The compressed portion of breast volume was treated as an even thickness plane making uniform contact with both surfaces. Remaining breast periphery volume was defined as tissue which did not make contact with either compression surface. This periphery portion of the breast volume was modeled as a half-circle cross-section whose circumference made contact with the top compression paddle, the bottom detector platform, and the breast boundary on the image at every point along the breast edge. An example thickness map can be seen in Figure 3.3.

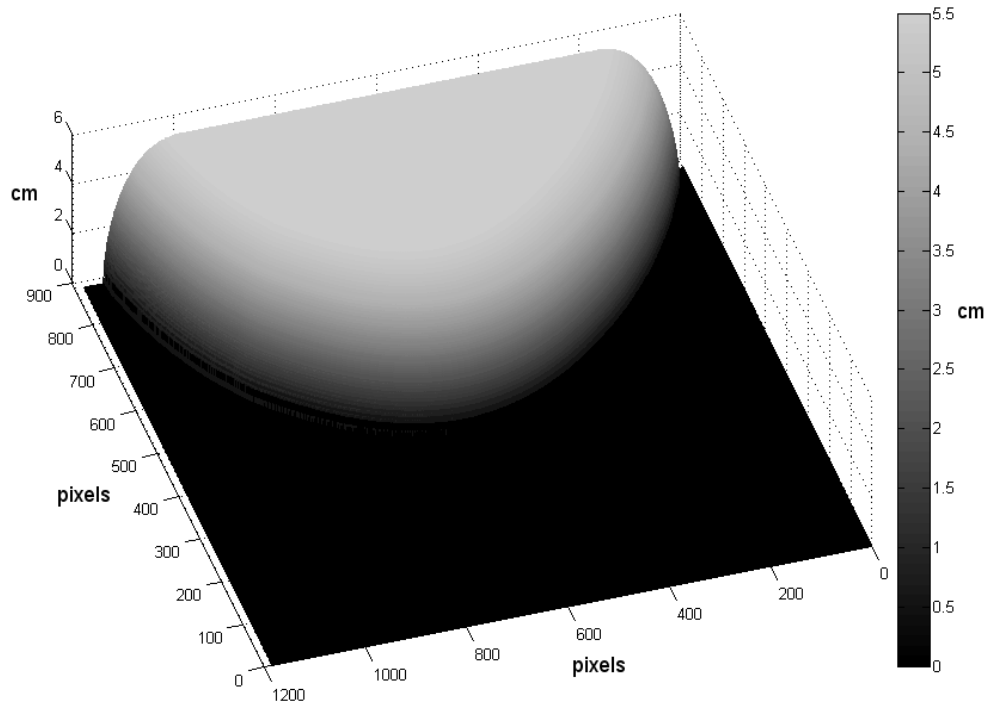

Figure 3.3: Breast thickness modeling. Example breast tissue thickness map, modeled using compression paddle thickness and breast edge as described in methods. Breast tissue thickness (in z-direction, per color bar) is projected atop the bottom detector platform (0cm, represented as black). The sum of thicknesses at all pixels make up total breast volume (TBV).

Total breast volume (TBV) was defined as the sum of volumes at each breast pixel (pixel area * breast thickness at pixel). Fibroglandular density (VBD) at each breast pixel was estimated with the ANN model learned using the pixel signal intensity, calculated breast thickness, and five imaging parameters described above read from the FFDM image meta-data (Table 3.2). Absolute dense volume (DV) of the entire breast was next calculated by totaling the product of breast thickness and VBD at each pixel of the breast. Subsequently, VBD of the entire breast was calculated as the ratio of DV to TBV. Figure 3.4 summarizes the study's ANN VBD modeling and analysis as described above. In addition to analysis of breast density by ANN, breast tissue composition scores were assigned to each mammogram of the patients using a standard for categorical breast

density in the medical community. According to the current American College of Radiology (ACR) Breast Imaging-Reporting and Data System (BI-RADS) protocol, mammograms can be classified into four categories: a (almost entirely fatty), b (scattered fibroglandular densities), c (heterogeneously dense), and d (extremely dense) (42). Consensus readings were performed by a radiologist specializing in breast imaging with 16 years of experience (F.K.) and a breast surgeon with 13 years of experience (M.B.) to visually estimate content of fibroglandular-density tissue within the breasts imaged.

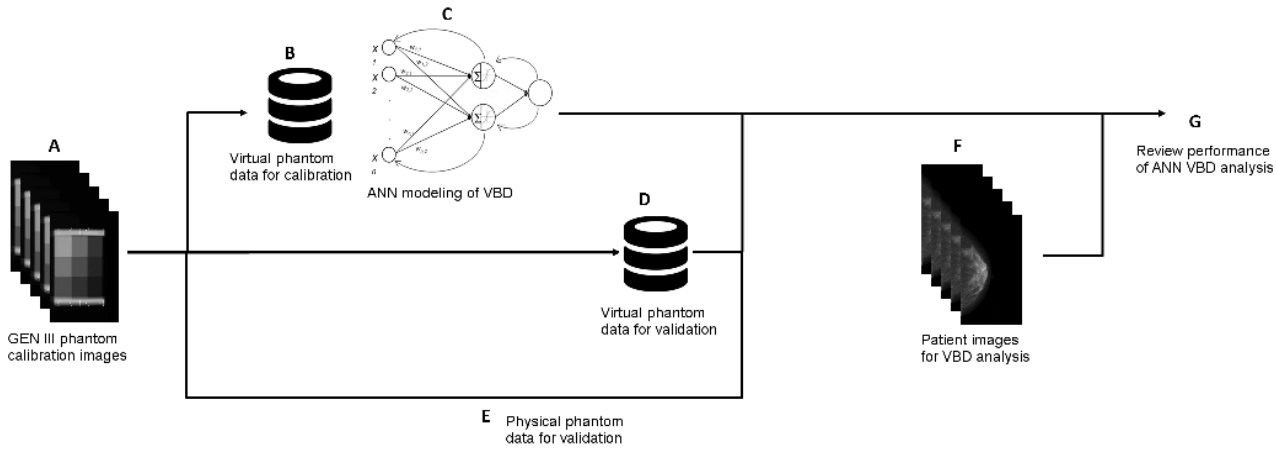

Figure 3.4: Summary of mammographic breast density modeling and analysis performed in this study. Repeated imaging of a GEN III Volumetric Breast Density (VBD) phantom was performed on one FFDM system over a broad range of imaging parameters, determined by a survey of screening exams acquired as part of standard clinical practice (A). Physical phantom images were used to derive a set of virtual phantom data for the purposes of calibrating the FFDM system (B). This set of virtual calibration phantom data was used to train an Artificial Neural Network (ANN), taking imaging parameters and data (anode and filter materials used, tube voltage, current exposure time, image exposure sensitivity, background detector signal, tissue thickness, and signal intensity) as input and outputting VBD (C). Physical phantom images were used to derive a second set of virtual phantom data for the purposes of validating the ANN calibrations of the FFDM system for calculating VBD (D). Furthermore, the original set of physical phantom images were used directly for validating the ANN calibrations (E). ANN VBD was calculated for a set of cancer and non-cancer patient's FFDM images (F). Statistical analysis was performed five-fold: by reviewing performance on two sets of validation phantom data, intra-patient, against BI-RADS breast composition scoring, against MRI VBD, and inter-patient (G).

VBD was also calculated quantitatively from MRI (as a standard of volumetric anatomical imaging) collected of the patients with cancer, using only the breast contralateral to the tumor. This was performed using a fuzzy *c*-means clustering technique, previously described (56). An unsupervised algorithm assigned membership to each image voxel and automatically determined a threshold best to separate parenchyma and adipose tissue. MRI VBD was then calculated of the parenchyma (DV) as a percentage of total breast volume (TBV).

Mammograms used for quantitative density analysis were raw attenuation ("For Processing") images normalized by the system's image exposure sensitivity optimization algorithm, while those used for qualitative density assessment (BI-RADS scoring) were post-processed ("For Presentation") images, as available clinically to radiologists. All image processing was performed using MATLAB R2012b (Mathworks, Inc., Natick, MA, USA).

3.2.4 Statistical analysis

Validation of the ANN VBD modeling was performed five-fold: by reviewing performance on two sets of phantom data, intra-patient, against BI-RADS breast composition scoring, against MRI VBD, and inter-patient.

Phantom validation of ANN VBD was performed using a second set of virtual phantom data, randomly generated as the ANN training set was above, as well as that of the physical phantom itself. ANN-estimated VBD values were compared to their actual VBD values (derived in case of virtual phantom, known in case of physical phantom) using linear regression. Intensity-saturated regions on the physical phantom images, a small minority of values observed at limits of image bit depth were excluded. This saturation occurred due to some tissue-equivalent regions being at extremes of normal breast tissue

composition by design (2cm of 0% VBD and 100% VBD) having been exposed to broad ranges of imaging parameters extending beyond that seen from normal clinical use. Intra-patient validation of VBD measures was performed using linear regression to determine the relationships between VBD measures calculated of non-cancer patients' left and right breasts.

Comparisons of ANN VBD estimated of mammograms assigned in each of the four BI-RADS categories was performed using Tukey's multiple comparisons test. Spearman's correlation coefficient was also used to evaluate the correlation between the visual assessments and ANN VBD. Comparisons between ANN VBD measures and MRI VBD measures were performed using linear regression.

Inter-patient validations were conducted by reviewing VBD measures for their association with classical risk factors of breast cancer - age, age at menarche, BMI, menopause status, parity status, in addition to cancer status itself. Comparisons with risk factors were performed with the cancer patient data only. Linear regression was performed to determine the relationships between quantitative risk factors and values of VBD. X^2 tests were used in comparing categorical risk factors, and Wilcoxon Mann-Whitney tests were used in comparing cancer status, with VBD measures. VBD measures used here of non-cancer patients were averages of left and right breasts.

Squared Pearson's correlation coefficients (R^2) were calculated from associations of linear regression analyses performed. Root-mean-square errors (RMSE) were calculated to aggregate the magnitude of individual differences between compared values. Regression equation fit coefficients were also calculated for comparisons and their significance were tested with *t* Tests. p-values less than 0.05 were interpreted as significant. All statistical analyses were performed using JMP 11.0 (SAS Institute Inc., Cary, NC, USA).

3.3 Results

3.3.1 Phantom validation

Results comparing ANN-estimated VBD against actual VBD values of the virtual and physical validation phantom data are shown in Figure 3.5. Very strong correlations were achieved ranging from 0.948 estimating physical phantom VBD to 0.993 estimating virtual phantom VBD. RSME of estimated VBD ranging from 9.2% with physical phantom data to 12.1% with virtual phantom data were also apparent.

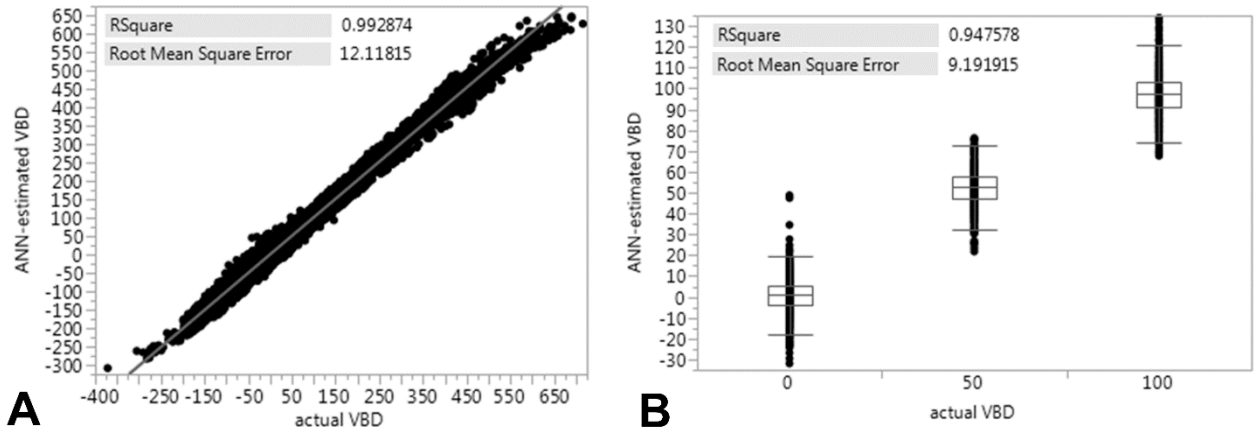

Figure 3.5: Phantom validation of ANN-VBD. Linear regression results comparing ANN-estimated VBD against actual VBD (derived with virtual phantom in A, known with physical phantom in B). In considering random combinations of imaging parameters within expected ranges seen clinically, virtual phantom data derivation produced VBD values beyond 0 and 100% VBD (A). Comparison against physical phantom data shown as a boxplot since values quantized at 0, 50, and 100% VBD (B). Plots are inlayed with squared Pearson's correlation coefficients and root-mean-square errors.

3.3.2 Patient validation

Example clinical images and ANN-VBD maps of low-, mid-, and high-VBD breast are shown in Figure 3.6. Comparisons of VBD results from non-cancer patients' left and right breasts are summarized in Figure 3.7. Excellent agreement between TBV values is seen

with an R^2=0.908 (Figure 3.7A), as well as between DV values with R^2=0.901 (Figure 3.7B). Very good agreement between VBD values is also apparent with R^2=0.789 (Figure 3.7C). *t* Tests reveal regression line intercepts to be significant with comparisons of TBV (43.3ml, p=0.040) and VBD (7.7%, p<0.001) between breasts.

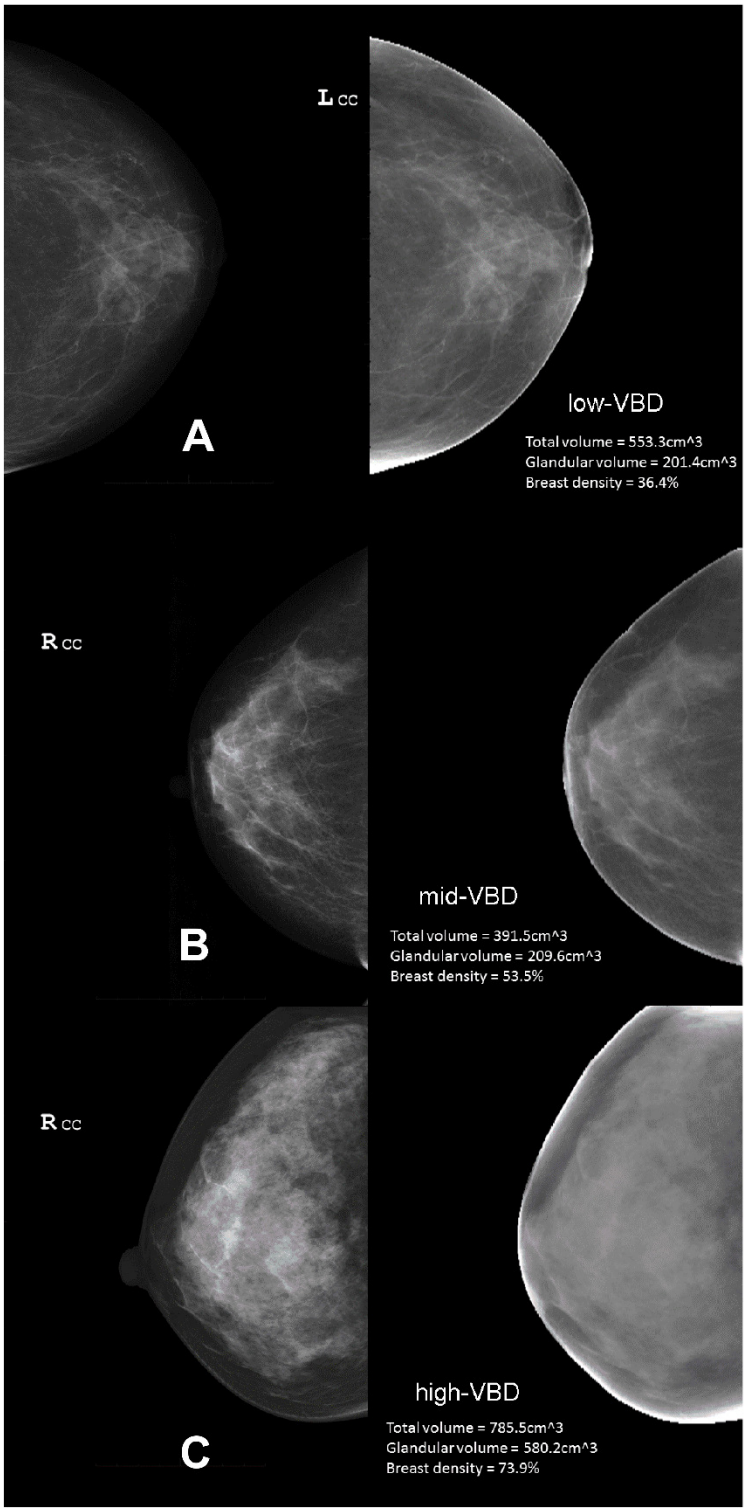

Figure 3.6: Clinical images and ANN-VBD maps. Example images of low- (A), mid- (B), and high-VBD (C) breasts. In each panel, the vendor post-processed image used by radiologists for diagnostic reading is shown at left (not used in study, but shown here to illustrate images used clinically and commonly used to estimate mammographic density) and the ANN-calculated VBD map is shown at right.

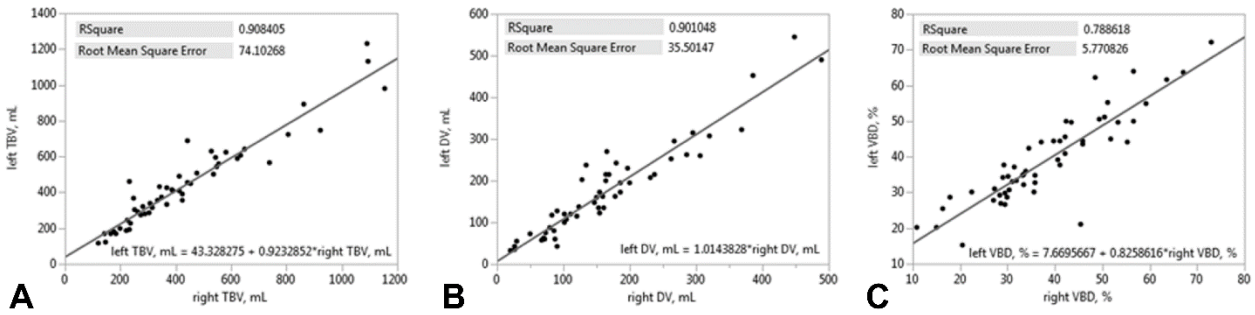

Figure 3.7: Intra-patient validation of ANN-VBD. Linear regression plots comparing ANN-estimated VBD measures of the left (y-axis) and right (x-axis) breast of non-cancer patients. Plots are inlayed with squared Pearson's correlation coefficients, root-mean-square errors, and fit equations for comparisons of VBD (A), DV (B), and TBV (C).

Comparison of ANN VBD against BI-RADS scoring is summarized in Figure 3.8. Number of women falling into categories 'a', 'b', 'c', and 'd' proved to be 6, 23, 68, and 2, respectively, with mean and standard deviation for ANN VBD distributions in each category being 30.7% (8.0%), 35.4% (9.1%), 43.7% (13.1%), and 71.9% (13.6%), respectively. Tukey's multiple comparisons test revealed VBD in all categories as significantly different from each other ('a' and 'd' $p<0.001$, 'b' and 'c' $p=0.029$, 'b' and 'd' $p<0.001$, 'c' and 'd' $p=0.008$), except between categories 'a' and 'b' ($p=0.825$) as well as 'a' and 'c' ($p=0.062$). ANN VBD and BI-RADS categories showed significant moderate association ($r=0.405$, $p<0.001$).

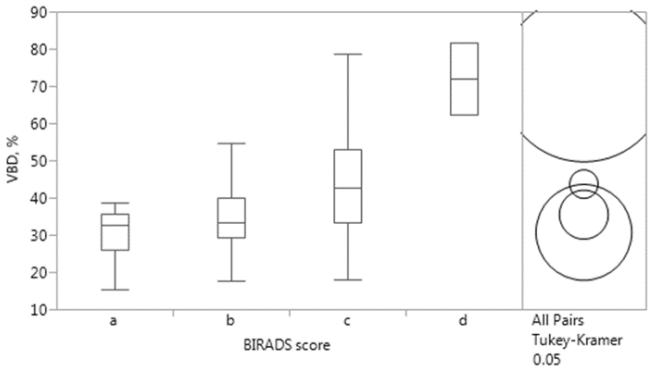

Figure 3.8: Comparison of BI-RADS breast tissue composition scoring and ANN VBD. Tukey's multiple comparisons test reveal VBD in all categories as significantly different from each other, except between 'a' and 'b' as well as 'a' and 'c' (circles not overlapping at right, p-values detailed in text).

Comparison of ANN VBD results against MRI VBD is summarized in Figure 3.9. Very good agreement between TBV is apparent with R^2=0.852 (Figure 3.9A), moderate agreement between DV is apparent with R^2=0.751 (Figure 3.9B), and good agreement between VBD is apparent with R^2=0.665 (Figure 3.9C).

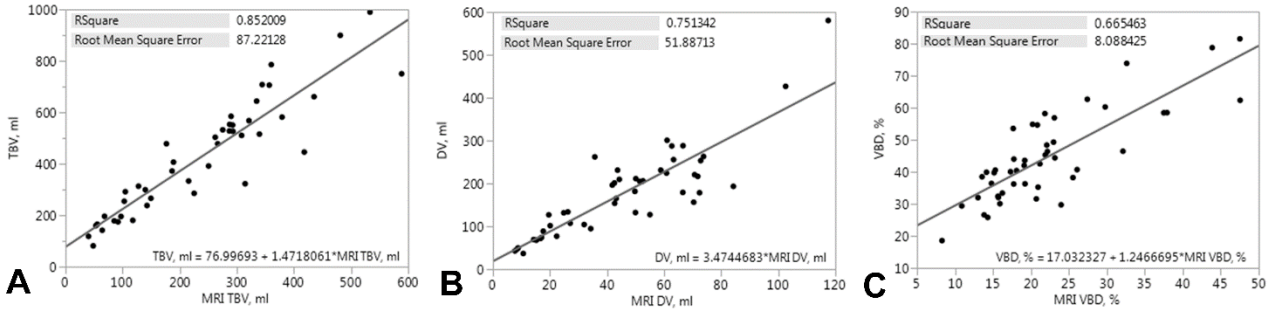

Figure 3.9: Linear regression plots comparing ANN-estimated measures of VBD (y-axis) with those measured on MRI (x-axis); total breast volume (A), dense volume (B), and volumetric breast density (C). Plots are inlayed with squared Pearson's correlation coefficients, root-mean-square errors, and fit equations for comparisons.

Comparisons of VBD results against classical risk factors of breast cancer that showed significant differences are summarized in Figure 3.10. VBD shows a negative association with age, where slope of the regression line is significantly less than 0 with

p<0.001 (Figure 3.10A). VBD also shows a negative association with BMI (Figure 3.10B), where regression slope is significantly less than 0 with p=0.027. The negative relationship of VBD with age seen above is also reflected in the comparison with respect to menopause status (Figure 3.10C), where VBD in post-menopausal women is significantly lower than that of pre-menopausal women (p<0.001). Comparing TBV and DV against the same risk factors revealed only that both are positively associated with BMI (p<0.001, Figure 3.10D, 3.8E). Though VBD was not significantly different given parity status in this population (p=0.068), it is also shown in Figure 3.10F and discussed below.

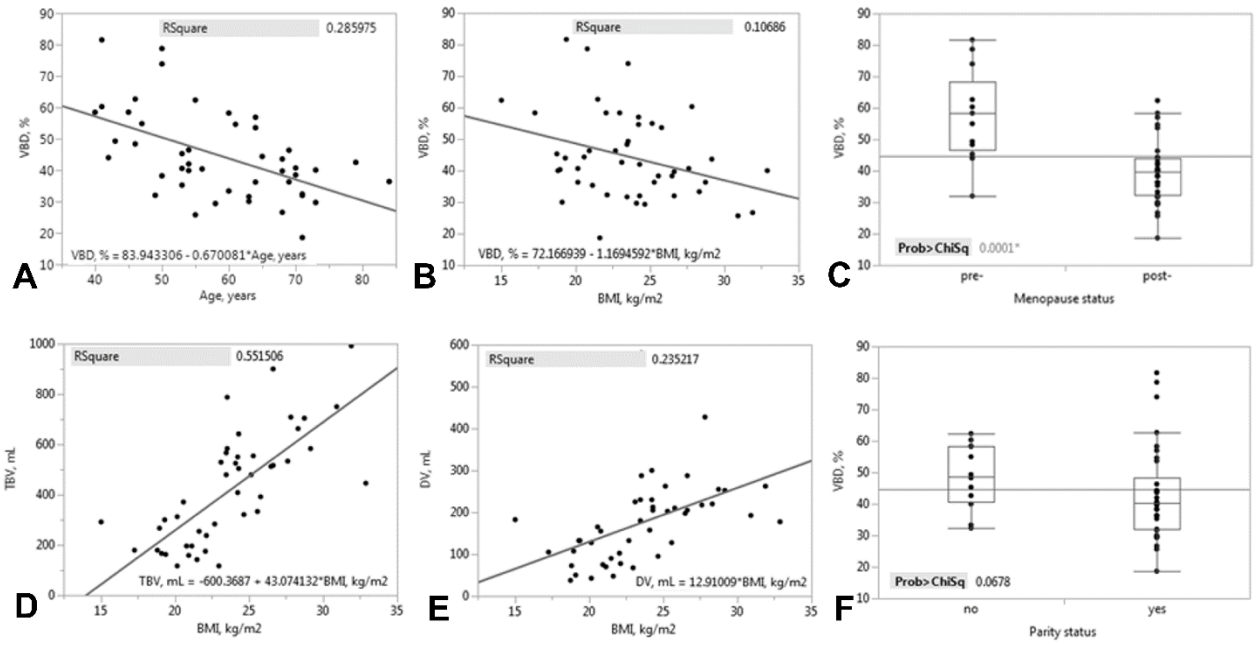

Figure 3.10: Inter-patient validation of ANN-VBD. Comparison plots of ANN-estimated measures of VBD with significant difference against classical risk factors of breast cancer; age (A), BMI (B, D, E), and menopause status (C). Though not reaching statistical significance, VBD compared with parity status (F) is also shown. Linear regression plots are inlayed with squared Pearson's correlation coefficients and fit equations. Boxplots are inlayed with X^2 test statistics.

Comparisons of VBD results against cancer status are shown in Table 3.3. Though no significant difference was found of age, BMI (Table 3.1, p=0.516 and 0.189

respectively), TBV, or DV (Table 3.3, p=0.707 and 0.539 respectively) between the two groups, VBD was found to be significantly higher in the cancer group in relation with the non-cancer group (median of 41.3% vs 34.9%, respectively, p=0.046).

Table 3.3: Comparisons of ANN-estimated measures of VBD between cancer and non-cancer patients

Parameter	Cancer			Non-cancer			p-value, by Wilcoxon test
	Median (IQR)	Min	Max	Median (IQR)	Min	Max	
TBV, mL	399.0 (361.0)	81.1	989.7	390.3 (290.8)	116.2	1161.0	0.707
DV, mL	178.8 (126.2)	36.7	580.2	115.8 (123.6)	26.9	497.0	0.539
VBD, %	41.3 (19.9)	18.6	81.5	34.9 (17.4)	15.5	72.5	0.046

3.4 Discussion

A novel automatic method of measuring mammographic density has been presented and evaluated in this work. Quantifying breast tissue density by ANN modeling of FFDM appears feasible, producing reasonable VBD estimates of breast tissue-equivalent phantoms, as well as of a patient population of Japanese women with and without cancer.

In our phantom validation studies of the method, excellent agreement of ANN-estimated VBD against actual values were seen reaching an R^2 of 0.993 with virtual phantom data and an R^2 of 0.948 with physical phantom data. In our intra-patient validation study, good to excellent agreement between VBD measures of left and right breasts of patient were apparent ranging from an R^2 of 0.789 with VBD itself to an R^2 of 0.908 with TBV. Comparisons of ANN VBD against two breast density metrics being used in the industry (one qualitative standard and one quantitative) also showed promising results. ANN VBD agreed and trended well with BI-RADS scoring, with categorizations having been shown as significantly different from each other (except between the lowest

category with the adjacent two) and as significantly associated with ANN VBD. ANN VBD also agreed well with MRI VBD, where R^2 ranging from 0.665 with VBD itself to an R^2 of 0.852 with TBV were seen, in line with findings previously reported comparing mammographic VBD with MRI VBD (95). In our inter-patient validation study, confirmations of relationships between mammographic density and several classical risk factors of breast cancer as previously reported were seen, with ANN VBD showing a negative association with age, BMI, and menopause status (48,72,96). The reported association between mammographic density and cancer status (18,19,48,72) was also reflected in our findings. Despite there being no difference in age, BMI (Table 3.1, p=0.516 and 0.189 respectively), or breast size (TBV, Table 3.3, p=0.707) between the two groups of women, a significant difference in their VBD was seen. Women with cancer tended to have breasts higher in fibroglandular density than women without (median of 41.3% vs 34.9%, respectively, p=0.046), as expected.

It is our understanding that the clinical implementation of the methods developed in this study is attractive and feasible due to its ability, versatility, and simplicity. ANN, by their nature, can adapt to any measurable mathematical function with remarkable ease and we consider this study's findings demonstrates ANN-modeling has the potential to estimate VBD both accurately (against phantom references and existing metrics) and precisely (within patients). Contrasting with several other recently developed methods for quantifying breast density from mammography further highlights the uniqueness of our proposed method. It appears new works published still primarily take a 2D approach to either density- (97,98) or parenchymal pattern-based segmentations (83,85), though our method takes advantage of the third dimension. Other 3D approaches using mammography exist and also become more common, but are often proprietary (20,99) or may require in-image phantoms (19). Studies using 3D imaging modalities also continue, with MRI (100,101) as well as 3D digital breast tomosynthesis (102), though neither of

these modalities are standard for breast cancer screening. Furthermore, with the exception of Seo et al. and Japanese references mentioned in the introduction above, all aforementioned studies have looked at American, European, and Australian populations only, making our study distinctive in terms of population as well.

We note several limitations to the current study that highlight opportunities for future investigation. First, as summarized above, though previously reported relationships of mammographic density with clinical indicators of breast cancer and cancer status itself were largely confirmed with ANN VBD measures, such was not the case with parity. In both pre- and post-menopausal women, parity has been shown to strongly associated with breast cancer risk and inversely related to density (48,72,96). Though it appears our comparison against parity trends in the same direction as that established (Figure 3.10F), statistical significance was not reached with this population ($p=0.068$). Sampling a larger population and further development of the ANN-VBD method itself may allow for confirming the same relationship as well in the future, however.

Next, in looking more closely at patient analyses again, a few questions are raised. In intra-patient validation experiments, the correlation coefficient of comparing left and right measures of VBD was not as high as that of TBV and DV comparisons, though the reason for this is not completely clear. There also appears to be a trend where VBD of the left breast in non-cancer women were higher than that of their right breast, in general (+7.67%, $p=0.003$). This was also true of TBV, though less pronounced (+43.3ml, $p=0.040$). Clinical protocol at the study institution specifies mammograms be acquired to focus on glandular tissue, sometimes at the cost of dropped adipose tissue near the chest wall, though this does not explain these trends fully. Further investigation is needed as the method continues development and validation.

Furthermore, comparisons of ANN VBD with BI-RADS scoring did not show significant differences between categories ‘a’ and the next two adjacent categories. We

attribute this to there being a relatively small number of women categorized as ‘a’ (n=6), possibly due to the distribution of density in Japanese women being relatively high in general, which would be consistent with previous findings comparing Japanese women with others (103,104). Of course, the subjective nature of the manual assessment itself may have also had an effect.

Lastly, though measures were taken to exclude erroneous phantom data due to signal intensity saturation seen at imaging extremes, it was necessary to retain and use some saturated phantom regions in order to generate virtual phantom data. Inherently, calibration methods may have such limitations, and fitted functions ultimately may not agree exactly with experimental data. Further investigation is needed concerning these issues as well.

Though breast cancer incidence in Japan is only one-third of that seen in Western countries, local numbers have doubled in all age groups over the past two decades. Moreover, incidence in women <50 in Japan appears similar to that in the United States and the United Kingdom, peaking at ~45 years old (105). Given the consistent findings in association of breast density to breast cancer risk and its potential to further personalize breast cancer care, it is very likely efforts toward capturing the measure quantitatively may permeate throughout Japan and elsewhere to identify those at high risk. We consider the conclusions of this research support the growing amount of compelling evidence that further investigation of methods to quantify VBD automatically are necessary with larger datasets and against outcomes such as performance of human and computerized screening, as well as risk of developing breast cancer.

Enabled by major advances in technology and necessitated by the rapid generation of large datasets, advances in statistical methods and machine learning are fundamentally changing the way biomedical research and image analysis is conducted. These methods have become essential not only in terms of confirmatory testing, but as well a data-driven

computational knowledge discovery process. We demonstrate here the adaptability of ANN in estimating complex functions to quantify radiographic markers of VBD from FFDM.

3.5 Conclusions

The proposed ANN calibrated model appears to produce reasonable measures of mammographic density that are validated with breast tissue composition phantoms, associated with existing qualitative and quantitative measures of breast density, and associated with classical biomarkers of breast cancer as previously reported. VBD calculated from FFDM of Japanese women in this feasibility study appears to be significantly higher in those with cancer in comparison to those without. Further studies are warranted to confirm these findings and determine potential implications.

Chapter 4 Identifying Triple-Negative Breast Cancer Using Background Parenchymal Enhancement Heterogeneity on Dynamic Contrast-Enhanced MRI: A Pilot Radiomics Study

4.1 Introduction

Breast cancer is a disease with several distinct biological subgroups (106,107). Gene expression-based molecular subtyping is used clinically in the selection of the most appropriate therapy and has proved valuable for individualized management (108). In particular, breast cancers that overexpress the estrogen receptor (ER), progesterone receptor (PR), and/or human epidermal growth factor 2 receptor (HER2) can be specifically targeted with hormonal and/or anti-HER2 therapies. Triple-negative (TN) breast cancers, however, lack expression of these three receptors, so currently have no targeted therapy available and are limited to general cytotoxic chemotherapies. TN cancers tend to be larger in size, are of higher grade, have lymph node involvement at diagnosis, and have the poorest prognosis (109–112). The ability to differentiate TN cancers from other less aggressive subtypes using diagnostic imaging, could help identify and stratify patients with this rare and particularly difficult subtype for the appropriate therapy earlier than biopsy in the future.

Current methods of biopsy have limitations considering more than small samples of tissue, hence meet some issues with large and/or heterogeneous cancers (113). MRI, however, provides anatomical and functional properties of whole tissues. Findings on MRI such as tumor size, morphology, shape, and enhancement characteristics (such as rim enhancement) have been shown as significant in differentiating breast cancer subtypes including TN breast cancers (114–117), though such manual annotation of tumor characteristics are generally limited to a few qualitative descriptors and are dependent on the operator (118). On the other hand, computer-aided diagnosis (CAD) has paved the way to improve diagnostic specificity by computing quantitative information about the

entire tumor non-invasively in an objective manner and reducing inter-reader variability (119–122). More recently, the radiomics approach of CAD has emerged with the central premise that cancer imaging phenotypes reflect underlying gene expression patterns and combining these sources of information will improve individualized treatment selection and monitoring (25,26,123). The approach has shown great promise considering whole tissues relatively comprehensively by automatically extracting and evaluating large sets of advanced quantitative imaging features, including texture heterogeneity patterns (124).

To date, most breast cancer studies have focused on characterization of the tumor itself (118–122,125–129). Relatively little is known about the diagnostic and prognostic significance of its surrounding parenchyma tissue on MRI. Initial studies indicate that increased background parenchymal enhancement (BPE) on dynamic contrast-enhanced MRI (DCE-MRI) could lead to higher rates of misinterpreting benign tissues as suspicious (130,131), though the precise reasons for this enhancement are not clear. There is also evidence suggesting that tumor microenvironment may help define and regulate breast cancer progression (12–15), as well as predict disease recurrence following therapy (132). Additionally, the appearance of BPE on MRI and parenchyma on mammography have been associated with risk of developing breast cancer independently (16–21). We hypothesize BPE may also have prognostic significance with breast cancer subtype.

The purpose of this study was to determine the added discriminative value of detailed quantitative characterization of BPE in addition to the tumor itself on DCE-MRI at 3.0 Tesla in identifying TN breast cancers. Our work was based on semi-automated, volumetric segmentation of the tumor and its surrounding parenchyma. We extracted a variety of quantitative imaging features of both regions in 3D on DCE-MRI in addition to standard radiologist-evaluated clinical features and combined them with machine learning tools to obtain the optimal subtype classification.

4.2 Materials and Methods

4.2.1 Study Population

Eighty-four women, presenting 88 lesions pathologically proven as invasive carcinoma, were enrolled in the study. All underwent DCE-MRI before their surgical procedure in the period of February 2012 to May 2013 and had pathology reports with molecular subtype results available. Four women with multiple lesions (one in each breast, treated as separate cases) were included. Those found with multiple unilateral lesions, however, were excluded. Additional criteria for exclusion from the study included having received neoadjuvant chemotherapy, hormonal therapy, or having artifacts on MRI exams. Patient demographics are summarized in Table 4.1. This retrospective study was approved by the Institutional Review Board of Hokkaido University Hospital and informed consent was waived according to Ethical Guidelines for Clinical Studies of the Japanese Ministry of Health, Labour, and Welfare. All patient data were anonymized and de-identified prior to analysis.

4.2.2 Pathological Subtyping

Expression of ER, PR, HER2 and Ki67, a marker of cellular proliferation and subtype (133,134), were determined by immunohistochemical analysis of tumor specimens. Each tumor sample was classified as ER+, PR+, and/or HER2+, or being triple-negative (TN) if negative for all three. Additionally, cancers were classified as Luminal A (LumA, if ER+ and/or PR+, HER2-, Ki67<14%) or Luminal B (LumB, if ER+ and/or PR+, HER2 over-expressed or Ki67≧14%), as also clinically relevant in individualized therapy (108). Tumor subtype demographics are also summarized in Table 4.1.

Table 4.1: Patient and subtype demographics

Parameter	All	By presence of receptor			By St. Gallen consensus (108)		
		ER+	PR+	HER2+	TN	LumA	LumB
Patients, n	84	69	60	4	11	42	27
Lesions, n	88	73	63	4	11	45	28
Mean age, years	59.1 (11.0)	58.8 (10.7)	58.9 (11.0)	60.0 (10.2)	61.2 (12.3)	59.1 (9.9)	58.4 (12.1)
Mean BMI	23.2 (4.1)	23.5 (4.3)	23.9 (4.3)	22.5 (1.5)	22.3 (4.9)	23.3 (4.5)	23.9 (4.0)
Postmenopausal patients, n	61	51	42	4	10	33	18

Some overlap exists between subtypes defined and tumors may be represented in more than one category. Classification criteria are described in the Pathological Subtyping section of the methods. Standard deviations are displayed in parentheses with mean measures. BMI=body mass index, TN=Triple-negative, LumA=Luminal A, LumB=Luminal B.

4.2.3 Image Acquisition

MR imaging was performed using one Achieva 3.0T TX system (Philips Healthcare, Best, Netherlands) with a 7-channel breast coil while patients lied prone. The dynamic protocol used was in accordance with European Society of Breast Imaging (135) and American College of Radiology guidelines (136). In brief, 3D T1-weighted images were acquired bilaterally in the axial plane with a fat-suppressed gradient echo sequence (e-Thrive): Repetition time/echo time 4.9 ms/2.4 ms, flip angle 10°, field of view 320 × 320 mm, voxel size 0.8 × 0.8 × 1.6 mm (reconstructed 0.8 mm isovoxel), and SENSE parallel imaging factor 2.4. Images at four time points were acquired, each lasting one minute. The first image was taken immediately before injection of contrast material (Gadopentetic acid with diethylenetriaminepentacetate, 0.1 mmol/kg) and flushing with 20mL saline (t1), the second and third in the early phase at 1 and 2 minutes after injection (t2 and t3 respectively), and the last in the late phase at 6 minutes after injection (t4).

4.2.4 Image Segmentation and Feature Extraction

MR images were reviewed retrospectively by a board-certified radiologist specializing in breast MRI with 13 years of experience (F.K.), blinded of findings other than diagnosis as invasive breast cancer. Clinical features concerning tumor morphology were evaluated according to Breast Imaging Reporting and Data System (BI-RADS) MRI (42). For mass lesions: shape, margin, and internal enhancement characteristics were evaluated; and for non-mass lesions: distribution and internal enhancement characteristics were evaluated. Morphology and mass size (mass lesions via longest axis) were also included in the analysis.

Contouring of the affected breast was performed using images acquired at t3 (near max intensity for all tissues) with automatic detection of the skin edge by thresholding plus semi-automatic delineation of the chest wall and nipple at every slice by interactive placement of an expanding polygon mask. The breast tumor was also segmented at t3, to better distinguish it from background parenchyma, using a semi-automated gray-level intensity threshold 3D region-growing technique (137) that was manually modified as necessary. Subsequently, separation of the remaining non-tumor breast tissue into fibroglandular parenchyma and adipose tissue compartments was performed semi-automatically at t1 (pre-contrast, as recommended for assessment of fibroglandular tissue in latest edition of BI-RADS atlas) (42), using an adapted fuzzy *c*-means clustering technique (56). The unsupervised algorithm assigns a membership to each voxel initiated with user-seeding and automatically determines a threshold best to separate parenchyma and adipose tissue. An illustration of tissue compartment segmentation is shown in Figure 4.1. At this point, breast density was calculated as the percentage of breast volume that was made up of parenchyma and included in the analysis. All subsequent features were extracted of the tumor and parenchyma compartments.

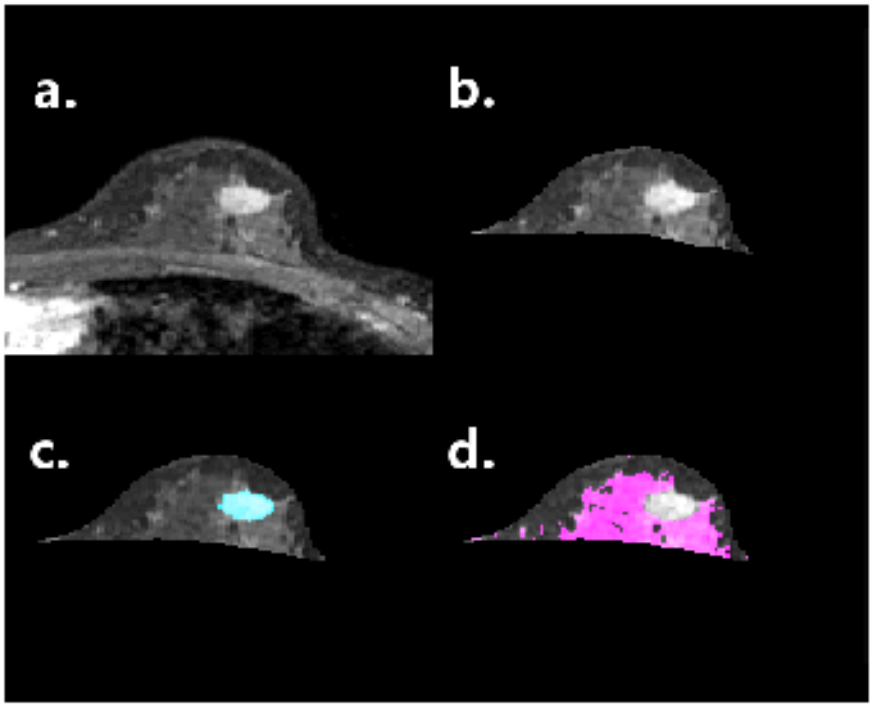

Figure 4.1: Example of tissue segmentation performed of all cancer patients' affected breast images. At top left (a), a dynamic contrast-enhanced MRI exam at t3 is seen in the axial plane, illustrating one slice of the view used for contouring the breast and tumor. At top right (b), the result of breast segmentation is shown. At bottom left (c), the segmented tumor is highlighted in blue. Finally at bottom right (d), the parenchyma segmented at t1 is highlighted in pink. Breast subcompartment segmentation was performed in 3-dimensions.

Three standard pharmacokinetic parametric maps were generated from each DCE-MRI to capture enhancement quantitatively: contrast material rate in (from t1 to maximum), percent enhancement (PE, from t1 to maximum), and signal enhancement ratio (SER, change from t1 to maximum relative to change from t1 to t4) (138).

Four first-order statistical features were calculated from the parameter maps of the tumor and parenchyma compartments: mean, standard deviation, skewness, and kurtosis. Nine second-order statistical features, also known as gray-level co-occurrence texture features (139), were calculated at a 1-voxel distance offset of the maps and averaged across the 26-directions of 3D space after rescaling to 8-gray level (3-bit) data: energy, contrast, correlation, variance, homogeneity, sum mean, entropy, inertia, and cluster shade. Thirteen statistical textures of 3 parametric maps resulted in 39 features capturing

enhancement heterogeneity for each tissue compartment studied, which were included in the analysis. Those of the parenchyma are defined here also as BPE texture or heterogeneity. All image processing was performed using MATLAB R2012b (Mathworks, Inc., Natick, MA, USA) software.

4.2.5 Predictive Modeling for Differentiation

Based on the extracted imaging features, we aimed to distinguish TN breast cancers from other subtypes using machine learning tools. In particular, we performed 5 classification tasks: differentiating TN cancers against non-TN, ER+, PR+, LumA, and LumB cancers. Differentiating TN from HER2+ cancers was not performed as the number of tumors combined from these groups was insufficient given techniques used (n=15).

To minimize bias in our evaluation, stratified 10-fold cross-validation was performed (140). 10 bootstrap repetitions of each were run, from which performance metrics (described below) were averaged and confidence intervals were estimated. A two-step feature selection technique was applied on imaging features before classification. First, features were ranked by the X^2 statistic (141) to identify strength of association with the subtype in question; second, a sequential forward floating search algorithm was used to identify a small subset of optimal features large enough to capture data complexity (142). Finally, a support vector machine (SVM) classifier was trained (143,144) from the selected feature subset of preceding search steps. SVM models are non-probabilistic binary linear classifiers, which represent the data in higher dimensionality spaces, mapped so as to separate the categories with a divide that is as wide as possible. Both steps of feature selection were encapsulated with the classifier within each training fold in order to avoid feature selection bias and overfitting (144,145). Feature selection and classification were also performed with regularization, which served to penalize model complexity, as another measure to avoid overfitting (144). All predictive modeling was performed using Waikato

Environment for Knowledge Analysis (WEKA) 3.6.12 (University of Waikato, Hamilton, New Zealand) (146).

4.2.6 Cluster Analysis

In addition to the supervised approach (differentiating subtype-labeled data) in the classification modeling described above, an unsupervised approach was also taken with learning tasks. Clustering of BPE texture features in a *k*-means manner, as used with gene analysis to reveal groups with similar expression patterns (147,148), was performed without using knowledge of subtype. All BPE texture features of included cases were normalized as *z*-scores and clustered into two partitions using genomic data analysis framework Gitools 2.2.1 (Universitat Pompeu Fabra, Barcelona, Spain) (149).

4.2.7 Statistical Analysis

Figure 4.2 summarizes the study's radiomic analysis performed as described above. Classification performance of predictive modeling was evaluated using accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) values averaged over all bootstrap folds of cross-validation. Wilcoxon signed-rank tests were used to test significance of paired difference between classification models' performance without inclusion of BPE-derived features against those with in a non-parametric manner. *p*-values of less than 0.05 were interpreted as significant. All statistical analyses were performed using JMP 11.0.0 (SAS Institute Inc., Cary, NC, USA).

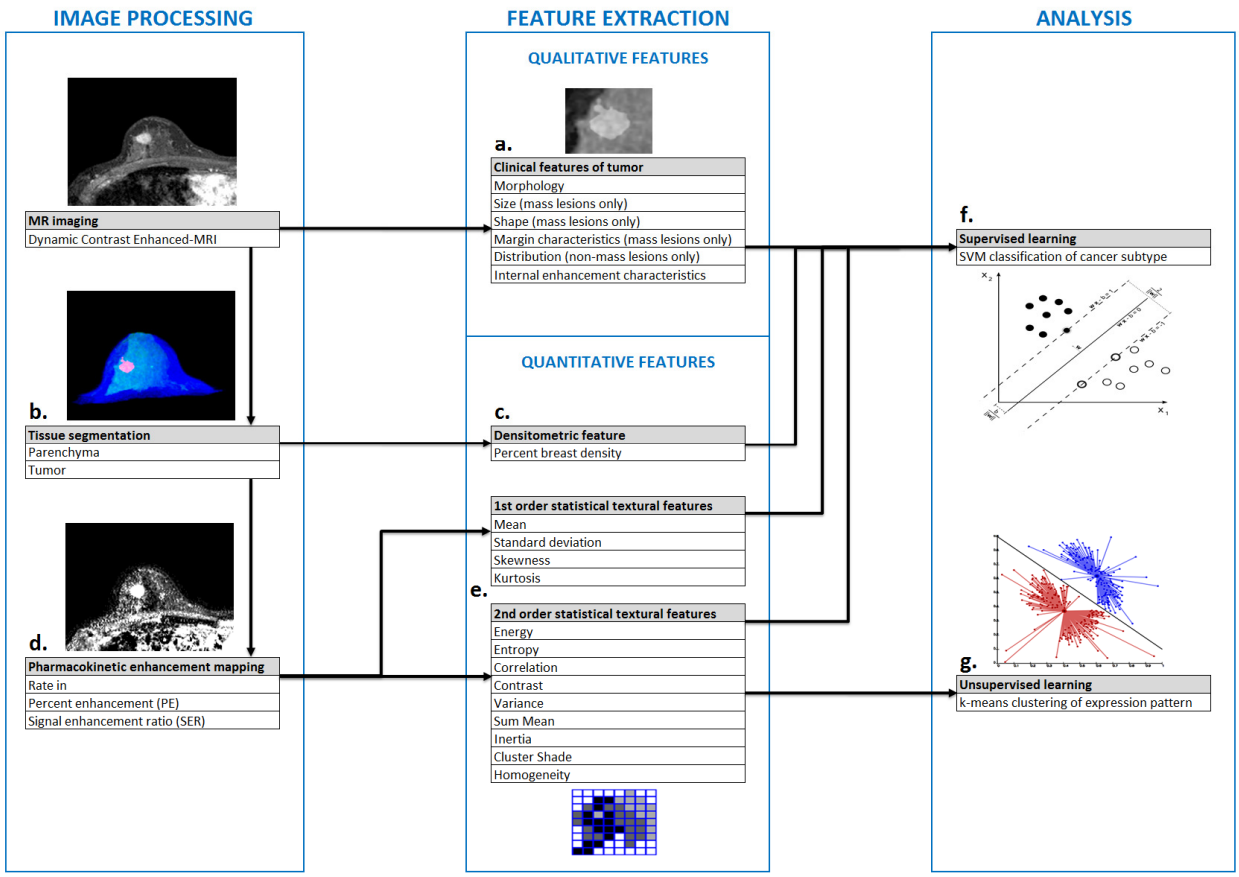

Figure 4.2: Summary of radiomic analysis performed in this study. Clinical features were evaluated by a radiologist according to Breast Imaging Reporting and Data System directly from dynamic contrast-enhanced MRI (a). 3-Dimensional tumor (red) and parenchyma (light blue) compartments were segmented (b), from which volumetric breast density was immediately estimated (c). Enhancement maps were then generated (d), from which textural features of tissue compartments were extracted and defined as enhancement heterogeneity (e). Subsequently, two analyses were conducted using extracted features: supervised learning of breast cancer subtype was performed with a support vector machine classifier (f) and unsupervised learning of background parenchymal enhancement feature expression pattern was performed with *k*-means clustering (g).

4.3 Results

4.3.1 Differentiation of Molecular Subtypes

Performance metrics for subtype differentiation tasks are detailed in Table 4.2.

Classification models using both tumor and parenchyma features generally outperformed those based only on tumor features, most notably in terms of accuracy, sensitivity, and

AUC. The most remarkable discriminative performance was seen in the classification of TN against ER+ cancers. While the conventional model based on tumor features achieved an AUC of 0.780, the model using both parenchyma and tumor features improved AUC significantly to 0.883 ($p<0.01$). Accuracy and sensitivity improved from 86.3% to 89.4% ($p<0.01$) and 35.5% to 62.0% ($p<0.01$) respectively with this task. Performance classifying TN against all other cancers improved similarly from an AUC of 0.782 to 0.878 ($p<0.01$) by including BPE texture features, with accuracy and sensitivity improved from 86.9% to 90.0% ($p<0.01$) and 33.0% to 57.0% ($p<0.01$) respectively. Classifying TN against PR+ cancers improved from an AUC of 0.731 to 0.859 ($p<0.01$), with accuracy and sensitivity improved from 83.5% to 87.8% ($p<0.01$) and 28.5% to 53.0% ($p<0.01$) respectively.

Table 4.2: Performance results of predictive modeling

Differentiation task	n	Using tumor features				Using both tumor & BPE features			
		Accuracy, %	Sensitivity, %	Specificity, %	AUC	Accuracy, %	Sensitivity, %	Specificity, %	AUC
TN vs others	**88**	86.9 (85.1, 88.7)	33.0 (23.8, 42.2)	94.7 (92.9, 96.4)	0.782 (0.730, 0.833)	90.0* (88.1, 91.8)	57.0* (47.4, 66.6)	94.7 (93.0, 96.2)	0.878* (0.838, 0.918)
TN vs ER+	**84**	86.3 (84.2, 88.3)	35.5 (26.1, 44.9)	94.1 (92.3, 95.8)	0.780 (0.730, 0.830)	89.4* (87.5, 91.3)	62.0* (52.7, 71.3)	93.6 (91.7, 95.4)	0.883* (0.843, 0.923)
TN vs PR+	**74**	83.5 (81.4, 85.6)	28.5 (19.8, 37.2)	93.0 (90.7, 95.3)	0.731 (0.674, 0.788)	87.8* (85.7, 90.0)	53.0* (43.2, 62.8)	94.1 (92.2, 95.9)	0.859* (0.817, 0.901)
TN vs LumA	**56**	79.6 (76.9, 82.3)	40.5 (30.9, 50.1)	88.8 (85.8, 91.8)	0.795 (0.745, 0.844)	81.8 (78.6, 85.1)	49.5 (39.8, 59.2)	89.8 (86.8, 92.8)	0.814 (0.756, 0.872)
TN vs LumB	**39**	61.3 (57.5, 65.2)	29.0 (20.3, 37.7)	73.8 (68.5, 79.2)	0.635 (0.577, 0.693)	84.3* (80.7, 87.8)	69.5* (60.7, 78.3)	90.0* (86.2, 93.8)	0.789* (0.728, 0.850)

Metrics displayed as: mean (95% confidence interval). TN=Triple-negative, ER=estrogen receptor, PR=progesterone receptor, HER2=human epidermal growth factor 2 receptor, LumA=Luminal A, LumB=Luminal B, BPE=background parenchymal enhancement, AUC=area under receiver operating characteristic curve.

*$p<0.01$ by Wilcoxon signed-rank test in comparing models including use of both tumor and BPE features against those using only tumor features.

In differentiating TN against LumB cancers, a significant improvement by including BPE texture features in terms of specificity from 73.8% to 90.0% ($p<0.01$) was also

apparent in addition to improvements in accuracy, sensitivity, and AUC from 61.3% to 84.3% ($p<0.01$), 29.0% to 69.5% ($p<0.01$), and 0.635 to 0.789 ($p<0.01$) respectively. Including parenchyma features did not significantly improve the ability to differentiate TN against LumA cancers, where performance using only tumor features already achieved an AUC of 0.795.

4.3.2 Optimal Imaging Features

The most discriminating features selected in each classification task's cross-validation process are summarized in Table 4.3, indicated with percentage of cross-validation folds in which they were selected as a simple indicator of significance. The Supplemental Table (Appendix.S1) elaborates on this further by detailing distributions of selected feature values as well as their SVM weights, whose magnitude indicates prognostic value. In tumor feature-based models, standard clinical features such as mass size and mass shape proved to be most prevalent in differentiating TN from all other subtype groups, having been selected in all tasks as the top features. Mass margin characteristics also proved to be prevalent in several models, as did internal enhancement characteristics. Tumor morphology and breast density were discriminative in a couple of models each. Tumor enhancement texture features also proved to be prevalent across many differentiation tasks performed. Specifically, 'mean of tumor SER' and 'inertia of tumor rate in' proved to be effective discriminators of TN against all other, ER+, and PR+ subtypes. Several other enhancement texture features, derived from tumor PE, proved discriminative across each tumor-based models performed. Enhancement textures derived from tumor rate in were also selected classifying all other and ER+ cancers from the TN subtype.

Table 4.3: Optimal imaging features selected

Imaging features most discriminative in prediction models, using tumor features (first 5 rows) or using both tumor and parenchyma features (last 5 rows). Features that survived the selection algorithm in the majority of cross-validation folds of the given task are shown in highlighted in gray, with numbers and gradation indicating percentage of folds in which the feature was selected as a simplistic indicator of significance (see Supplemental Table, Appendix.S1, for further elaboration of feature values and their significance) .

		Feature type																						
Model type	Differentiation task	Tumor					Breast	Tumor enhancement									BPE							
		Clinical					Density metric	1st order texture				2nd order texture					1st order texture		2nd order texture					
		morphology	mass size	mass shape	mass margin characteristics	internal enhancement characteristics	breast density	tumor PE mean	tumor PE std	tumor PE skewness	tumor SER mean	tumor rate in Energy	tumor rate in Homogeneity	tumor rate in Inertia	tumor PE Variance	tumor PE Homogeneity	tumor PE Inertia	parenchyma rate in std	parenchyma SER skewness	parenchyma rate in Energy	parenchyma PE Homogeneity	parenchyma PE Entropy	parenchyma PE ClusterShade	parenchyma SER Variance
tumor	TN vs others		100 %	99 %	89 %	85 %			64 %		95 %	59 %		76 %										
	TN vs ER+		100 %	99 %	86 %	82 %	75%			55 %	93 %		62 %	70 %			61 %							
	TN vs PR+	62 %	100 %	97 %			91%	64 %			88 %													
	TN vs LumA	52 %	100 %	93 %		96 %										75 %								
	TN vs LumB		79 %	79 %	51 %										54 %									
tumor + BPE	TN vs others			98 %														100 %	100 %		69 %		69 %	67 %
	TN vs ER+			96 %														100 %	100 %		62 %		62 %	58 %
	TN vs PR+			92 %				72 %										96 %	100 %			66 %		
	TN vs LumA			79 %														89 %	98 %					
	TN vs LumB														87 %				100 %	74 %				95 %

TN=Triple-negative, LumA=Luminal A, LumB=Luminal B, BPE=background parenchymal enhancement, PE=percent enhancement, SER=signal enhancement ratio.

When parenchyma features were used in addition to those of the tumor in modeling cancer subtype, BPE texture features overshadowed standard clinical, breast density, and

tumor enhancement features almost completely as most prevalent in nearly all differentiation tasks performed. As seen in Table 4.3, though mass shape remained an effective discriminator of TN against all other, ER+, and PR+ cancers, it appears BPE features ‘skewness of parenchyma SER’ and ‘standard deviation of parenchyma rate in’ were by far the most prevalent predictors in all differentiation tasks. ‘Variance of parenchyma SER’, ‘homogeneity of parenchyma PE’, and ‘cluster shade of parenchyma PE’ proved to be effective discriminators of TN vs all other, ER+, and LumB cancers. Two texture features of tumor enhancement, ‘mean’ and ‘variance of tumor PE’, remained in models differentiating TN against PR+ and LumB cancers respectively, as were previously selected in tumor-based models differentiating the same.

Three of the most discriminative tumor and parenchyma features selected in comparisons between TN and non-TN cases are presented in Figure 4.3. Box plots illustrate the distributions of the three most predictive quantitative features found in differentiation tasks, and the differences in these distributions between the TN and non-TN group. Figure 4.4 presents visualizations of the BPE feature ‘standard deviation of parenchyma rate in’ (also Figure 4.3b), in the form of pharmacokinetic enhancement maps, illustrating the difference of this BPE texture between examples of one non-TN (left) and one TN (right) patient.

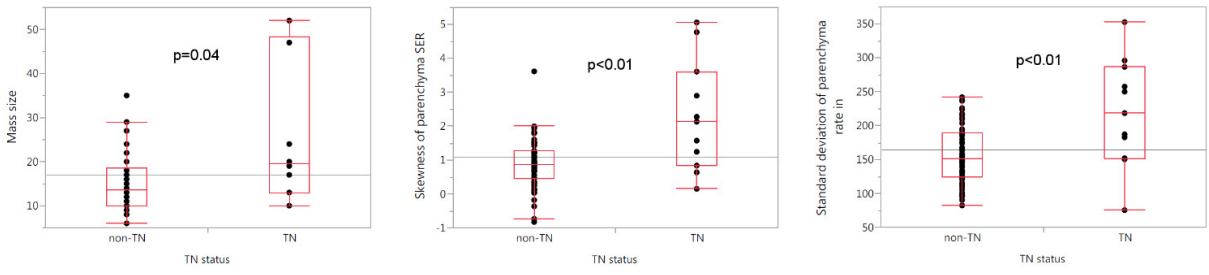

Figure 4.3: Box plots illustrating differences in distributions (quartiles as red boxes, grand mean indicated as spanning line) of the three most predictive quantitative features found in differentiation tasks: the lesion’s ‘mass size’ feature (a), parenchyma’s ‘skewness of Signal Enhancement Ratio’ feature (b), and parenchyma’s ‘standard deviation of rate in’ feature (c)

compared between the triple-negative (TN) and non-TN groups. p-values were calculated by Wilcoxon Mann-Whitney tests.

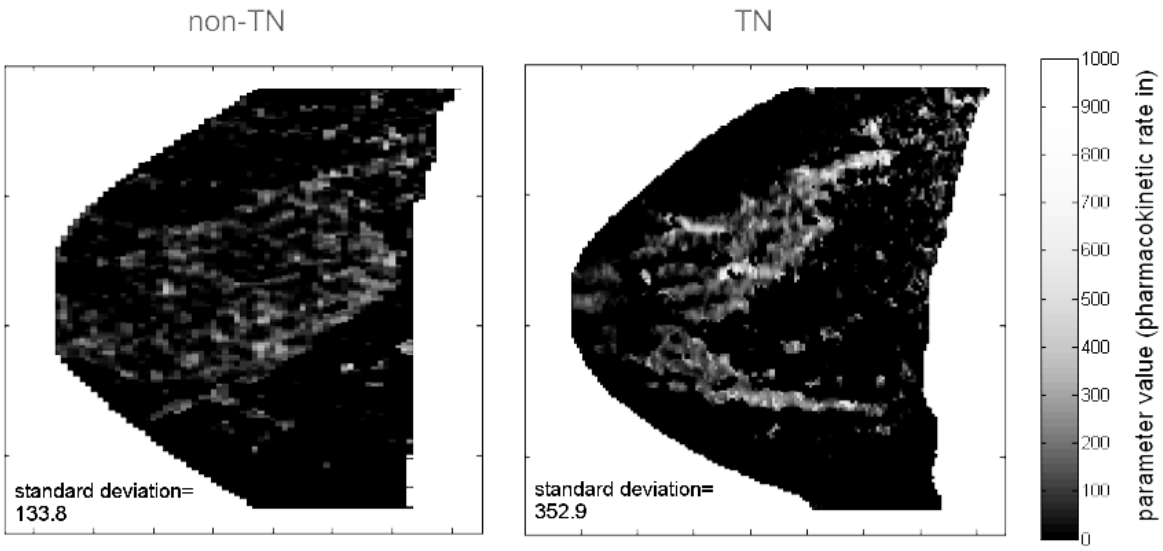

Figure 4.4: Examples of 'parenchyma rate in' parameter (also Figure 4.3b) maps from a non-triple-negative (non-TN) patient (left) and a TN patient (right) illustrating the difference of a statistical texture feature between members of the two groups in image form. Slices of the 'parenchyma rate in' parameter map void of tumor tissue are presented in the sagittal plane. It is evident the variation of this background parenchymal enhancement texture feature's value is greater in TN cancers, where standard deviation is markedly higher at 352.9 as opposed to 133.8 in the non-TN patient.

4.3.3 Clustering

A graphical heatmap representation of clustering results can be seen in Figure 4.5. Unsupervised *k*-means clustering of cases into two partitions (highlighted orange and turquoise) based on BPE texture features resulted in TN breast cancers showing a much higher presence in one partition than the other, with 9 of 11 cases clustering together (left, highlighted orange bar). Values are represented as *z*-scores illustrating the distributions of each feature and the signatures of features across cases.

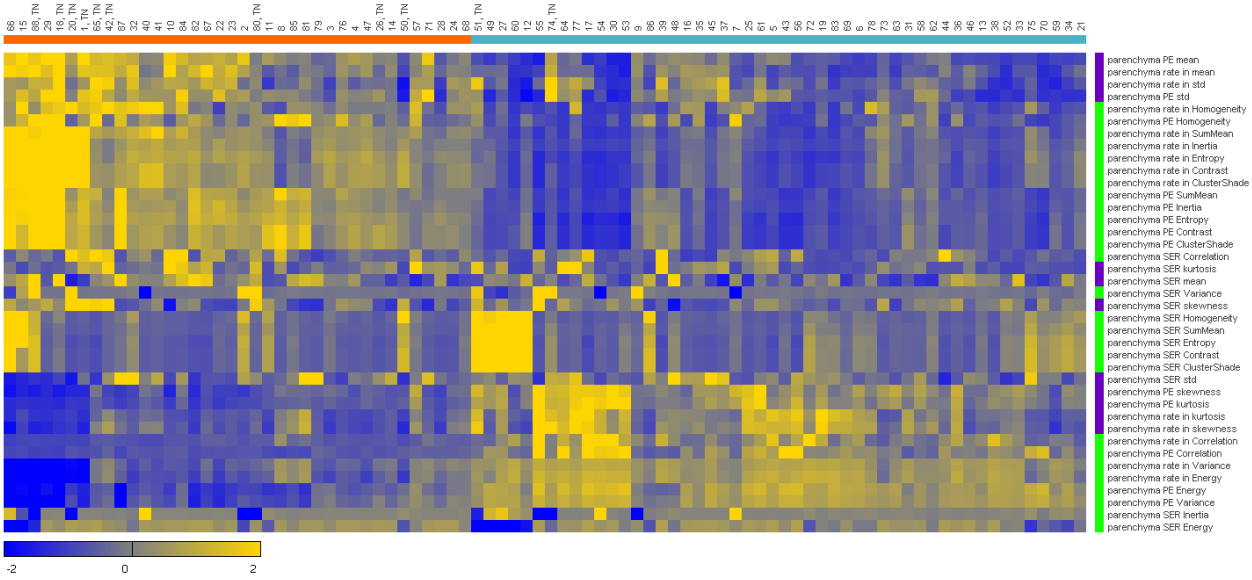

Figure 4.5: Unsupervised *k*-means clustering of breast cancer patients (n=88) on the x-axis and quantitative background parenchymal enhancement (BPE) feature expression (n=39) on y-axis (as z-scores, with scale at bottom left. std=standard deviation). Correspondence of patient groups with similar radiomic expression patterns can be seen where the majority of triple-negative (TN) breast cancers have grouped together in the left cluster (9 of 11 TN in partition highlighted orange at top left) due to association of the BPE heterogeneity feature signatures. 1st order statistical texture features are highlighted as purple and similarly 2nd order statistical texture features are green at right indicating correspondence of feature groups with clustered expression patterns.

4.4 Discussion

Our study demonstrates that differentiation of breast cancer subtype with DCE-MRI can be improved by CAD systems exploiting features of the surrounding parenchyma tissue. In predictive classification models based on imaging features of the tumor, we see performance metrics on the order with the current state of the art in the 0.7-0.8 range (128,129). Based on our results, adding quantitative imaging features of BPE greatly improves the discriminative ability of such prediction models, bringing performance up to 90.0% accuracy and AUC up to 0.883. Unsupervised clustering of BPE texture features into two partitions also revealed a significant association between BPE heterogeneity and TN status. 9 of 11 TN cancers in our study grouped together in one partition based only on

BPE features, reinforcing the notion that BPE heterogeneity on DCE-MRI has a strong relationship with TN breast cancers

To take the functional portion of breast tissue into consideration with the radiomics approach of extracting comprehensive amounts of imaging features appears to be an improvement over conventional methods of tumor-based image phenotyping. While the standard tool to prove subtype in breast cancer remains to be tissue biopsy, we add to the growing understanding of imaging's clinical importance and expand the role of MRI in more personal approaches of breast cancer diagnosis. Beyond considering information from a limited portion of the tumor with its inherent issues with false negatives especially when it comes to large or heterogeneous targets (113), imaging has the ability to provide information on entire tissues and their heterogeneity in a non-invasive manner. This ability appears valuable clinically in differentiating molecular subtypes of breast cancer. To our knowledge, this is the first study to demonstrate that quantitative texture features of BPE extracted from routine MRI are strongly predictive of TN breast cancers.

Our findings appear consistent with recent developments toward the tumor's local environment being gradually recognized as a key contributor in breast cancer progression and aggressiveness (12,13). Likewise, Pathak et al. (22) demonstrated that *in vivo* MRI specifically, could non-invasively monitor changes in tumor microenvironment, which could predict the cancer's ability to metastasize.

Our findings also appear consistent with recent works showing an association between BPE and breast cancer diagnoses (16,21). Existing evidence has linked increased BPE levels with greater hormonal activity, particularly estrogen (150–152), and these two studies suggest BPE could be a stronger predictor of breast cancer risk and potentially serve as an imaging biomarker of estrogen responsive malignant transformation. King et al. allude to presence of cancer having some systemic effects causing increased BPE and Dontchos et al. elaborate on this further, acknowledging the

possibility BPE is a marker of physiologically active tissue more prone to tumorigenesis. Besides being consistent with associations found in our study, such concepts appear to be supported by work linking local inflammation and breast cancer transformation (153). These findings, in support of our own, suggest that BPE could potentially help physicians better tailor screening and management strategies with breast cancer. This is important as we move into an era of more personalized approaches to medicine.

Furthermore, in a 2014 radiogenomic study, Mazurowski et al. reported on the relationship between MRI enhancement dynamics of the tumor and parenchyma to LumB cancer (154). Though we did not include the same measure explicitly in our analysis, our findings reflect tumor and parenchyma enhancement characteristics both play significant roles in differentiating LumB cancers, in our case against TN cancer.

Our findings appear to be inconsistent with those of Ahn et al. (155), who concluded no association between BPE and aggressiveness of the primary cancer. Though in their study, the method of BPE quantification differed and only postmenopausal woman were enrolled, our findings indicate BPE is predictive of the more aggressive TN cancers against other subtypes. We attribute our discovery to having been able to capture functional activity of a tumor's active microenvironment related to tumor progression (12–15) more broadly via extensive measures of enhancement texture.

It is perhaps interesting to note the presence of 'mean of tumor SER' as important in differentiating all other, ER+, and PR+ cancers against TN in tumor feature-based models. In 2011, Arasu et al. showed use of SER volume parameters on MRI were significantly associated with malignancy and improved diagnostic specificity without affecting sensitivity (138). The following year Hylton et al. showed tumor SER as a stronger predictor of pathologic response to neoadjuvant chemotherapy than clinical assessment (156). We elaborate on the utility of volumetric tumor SER in predicting molecular subtype of breast cancer.

It is also perhaps interesting to discuss the presence of breast density as important in differentiating ER+ and PR+ cancers against TN in tumor feature-based models. As it has been shown to be largely in agreement with mammographically-derived density measures (95), MRI-derived density also appears to be associated with breast cancer subtypes and tumor aggressiveness (157). Though the predictive value of this specific feature in our study appears to be outweighed by that of BPE texture when available (breast density was no longer selected in models which also made use of parenchyma enhancement features in our study), it is apparent quantitative measures relating to parenchyma tissue may have prognostic value with breast cancer subtype on DCE-MRI. Bearing in mind that breast density is a limiting factor for cancer detection using mammography (18,73) and dense-breasted women are currently recommended to be stratified for supplemental screening with MRI to compensate for this (158,159) further emphasizes the potential value of considering the parenchyma in diagnostic imaging and predictive modeling of breast cancer.

Compared with another recent study on identifying TN cancers using quantitative image analysis by Agner et al. (129), our work has several strengths. First, we investigated not only tumor features but also BPE features for predicting TN breast cancers. Second, our imaging features were obtained by 3D semi-automated segmentation of the tumor while Agner's approach was based on 2D manual segmentation. Semi-automated image analysis can reduce inter-observer variations and be scaled up relatively easily (160). Both of which are critical components of the radiomics approach. Finally, using the increased signal-to-noise ratio of 3.0T MRI exclusively may have also contributed to the improved classification performance of our study.

Our study has several limitations as well. First, it is a relatively small, retrospective study. Larger prospective validation studies are warranted to confirm these findings and determine potential implications. Also, our method for semi-automated tumor segmentation

was at times imperfect. Expert review of each MRI slice and appropriate correction was performed to best separate tumor from the background parenchyma, but the technique leaves some room for improvement. Lastly, due to the relatively large number of image features and exponentially increasing computing constraints involved, it was not within the scope of this study to perform exhaustive searches for optimal combinations of features. Here we adopted an efficient, locally optimal selection technique (forward floating search), that takes interaction between variables into account.

4.5 Conclusion

In conclusion, we demonstrated that quantitative image phenotyping of breast tumors and their surrounding parenchyma on DCE-MRI could distinguish TN breast cancers from other subtypes with higher accuracy than considering characteristics of the tumor alone. This is due to heterogeneity of background parenchymal enhancement characterized by texture on DCE-MRI being strongly associated with TN cancers. Considering heterogeneity of the tissue surrounding cancer in addition to the cancer itself could make for more sensitive and comprehensive differentiation of breast cancer subtype.

Chapter 5 Summary

It appears key characteristics of breast parenchyma tissue imaging act as clinically meaningful biomarkers for breast cancer. We reinforce previous findings describing the importance of considering the microenvironment in which cancer may develop for predicting cancer outcomes with our own. During the course of this research, the following has been achieved:

* Automatic measures of breast parenchyma tissue volume from 2-dimensional imaging (FFDM) are shown to be in substantial agreement with MRI as a standard for 3-dimensional anatomical imaging and could be used in clinical practice to enhance risk assessment and prevention of breast cancer.
* ANN modeling is shown to extract reasonable measures of breast parenchyma tissue volume from FFDM, as validated with phantoms, with existing measures of breast tissue density, and with classical biomarkers of breast cancer risk.
* Considering a breast tumor's surrounding parenchyma on DCE-MRI for radiomic image phenotyping is shown to provide useful information for identifying triple-negative breast cancers. Heterogeneity of Breast Parenchymal Enhancement (BPE) appears to add value to differentiation models of breast cancer subtype.

Exploiting radiomics and machine learning methods is also shown to be beneficial. Accurate and reproducible discovery and measurement of cancer imaging biomarkers are facilitated with their use. Hence, the presence of such methods and other bioinformatics tools should only increase in future practice of radiology and medicine. As computing and imaging technologies advance over time, new approaches that are able to tease apart the complicated relationships between imaging surrogates should elucidate better predictors of clinical outcomes. Data mining and knowledge extraction of imaging and medical

information will enable doctors and researchers to view the ever growing amount and richness of data in an increasingly meaningful way. Standardizations such as quantitative imaging protocols and analysis algorithms will surely enable more accurate predictions made by computer systems in the effort toward supporting decisions made by clinicians as medical domain experts. Advances in these fields themselves will undoubtedly further enable new capabilities. Such efforts will inevitably find value toward precision medicine, and their presence should only increase in radiology as is already taking place.

Though meaningfully changing healthcare is difficult, progress facilitated by new techniques, subsequently experienced by patients, is being gradually realized and is tremendously exciting.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359–86.
2. Peter, B. Bernard L. World cancer report. 2008.
3. Chu KC, Smart CR, Tarone RE. Analysis of Breast Cancer Mortality and Stage Distribution by Age for the Health Insurance Plan Clinical Trial. JNCI J Natl Cancer Inst. 1988 Sep 21;80(14):1125–32.
4. Alexander FE, Anderson TJ, Brown HK, Forrest APM, Hepburn W, Kirkpatrick AE, Muir BB, Prescott RJ, Smith A. 14 years of follow-up from the Edinburgh randomised trial of breast- cancer screening. Lancet. 1999;353(9168):1903–8.
5. Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: Updated overview of the Swedish randomised trials. Lancet. 2002;359(9310):909–19.
6. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AMF, Chen THH. The randomized trials of breast cancer screening: What have we learned? Vol. 42, Radiologic Clinics of North America. WB Saunders Company; 2004. p. 793–806.
7. Elmore J, Armstrong K, Lehman C, Fletcher S. Screening for breast cancer. J Am Med Assoc. 2005;293(10):1245–56.
8. Peters NHGM, Borel Rinkes IHM, Zuithoff NP a, Mali WPTM, Moons KGM, Peeters PHM. Meta-analysis of MR imaging in the diagnosis of breast lesions. Radiology. 2008;246(1):116–24.
9. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mutvihill JJ. Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. J Natl Cancer Inst. 1989;1879–86.
10. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med. 2004 Apr 15;23(7):1111–30.
11. Engmann N, Golmakani M, Miglioretti D, Sprague B, Kerlikowske K, Consortium for the Breast Cancer Surveillance. Population-attributable risk proportion of clinical risk factors for breast cancer. JAMA Oncol. 2017 Feb 2;
12. Polyak K. Breast cancer: origins and evolution. J Clin Invest. 2007;17:3155–63.
13. Quail D, Joyce J. Microenvironmental regulation of tumor progression and metastasis. Nat Med. 2013;19(11):1423–37.
14. Bissell MJ, Radisky DC, Rizki A, Weaver VM, Petersen OW. The organizing principle: Microenvironmental influences in the normal and malignant breast. Vol. 70, Differentiation. 2002. p. 537–46.

15. Chaudhuri O, Koshy ST, Branco da Cunha C, Shin J-W, Verbeke CS, Allison KH, Mooney DJ. Extracellular matrix stiffness and composition jointly regulate the induction of malignant phenotypes in mammary epithelium. Nat Mater. 2014;

16. King V, Brooks J, Bernstein J, Reiner A, Pike M, Morris E. Background Parenchymal Enhancement at Breast MR Imaging and Breast Cancer Risk. Radiology. 2011;260(1):50–60.

17. Giger ML, Karssemeijer N, Schnabel JA. Breast Image Analysis for Risk Assessment, Detection, Diagnosis, and Treatment of Cancer. Annu Rev Biomed Eng. 2013 Jan;15(1):327–57.

18. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, Jong RA, Hislop G, Chiarelli A, Minkin S, Yaffe MJ. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007 Jan 18;356(3):227–36.

19. Shepherd JA, Kerlikowske K, Ma L, Duewer F, Fan B, Wang J, Malkov S, Vittinghoff E, Cummings SR. Volume of mammographic density and risk of breast cancer. Cancer Epidemiol Biomarkers Prev. 2011;20(7):1473–82.

20. Eng A, Gallant Z, Shepherd J, McCormack V, Li J, Dowsett M, Vinnicombe S, Allen S, Dos-Santos-Silva I. Digital mammographic density and breast cancer risk: a case-control study of six alternative density assessment methods. Breast Cancer Res. 2014;16(5):439.

21. Dontchos BN, Lam DL, Scheel JR. Are Qualitative Assessments of Background Parenchymal Enhancement, Amount of Fibroglandular Tissue on MR Images, and Mammographic Density Associated with Breast Cancer Risk?1. Radiology. 2015;276(2):371–80.

22. Pathak AP, McNutt S, Shah T, Wildes F, Raman V, Bhujwalla ZM. In Vivo “MRI Phenotyping” Reveals Changes in Extracellular Matrix Transport and Vascularization That Mediate VEGF-Driven Increase in Breast Cancer Metastasis. PLoS One. 2013;8(5):1–11.

23. Dromain C, Boyer B, Ferre R. Computed-aided diagnosis (CAD) in the detection of breast cancer. Eur J Radiol. 2013;

24. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2015;0(0):151169.

25. Lambin P, Rios-Velazquez E. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;

26. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

27. Wang S, Summers RM. Machine learning and radiology. Med Image Anal. Elsevier B.V.; 2012;16(5):933–51.

28. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep. Nature Publishing Group; 2015;5:13087.

29. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521.

30. Court LE, Fave X, Mackin D, Lee J, Yang J, Zhang L. Computational resources for radiomics. 2016;5(I):340–8.

31. Kansagra AP, Yu J-PJ, Chatterjee AR, Lenchik L, Chow DS, Prater AB, Yeh J, Doshi AM, Hawkins CM, Heilbrun ME, Smith SE, Oselkin M, Gupta P, Ali S. Big Data and the Future of Radiology Informatics. Acad Radiol. The Association of University Radiologists; :1–13.

32. Wolfe JN. Breast patterns as an index of risk for developing breast cancer. Am J Roentgenol. American Roentgen Ray Society; 1976 Jun 1;126(6):1130–7.

33. Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer. Wiley Subscription Services, Inc., A Wiley Company; 1976;37(5):2486–92.

34. Carney P, Miglioretti D, Yankaskas B. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med. 2003 Feb 4;138(3):168–75.

35. Kerlikowske K, Hubbard R, Miglioretti D. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: A cohort study. Ann Intern Med. 2011 Oct 18;155(8):493–502.

36. Chen JH, Chang YC, Chang D, Wang YT, Nie K, Chang RF, Nalcioglu O, Huang CS, Su MY. Reduction of breast density following tamoxifen treatment evaluated by 3-D MRI: Preliminary study. Magn Reson Imaging. 2011;29(1):91–8.

37. Chlebowski RT, McTiernan A. Biological Significance of Interventions That Change Breast Density. JNCI J Natl Cancer Inst. 2003 Jan 1;95(1):4–5.

38. Chow CK, Venzon D, Jones EC, Premkumar A, O'Shaughnessy J, Zujewski J. Effect of tamoxifen on mammographic density. Cancer Epidemiol Biomarkers Prev. 2000;9(9):917–21.

39. Cuzick J, Warwick J, Pinney E, Warren RML, Duffy SW. Tamoxifen and Breast Density in Women at Increased Risk of Breast Cancer. JNCI J Natl Cancer Inst. 2004;96(8):621–8.

40. Woolcott CG, Conroy SM, Nagata C, Ursin G, Vachon CM, Yaffe MJ, Pagano IS, Byrne C, Maskarinec G. Methods for assessing and representing mammographic density: an analysis of 4 case-control studies. Am J Epidemiol. 2014 Jan 15;179(2):236–44.

41. Cuzick J, Warwick J, Pinney E, Duffy SW, Cawthorn S, Howell A, Forbes JF, Warren RML. Tamoxifen-induced reduction in mammographic density and breast cancer risk reduction: A nested case-control study. J Natl Cancer Inst. 2011;103(9):744–52.

42. American College of Radiology. ACR Breast Imaging Reporting and Data System (BI-RADS) Atlas. 5th Editio. Reston, VA: American College of Radiology; 2013.

43. Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA, Tice JA, Buist DS, Geller BM, Rosenberg R, Yankaskas BC, Kerlikowske K. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer

Inst. 2006 Sep 6;98(17):1204–14.

44. Tice J, Cummings S, Smith-Bindman R, Ichikawa L, Barlow W, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med. 2008;148:337–47.

45. Leung J, Fan B, Kerlikowske K, Chen J, Shepherd J. Single X-ray absorptiometry for measuring breast density compared with visual inspection by radiologists: inter- and intraobserver variability. Am J Roentgenol. 2006;186(A29).

46. Kerlikowske K, Ichikawa L, Miglioretti DL, Buist DSM, Vacek PM, Smith-Bindman R, Yankaskas B, Carney PA, Ballard-Barbash R. Longitudinal measurement of clinical mammographic breast density to improve estimation of breast cancer risk. J Natl Cancer Inst. 2007;99(5):386–95.

47. Cummings SR, Tice JA, Bauer S, Browner WS, Cuzick J, Ziv E, Vogel V, Shepherd J, Vachon C, Smith-Bindman R, Kerlikowske K. Prevention of breast cancer in postmenopausal women: Approaches to estimating and reducing risk. J Natl Cancer Inst. 2009;101(6):384–98.

48. McCormack V, Silva I dos S. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2006;

49. Kopans DB. Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. Radiology. 2008;246(2):348–53.

50. A.Pietrobelli, C. F, Z. W, B. HS, Pietrobelli A, Formica C, Wang Z, Heymsfield SB. Dual-energy X-ray absorptiometry body composition model: review of physical concepts. Am J Physiol. 1996;271(6):E941-51.

51. Kelly TL, Berger N, Richardson TL. DXA body composition: Theory and practice. In: Applied Radiation and Isotopes. 1998. p. 511–3.

52. Highnam R, Brady M, Yaffe M. Robust breast composition measurement-VolparaTM. Lect Notes Comput Sci Digit Mammogr. 2010;342–9.

53. Shepherd JA, Herve L, Landau J, Fan B, Kerlikowske K, Cummings SR. Novel use of single X-ray absorptiometry for measuring breast density. Technol Cancer Res Treat. 2005;4(2):173–82.

54. Pawluczyk O, Augustine BJ, Yaffe MJ, Rico D, Yang J, Mawdsley GE, Boyd NF. A volumetric method for estimation of breast density on digitized screen-film mammograms. Med Phys. 2003;30(3):352.

55. van Engeland S, Snoeren PR, Huisman H, Boetes C, Karssemeijer N. Volumetric breast density estimation from full-field digital mammograms. IEEE transactions on medical imaging. 2006.

56. Klifa C, Carballido-Gamio J, Wilmes L, Laprie A, Lobo C, Demicco E, Watkins M, Shepherd J, Gibbs J, Hylton N. Quantification of breast tissue index from MR data using fuzzy clustering. Conf Proc IEEE Eng Med Biol Soc. 2004;

57. Wu S, Weinstein S, Keller B, Conant E, Kontos D. Fully-automated fibroglandular tissue segmentation in breast MRI. Breast Imaging. 2012;244–51.

58. Klifa C, Carballido-Gamio J, Wilmes L, Laprie A, Shepherd J, Gibbs J, Fan B, Noworolski S, Hylton N. Magnetic resonance imaging for secondary assessment of breast density in a high-risk cohort. Magn Reson Imaging. 2010;28:8–15.

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

60. Hartman K, Highnam R, Warren R, Jackson V. Volumetric assessment of breast tissue composition from FFDM images. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2008. p. 33–9.

61. Aitken Z, McCormack VA, Highnam RP, Martin L, Gunasekara A, Melnichouk O, Mawdsley G, Peressotti C, Yaffe M, Boyd NF, Silva IDS. Screen-film mammographic density and breast cancer risk: A comparison of the volumetric standard mammogram form and the interactive threshold measurement methods. Cancer Epidemiol Biomarkers Prev. 2010;19(2):418–28.

62. Heine JJ, Cao K, Rollison DE. Calibrated Measures for Breast Density Estimation. Acad Radiol. 2011;18(5):547–55.

63. Shepherd JA, Herve L, Landau J, Fan B, Kerlikowske K, Cummings SR. Clinical comparison of a novel breast DXA technique to mammographic density. Med Phys. 2006;33(5):1490.

64. Boyd N, Martin L, Chavez S, Gunasekara A, Salleh A, Melnichouk O, Yaffe M, Friedenreich C, Minkin S, Bronskill M. Breast-tissue composition and other risk factors for breast cancer in young women: a cross-sectional study. Lancet Oncol. 2009;10(6):569–80.

65. Khazen M, Warren RML, Boggis CRM, Bryant EC, Reed S, Warsi I, Pointon LJ, Kwan-Lim GE, Thompson D, Eeles R, Easton D, Evans DG, Leach MO. A pilot study of compositional analysis of the breast and estimation of breast mammographic density using three-dimensional T1-weighted magnetic resonance imaging. Cancer Epidemiol biomarkers Prev. 2008;17(9):2268–74.

66. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.

67. Kallenberg MGJ, van Gils CH, Lokate M, den Heeten GJ, Karssemeijer N. Effect of compression paddle tilt correction on volumetric breast density estimation. Phys Med Biol. 2012;57:5155–68.

68. Kontos D, Xing YY, Bakic PR, Conant EF, Maidment ADA. A comparative study of volumetric breast density estimation in digital mammography and magnetic resonance imaging: results from a high-risk population. In: SPIE 7624 Medical Imaging 2010: Computer-Aided Diagnosis. 2010. p. 762409.

69. Katanoda K, Hori M, Matsuda T, Shibata A, Nishino Y, Hattori M, Soda M, Ioka A, Sobue T, Nishimoto H. An updated report on the trends in cancer incidence and mortality in Japan, 1958-2013. Jpn J Clin Oncol. 2015 Apr;45(4):390–401.

70. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Vol. 132, Breast Cancer Research and Treatment. 2012. p. 365–77.

71. Santen RJ, Boyd NF, Chlebowski RT, Cummings S, Cuzick J, Dowsett M, Easton D, Forbes JF, Key T, Hankinson SE, Howell A, Ingle J. Critical assessment of new risk factors for breast cancer: Considerations for development of an improved risk prediction model. Vol. 14, Endocrine-Related Cancer. 2007. p. 169–87.

72. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ. Mammographic densities and breast cancer risk. Vol. 7, Cancer Epidemiology Biomarkers and Prevention. 1998. p. 1133–44.

73. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, White E. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst. 2000;

74. Obenauer S, Sohns C, Werner C, Grabbe E. Impact of breast density on computer-aided detection in full-field digital mammography. J Digit Imaging. 2006;19(3):258–63.

75. Brem RF, Hoffmeister JW, Rapelyea JA, Zisman G, Mohtashemi K, Jindal G, Disimio MP, Rogers SK. Impact of breast density on computer-aided detection for breast cancer. AJR Am J Roentgenol. 2005;184(2):439–44.

76. Masarwah A, Auvinen P, Sudah M. Very low mammographic breast density predicts poorer outcome in patients with invasive breast cancer. Eur Radiol. 2015;1875–82.

77. Ginsburg OM, Martin LJ, Boyd NF. Mammographic density, lobular involution, and risk of breast cancer. Br J Cancer. 2008;99(9):1369–74.

78. McCormack VA, Perry NM, Vinnicombe SJ, Dos Santos Silva I. Changes and tracking of mammographic density in relation to Pike's model of breast tissue aging: A UK longitudinal study. Int J Cancer. 2010;127(2):452–61.

79. Cuzick J, Warwick J, Pinney E, Duffy SW, Cawthorn S, Howell A, Forbes JF, Warren RML. Tamoxifen-induced reduction in mammographic density and breast cancer risk reduction: a nested case-control study. J Natl Cancer Inst. 2011 May 4;103(9):744–52.

80. Byng J, Boyd N, Fishell E, Jong R, Yaffe M. The quantitative analysis of mammographic densities. Phys Med Biol. 1994;1629.

81. Heine JJ, Carston MJ, Scott CG, Brandt KR, Wu F-F, Pankratz VS, Sellers TA, Vachon CM. An automated approach for estimation of breast density. Cancer Epidemiol Biomarkers Prev. 2008 Nov;17(11):3090–7.

82. Alonzo-Proulx O, Jong RA, Yaffe MJ. Volumetric breast density characteristics as determined from digital mammograms. Phys Med Biol. 2012 Nov 21;57(22):7443–57.

83. Oliver A, Tortajada M, Llado X, Freixenet J, Ganau S, Tortajada L, Vilagran M, Sentis M, Marti R. Breast Density Analysis Using an Automatic Density Segmentation Algorithm. J Digit Imaging. 2015;28(5):604–12.

84. Dong M, Lu X, Ma Y, Guo Y, Ma Y, Wang K. An Efficient Approach for Automated Mass Segmentation and Classification in Mammograms. J Digit Imaging. 2015;28(5):613–25.

85. Abdel-Nasser M, Rashwan HA, Puig D, Moreno A. Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern. Expert Syst Appl. Elsevier Ltd.; 2015;42(24):9499–511.

86. Jiang J, Trundle P, Ren J. Medical image analysis with artificial neural networks. Comput Med Imaging Graph. Elsevier Ltd; 2010;34(8):617–31.

87. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks. 1989;2(5):359–66.

88. Markey MK, Tourassi GD, Margolis M, DeLong DM. Impact of missing data in evaluating artificial neural networks trained on complete data. Comput Biol Med. 2006 May;36(5):516–25.

89. Machida Y, Tozaki M, Shimauchi A, Yoshida T. Breast density: the trend in breast cancer screening. Breast Cancer. 2015;253–61.

90. Nagao Y, Kawaguchi Y, Sugiyama Y, Saji S, Kashiki Y. Relationship between mammographic density and the risk of breast cancer in Japanese women: a case-control study. Breast Cancer. 2003;10(3).

91. Nagata C, Matsubara T, Fujita H, Nagao Y, Shibuya C, Kashiki Y, Shimizu H. Mammographic density and the risk of breast cancer in Japanese women. Br J Cancer. 2005;92(12):2102–6.

92. Kotsuma Y, Tamaki Y, Nishimura T, Tsubai M, Ueda S, Shimazu K, Jin Kim S, Miyoshi Y, Tanji Y, Taguchi T, Noguchi S. Quantitative assessment of mammographic density and breast cancer risk for Japanese women. Breast. 2008 Feb;17(1):27–35.

93. Malkov S, Wang J, Duewer F, Shepherd JA. A calibration approach for single-energy x-ray absorptiometry method to provide absolute breast tissue composition accuracy for the long term. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2012. p. 769–74.

94. Hauge IHR, Hogg P, Szczepura K, Connolly P, McGill G, Mercer C. The readout thickness versus the measured thickness for a range of screen film mammography and full-field digital mammography units. Med Phys. 2012;39(1):263–71.

95. Wang J, Azziz A, Fan B, Malkov S, Klifa C, Newitt D, Yitta S, Hylton N, Kerlikowske K, Shepherd JA. Agreement of mammographic measures of volumetric breast density to MRI. PLoS One. 2013 Jan;8(12):e81653.

96. Vachon CM, Vachon CM, Kuni CC, Kuni CC, Anderson K, Anderson K, Anderson E, Anderson E, Sellers T a, Sellers T a, Foundation M, Foundation M, Research HS, Research

HS, Sw FS, Sw FS, Clinic M, Clinic M, Health P, et al. Association of mammographically defined percent breast density with epidemiologic risk factors for breast cancer (United States). Cancer Cause Control. 2000;11:1–10.

97. Li J, Szekely L, Eriksson L. High-throughput mammographic-density measurement: A tool for risk prediction of breast cancer. Breast Cancer Res. 2012;

98. Nickson C, Arzhaeva Y, Aitken Z, Elgindy T, Buckley M, Li M, English DR, Kavanagh AM. AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes. Breast Cancer Res. 2013 Sep 11;15(5):R80.

99. Seo JM, Ko ES, Han B-K, Ko EY, Shin JH, Hahn SY. Automated volumetric breast density estimation: a comparison with visual assessment. Clin Radiol. 2013 Jul;68(7):690–5.

100. Ding H, Johnson T, Lin M, Le HQ, Ducote JL, Su M-Y, Molloi S. Breast density quantification using magnetic resonance imaging (MRI) with bias field correction: a postmortem study. Med Phys. 2013;40(12):122305.

101. Wu S, Weinstein SP, Conant EF, Kontos D. Automated fibroglandular tissue segmentation and volumetric density estimation in breast MRI using an atlas-aided fuzzy C-means method. Med Phys. 2013 Dec;40(12):122302.

102. Pertuz S, McDonald ES, Weinstein SP, Conant EF, Kontos D. Fully Automated Quantitaive Estimation of Volumetric Breast Density from Digital Breast Tomosynthesis Images : Preliminary Results and Comparison with Digital Mammography and MR Imaging. Radiology. 2016;0(0):1–10.

103. Maskarinec G, Nagata C, Shimizu H, Kashiki Y. Comparison of mammographic densities and their determinants in women from Japan and Hawaii. Int J Cancer. 2002;102(1):29–33.

104. Chen Z. Does Mammographic Density Reflect Ethnic Differences in Breast Cancer Incidence Rates? Am J Epidemiol. 2004 Jan 15;159(2):140–7.

105. Tamaki Y, Kotsuma Y, Miyoshi Y, Noguchi S. Breast cancer risk assessment for possible tailored screening for Japanese women. In: Breast Cancer. 2009. p. 243–7.

106. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012 Oct 4;490(7418):61–70.

107. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Langerød A, Green A, Provenzano E, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012 Jun 21;486(7403):346–52.

108. Goldhirsch A, Winer E, Coates A, Gelber R, Piccart-Gebhart M, Thurlimann B, Senn H. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol. 2013;24:2206–23.

109. Haffty BG, Yang Q, Reiss M, Kearney T, Higgins S a, Weidhaas J, Harris L, Hait W,

Toppmeyer D. Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. J Clin Oncol. 2006 Dec 20;24(36):5652–7.

110. Reis-Filho JS, Tutt ANJ. Triple negative tumours: a critical review. Histopathology. 2008 Jan;52(1):108–18.

111. Foulkes W, Smith I, Reis-Filho JS. Triple-negative breast cancer. N Engl J Med. 2010;363(20):1938–48.

112. Voduc KD, Cheang MCU, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. J Clin Oncol. 2010 Apr 1;28(10):1684–91.

113. Harris JR, Lippman ME, Morrow M, Osborne CK. Diseases of the breast. 5th ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health; 2014.

114. Lee SH, Cho N, Kim SJ, Cha JH, Cho KS, Ko ES, Moon WK. Correlation between high resolution dynamic MR features and prognostic factors in breast cancer. Korean J Radiol. 2008;9(1):10–8.

115. Uematsu T, Kasami M, Yuen S. Triple-Negative Breast Cancer: Correlation between MR Imaging and Pathologic Findings. Radiology. 2009;250(3):638–47.

116. Youk JH, Son EJ, Chung J, Kim J-A, Kim E-K. Triple-negative invasive breast cancer on dynamic contrast-enhanced and diffusion-weighted MR imaging: comparison with other breast cancer subtypes. Eur Radiol. 2012 Aug;22(8):1724–34.

117. Schmitz AMT, Loo CE, Wesseling J, Pijnappel RM, Gilhuijs KG a. Association between rim enhancement of breast cancer on dynamic contrast-enhanced MRI and patient outcome: impact of subtype. Breast Cancer Res Treat. 2014;148(3):541–51.

118. Goto M, Ito H, Akazawa K, Kubota T, Kizu O, Yamada K, Nishimura T. Diagnosis of breast tumors by contrast-enhanced MR imaging: Comparison between the diagnostic performance of dynamic enhancement patterns and morphologic. J Magn Reson Imaging. 2007 Jan;25(1):104–12.

119. Deurloo E, Muller S, Peterse J, Besnard A, Gilhuijs K. Clinically and mammographically occult breast lesions at MR images: Potential effect of computerized assessment on clinical reading. Radiology. 2005;234:693–701.

120. Meinel L, Stolpen A, Berbaum K, Fajardo L, Reinhardt J. Breast MRI lesion classification: Improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. J Magn Reson Imaging. 2007;25:89–95.

121. Williams T, DeMartini W, Partridge S, Peakcock S, Lehman C. Breast MR Imaging: Computer-aided Evaluation Program for Discriminating Benign from Malignant Lesions. Radiology. 2007;244(1):94–103.

122. Meeuwis C, van de Ven S, Fernandez Gallardo A, van den Bosch M, Mali W, Veldhuis W. Computer-aided detection (CAD) for breast MRI: evaluation of efficacy at 3.0 T. Eur Radiol. 2010;20:522–8.

123. Gillies RJ, Anderson a R, Gatenby R a, Morse DL. The biology underlying molecular imaging in oncology: from genome to anatome and back again. Clin Radiol. The Royal College of Radiologists; 2010 Jul;65(7):517–21.

124. Alic L, Niessen WJ, Veenland JF. Quantification of Heterogeneity as a Biomarker in Tumor Imaging: A Systematic Review. PLoS One. 2014;9(10):e110300.

125. Bhooshan N, Giger M, Jansen S, Li H. Cancerous breast lesions on dynamic contrast-enhanced MR images: computerized characterization for image-based prognostic markers. Radiology. 2010;254(3):680–90.

126. Lee SH, Kim JH, Cho N, Park JS, Yang Z, Jung YS, Moon WK. Multilevel analysis of spatiotemporal association features for differentiation of tumor enhancement patterns in breast DCE-MRI. Med Phys. 2010;37(8):3940–56.

127. Agliozzo S, De Luca M, Bracco C, Vignati A, Giannini V, Martincich L, Carbonaro L, Bert A, Sardanelli F, Regge D. Computer-aided diagnosis for dynamic contrast-enhanced breast MRI of mass-like lesions using a multiparametric model combining a selection of morphological, kinetic, and spatiotemporal features. Med Phys. 2012;39(4):1704–15.

128. Giger M, Li H, Lan L, Abe H, Newstead G. Quantitative MRI Phenotyping of Breast Cancer across Molecular Classification Subtypes. Breast Imaging. 2014;8539:195–200.

129. Agner S, Rosen M, Englander S, Tomaszewski J, Feldman M, Zhang P, Mies C, Schnall M, Madabhushi A. Computerized Image Analysis for Identifying Triple-Negative Breast Cancers and Differentiating Them from Other Molecular Subtypes of Breast Cancer on Dynamic Contrast-enhanced MRI: A Feasibility Study. Radiology. 2014;272(1):91–9.

130. Hambly NM, Liberman L, Dershaw DD, Brennan S, Morris EA. Background parenchymal enhancement on baseline screening breast MRI: Impact on biopsy rate and short-interval follow-up. Am J Roentgenol. 2011;218–24.

131. DeMartini WB, Liu F, Peacock S, Eby PR, Gutierrez RL, Lehman CD. Background parenchymal enhancement on breast MRI: impact on diagnostic performance. AJR Am J Roentgenol. 2012 Apr;198(4):W373-380.

132. Jones EF, Sinha SP, Newitt DC, Klifa C, Kornak J, Park CC, Hylton NM. MRI Enhancement in Stromal Tissue Surrounding Breast Tumors: Association with Recurrence Free Survival following Neoadjuvant Chemotherapy. PLoS One. 2013;8(5).

133. Yoshioka T, Hosoda M, Yamamoto M, Taguchi K, Hatanaka KC, Takakuwa E, Hatanaka Y, Matsuno Y, Yamashita H. Prognostic significance of pathologic complete response and Ki67 expression after neoadjuvant chemotherapy in breast cancer. Breast Cancer. 2013 May 5;

134. Yerushalmi R, Woods R, Ravdin P, Hayes M, Gelmon K. Ki67 in breast cancer: prognostic and predictive potential. Lancet Oncol. 2010;11:174–83.

135. Mann R, Kuhl C, Kinkel K, Boetes C. Breast MRI: guidelines from the European society of breast imaging. Eur Radiol. 2008 Jul;18(7):1307–18.

136. American College of Radiology. ACR practice guideline for the performance of contrast-

enhanced magnetic resonance imaging (MRI) of the breast. 2010;1–8.

137. Ercan G, Whyte P. Digital image processing. 2007.

138. Arasu V, Chen R, Newitt D, Chang C, Tso H, Hylton N, Joe B. Can signal enhancement ratio (SER) reduce the number of recommended biopsies without affecting cancer yield in occult MRI-detected lesions? Acad Radiol. 2011;18(6):716–21.

139. Haralick R, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern. 1973;3(6):610–21.

140. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI. 1995;

141. Liu H, Setiono R. Chi2: Feature selection and discretization of numeric attributes. Proc 7th Int Conf Tools with Artif Intell. 1995;88.

142. Pearl J. Heuristics: intelligent search strategies for computer problem solving. Addison-Wesley; 1985;

143. Platt J. Fast training of support vector machines using sequential minimal optimization. Adv kernel methods. 1999;41–65.

144. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2001.

145. Smialowski P, Frishman D, Kramer S. Pitfalls of supervised feature selection. Bioinformatics. 2010;26(3):440–3.

146. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM Sigkdd Explor Newsl. 2009;11(1):10–8.

147. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognit Lett. Elsevier B.V.; 2010;31(8):651–66.

148. Easton DF, Pooley K a, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Meyer KB, Haiman C a, Kolonel LK, Henderson BE, Le Marchand L, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007;447(7148):1087–93.

149. Perez-Llamas C, Lopez-Bigas N. Gitools: analysis and visualisation of genomic data using interactive heat-maps. PLoS One. 2011;6(5).

150. Kuhl CK, Bieling HB, Gieseke J, Kreft BP, Sommer T, Lutterbey G, Schild HH. Healthy premenopausal breast parenchyma in dynamic contrast-enhanced MR imaging of the breast: normal contrast medium enhancement and cyclical-phase dependency. Radiology. Radiological Society of North America; 1997 Apr 1;203(1):137–44.

151. Müller-Schimpfle M, Ohmenhaüser K, Stoll P, Dietz K, Claussen CD. Menstrual cycle and age: influence on parenchymal contrast medium enhancement in MR imaging of the breast. Radiology. Radiological Society of North America; 1997 Apr 1;203(1):145–9.

152. Delille JP, Slanetz PJ, Yeh ED, Kopans DB, Garrido L. Physiologic changes in breast magnetic resonance imaging during the menstrual cycle: Perfusion imaging, signal enhancement, and influence of the T1 relaxation time of breast tissue. Breast J. 2005;11(4):236–41.

153. Bhatelia K, Singh K, Singh R. TLRs: Linking inflammation and breast cancer. Cell Signal. 2014;26(11):2350–7.

154. Mazurowski M, Zhang J, Grimm L, Yoon S, Silber J. Radiogenomic Analysis of Breast Cancer: Luminal B Molecular Subtype Is Associated with Enhancement Dynamics at MR Imaging. Radiology. 2014;273(2):365–72.

155. Ahn HS, Kim SM, Jang M, Yun B. Quantitative analysis of breast parenchymal background enhancement (BPE) on magnetic resonance (MR) imaging: Association with mammographic breast density and aggressiveness of the primary cancer in postmenopausal women. In 2013.

156. Hylton NM, Blume JD, Bernreuter WK, Pisano ED, Rosen M a., Morris E a., Weatherall PT, Lehman CD, Newstead GM, Polin S, Marques HS, Esserman LJ, Schnall MD. Locally Advanced Breast Cancer: MR Imaging for Prediction of Response to Neoadjuvant Chemotherapy--Results from ACRIN 6657/I-SPY TRIAL. Radiology. 2012;263(3):663–72.

157. Bertrand K a, Tamimi RM, Scott CG, Jensen MR, Pankratz VS, Visscher D, Norman A, Couch F, Shepherd J, Fan B, Chen Y-Y, Ma L, Beck AH, Cummings SR, Kerlikowske K, Vachon CM. Mammographic density and risk of breast cancer by age and tumor characteristics. Breast Cancer Res. 2013 Nov 4;15(6):R104.

158. Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, Morris E, Pisano E, Schnall M, Sener S, Smith RA, Warner E, Yaffe M, Andrews KS, Russell CA. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin. 2007;57(2):75–89.

159. Pediconi F, Catalano C, Roselli A, Dominelli V, Cagioli S, Karatasiou A, Pronio A, Kirchin MA, Passariello R. The Challenge of Imaging Dense Breast Parenchyma: Is Magnetic Resonance Mammography the Technique of Choice? A Comparative Study With X-Ray Mammography and Whole-Breast Ultrasound. Invest Radiol. 2009;44(7).

160. Parmar C, Velazquez ER, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, Mitra S, Shankar BU, Kikinis R, Haibe-Kains B, Lambin P, Aerts HJWL. Robust radiomics feature quantification using semiautomatic volumetric segmentation. PLoS One. 2014;

Selected List of Relevant Works

Book Chapters

- Wang J, Malkov S, Fan B, Shepherd J. Cross-Calibration and Longitudinal Quality Control of Hologic Selenia Full-Field Digital Mammography Systems for Volumetric Breast Density Measurements. Digital Mammography. Springer Berlin/Heidelberg; 2008. p. 740-747.

- Wang J, Azziz A, Newitt D, Joe B, Hylton N, Shepherd J. Comparison of Hologic's Quantra Volumetric Assessment to MRI Breast Density. Breast Imaging. Springer Berlin/Heidelberg; 2012. p. 619-626.

- Wang J, Kato F, Kudo K, Yamashita H, Shirato H. Predicting Triple-Negative Breast Cancer and Axillary Lymph Node Metastasis Using Diagnostic MRI. Breast Imaging. Springer International; 2014. p. 334-340.

Publications

- Wang J, Azziz A, Fan B, Malkov S, Klifa C, Newitt D, Yitta S, Hylton N, Kerlikowske K, Shepherd J. Agreement of Mammographic Measures of Volumetric Breast Density to MRI. PLoS One 2013; (8) 12:e81653.

- Wang J, Kato F, Oyama-Manabe N, Li R, Cui Y, Tha K, Yamashita H, Kudo K, Shirato H. Identifying Triple-Negative Breast Cancer Using Background Parenchymal Enhancement Heterogeneity on Dynamic Contrast-Enhanced MRI: A Pilot Radiomics Study. PloS one 2015; 10(11):e0143308.

- Wang J, Kato F, Yamashita H, Baba M, Cui Y, Li R, Oyama-Manabe N, Shirato H. Automatic Estimation of Volumetric Breast Density Using Artificial Neural Network-Based Calibration of Full-Field Digital Mammography: Feasibility on Japanese Women With and Without Cancer. Journal of Digital Imaging 2016; 30(2):215–227.

Oral Presentations

- Wang J, Malkov S, Fan B, Shepherd J. Cross-Calibration and Longitudinal Quality Control of Hologic Selenia Full-Field Digital Mammography Systems for Volumetric Breast Density Measurements. International Workshop on Digital Mammography 2008, Tucson, AZ.

- Wang J, Kato F, Kudo K, Yamashita H, Shirato H. Heterogeneity of Background Parenchymal Enhancement of MRI Strongly Predictive of Breast Cancer Molecular Subtypes. 100th Radiological Society of North America Annual Meeting 2014, Chicago, IL.

Awards

- Wang J, Kato F, Oyama-Manabe N, Kudo K, Shirato H. Alternative Image Representations of Dynamic Contrast-Enhanced MRI in Breast Cancer. GI-CoRE Medical Science and Engineering Symposium on Radiation Oncology, Biology, and Interventional Radiology 2016 Poster Award, Sapporo, Japan.

Acknowledgements

Throughout this research, my mentors, colleagues, as well as family and friends have been a great guidance and support.

My heartfelt thanks go to my thesis advisor, Professor Hiroyuki Date and my mentor Professor Hiroki Shirato.

Thanks are also due to my other committee members Professors Katsuhiko Ogasawara, Tamotsu Kamishima, and Toru Yamamoto.

This research was begun at the University of California, San Francisco and UCSF Medical Center, where I collaborated with many amazing researchers in the Departments of Radiology and Biomedical Imaging, Medicine and Epidemiology/Biostatistics, as well as Graduate School of Bioengineering:
John Shepherd, Bo Fan, Serghei Malkov, Lionel Herve, Aurelie Laidevant, Joseph Wilson, Fred Duewer, Lucy Zhang, Song Ge, Li Wang, Mary Sherman, Laura Fairbanks, Jessie Landau Goodman, Lorena Marquez, Chyi Huang, Eric Lee, Cassidy Powers, Jenny Sherman, Keyna Chow, Catherine Klifa, David Newitt, Nola Hylton, Ania Aziz, Silaja Yitta, Karla Kerlikowske, and Lin Ma, and other university and hospital staff.

Continuing in Japan, I have also collaborated with many amazing researchers at the Hokkaido University School of Medicine, Hokkaido University Hospital, and Stanford University:
Kohsuke Kudo, Noriko Oyama-Manabe, Fumi Kato, Khin Khin Tha, Noriyuki Fujima, Kenneth Sutherland, Satoshi Yabusaki, Rie Mimura, Yousuke Sakuhara, Takeshi Soyama, Hiroyuki Sugimori, Ayako Yagahara, Makoto Oda, Kinya Ishizaka, Suzuko Mito, Hiroko Yamashita, Motoi Baba, Ruijiang Li, Yi Cui, and other university and hospital staff.

I am forever grateful for all of your encouragement in pursing my research interests and professional growth.

Appendix

Supplemental Table Appendix.S1: Optimal imaging feature distribution and prognostic significance

Details of imaging features selected as most discriminative in prediction models. Feature values (or numbers, with categorical features) are presented with their support vector machine (SVM) weights. Bootstrap estimates of SVM weights indicate prognostic significance, where the larger in absolute magnitude the weight, the more significant the feature (features were normalized with use of SVM classifier, though are presented here in their original scale).

Differentiation task	Model type	Feature	Feature value				SVM weight	
			TN		Alternate subtype			
			mean (n)	std	mean (n)	std	mean	std
TN vs others	tumor	mass enhancement characteristics = heterogenous	2		29		0.110	0.063
		mass enhancement characteristics = n/a	1		10		-0.047	0.098
		mass enhancement characteristics = rim enhancement	7		36		-0.086	0.046
		mass margin = irregular (or spiculated)	9 (0)		35 (30)		0.443	0.147
		mass shape = irregular	0		37		0.237	0.151
		mass shape = lobulated	6		15		-0.074	0.086
		mass shape = oval	3		8		-0.359	0.145
		mass shape = round	0		5		0.218	0.084
		mass size	27.556	17.629	15.538	6.534	-1.377	0.338
		tumor PE std	44.173	20.093	44.043	27.690	0.027	0.098
		tumor rate in Energy	0.133	0.060	0.129	0.069	-0.148	0.139
		tumor rate in Inertia	3.210	0.557	3.190	0.563	0.254	0.188
		tumor SER mean	0.992	0.213	1.027	0.223	0.648	0.140
	tumor + BPE	mass shape = irregular	0		37		0.363	0.120
		mass shape = lobulated	6		15		-0.458	0.114
		mass shape = oval	3		8		-0.589	0.068
		mass shape = round	0		5		0.760	0.076
		parenchyma PE ClusterShade	0.271	0.126	0.166	0.099	-1.096	0.130
		parenchyma PE Homogeneity	13.555	9.716	8.764	7.463	-0.216	0.110
		parenchyma rate in std	218.423	79.792	155.916	39.941	-1.052	0.263
		parenchyma SER skewness	2.479	1.576	0.824	0.704	-2.044	0.176
		parenchyma SER Variance	0.283	0.039	0.251	0.020	-1.261	0.174
TN vs ER+	tumor	breast density	38.095	19.487	29.126	12.121	-0.704	0.268
		mass enhancement characteristics = heterogenous	2		29		0.102	0.061
		mass enhancement characteristics = n/a	1		10		-0.029	0.067

		mass enhancement characteristics = rim enhancement	7		32		-0.083	0.036
		mass margin = irregular (or spiculated)	9 (0)		32 (29)		0.681	0.189
		mass shape = irregular	0		33		0.352	0.089
		mass shape = lobulated	6		15		-0.114	0.086
		mass shape = oval	3		8		-0.374	0.092
		mass shape = round	0		5		0.136	0.081
		mass size	27.556	17.629	15.590	6.484	-1.498	0.235
		tumor PE Inertia	2.858	1.348	3.224	0.978	0.038	0.127
		tumor PE skewness	0.881	1.199	1.147	3.111	0.131	0.118
		tumor rate in Homogeneity	71.528	32.619	103.464	63.964	0.313	0.124
		tumor rate in Inertia	3.210	0.557	3.185	0.570	0.434	0.207
		tumor SER mean	0.992	0.213	1.027	0.202	0.710	0.139
	tumor + BPE	mass shape = irregular	0		33		0.388	0.135
		mass shape = lobulated	6		15		-0.472	0.085
		mass shape = oval	3		8		-0.614	0.100
		mass shape = round	0		5		0.775	0.057
		parenchyma PE ClusterShade	0.271	0.126	0.164	0.097	-1.092	0.120
		parenchyma PE Homogeneity	13.555	9.716	8.544	7.360	-0.193	0.143
		parenchyma rate in std	218.423	79.792	154.280	39.607	-1.096	0.189
		parenchyma SER skewness	2.479	1.576	0.821	0.708	-2.052	0.168
		parenchyma SER Variance	0.283	0.039	0.251	0.020	-1.284	0.172
TN vs PR+	**tumor**	breast density	38.095	19.487	27.856	11.274	-0.754	0.259
		mass shape = irregular	0		29		0.471	0.100
		mass shape = lobulated	6		14		-0.106	0.064
		mass shape = oval	3		7		-0.366	0.061
		mass shape = round	0		5		0.002	0.004
		mass size	27.556	17.629	15.709	6.652	-1.360	0.258
		morphology = mass (or non-mass)	9 (2)		55 (8)		-0.345	0.029
		tumor PE mean	149.628	33.864	156.046	36.605	0.278	0.149
		tumor SER mean	0.992	0.213	1.033	0.203	0.536	0.177
	tumor + BPE	mass shape = irregular	0		29		0.333	0.144
		mass shape = lobulated	6		14		-0.305	0.091
		mass shape = oval	3		7		-0.249	0.107
		mass shape = round	0		5		0.221	0.090
		parenchyma PE Entropy	0.488	0.185	0.317	0.163	-0.774	0.199
		parenchyma rate in std	218.423	79.792	154.959	41.188	-1.052	0.216
		parenchyma SER skewness	2.479	1.576	0.856	0.696	-1.931	0.168
		tumor PE mean	149.628	33.864	156.046	36.605	0.485	0.166
TN vs LumA	**tumor**	mass enhancement characteristics = heterogenous	2		24		0.192	0.137
		mass enhancement characteristics = n/a	1		7		0.238	0.169
		mass enhancement characteristics = rim enhancement	7		13		-0.450	0.107
		mass shape = irregular	0		21		0.386	0.059

		mass shape = lobulated	6		7		-0.358	0.128
		mass shape = oval	3		6		-0.273	0.131
		mass shape = round	0		3		0.245	0.110
		mass size	27.556	17.629	14.568	5.305	-1.428	0.232
		morphology = mass (or non-mass)	9 (2)		37 (8)		-0.534	0.243
		tumor PE Homogeneity	13.555	9.716	34.020	26.408	0.851	0.200
	tumor + BPE	mass shape = irregular	0		21		0.077	0.050
		mass shape = lobulated	6		7		-0.404	0.030
		mass shape = oval	3		6		-0.211	0.097
		mass shape = round	0		3		0.538	0.119
		parenchyma rate in std	218.423	79.792	147.327	40.114	-1.681	0.237
		parenchyma SER skewness	2.479	1.576	0.906	0.689	-1.604	0.181
TN vs LumB	**tumor**	mass margin = irregular (or spiculated)	9 (0)		16 (8)		0.491	0.198
		mass shape = irregular	0		12		0.754	0.240
		mass shape = lobulated	6		8		0.034	0.253
		mass shape = oval	3		2		-0.904	0.437
		mass shape = round	0		2		0.167	0.194
		mass size	27.556	17.629	17.167	7.833	-1.389	0.465
		tumor PE Variance	0.483	0.243	0.395	0.200	-0.505	0.331
	tumor + BPE	parenchyma rate in Energy	0.671	0.173	0.824	0.143	0.637	0.187
		parenchyma SER skewness	2.479	1.576	0.683	0.730	-1.574	0.239
		parenchyma SER Variance	0.283	0.039	0.249	0.005	-1.887	0.190
		tumor PE Variance	0.483	0.243	0.395	0.200	-0.447	0.062

Features are listed alphabetically by differentiation task and model type. TN=Triple-negative, LumA=Luminal A, LumB=Luminal B, SVM=support vector machine, BPE=background parenchymal enhancement, std=standard deviation, n/a=not applicable, PE=percent enhancement, SER=signal enhancement ratio.