Title

Developing Measures of Interval Breast Cancer Risk

Author

Hinton, Benjamin J

Publication Date

2018

Developing Measures of Interval Breast Cancer Risk

by

Benjamin John Hinton

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOENGINEERING

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

AND

UNIVERSITY OF CALIFORNIA, BERKELEY

[redacted]

Dedication

In August 2014, I began this PhD program with a mix of excitement and apprehension. I had just been teaching for two years and the idea of transitioning from teaching Punnett Squares to innovating at a top research institution was intimidating to say the least. Four years later and through a lot of successes, challenges, and memories, I can proudly say that I've both helped spread science to others and help develop it further. This wouldn't have been remotely possible without the various support structures in my life: the people who have taught me the skills and passion for science, the people who have encouraged me through successes and failures, and the people who have given me the personality I needed to be successful. To those people listed below and the innumerable others that have helped, thank you:

- **To my PI, Professor John Shepherd**: Thank you for supporting, challenging, and encouraging me to grow as a researcher and engineer. You've helped me understand how to ask innovative questions, plan and execute long term projects, enjoy the successes, and roll with the setbacks of science.

- **To my wife, Grace:** You are the best thing to happen to me. You helped me enjoy our journey through life and through these four years, and you've always been there for celebrations, adventures and encouragement. I wouldn't have achieved this without you.

- **To my various advisors, Steven Conolly, Bonnie Joe, Karla Kerlikowske, Christian Diederich, Serge Muller, and more:** Thank you for guiding me throughout this time. I learned the value of outside perspective, and conversations with you all taught me how important it is to check in with others in order to ensure the project will be feasible and valuable to the community.

- **To my labmates, Bennett, Jesus, Lisa, Markus, Bo, Leila, En, Nisa, et al.:** Thanks for joining me, teaching me, laughing with me, and supporting me through all the ups and downs through my PhD.

- **To Gary Knoke, Dave Reich, Pat Arndt, Jean Anderegg, Robert Tranquillo:** Thank you for teaching and encouraging me to love learning and love science. You helped transform my curiosity into a career.

- **To my family and the Pearsons:** Thank you for instilling in me the personality, curiosity, and perseverance that I later learned scientists need so much to succeed.

- **To my friends, Tyler, Max, Tylar, Paul, Steve, Bryan et al.:** Thank you for supporting me, laughing with me, and being a part of my life through good and bad.

- **To the various sources of financial support throughout my education: Berlin High School, The National Science Foundation, The National Institutes of Health, The Goldwater Foundation, The University of Minnesota, et al:** Thank you for helping provide a safety net and financial support that I didn't have. You allowed me to focus on school and gave me the space to take a risk on and invest in myself.

Acknowledgements

Parts of this dissertation are reprints of materials that appear in several publications listed below. The co-authors listed in this publication directed and supervised the research that forms the basis for the dissertation. Specific author contributions for each publication are provided.

Publications:

1. B. J. Hinton, B. Fan, B. K. Ng, and J. A. Shepherd, "Dual energy X-ray absorptiometry body composition reference values of limbs and trunk from NHANES 1999–2004 with additional visualization methods," *PLOS ONE*, vol. 12, no. 3, p. e0174180, Mar. 2017.

   JAS, BKN, and BJH designed and conducted the research; BJH, BF, BKN and JAS analyzed data; BJH and JAS drafted the manuscript and had primary responsibility for final content. All authors reviewed and approved the final manuscript.

2. B. J. Hinton, L. Ma, A. P. Mahmoudzadeh, S. Malkov, B. Fan, H. Greenwood, B. Joe, V. Lee, F. Strand, K. Kerlikowske, J. A. Shepherd, "Derived mammographic masking measures based on simulated lesions predict the risk of interval cancer," Medical Physics, *In Review*.

   JAS, KK, and BJH designed and conducted the research. JAS, BJH, SM, and APM developed software to perform analysis. FS, HG, BJ, VL provided feedback on results and potential outputs. LM, BF, APM, KK, and BJH worked to organize data and images. BJH and JAS wrote and edited the manuscript. All authors reviewed and approved the final manuscript.

3. B. J. Hinton, L. Ma, A. P. Mahmoudzadeh, S. Malkov, B. Fan, H. Greenwood, B. Joe, V. Lee, K. Kerlikowske, J. A. Shepherd, "Deep learning networks find unique mammographic differences in previous negative mammograms between interval and screen-detected cancers," Breast Cancer Research, *In Review*.

   JAS, KK, and BJH designed and conducted the research. JAS, BJH, SM, and APM developed software to perform analysis. BJH and JAS worked to develop and optimize the deep learning networks. HG, BJ, KK VL provided feedback on results and potential outputs. LM, BF, APM, KK, and BJH worked to organize data and images. BJH and JAS wrote and edited the manuscript. All authors reviewed and approved the final manuscript.

4. B. J. Hinton, L. Ma, C Hernandez, A. P. Mahmoudzadeh, C. Vachon, C. G. Scott, B. Fan, H. Greenwood, B. Joe, V. Lee, S.J. Winham, K. Kerlikowske, J. A. Shepherd, "Deep learning networks find unique mammographic interval risk and screen-detected risk measures," Breast Cancer Research, *In Review*.

   JAS, and BJH designed and conducted the research. JAS, BJH, SM, and APM developed software to perform analysis. BJH and JAS worked to develop and optimize the deep learning networks. HG, BJ, KK VL, CS, CV provided feedback on results and potential outputs. LM, CH, BF, APM, KK, BJH worked to organize data and images. BJH and JAS wrote and edited the manuscript. All authors reviewed and approved the final manuscript.

5. B. K. Ng, B. J. Hinton, B. Fan, A. Kanaya, J. A. Shepherd, "Clinical anthropometrics and body composition from 3D whole-body surface scans," Eur J Clin Nutr, vol. 70(11), pp. 1265–70, Nov. 2016.

JAS and BKN designed and conducted the research; BKN, BJH and JAS analyzed data; BKN and JAS drafted the manuscript and had primary responsibility for final content. All authors reviewed and approved the final manuscript.

Conference Abstracts:

6. B. J. Hinton, S. Malkov, J. Avila, B. Fan, B. Joe, K. Kerlikowske, L. Ma, A. Mahmoudzadeh, J. Shepherd, "A Measure of Regional Mammographic Masking Based on the CDMAM Phantom. In *International Workshop on Digital Mammography* (pp. 525-531), Jun. 2016.

   JAS, KK, and BJH designed and conducted the research. JAS, BJH, SM, and APM developed software to perform analysis. BJ provided feedback on results and potential outputs. LM, BF, APM, KK, and BJH worked to organize data and images. BJH and JAS wrote and edited the manuscript. All authors reviewed and approved the final manuscript.

7. B. J. Hinton, S. Malkov, J. Avila, B. Fan, B. Joe, K. Kerlikowske, L. Ma, A. Mahmoudzadeh, J. Shepherd, "Deep Learning Methods aid in predicting risk of interval cancer. *14th International Workshop on Breast Imaging (IWBI 2018)*. Vol. 10718. International Society for Optics and Photonics, 2018.

   JAS, KK, and BJH designed and conducted the research. JAS, BJH, SM, and APM developed software to perform analysis. BJH and JAS worked to develop and optimize the deep learning networks. BJ, JIA, KK provided feedback on results and potential outputs. LM, BF, APM, KK, and BJH worked to organize data and images. BJH and JAS wrote and edited the manuscript. All authors reviewed and approved the final manuscript.

8. A.P. Mahmoudzadeh, S. Malkov, B. J. Hinton, B. Sprague, K. Kerlikowske, J.A. Shepherd. (2016, June). Do Women with Low Breast Density Have Regionally High Breast Density?. In *International Workshop on Digital Mammography* (pp. 548-553). Springer, Cham.

   JAS, KK, and APM designed and conducted the research; SM, APM and BS organized data; BJH, SM, APM, and JAS consulted on results and analyses; APM and JAS drafted the manuscript and had primary responsibility for final content. All authors reviewed and approved the final manuscript.

Developing Measures of Interval Breast Cancer Risk

Benjamin John Hinton

Abstract

One in eight women experiences breast cancer in their lives, making it the number one cancer affecting women. Interval cancers are cancers that are found by women in between normal screening mammogram intervals, and often represent a failure mode of mammography screening where it was unable to detect a lesion. While supplemental screening can help prevent these interval cancers, referring all women with interval cancer risk to supplemental screening would lead additional biopsies, stress, and false positives. A need exists to create a specific measure of interval risk that could help identify women that would truly benefit from supplemental screening.

The goal of this dissertation was to develop that measure of interval risk and is centered on the hypothesis that advanced computer vision methods can identify metrics of detectability that quantify risk of interval cancer more effectively than currently used metrics of breast density. I first present methods to better quantify composition of limbs, and trunk, as compositional measures have been shown to relate to cancer risk. I then present an algorithm that directly quantifies mammogram detectability by creating and inserting pseudo-lesions in digital mammograms and summarize its ability to quantify interval risk. I lastly present a method using deep learning to create and train classifier to identify risk of interval cancer. I describe the underlying hypotheses behind these methods, their effectiveness of identifying interval risk compared to current gold standards, strategies for improving their effectiveness, and future steps.

In conclusion, we have developed several methods that help to improve upon current measures of interval cancer risk. Further work to refine and develop these methods could be applied to improve risk models, identify groups of women at high risk of interval cancers, develop software to

aid radiologists, and help radiologists identify women who would benefit from supplemental screening. As a result, these methods may be able to help prevent interval cancers, improve sensitivity of screening, and save lives. The following chapters outline these tools and methods, as well as ways they can be applied.

Abbreviations

BMI	Body Mass Index
BI-RADS	Breast Imaging – Reporting and Data System
CC	Craniocaudal
MLO	Mediolateral Oblique
US	Ultrasound
MRI	Magnetic Resonance Imaging
ACR	American College of Radiology
USPSTF	United States Preventative Services Task Force
OR	Odds Ratio
HRT	Hormone Replacement Therapy
BCSC	Breast Cancer Surveillance Consortium
DCIS	Ductal Carcinoma *In Situ*
SXA	Single X-ray Absorptiometry
CI	Confidence Interval
GLCM	Grey Level Co-occurrence Matrix
SVM	Support Vector Machine
QIA	Quantitative Image Analysis
3CB	3-Compartment Breast
CAD	Computer Aided Diagnosis
ROI	Region of Interest
TPF	True Positive Fraction
FPF	False Positive Fraction
OCR	Optical Character Recognition
NLP	Natural Language Processing
FMI	Fat Mass Index
FFMI	Fat Free Mass Index
LMI	Lean Mass Index
NHANES	National Health and Nutrition Examination Survey
ALMI	Appendicular Lean Mass Index
DXA	Dual-energy X-ray Absorptiometry
LMS	Lambda-Mu-Sigma
NPW	Non Pre-Whitening
NPWE	Non Pre-Whitening Eye
2AFC	2-Alternative Forced Choice
4AFC	4-Alternative Forced Choice
kVp	Peak Kilovoltage
CDMAM	Contrast Detail in Mammography
CD	Contrast Detail
QA	Quality Assurance
IQF	Image Quality Factor
HIPAA	Health Insurance Portability and Accountability Act
DICOM	Digital Imaging and Communications in Medicine
mAs	Milliamp Seconds
FWHM	Full Width Half Max
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
FDA	Food and Drug Administration
CT	Computed Tomography
FC	Fully Connected
SGD	Stochastic Gradient Descent
ReLU	Rectified Linear Units
CPU	Central Processing Unit
GPU	Graphics Processing Unit
CNN	Convolutional Neural Network

Table of Contents

List of Tables

List of Figures

1 | Background

1.1: Motivation: Breast Cancer

Breast cancer is a common cancer with 1 in 8 women experiencing some form of malignant breast cancer in their lifetime and is the number one killer of women in their 50s (1,2). It accounts for nearly 1 in 3 cancers in women and is the second leading cause of cancer death in women after lung cancer (1). Worldwide there are roughly 1 million yearly diagnoses and over 400,000 deaths annually (3), and costs related to breast cancer are in the tens of billions (4,5).

There are several internal and external factors that can increase the long-term risk of breast cancer. Internal risk factors such as age, specific genetic mutations such as BRCA1 and BRCA2, late pregnancies, late menopause, and breast density all increase the risk of breast cancer. External risk factors such as alcohol consumption, family history of breast cancer, use of hormone replacement therapy, use of oral contraceptives, and Body Mass Index (BMI) are all additional risk factors for breast cancer (6).

Several studies have shown that early detection is one of the most important factors in long term breast cancer survival. Women diagnosed with earlier stage cancers have much higher survival rates (2), and regardless of node positive or negative status, the size of the detected tumor has been shown to be a key factor in the 15 year survival rate (7). Precision medicine has also helped to identify certain subgroups of women that early detection would not benefit. Esserman et al determined that early detection does not benefit or improve survival rates in women with fast growing and aggressive tumors. In women with moderately aggressive and slow growing tumors, however, it was shown that early detection did improve survival rates (8,9). These studies have shown that making subclassifications of women based on various factors have helped improve mortality rates in breast cancer, and that improving and optimizing screening for breast cancer is hugely important.

Digital mammography is the clinical standard in screening for breast cancer in average-risk women (10). Clinically, screening mammography is defined as routine scheduled mammograms that simply look for signs of cancer in asymptomatic women, while diagnostic mammograms are mammograms that are specifically looking to investigate a specific symptom, such as a lump, breast pain, or other concerning issue (10). In brief, mammography works by sending low energy X-rays through breast tissue to produce an image that can distinguish between tumors and healthy breast tissue (11). The details of X-ray mammography, its imaging process, and its limitations are discussed in detail in Section 1.4: . Overall performance of digital mammography in one study observed a sensitivity of cancer detection in screening mammography of 87.8% and specificity of 90.5% (12). One study showed that radiologically dense breast tissue can reduce the screening sensitivity of detection down to 68% in high density mammograms, as rated by the Breast Imaging - Reporting and Data System (BI-RADS) breast density, leading to cancers missed by screening mammography (13,14). A detailed discussion on what breast density is and how it affects tumor detection is in Section 1.3: and 1.5: . These cancers that go undetected by mammography are called interval cancers (13), and they are of great interest to researchers and the focus of this dissertation.

Interval cancers are commonly defined as invasive breast cancers diagnosed within 12 months of a normal screening mammogram (13). Interval cancers represent a failure mode of screening mammography, making them a large research topic and area to improve. Women usually detect these interval cancers through self-examinations or experiencing various symptoms that lead to a clinical check-up or diagnostic mammogram (1,10). Further, interval cancers have characteristics that differentiate them from screen-detected cancers. Interval cancers are on average larger than screen detected cancers when found, are at a higher histologic grade, and have higher expression of markers for cancer aggressiveness, such as bcl-2, p53, and Ki-67 (15), making these cancers more deadly and dangerous. A significant goal in breast cancer research is to improve detection capabilities, allow for earlier detection of cancers, and reduce the number of breast cancer deaths by reducing the number of interval cancers (13).

In the following sections of this chapter, properties and causes behind interval cancers are discussed along with currently used measures to quantify interval cancer risk. Additionally, the imaging physics behind mammography is detailed to better understand the limitations of mammography that leads to interval cancers. Lastly, advances in technology are discussed that could potentially improve quantifying risk of interval cancer, such as computer vision methods and deep learning methods.

1.2: Interval Cancers

In order to understand how to identify interval cancer risk, we need to understand interval cancers. Roughly 13 percent of the breast cancers diagnosed in the U.S. are interval cancers based on initial assessment (12), making it a significant proportion of breast cancers. The American College of Radiology (ACR) defines an interval cancer as a cancer diagnosed in between normal screening mammograms, usually within 12 or 24 months of a normal result (10,16). These interval cancers can occur for several different reasons:

- **True interval or mammographically occult**: These are cancers where the prior screening mammogram showed no reason for assessment and appeared normal. This includes interval cancers that were not mammographically visible even at the time of diagnosis, and is roughly 42% of interval cancers (17,18).

- **Minimal sign of cancer:** There are some interval cancers where, in hindsight, there was a possible abnormality but it would not have necessitated further assessment at the time. This is roughly 23% of interval cancers (17,18).
- **Radiologist Error / False Negative**: These are interval cancers where an abnormality was clearly visible in the prior mammogram and should have warranted additional assessment. This is roughly 35% of interval cancers (17,18).

A major goal is to improve and reduce each type of these interval cancers, if possible. Interval cancers that were not present at the time of the mammogram are an unimprovable category without increasing the screening rate of mammography. The United States Preventive Services Task Force (USPSTF) recommends biennial screening mammography for women aged 50 to 74 (10,19). While increasing the rate of screening could potentially catch these fast growing cancers, that benefit is offset by the risk of additional biopsies or false positives (19).

The subset of interval cancers where the tumor was present in the mammogram is of most importance to researchers. A large fraction of these interval cancers have tumors that are occult and undetectable in the image (17). This effect is commonly called mammographic masking, and can lead to the type of interval cancer where lesions are missed by screening mammography due to dense tissue masking the presence of a lesion (15,20,21). Magnetic resonance imaging (MRI) or ultrasound (US) in addition to screening mammography has been shown to reduce interval cancer rates in high risk populations (13,22,23). If we can identify mammograms that are likely to mask any future tumors and make them invisible, we can refer those women at high risk of a missed cancer to supplemental screening like ultrasound or MRI. In order to accomplish this, we need to identify the groups of women that are at highest risk of interval cancer.

Radiologist fatigue is a factor in interval cancers where the cancer is present and visible but undetected due to user error (17). Without altering the workload or work flow of radiologists, these interval cancers are unlikely to be helped. However, if we are able to identify a subset of women with a high risk of interval cancer or low detectability in their mammograms, we may be able to subdivide the screening population into a higher interval risk group and a lower interval risk group, allowing for a more optimized workflow.

Regardless of the type of interval cancer, on average interval cancers have several different, and more dangerous, characteristics compared to screen detected cancers. First, interval cancers are on average larger than screen detected cancers when detected, with tumors 2-4 cm having an interval odds ratio (OR) of 2.2 and tumors larger than 4 cm having an interval odds ratio of 2.57 (20). This is significant, as larger tumor sizes directly lower the 15-year survival rate of breast cancers (7). Tumors that were lymph node positive had an 80.4%, 70.1%, and 47.1% 15-year survival rate when found at sizes between 0.1 - 1 cm, 1.1 – 2 cm, and 2.1 – 5 cm respectively. Further, interval cancers on average tend to be more aggressive, with a higher percentage of interval cancers having p53 and Ki-67 expression compared to screen detected cancers, which are markers for tumor aggressivity (15). 35 percent of interval cancers had p53 expression compared to only 16 percent of screen detected cancers, and 45 percent of screen detected cancers had 0-4% cell expression of Ki-67, while only 25 percent of interval cancers had the same expression level. Only 13.1 percent of screen detected cancers had over 20 percent cell expression of Ki-67, while 26.6 percent of interval cancers had the same expression level. Additionally, when interval cancers are found they are more likely to exhibit lymph node involvement and are of a more severe histological grade, with 43 percent of interval cancers being at histological grade 3 compared to only 19 percent of screen detected cancers. (15,20). These factors, in addition to the known evidence that earlier detection of breast cancer leads to higher survival rates (2,7), provides strong motivation to provide analyses that help identify risk of interval cancer in order to help those women get their cancers detected earlier if they do happen.

It has been shown that additional screening methods can better detect interval cancers compared to using mammography alone. Adding ultrasound to mammography was shown to detect an additional 4 interval cancers per 1000 mammograms, and adding magnetic resonance imaging to mammography was shown to detect an additional 14.7 interval cancers per 1000 mammograms (13,22,23). This is because in true interval cancers, breast density makes it difficult to distinguish between lesions and other healthy types of tissues, and different imaging modalities are able to see through and distinguish between tissues differently. This will be discussed further future sections. Because it is more difficult to detect lesions in certain types of mammograms and supplemental screening could aid in helping identify lesions, legislation has been passed or is in process to be passed in 39 states to mandate notifying women when they have dense breast tissue, including the recommendation to consider additional supplementary screening methods (24).

Unfortunately, recommending all women with dense breasts get supplemental screening is both impractical and not beneficial, as it would cause roughly 50% of women to get supplemental screening and would increase the false positive rate (19). Currently, a more specific and stronger measure of interval cancer risk is required before supplemental screening is reasonable and practical to perform in those women. If we can better identify women at high risk of interval cancer, we can identify the women that would most benefit from supplemental screening, while maintaining regular screening methods for the women that it benefits best. This would be the optimum method to both prevent the stress, cost, and false positives inherent in additional screening, but also give the additional screening to the women who benefit from it.

1.3: Current Methods of Interval Risk

To understand how to make a more effective measure of interval risk, we need to understand the current methods that exist of quantifying interval cancer risk. Over the years there have been several measures that have served as proxies for interval risk, some more sophisticated than others.

First, many demographic and environmental factors are not only risk factors for breast cancer, but also risk factors for interval cancer. In particular, a young age, use of hormone replacement therapy (HRT), and oral contraceptive use, and low BMI are all known to be associated with higher interval cancer rate (25). Being older was shown to be protective of interval risk, as women aged 60-69 had roughly half the odds ratio for interval cancer compared to those aged 50-59 (26). At least some of the interval risk associated with these measures is due to the fact that these measures intrinsically affect breast density.

One of the more established basic measurements known to be a risk of interval cancer is high breast density. A plethora of studies have shown increased interval rate with higher breast density (12,13,16,20,25,27), with odds ratios for interval cancer risk being up to 5 times higher in the highest density groups compared to moderate and low density groups. Given that breast density is a known risk factor for interval cancer, clinicians have spent considerable time to quantify and standardize breast density measurements. Currently, radiologists rate the breast density of the images on a 4 point categorical scale called the BI-RADS density scale (10). This scale places women into either category A: Entire Fatty, B: Scattered Fibroglandularities, C: Heterogeneously Dense, D: Almost entirely Dense. Representative images of each level are shown in Figure 1-1. Roughly 8% of the population is in category A, 40% in B, 44% in C, and 8% in D (16). Mammograms are categorized on this scale based on the judgement of the radiologist, so this is a subjective scale, with a moderate interobserver agreement with a kappa score of roughly 0.44-0.54 (28,29).

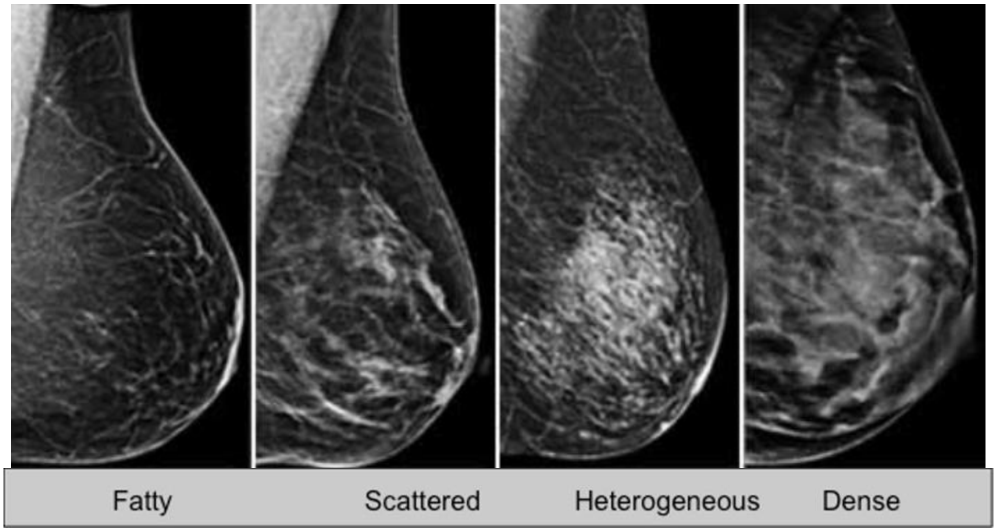
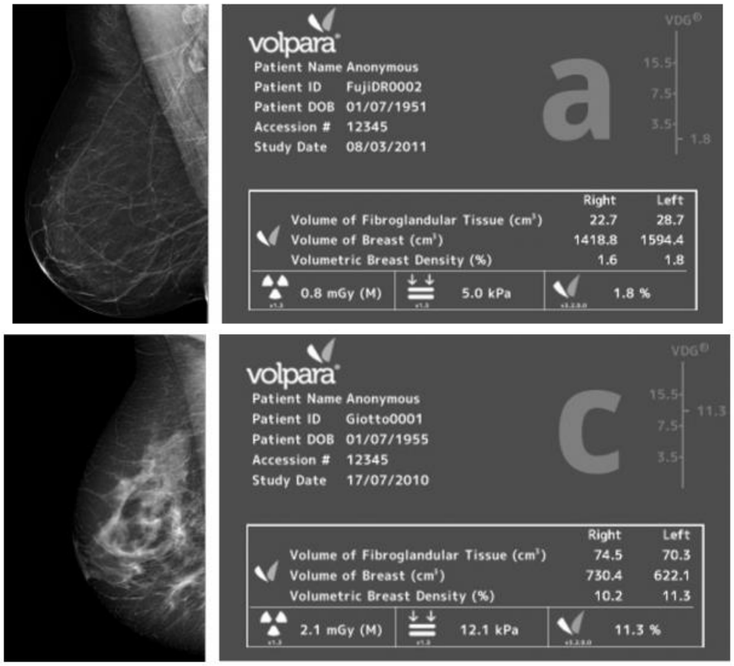


Figure 1-1: Representative images of the four BI-RADS density categories: A-fatty, B-scattered fibroglandularities, C-heterogeneously dense, D-almost entirely dense. The dense tissue seen in categories C and D can mask tumors and increase the risk of interval cancer. Source: (30)

Even though BI-RADS density is on a subjective scale and has reproducibility issues, it still is one of the better indicators we currently have for risk of interval cancer. Women with BI-RADS density D breasts have a 5.65 times higher interval cancer risk compared to women with BI-RADS density B breasts (13,16). Individuals with dense breasts had nearly 3 times higher interval cancer rate than individuals with non-dense breasts (13). Luckily it has been shown that women with dense breasts have a reduced interval cancer rate with additional or supplemental screening, such as ultrasound (US) or magnetic resonance imaging (MRI) (13). If we can more effectively identify women at higher risk of interval cancer, we can help to lead conversations and better identify women that would benefit from supplemental screening, helping to reduce mortality of the disease.

In addition to using BI-RADS breast density, several groups have created risk models by incorporating information from both demographic information and breast density information. These risk models have also shown significant associations with interval breast cancer (13). The Gail risk model combines information about age, age at first menstruation, age at first live birth, first degree relatives with breast cancer, previous breast biopsies, and race/ethnicity. This model was shown to be significant in projecting 10 year risk of breast cancer (31,32). Also, the Breast Cancer Surveillance Consortium (BCSC) Model combines information about age, race/ethnicity, first degree relatives with breast cancer, prior biopsies, history of Ductal Carcinoma *In Situ* (DCIS), breast augmentation, or mastectomy, and BI-RADS Density. This BCSC Model has been shown to improve 5 and 10 year risk projections of breast cancer (33). In addition to being risk factors for breast cancer in general, these have been studied to see if they are indicative of interval cancer risk. In addition to breast density, the BCSC five-year risk has been shown to be predictive of interval cancer risk and help stratify individuals in to higher risk categories of interval risk (13).

While risk models and BI-RADS density have helped identify individuals at higher risk for interval cancer, there is a flaw inherent to BI-RADS density that likely reduces its strength. BI-RADS density is a subjective measure, with inherent variance and interobserver disagreement (34). To try to combat this subjectivity and variance, several companies hypothesized that if they could standardize the measure and remove the variance it would increase its strength as a predictor of interval risk. Several companies and automated programs exist to automate breast density measures and categorize mammograms into BI-RADS density categories such Cumulus (University of Toronto, Toronto, CA) (35,36), Volpara (Wellington, New Zealand) (37,38), Quantra (Bedford, MA, USA) (37,39), and Single X-ray Absorptiometry (SXA) (40,41). Figure 1-2 shows a sample output from the Volpara program for both a non-dense and a dense breast. Many of these programs have been found to reproducibly automate the process of breast density measures (41). While these programs have high consistency, one study showed that using these automated measures did not increase the interval risk compared to using BI-RADS density, as seen in Table 1-1 (16). This indicates that while these measures removed the subjectivity, they did not add any additional information to help identify interval cancer risk.



Patient Name Anonymous
Patient ID FujiDR0002
Patient DOB 01/07/1951
Accession # 12345
Study Date 08/03/2011

	Right	Left
Volume of Fibroglandular Tissue (cm³)	22.7	28.7
Volume of Breast (cm³)	1418.8	1594.4
Volumetric Breast Density (%)	1.6	1.8

0.8 mGy (M) | 5.0 kPa | 1.8 %

Patient Name Anonymous
Patient ID Giotto0001
Patient DOB 01/07/1955
Accession # 12345
Study Date 17/07/2010

	Right	Left
Volume of Fibroglandular Tissue (cm³)	74.5	70.3
Volume of Breast (cm³)	730.4	622.1
Volumetric Breast Density (%)	10.2	11.3

2.1 mGy (M) | 12.1 kPa | 11.3 %

Figure 1-2: Sample images showing the automated categorization of breast density for a low-density breast (top, BI-RADS density A), and a high-density breast (bottom, BI-RADS density C) from Volpara Software. This turns the subjective measure of BI-RADS density into a more objective and repeatable measure. Source: (42)

Table 1-1: Interval cancer odds ratios using BI-RADS density as a predictor versus using automated BI-RADS density measures. One can see that the interval cancer odds ratios do not differ significantly even though automated measures remove some level of subjectivity. Source: (16)

Clinical BI-RADS Density	Interval Cancer Odds Ratio (95% CI)	Automated BI-RADS Density	Interval Cancer Odds Ratio (95% CI)
A - Fatty	0.74 (0.41-1.36)	A - Fatty	0.73 (0.42-129)
B - Scattered	1.00 (REF)	B - Scattered	1.00 (REF)
C – Heterogeneous	2.51 (1.74-3.61)	C – Heterogeneous	2.22 (1.44-3.43)
D - Dense	5.09 (3.11-8.35)	D - Dense	5.65 (3.33-9.60)

Other researchers have taken a more analytical approach to identifying interval risk and attempted to identify specific image properties of the mammogram that correlate to interval risk. These researchers have analyzed the properties of the mammograms themselves, rather than the characteristics of the women, and generated different measures of image statistics and texture that could correlate to interval risk (43). These features include everything from first order image statistics like means and standard deviations, second order statistics like gray level co-occurrence matrix (GLCM) measures, and other novel features to summarize image data, and many have found these measures to be significant for interval risk (43,44). Strand et al found several features: first that eccentricity and skewness of intensity gradient measures differentiates between interval and screen detected cases with an odds ratio of 1.32 and 1.20 (45), and second that longitudinal fluctuations of breast density could also indicate risk of interval cancer, with an interval cancer odds ratio of 1.17 (46). Byrne et all found various subjective parenchymal patterns had significant odds ratios for cancer (47). Another study found that a support vector machine (SVM) of 158 different mammographic density, similarity, and texture based features were able to significantly add to risk predictions of near-term cancer (48). Another study found that various features Quantitative Image Analysis (QIA) features and several 3 Component Breast (3CB) QIA features were able to help distinguish between benign and malignant tumors (49). Lastly, Malkov et al analyzed 46 different mammographic image features and found several were significantly associated with breast cancer, like the fractal dimensions and second order statistics (50). Unfortunately, these measures have not broken into the mainstream of been significantly better than other interval measures like the current risk models.

Other companies have tried to make automated programs to detect or help diagnose various cancers, called Computer Aided Diagnosis (CAD) (51–56). These programs usually use model observers, feature detectors, or some other means to identify regions of interest (ROIs) where a tumor is possibly located, and a radiologist can then investigate those regions more rigorously (51,57). These systems have been shown to be highly sensitive in lesion detection, but they have several false positive per case as well (53). Overall, these CAD methods have a true positive fraction (TPF) of 80-95%, but their false positive fraction (FPF) leads to many false positives that radiologists need to ignore (54). However, most of these programs aim to aid radiologists in determining and detecting regions of interest (ROI) for potential cancers, and not on determining the actual ability of mammograms to detect an image or risk of interval cancer. While these programs and pieces of software are interesting and helpful, they are asking a fundamentally different question from the question.

While many systems have been developed to improve on BI-RADS breast density, develop QIA methods to analyze images, and use CAD systems to identify cancer locations, these methods have not generated significant and actionable results for interval risk. As a result of this and to expand the effort to identify interval risk predictors, the ACR has called for direct measures of mammographic masking and detectability and interval risk. We aim to use new methods to identify and predict interval cancer risk in order to refer women to supplemental imaging methods that can help reduce deaths and other injuries due to interval cancer.

1.4: X-Ray Imaging

In order to create better measures of interval risk, we need to understand why current measures of interval risk, like BI-RADS density, correlate with interval risk. In order to understand why BI-RADS density is one of the stronger risk factors for interval cancer, we need to understand how X-ray imaging systems such as mammography work, and fail, as imaging systems.

Fundamentally, mammography is a form of X-ray imaging specialized for imaging breast tissue. X-rays were first discovered in 1895, and they were first applied to try to identify breast cancers in 1913 (11,58). All X-ray imaging technologies, like X-Ray, computed tomography (CT), and mammography, create X-rays and images using the same fundamental imaging physics and methods, as seen in the schematic of the X-ray emitter in Figure 1-3. In the emitter, energy is passed through the filament or cathode, producing high energy electrons. These electrons then travel towards the target on the anode at extremely high speeds due to a voltage difference between cathode and anode. When these electrons strike the target, X-rays are created that fly in a variety of directions. The X-rays that make it through the window are the X-rays that eventually go through the tissue and create an image (11). Depending on the composition of the anode, cathode, and energy (kVp), different X-ray spectra will be produced.

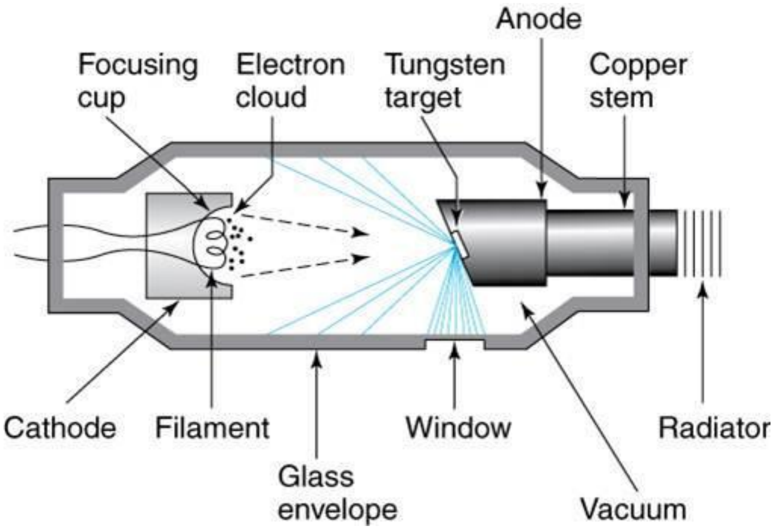


Figure 1-3: Diagram of an X-ray tube with labeled components. The parameters surrounding the cathode, anode, and the voltage affect the spectrum of X-rays produced. Source (59)

Further, in almost all cases the camera unit contains a filter at the window, which is a thin piece of metal that absorbs the lower energy X-rays that do not aid in the imaging process, which helps to improve image quality and reduce the amount of radiation that the patient receives (11,58). Most of the lower energy X-rays that exist are just absorbed by tissues and contribute to patient radiation exposure, but do not aid in the imaging process. To expose patients to less dose and improve image quality, after the X-rays are produced they go through a filter to reduce these lower energy, more harmful X-rays. These filters are usually thin sheets of various metals, from bronze to aluminum to tungsten (11). Figure 1-4 shows the difference in X-ray spectrum before and after filtration.

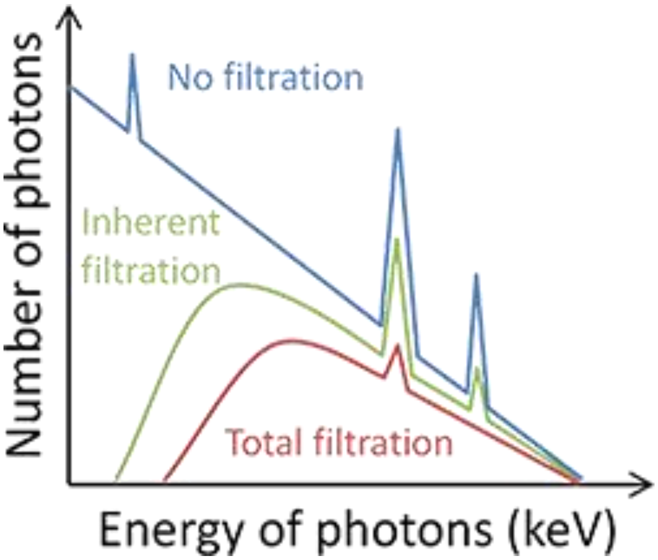


Figure 1-4: X-ray spectrum before and after filtration. Inherent and additional filtration sources help to narrow the energy range of the spectrum and reduce low-energy photons that do not contribute to imaging. Source: (60)

Once X-rays are produced, different tissues will absorb, or attenuate, X-rays at different amounts, depending on a number of factors. Most importantly, higher density materials attenuate X-rays more strongly (11). A film or detector is placed on the opposite side of the tissue as the X-ray emitter and detects the amount of X-ray photons that make it through the tissue at each point, with the differences in attenuation highlighting different tissue areas. For example, Figure 1-5 demonstrates that bone tissue attenuates X-rays much more than surrounding tissue, leading to the contrast between tissues that allows clinicians to detect broken bones or other anomalies (11). Through further research, clinicians discovered that certain image processing algorithms were able to improve diagnostic quality of images. Many of these algorithms involved enhancing contrast and inverting the image scale, so that the image regions that attenuated more photons actually appear brighter, rather than darker like they would appear on the actual raw data image (61).

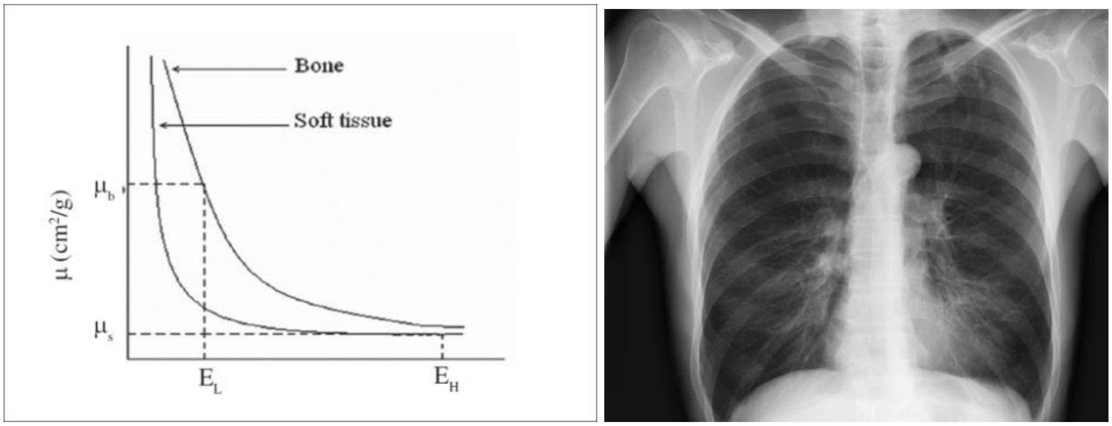


Figure 1-5: graph showing the relative difference in attenuation between bone and soft tissue (left), and the resulting image that forms due to that attenuation difference in a chest X-ray (right) Source: (62,63)

There are several geometric factors that affect the image quality in X-ray imaging. An important factor is focal spot size and focal spot distance from the object. Just as in working with a camera or human eye, The focal spot size is the minimum possible size that is able to be distinguished, and scientists have worked to reduce focal spot sizes in order to improve imaging quality (11).

There are also several factors inherent to X-ray imaging that contribute to blur. If an X-ray gets close to the nucleus of an atom during its travel or gets too close to electrons in an electron cloud, several types of scattering can occur, as seen in Figure 1-6 (11). In addition to the scatter types seen in the figure, Bremsstrahlung is another type of scatter where the electron changes directions after coming too close to the nucleus of an atom. All of these forms of scatter change the path of photons that traveled through one part of tissue to hit the detector and appear as though they traveled through a different part of the tissue, lowering image quality and creating blur (11). These factors partially contribute to blur but adding collimation at the end to block these deflected X-rays at the source helps to reduce noise from this (11,61).

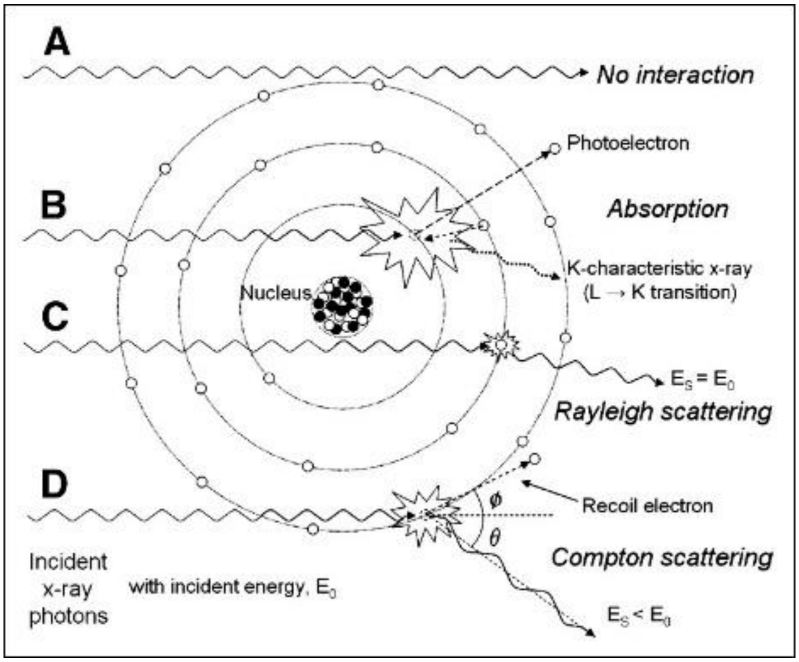


Figure 1-6: Schematic showing various forms of X-ray scatter that contribute to imaging blur. Absorption, Rayleigh scattering, and Compton scattering can all affect the energy of incoming photons and adjust the photon path, contributing to blur. Source: (64)

Lastly, a common factor that hinders image quality is movement artifacts. If the tissue moves as the X-rays are being produced, a blur will be introduced just as in traditional cameras (11). This source of errors gets more significant as one is looking for smaller objects (Such as small bones in the wrist) and objects with lower contrast (such as soft tissue lesions within other soft tissue), and radiologists often implement restraints to immobilize the region being imaged and into improve image quality.

1.5: Mammography Imaging Chain and Breast Density

Mammography uses these same methods to create and image tissues, just specialized and optimized for breast imaging. A key difference is that in mammography the X-rays created are of a much lower energy, in order to help differentiate between different types of soft tissues in the breast (11). Several other changes exist in mammography that do not exist in conventional X-ray imaging, such as introducing breast compression to help stabilize the breast and make the breast tissue thinner, which produces a higher quality image with less movement artifacts (14,58). A sample mammography unit displaying these differences in shown in Figure 1-7.

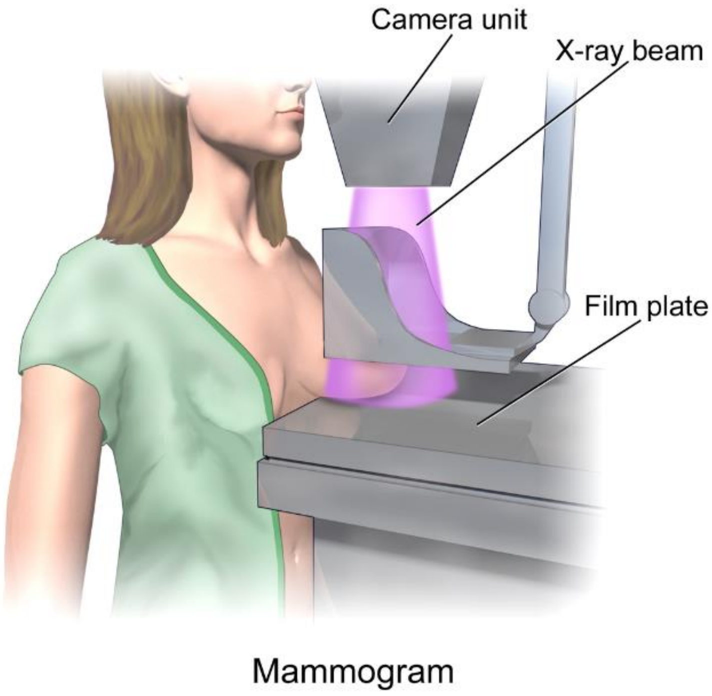


Figure 1-7: Diagram of a mammography unit with various components. Key features include tissue compression and adjustment of imaging parameters. Source (65)

While a variety of choices exist, typical anode/cathode combinations in mammography use Molybdenum, Tungsten, and Rhodium as they produce characteristic spectra that produce the best possible images. While the kVp of imaging can be from 100-150 kVp for traditional X-rays, they are usually only done at 25-35 kVp in mammography (66,67). All X-ray imaging modalities would use lower energy X-rays if possible, because it helps to reduce radioactive dose to patients. Because mammograms have less tissue to go through and that breast tissue attenuates less than other tissues in the body, mammography has the capability to use lower energy X-rays and still generate high quality images (11).

The standard procedure when getting a mammogram is to get two different views for each breast, and is shown in Figure 1-8 (10). This helps because it allows you to see any potential tumors from a variety of angles and can avoid any situations that might obstruct some tissue in one view. There are a variety of views that exist for specific cases, but the standard views are craniocaudal (CC) views and mediolateral oblique (MLO) views because they are nearly perpendicular to each other and show vital information (10). A craniocaudal view is an image from the top down and helps distinguish lateral and medial tissue, while the MLO view looks at a slightly oblique angle from the lateral side inward and helps to differentiate superior and inferior tissue, as well as gives the ability to see closer to the chest wall.

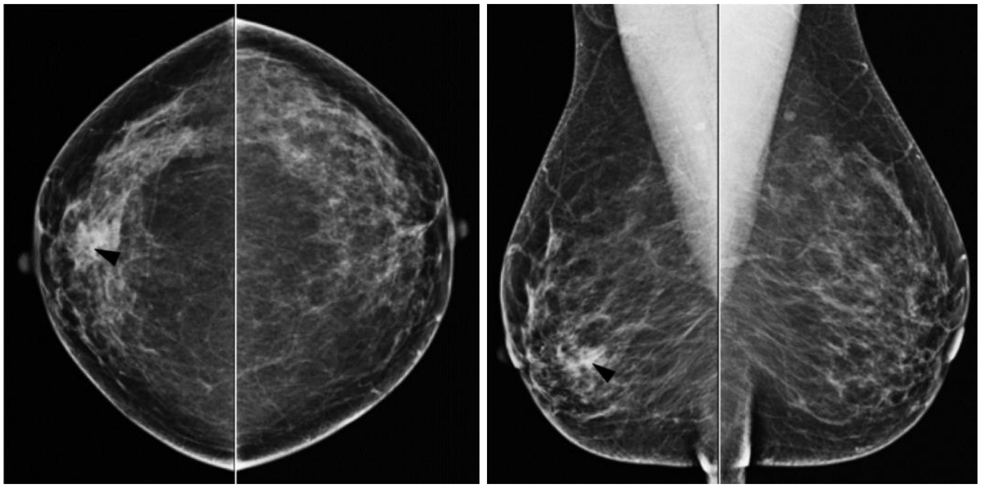


Figure 1-8: Standard viewing all four views of a mammogram. The views are Right CC (Far-Left), Left CC (Middle-Left), Right MLO (Middle-Right), and Left MLO (Far-Right). These four views help to see additional tissue and identify possible tumor location in 3D space by combining information from both view (such as with the tumor highlighted by the black arrow). Source: (68)

A radiologist then examines these images simultaneously to determine if there is a tumor or suspicious findings. They look for tumor-like objects, odd-texture patterns like spiculations, looking for differences in shape or texture of left and right breasts, or differences in the breast image over time (10). If a tumor or suspicious area is found, a biopsy or additional screening will next be performed in order to properly identify whether it is a tumor, the characteristics of the tumor, and inform doctors about best next steps for treatment (10).

This procedure is normally effective and helps detect and treat cancers. Unfortunately, this process is not perfect due to some inherent limitations in mammography and breast tissue composition. Breast tissue is mostly composed of two tissue types: fatty tissue and fibroglandular tissue (43,46). Fibroglandular tissue contains most of the connective tissue and functional tissue within the breast and attenuates X-rays at a higher rate than fatty tissue, as seen in Figure 1-9. Because of this, breasts with more fibroglandular tissue have more signal attenuated and more signal attenuated at the same amount as tumors, making it harder to distinguish other object such as tumors. This means that if a tumor is within fibroglandular tissue, it can be difficult to distinguish the cancerous tumor from actual healthy fibroglandular tissue, as seen in Figure 1-10. This is a fundamental problem that contributes to mammography not always detecting breast cancers. breasts with a high ratio of fibroglandular tissue are called radiographically dense breasts (or just dense breasts).

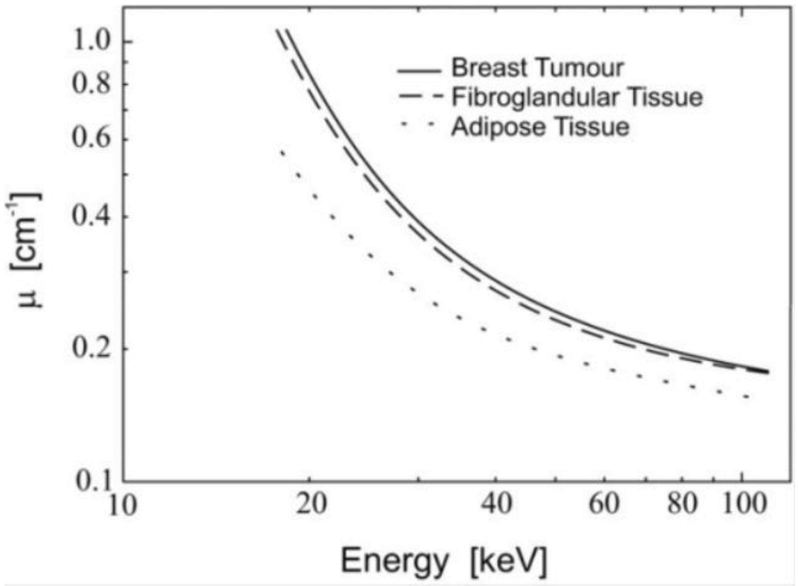


Figure 1-9: Attenuation plot comparing attenuation rates of adipose tissue, fibroglandular tissue, and breast tumors. One can see that fibroglandular tissue closely matches the attenuation curve of tumor tissue, meaning there is little differentiation between those two tissue types. Source: (69)

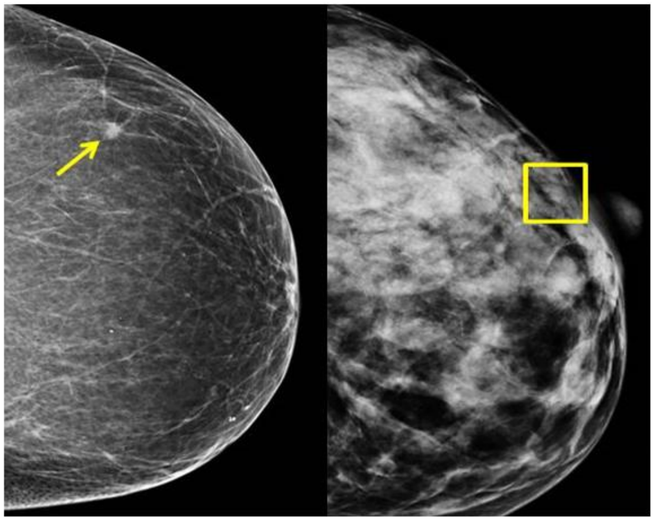


Figure 1-10: At left, a mostly fatty breast with an easily distinguishable tumor. At right, a mostly fibroglandular breast with a tumor that is hard to distinguish from the fibroglandular tissue. This masking of tumors is part of why dense breast tissue is a risk factor for interval cancers. Source: (70)

1.6: Computer Vision and Deep Learning as Potential Tools

In the past decades, there have been several advances and developments in the field of computer vision – the study to help computers perform the same tasks that human vision can perform. This field involves designing algorithms and software to automatically take an image and detect objects, make object measurements, identify unique features, define edges, and segment portions of a scene (44). This field has been applied in order to drive technologies such as facial detection, automated object detection and measurement, self-driving cars, and many other technologies in the military, private companies, and researchers (71,72). This field of computer vision has been a cutting-edge tool helping to make improvements to a variety of fields.

In recent decades a new sub-field of computer vision has opened up called deep learning (44). This field had been theorized for many years (73,74), but implementation was not practically possible until advances were made in computing ability and the amount of data available to researchers in the early 2010s (44). The key development is that instead of looking for a specific feature or object that a researcher pre-defines like in tradition computer vision, deep learning develops its own features to look for over time in order to achieve an optimal decision outcome, whether that decision is object identification, optical character recognition (OCR), natural language processing (NLP), or image segmentation (44,75). Deep learning has made advances and improvements to the field of computer vision in past decades, and even influenced medical fields (76,77).

Because interval cancers are fundamentally tied to an inability to detect an object (a tumor), it is reasonable to hypothesize that computer vision may also be able to better identify risk of interval cancer or quantify mammograms with low detectability than current measures of breast density. In future sections the history and methods of both traditional computer vision as well as deep learning will be detailed in depth, and my personal work in applying these methods will be detailed in my effort to quantify measures of interval cancer risk.

1.7: Unmet Needs

Currently, gold standard measurements for interval risk identify demographic factors, subjective breast density, and different rudimentary textural qualities to best identify risk of interval cancer. These methods have been able to identify that some groups of women, such as those with high breast density, are significantly more likely to experience interval cancer compared to other groups. Unfortunately, these measures of interval risk are not specific enough or strong enough to implement supplemental screening clinically. Currently, one of the strongest measures of interval risk, BI-RADS breast density, identifies roughly 50 percent of women as having high density breasts, and therefore a higher risk of interval cancer. Opening this subset of women to supplemental screening would prevent several interval cancers. Unfortunately, this level of additional testing would be expensive, impractical and liable to false positives and additional biopsies, and these considerations already lead to other countries, such as those in the European Union, to introduce screening later in life and at longer screening intervals compared to the U.S. based on their prioritization of identifying cancer risk, reducing dose, and avoiding false positives (21). Because of this, there is a need to better identify and subset the group of women at highest risk of interval breast cancer in order to maximize the ability to detect cancers and minimize the risk of false positives.

Many advances have been made to improve and apply computer vision methods to a variety of fields, but this trend has been slower to come into medicine and have not been applied to quantify risk of interval cancer. This dissertation addresses the unmet need of applying these current technologies to help address interval cancer risk, helping to mitigate an important problem and failure mode of mammography screening. Throughout this dissertation I detail, develop, and test a variety of computer vision techniques that improve on interval risk measures.

2 | Standards in Composition and Cancer Risk

2.1: Body Composition as a Measure of Cancer Risk

Body composition measures have been shown to be extremely valuable tools to measure health outcomes, risks, and recovery from various treatments. Additionally, compositional measures have been shown to be associated with cancer risk, including risk of breast cancer (25). Mandelson et al showed that individuals with low BMI were at higher risk for interval breast cancer, and high BMI was a risk factor for overall breast cancer compared to average BMI/FMI levels. Unfortunately, the most common compositional metric, BMI, is a non-specific measure that combines both lean and fat mass (78).

Measures such as fat mass index (FMI) and fat free mass index/lean mass index (FFMI/LMI) are able to better identify those tissues, and as a result have been shown to be more predictive of certain conditions, especially when broken down into composition of different regions of the body (78–81). Equations for FMI and LMI are defined as:

$$FMI = \frac{Fat\ Mass}{Height^2} \qquad LMI = \frac{Lean\ Mass}{Height^2}$$

Further, breaking down FMI and LMI to quantify the fat and lean mass of certain limbs or regions of the body has been shown to be predictive of cardiovascular disease, regional lipolysis, blood pressure, and other conditions (80,82–88). Caan et al have shown that muscle area and muscle density relate to breast cancer survival, as well as functional and movement limitations in breast cancer patients. They showed that women with sarcopenia, defined as a lower quantity of lean mass compared to standard levels, had a higher mortality with a Hazard Ratio (HR) of 1.41 compared to those without sarcopenia. Further, they showed that BMI alone was not significant for mortality.

They also showed that functional limitations were significantly associated with mortality (HR = 1.40) (89,90). It is reasonable to believe that improving regional composition measures could better identify risk of various diseases and outcomes, including breast cancer or interval breast cancer risk.

While regional measures of composition exist, there have not been standards or percentile charts produced in order to compare regional measures to some sort of standard or average measure. In body composition, the Lambda-Mu-Sigma (LMS) method is often utilized to produce standardized measures and percentiles as a function of age (91). Generating and providing this data will be a useful tool for groups investigating the effects of regional distribution, body shape, and composition on metabolic conditions and could lead to better measures of cancer and interval cancer risk based on regional composition.

2.2: NHANES

The National Health and Nutritional Examination Survey, or NHANES, is an effective sample to generate U.S. body composition standards of the limbs and trunk that could be useful in identifying individuals at risk of various diseases, from sarcopenia, to cachexia, to breast cancer. NHANES is a national survey of Americans that uses a rigorous sampling method and has been used many times to provide an accurate representative sample of health and body composition in the U.S. (92,93). This is by far one of the most comprehensive measures of body composition and outcomes that exists in the U.S., and is a comprehensive survey to generate standards of body compositions. These standards could later be used for a number of applications, from identifying disease risk to monitoring recovery from surgery to quantifying the negative effects on body composition that cancer has such as cachexia (94,95). Previous work from our group has used NHANES data to quantify risk of fracture, osteoporosis, osteopenia, and several other measures of composition (96–101).

2.3: Study: Dual energy X-ray absorptiometry body composition reference values of limbs and trunk from NHANES 1999-2004 with additional visualization methods (95)

In this study, we aimed to generate reference measures of regional body composition in order to help advance more specific compositional measures of various mortality risk markers, breast cancer and interval cancer risk factors, as well as produce measures that may be useful in tracking recovery from breast cancer treatments such as cachexia.

2.3.1 Introduction

Body composition is a known risk factor for a number of conditions such as diabetes and heart disease that contribute to higher healthcare costs and reduced lifespan (102,103). Body mass index (BMI, total mass/height^2) and waist circumference have long been used as indicators of body shape and adiposity and as crude measures of health risk (104,105), but these measures are not specific to lean or fat mass. Fat Mass Index (FMI, fat mass/height^2) and Lean Mass Index (LMI, lean mass/height^2) have been introduced as more specific composition measures than BMI (106–109), but even these measures are not specific to the composition of each region (arms, legs, trunk) of the body.

In many studies regional fat mass and composition has been shown to be predictive of cardiovascular disease, regional lipolysis, blood pressure, and other conditions. (80,82,83,110–114). Wilson et al. showed that the volume ratio of trunk to leg had a strong association to diabetes and mortality that was independent of total fat distribution (115). Prado et al. used regional composition of the limbs to calculate Appendicular Lean Mass Index (ALMI) and proposed new body shape and composition phenotypes to study along with ways to diagnose sarcopenia and sarcopenic obesity (116). Regional composition and volume measurements play an important role in both direct associations to disease states and in developing an improved understanding of healthy compositional phenotypes.

Performing studies with standardized reference curves of regional composition provides advantages over using raw regional FMI and LMI values. First, reference curves inherently control for differences in sex, age, and ethnicity (91). Second, Z-scores and T-scores are more interpretable than raw FMI and LMI values or ratios in many cases. Lastly, conditions such as sarcopenia and sarcopenic obesity rely on Z-score or T-score cutoff values for diagnosis (117–119). Reference curves have been made for total body FMI and LMI (94,120), but as of yet no reference curves have been produced for regional composition of the U.S. population. Deriving such reference curves would prove useful for groups studying how regional body composition varies across demographic groups and how it affects different health outcomes.

In this study, we produced FMI and LMI reference curves and LMS tables for the legs, arms, and trunk by sex and ethnicity in a representative U.S. sample. These LMS tables will allow researchers to determine when individuals have higher or lower fat or lean mass in different regions of the body for a given age, sex, and ethnicity by calculating Z-scores in each of those regions. We further produced software to visualize an individual's regional distribution of FMI and LMI Z-scores using radar charts. We do not aim to explain many of the differences found between demographics, but to provide this data as a useful tool for groups investigating the effects of regional distribution, body shape, and composition on metabolic conditions such as sarcopenia, sarcopenic obesity, and many other conditions.

2.3.2 Methods

Our study aimed to produce regional reference values for FMI and LMI of the arm, leg, and trunk for by sex and ethnicity in the cross sectional dual-energy X-ray (DXA) measurements from the 1999-2004 National Health and Nutrition Examination Survey (NHANES). NHANES uses a rigorous sampling method and has been used many times to provide an accurate representative sample of health and body composition in the U.S.(92,93).

Subjects

NHANES DXA scans report whole body and regional measures of fat mass, lean mass, bone mineral content, and bone mineral density (121). Measurements for our study were taken from 15,908 individuals from the NHANES reference database from 1999-2004 for all individuals aged 8-85 (92).

This survey used a multistage sampling method to enroll individuals in the study. Because reference compositional values are different in different ethnicities in the U.S., the survey provides statistics for different self-reported U.S. ethnic groups (Non-Hispanic Whites, Non-Hispanic Blacks, and Mexican Americans, Other Hispanics, and Other Minorities) (92,93,122). In order to provide more reliable estimates, Blacks, Mexican Americans, low-income Whites, individuals between 12-19 years old and above 60 years old were oversampled (92). Subjects were excluded if they were above the weight (136 kg) or height (196 cm) limit of the DXA table. Females were excluded if they reported they were pregnant or if a pregnancy test was positive at exam time (92). Approval for the study was obtained from the National Center for Health Statistics international review board.

DXA Measurement Protocol

Our analysis used the DXA data sets released by NHANES from 1999-2004 without imputation on the Center for Disease Control website (http://www.cdc.gov/nchs/about/major/nhanes/dxx/dxa.htm). DXA scans in NHANES were acquired per manufacturer recommendations of the QDR 4500A fan beam densitometer (Hologic, Inc., Bedford, MA). All subjects wore paper gowns and removed jewelry and other personal items capable of interfering with the DXA exam. These exams were reviewed and analyzed by the University of California-San Francisco Department of Radiology Bone Density Group. Prosthetics, implants and other regional devices capable of affecting results were listed as missing in the dataset and not included in our analysis (92).

Body composition results are calibration dependent and results provided by different instruments can vary. In 1999-2004 NHANES, the DXA scans were analyzed using the Hologic Discovery software version 12.1. NHANES calibration from Schoeller et al (123) were applied before results publicly released. The NHANES data sets contained whole body bone mineral content, bone mineral density, percent fat, lean mass, fat mass as well as with regional measurements (each arm and leg along with trunk) (92).

Producing Reference Curves

From the DXA measures, we calculated the FMI and LMI for the trunk, average arm, and average leg by dividing fat and lean mass of each region by the square of height (94,120,124). Next, we calculated the reference curves of these regional FMI and LMI values using a Lambda-Mu-Sigma (LMS) curve fitting method (lmsChartMaker Pro Version 2.54) (125). LMS is a mathematical method to produce reference curves for measures that corrects for skewed data by generating an "L" (power), "M" (Median), and "S" (Coefficient of Variation) curve across ages of interest. It has been used in the past to calculate reference curves and centiles for height, BMI, and total FMI/LMI (93,126–128). This method produces Z-scores via the following equation (91):

$$z = \frac{\left[\frac{y}{M(t)}\right]^{L(t)} - 1}{L(t)S(t)} \tag{1}$$

The centile curves of y (measure of interest) for a given t (age) are modeled by:

$$C_{100\alpha}(t) = M(t)(1 + L(t)S(t)Z_{\alpha})^{1/L(t)} \quad (2)$$

We developed these reference curves and LMS tables for the three major self-reported U.S. ethnic groups from NHANES: Non-Hispanic Whites, Non-Hispanic Blacks, and Mexican Americans/Other Hispanics (hereafter referred to as Hispanic). Mexican Americans and Other Hispanics were grouped to increase power of the model. There were not enough observations to develop reference data for other ethnic minorities.

The degrees of freedom of the model were increased for each LMS parameter in the order suggested by the developers of LMS (125), and were only increased if it improved the Bayesian Information Criterion more than ln(N) units (N = Sample Size of demographic group), as done in other work to minimize overfitting (124). As recommended by the LMS developers, we examined de-trended Q-Q plots and the fitted curves for smoothness of fit (125).

We used equation 1 to apply the LMS values for each individual based on their demographic and their FMI and LMI data to produce Z-scores for every limb and the trunk. We applied the LMS values and from the average arm and leg to the left and right limbs to produce Z-scores for each of the four limbs, which allowed us to compare symmetry of the left and right appendages of the body. These Z-scores can then be used to determine if an individual has high or low fat or lean mass in different regions of the body for their respective age, sex, and ethnicity.

Radar Charts

To visualize regional differences, we created software that outputs a pentagonal radar chart of regional body composition, where each spoke represents the Z-score FMI and LMI values of each region (each leg, each arm, and trunk) of the body. These radar charts were produced in R (Version 3.2.3) with the fmsb and shiny packages. We opted to plot the Z-score of FMI and LMI for each appendage as opposed to an absolute value because it provided better scaled images and provided more information about regional composition relative to people of the same age/sex/ethnicity.

2.3.3 Results

Subjects

The number of observations used in the reference database by age group, sex, and ethnicity is provided in Table 2-1. These data show the distribution of participants across a wide age range and set of ethnicities and an adequate number of individuals across the age distribution for each sex and ethnicity except for the oldest non-white individuals.

Table 2-1: Number of observations in the NHANES reference database:

Age Group	Sex	Whites	Blacks	Mexican Americans
8 to 9	Male	128	162	197
	Female	67	92	75
10 to 11	Male	132	169	166
	Female	52	63	66
12 to 13	Male	205	269	331
	Female	149	177	199
14 to 15	Male	197	244	284
	Female	144	153	187
16 to 17	Male	208	271	316
	Female	145	129	147
18 to 19	Male	188	212	276
	Female	166	163	257
20 to 24	Male	191	105	162
	Female	186	78	155
25 to 29	Male	202	74	160
	Female	165	64	115
30 to 34	Male	202	88	132
	Female	198	81	98
35 to 39	Male	199	85	133
	Female	204	81	115
40 to 44	Male	220	109	152
	Female	199	99	161
45 to 49	Male	186	97	125
	Female	196	96	128

Age Group	Sex	Whites	Blacks	Mexican Americans
50 to 54	Male	223	79	81
	Female	224	60	98
55 to 59	Male	158	44	64
	Female	140	47	56
60 to 64	Male	185	68	133
	Female	185	87	150
65 to 69	Male	178	67	107
	Female	179	59	119
70 to 74	Male	198	47	88
	Female	168	38	91
75 to 79	Male	149	30	56
	Female	127	36	40
80 to 84	Male	159	12	27
	Female	170	17	25
85+	Male	75	10	10
	Female	86	13	18
Total	Male	3583	2242	3000
	Female	3150	1633	2300
		6733	3875	5300

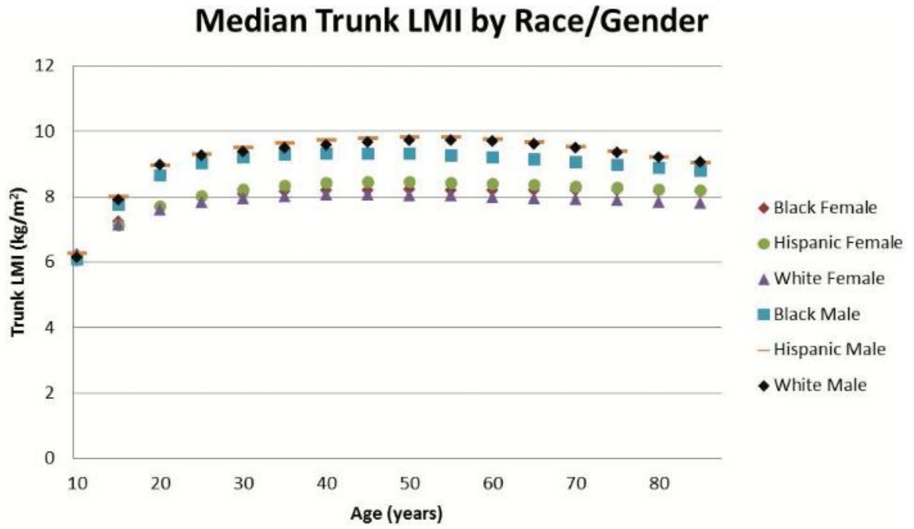
We created reference curves and tables of LMS values and included them as supplemental figures and tables. A list of the reference curves and tables is provided in

Table 2-2. For completeness, the tables for total FMI and total LMI were included. These centile curves show smooth transitions throughout the age range. De-trended Q-Q plots of the data affirmed the goodness of fit and our inclusion criterion for allowing extra degrees of freedom reduced overfitting. As expected, average Z-scores were very close to zero with standard deviations very close to one for all the fitted regional DXA measures.

Table 2-2: Table showing supplemental figures and tables of the LMS curves from this work. Source and figures located in Appendix B and available online at: (95)

DXA Measure	Supplemental Figure	Supplemental Tables (Female, Male)
Average Arm FMI	S1 Fig	S1-2(Black), S17-18(Hisp/Mex), S33-34(White) Tables
Average Arm LMI	S2 Fig	S3-4(Black), S19-20(Hisp/Mex), S35-36(White) Tables
Average Leg FMI	S3 Fig	S5-6(Black), S21-22(Hisp/Mex), S37-38(White) Tables
Average Leg LMI	S4 Fig	S7-8(Black), S23-24(Hisp/Mex), S39-40(White) Tables
Trunk FMI	S5 Fig	S9-10(Black), S25-26(Hisp/Mex), S41-42(White) Tables
Trunk LMI	S6 Fig	S11-12(Black), S27-28(Hisp/Mex), S43-44(White) Tables
Total FMI	S7 Fig	S13-14(Black), S29-30(Hisp/Mex), S45-46(White) Tables
Total LMI	S8 Fig	S15-16(Black), S31-32(Hisp/Mex), S47-48(White) Tables

There were noticeable differences observed across sex for the various measures, many of which varied with age. To help visualize some of these differences, we plotted the median (M) values across sex and ethnicity for the LMI and FMI of the trunk (Figure 2-1), the average leg (Figure 2-2), and the average arm (Figure 2-3). First, we noticed that in most cases and especially for regional LMI, differentiation occurred between males and females during the years of puberty and young adult development. Further in adults between 30 and 50, females had 39%, 83%, and 47% larger median arm, leg, and trunk FMI values than males. Males in this age range had 49%, 20%, and 15% higher regional LMI values than females for the arms, legs, and trunk respectively. Male median LMI values peaked in adulthood and decreased thereafter



especially in limbs, while female median LMI values peaked in adulthood and did not experience as much of a decrease as male values going into old age in the arm and trunk.

In this adult range of 30-50 years of age, there were also apparent differences in regional composition across ethnicity. The leg FMI and LMI of black females were 14% and 15% higher respectively than for Hispanic and white females. White and Hispanic males had 37% higher trunk FMI values than black males, while black males averaged 9% higher leg LMI than white and Hispanic males. Hispanic females had 20% higher trunk FMI than white and black females. Lastly, black and Hispanic females on average had 15% higher arm FMI than white females.

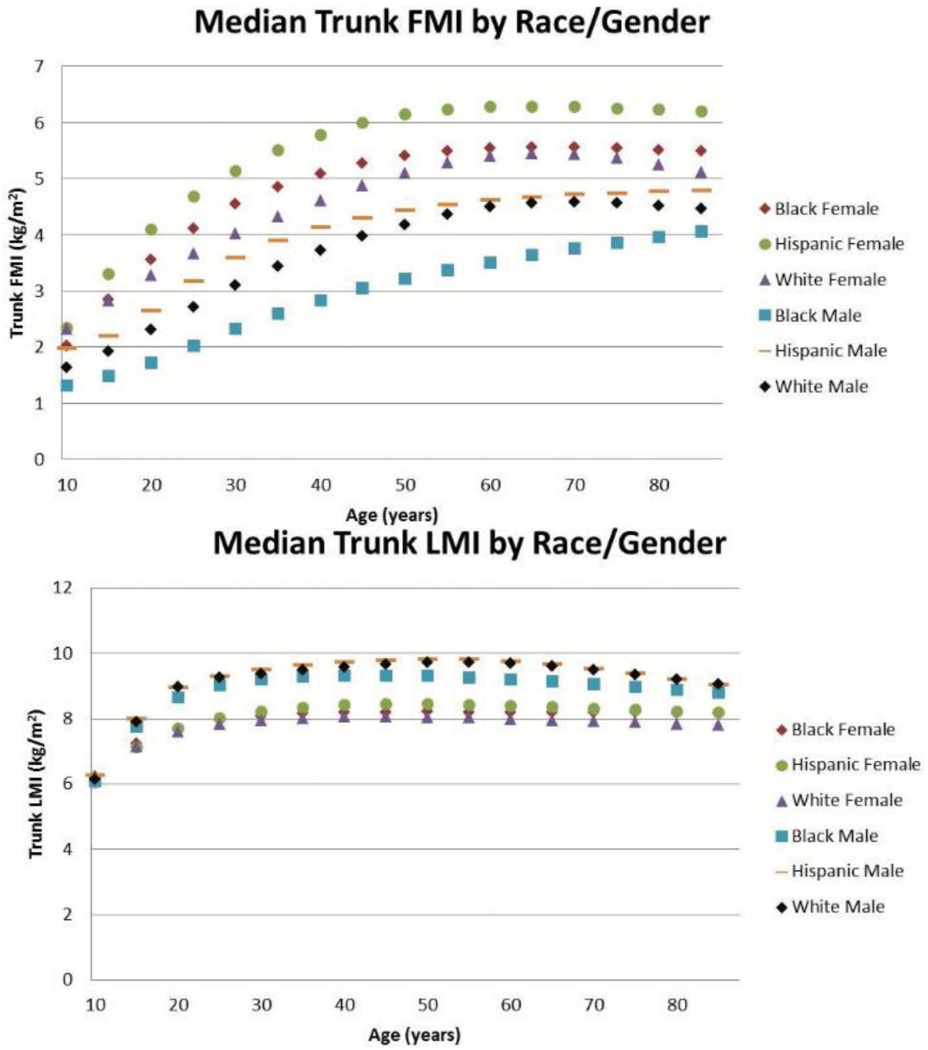


Figure 2-1: Median Trunk FMI and LMI values by ethnicity and sex. This comparison of the median trunk FMI values by ethnicity and sex (top) and median LMI values by ethnicity and sex (bottom). Females generally have larger trunk FMI and lower trunk LMI values than males, and males have a more pronounced drop off in trunk LMI values as they age compared to females. Deviations of each median measure not shown for figure clarity; consult supplemental figures S1-S8 to examine individual data points with percentiles shown or the LMS tables to further examine coefficient of variation

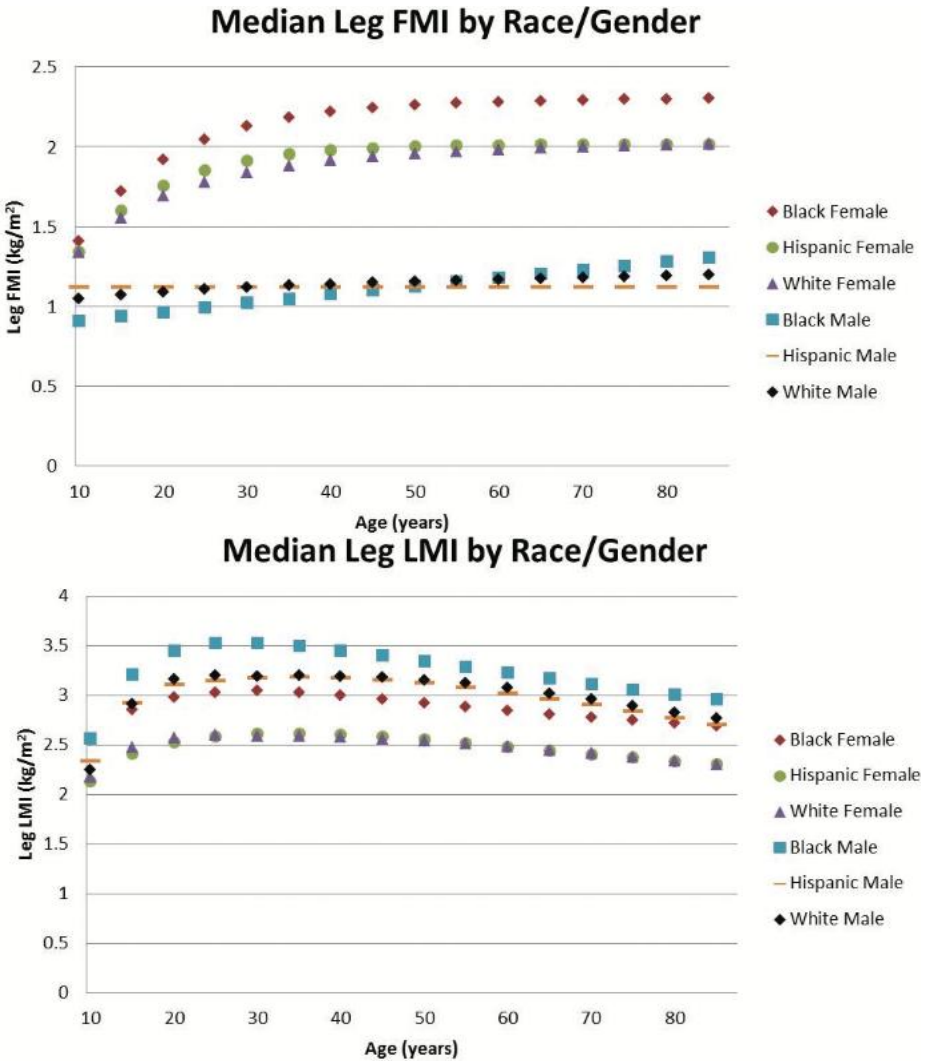


Figure 2-2: Median Leg FMI and LMI values by ethnicity and sex. This comparison of the median leg FMI values by ethnicity and sex (top) and median LMI values by ethnicity and sex (bottom). Females generally have larger leg FMI and lower leg LMI values than males, and black females tended to have larger FMI and LMI values in the legs compared to females of other ethnicities. Deviations of each median measure not shown for figure clarity; consult supplemental figures S1-S8 to examine individual data points with percentiles shown or the LMS tables to further examine coefficient of variation

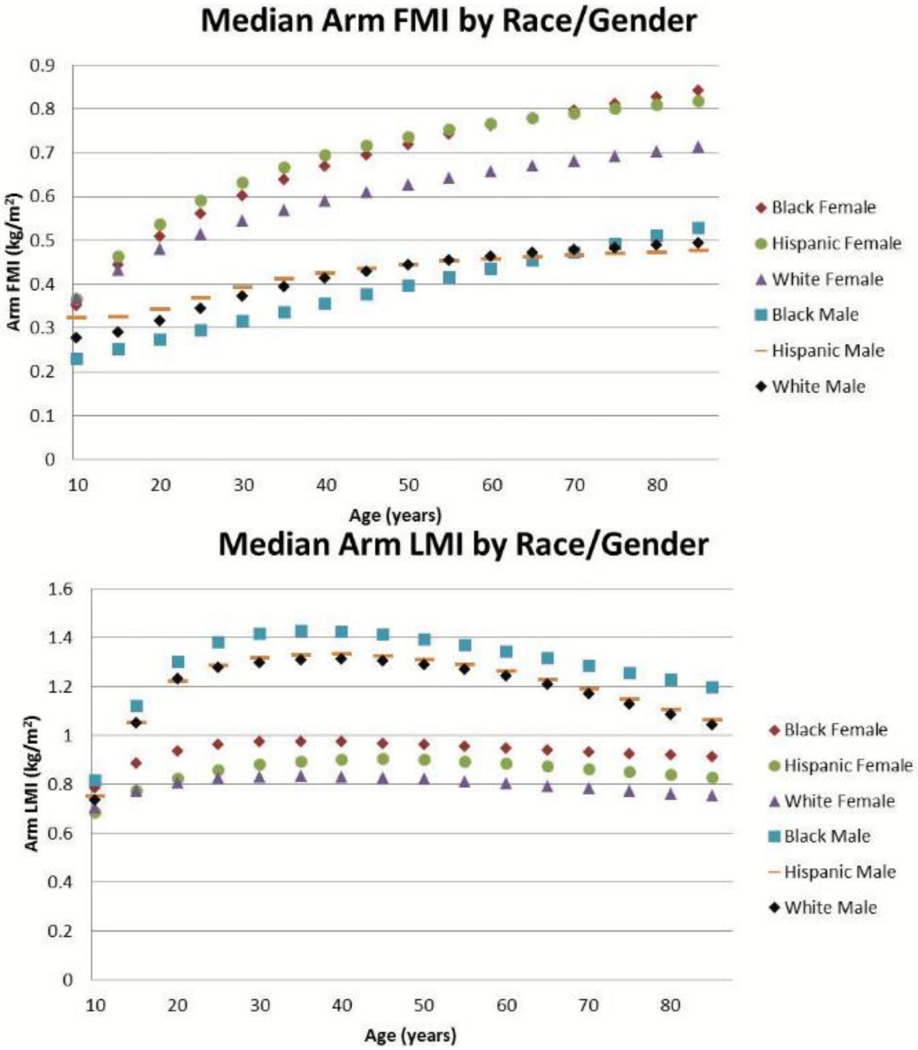


Figure 2-3: Median Arm FMI and LMI values by ethnicity and sex. This comparison of the median arm FMI values by ethnicity and sex (top) and median LMI values by ethnicity and sex (bottom). Females generally have larger arm FMI and lower arm LMI values than males. Males have a more pronounced drop off in their arm LMI values as they age compared to females. Deviations of each median measure not shown for figure clarity; consult supplemental figures S1-S8 to examine individual data points with percentiles shown or the LMS tables to further examine coefficient of variation

Radar Charts:

We developed software to produce radar charts of regional FMI and LMI based on the age, sex, and ethnicity of an individual and the regional fat and lean mass values. The software selects the appropriate LMS table based on demographic information and calculates and displays Z-scores based on the fat and lean mass entries. An example output of the software which displays demographic information, composition information, and the radar chart is included in Appendix B.

Figure 2-4 shows several of the generated radar charts (plots A-F) for 6 individuals and charts those same individuals on a scale of percentile total LMI vs. percentile total FMI (top chart) to show what the generated radar charts look like for individuals of varying overall levels of lean mass and fat mass. Below average LMI individuals are at the bottom half of this chart and low FMI individuals are at the left half of this chart. This top chart, inspired the chart produced in work from Prado et al. to identify compositional categories of individuals (116), shows that the 6 individuals chosen represent a wide variation of overall LMI and FMI.

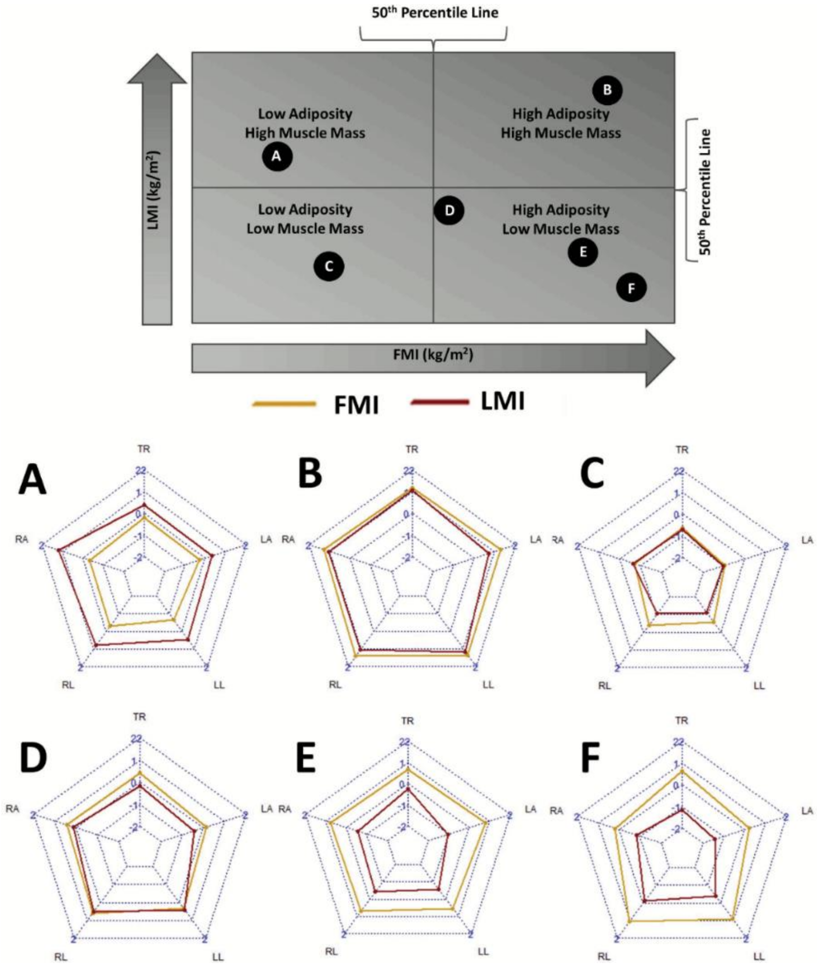


Figure 2-4: Sample radar charts of individuals in different quartiles of lean and fat mass indices. Radar charts of individuals as they fit into quadrants of adiposity and muscle mass. Each labeled circle in the above chart corresponds to an individual radar composition chart below. In the radar charts, each spoke represents: TR = Trunk, LA = Left Arm, LL = Left Leg, RL = Right Leg, RA = Right Arm.

Radar charts A-F in Fig 4 shows the radar charts that display the FMI and LMI Z-scores of subjects A-F that were plotted in the above chart. The top spoke represents the trunk, the lower spokes represent the legs, and the middle spokes represent the arms. An individual with median regional FMI and LMI values (Z-scores of zero) would have two regular pentagons with every spoke at zero. Subject A in Fig 4 shows a high lean mass-low adiposity individual with more lean mass in the right half of their body. Subject B in Fig 4 shows an individual with high lean mass and high adiposity, and their radar chart reflects this with all values regional FMI and LMI Z-scores being above zero. Subject C in Fig 4 shows a sarcopenic individual in the low muscle mass low adiposity category.

The second row of radar charts in Fig 4 depicts three levels of severity in the high adiposity low muscle mass category similar to those defined by Prado et al (116). This high FMI and low LMI quadrant of the top chart contains many high-risk groups including those with sarcopenic obesity. Subject D in Fig 4 shows someone with slightly higher than normal adiposity and slightly lower than normal muscle mass. Subject E shows an individual deeper in this high-risk quadrant of the top chart with above average adiposity and below average muscle mass. Subject F shows an individual severely in this high-risk quadrant of the top chart with much higher than normal adiposity and very low muscle mass relative to that.

We discovered a wide variety of different compositional shapes. We saw more asymmetry in the LMI Z-score distributions across regions than in the FMI Z-score distributions. Further, we found some individuals with distinct distributions, such as individuals who had relatively normal compositions in most regions but their legs, trunk, or arms contained more mass leading to a ‘spike’ in those regions in the radar chart.

2.3.4 Discussion

This study is the first to produce regional LMI and FMI curves and LMS tables representative of the US population, which will be useful in many body composition studies (124). This development of standard FMI and LMI LMS curves for each appendage and a method such as radar charts to visualize body symmetry will prove useful for doctors, researchers, therapists, athletes, and trainers.

These reference curves will help to identify and monitor abnormal regional body composition patterns that arise in childhood and adulthood including sarcopenia, cachexia, anorexia nervosa, female athlete triad, growth hormone deficiencies, cancers, endocrine disturbances, and many others (94,126). Sood et al. showed that trunk lean mass could be predictive of asthma in females (79). Another study showed that two weeks of inactivity specifically reduced the lean mass of the legs in older adults (82). Leg lean mass has been shown to be a predictor of femur BMD (129). Studies of cardiovascular health have shown that trunk fat mass is a risk factor of cardiovascular disease and leg fat mass had a protective effect (80,83). Studies have also shown that regional fat distribution affects the regional rate of lipolysis in obesity (130). It is clear that regional body composition can affect various health outcomes and is worthy of studying, and this research will help to perform studies on height-normalized regional FMI and LMI values to better understand the role composition plays in these conditions.

This work also enables identification and monitoring of the relative symmetry and asymmetry of the lean and fat mass of individuals, as well as research on the effects of symmetry on the body. We noticed several cases of handedness, where one leg or arm had significantly more lean mass than the other. Research has already shown limb and body symmetry plays a role in sports performance and injury prevention (81,131), and these tables and this software enables further research in the role regional symmetry plays in health and performance.

Analyzing the regional FMI and LMI median values highlights several trends that provide insight or warrant further investigation. We can see the clear effect of puberty in all regional LMI values, where males and females start out at similar values until adult development occurs. Once adult development occurs, we can see males have larger LMI values in every region while females have larger FMI values in every region, which is consistent with previous comparisons of body composition by gender and explained by endocrine differences (93,132,133). It is interesting to note the differences are most pronounced in the limbs.

Further, we can see in some cases certain ethnicities have a different trend from other ethnicities of the same sex. Black males had lower trunk FMI and higher leg LMI than their white or Hispanic counterparts. Hispanic females averaged a noticeably higher trunk FMI than black or white females, and white females had lower arm FMI values than black or Hispanic females. These differences in regional composition by sex and ethnicity could serve as avenues of future research for some investigators and highlight the importance in accounting for sex and ethnicity in future body composition studies.

The creation of the software to create radar charts that visualize regional composition will be useful for researchers to intuitively interpret these data and any future studies of regional composition. These charts could aid in interpreting regional composition and in tracking changes over time through interventions such as diet, exercise, or other means. While the radar charts provide a mostly qualitative sense of composition, they provide an excellent structure to start visualizing these data and examining abnormalities, asymmetries, and changes over time.

This paper has several strengths that contribute to the power of the study. First, the large sample size from the NHANES data set provides a wide and comprehensive variety of data that describes the U.S. population by sex and ethnicity. Next, we have used established methods in producing these regional FMI and LMI values and LMS curves and our total body FMI/LMI LMS measurements matched up well with previous studies. Lastly, providing the software to create radar charts will make studies by other researchers much more accessible.

While there are several strengths to this study there are several limitations that, if avoided, would improve the study. A larger sample size especially in the black and Hispanic groups would have allowed for even more accurate reference curves especially at the ends of the age spectrum. Further segmentation of our population into separate LMS curves for youth and adults may have provided slightly improved curves, but this would have caused a sharp transition in Z scores during this transition. Further, our large sample size in this transition period produced LMS curves and de-trended Q-Q plots with enough smoothness to warrant calculating curves for all ages combined. Further, it should be noted that the values reports are only valid to directly compare in new measurements that use the same procedure and same machines as the NHANES dataset. Further studies will need to be performed in order to elucidate the usefulness of these regional values and how to best use them in conjunction with full body composition measures for risk assessment.

From this study, we can conclude that these regional measures follow expected curves and already provide insight about compositional phenotypes by sex and ethnicity. Additionally, these data could be useful for stronger descriptions of risk of mortality and metabolic conditions. Implementing radar charts to visualize regional composition may enable patients to track their regional composition to avoid unhealthy or undesirable compositional shapes (e.g., larger fat mass centile than lean mass centile, larger trunk FMI centile than leg/arm FMI centile). In the future, we plan studies to further investigate the role that regional body composition plays in health outcomes.

3 | Computer Vision

Computer vision is the field that attempts to create software and algorithms that allows computers to interpret images in the various ways humans interpret images. Different applications include edge detection, object segmentation, object detection, language processing, and object identification and classification. This field has expanded wildly as technology has improved and has helped improve the workflow and accuracy of many industries, from automated mail sorting by zip code to crash avoidance systems in cars (44). Given its impact in various fields, we explored the ability of computer vision to identify poor detectability in mammograms, in order to quantify interval risk.

3.1: History, Cats, and V1 Neurons

The roots of computer vision are based in biology and have been refined over the years. Computer vision research was spearheaded by Hubel and Wiesel in the 60's (44,134,135). They examined the response of specific neurons in cats while showing cats an image moving in a specific direction, as seen in Figure 3-1. In this study, they discovered that specific neurons showed activity only when an object was moved along a certain direction, indicating that these neurons served as edge detectors along specific angles and orientations.

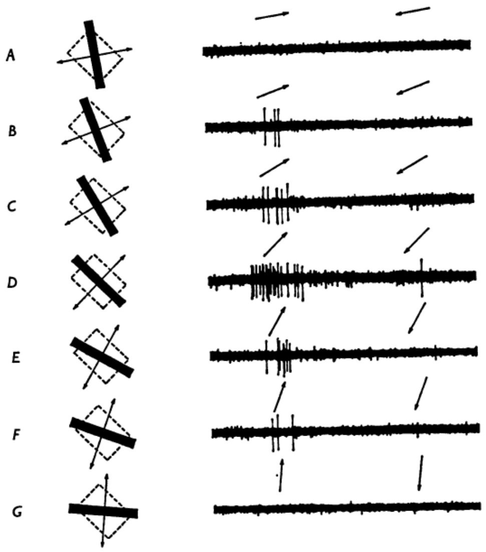


Figure 3-1: Figure from Hubel and Wiesel (134,135) showing a neuron's varying responses to object moving in different orientations, which was one of the foundational works that led to edge and feature detection in computer vision.

Scientists already knew of several functions that behaved similarly to this, such as derivatives, where a strong response is elicited when a large change happens along a given axis or variable. Further research merged these ideas by applying these algorithms to detect a change along an image direction, rather than along a specific function. These were the first edge detectors in computer vision, all inspired by our own biology.

3.2: Edge and Feature Detection

3.2.1 Edge Detection

The field of edge detection was immediately iterated and improved upon, as understanding and replicating robust and automated edge detection would be useful in a huge number of fields such as robotics, artificial intelligence, computer graphics, image processing, and neuroscience (44). Being able to properly define edges helps identify objects, segment different parts of an image, and understand and interpret orientation. Given most edges present themselves in images as a difference in intensity across a certain orientation, many designed filters to detect a sharp intensity gradient (44). The values in these filters look for an intensity gradient across a specific orientation and output a high value when that gradient is seen. We can see an example of convolving a vertical edge filter on an image of the number 7 in Figure 3-2. We can see a large output exists at the main vertical edge of the '7'. If one knows the specific objects you are looking for such as numbers, letters, or tracking a specific object, it is then possible to look at these edge responses at various orientations and make basic decisions. Even simple filters like this can be adapted and applied to help recognize digits for automating mail processing, determine collision risk or spatial orientation for different objects, or helping to segment objects within a scene (44).

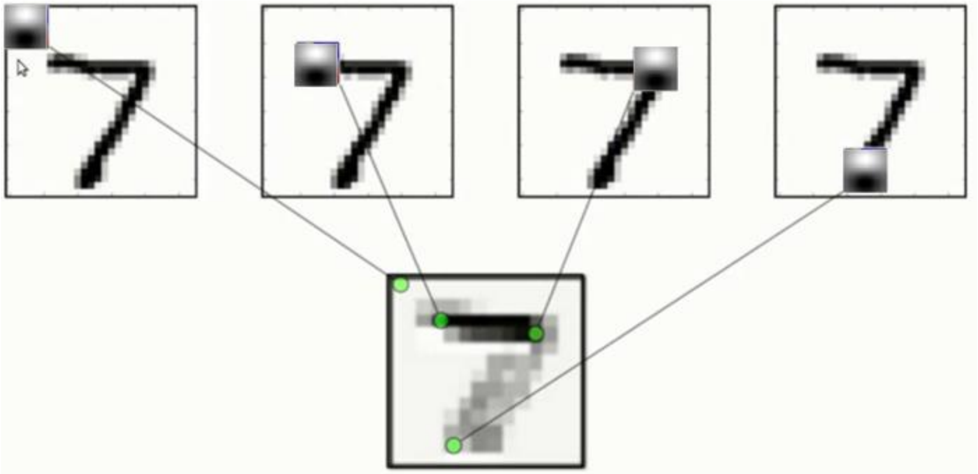


Figure 3-2: Convolution showing how a horizontal edge detector produces different responses in different regions. In regions with no variation (left) there is no output, with an edge there is a strong output (middle cases), and a medium output occurs at a medium intensity edge (right). Source: (136)

Various filters have been developed over years to help detect edges along a certain orientation through convolution of a specific function, including through the use of Laplacian functions, Sobel filters, Prewitt filters, and several others (137). Most of these functions look for a difference in intensity across a specific orientation, as seen in the Sobel filter shown in Figure 3-3.

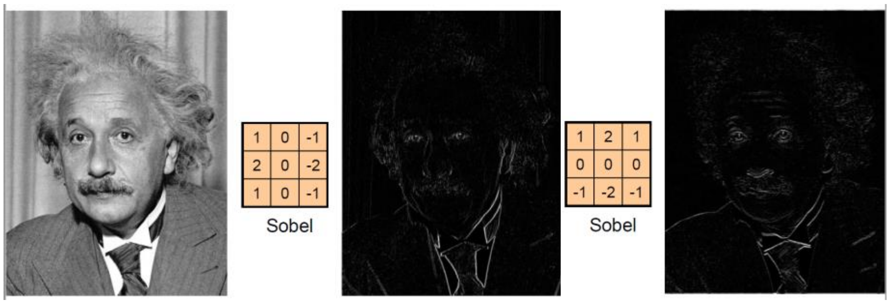


Figure 3-3: Example showing how a Sobel filter simulates the edge orientation neurons as a vertical edge detector (middle) and horizontal edge detector (right). Source: (44)

Unfortunately, edge detection alone is not enough to completely satisfy every need in computer vision, such as advanced object identification, object characterization, and scene interpretation. Scientists continued working to develop further methods of computer vision, including feature detection and direct object identification.

3.2.2 Feature and Object Identification

Feature Quantification

A next level of sophistication in computer vision was to identify specific features and objects within an image. ‘Features’ is a broad term that describes a quantifiable property of an object, image, or image region, and can include likelihood of specific objects like eyes, or various statistical or textural measures of an image (44). Identification and quantification of these features is useful in several aspects of computer vision. First it can help in object segmentation, as a single object will often maintain several image features throughout the object, such as image intensity, color, or texture. Further, this can help in identifying more complex objects, as an object with a region similar to an eye is likely to be a face of an animal or human. To better understand the usefulness of these features, we need to understand which features are often quantified and defined.

Features can include direct features on an object like the presence of an eye, arm, or face. More often, however, it is used to describe statistical measures. These features could include basic measures such as image minimum, maximum, variance, mode, percentile, and other basics statistics qualities (138,139). Statistics of the histogram of the image can also be considered, such as kurtosis, adherence to normality, or skewness (138,139). Because similar objects often maintain certain statistical measures throughout the object, this helps in object segmentation and scene interpretation (44).

Unfortunately, many of these basic statistics only give a sense of absolute values and distributions of values but cannot compare the positioning of these pixel values throughout the image. Images with the same basic statistics but different positioning of these pixel values are said to have different textures. Some examples of images with similar statistics but different textures are shown in Figure 3-4. Identifying texture is an important distinguishing characteristic of many human vision tasks, such as edge detection and image segmentation (44). Therefore, it is important to identify texture in computer vision tasks as well.

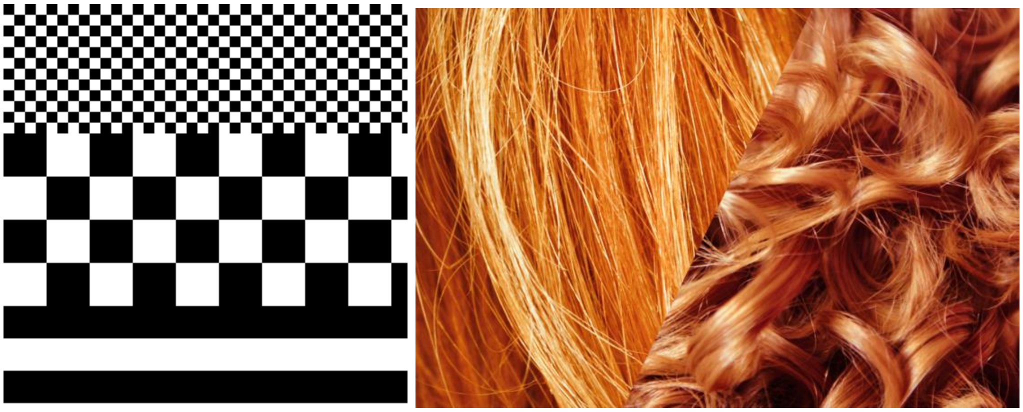


Figure 3-4: Sample image showing a variety of textures with similar image statistics at left, and images with similar color and statistics, but different texture qualities at right. This shows that texture analysis is necessary to fully understand and interpret an image. Source: (140)

Several algorithms and procedures have been developed to help identify and quantify textures. First, statistics of edge detection images can give a sense of frequency of edges or number of objects in an image (44,137). When locations of images or regions are identified, orientation, eccentricity, and object area can be quantified to differentiate objects of different shapes. Further, a gray-level co-occurrence matrix (GLCM) compares the values of connected pixels and can give a sense of various texture features such as image contrast, correlation, energy, and homogeneity. These texture differences help to identify changes in object type, orientation of an object, or other ways to interpret a scene that are useful to researchers.

Object Identification

In addition to identifying features within an image, scientists have worked to directly identify and detect specific objects. This is often a goal of computer vision applications, whether it is to detect a postage stamp on a letter, detecting eyes or faces in a crowd, or detecting different types of objects in satellite images.

Many methods surrounding object identification involves first creating a template of an object (or part of an object) that one is looking for. This can be as simple as a sample image of what you are looking for, like an eye if you are looking for eye-like objects. Figure 3-5 shows how having an eye template and then applying a cross-correlation of that template through an image will show spikes in value right where eyes are located in an image (44).

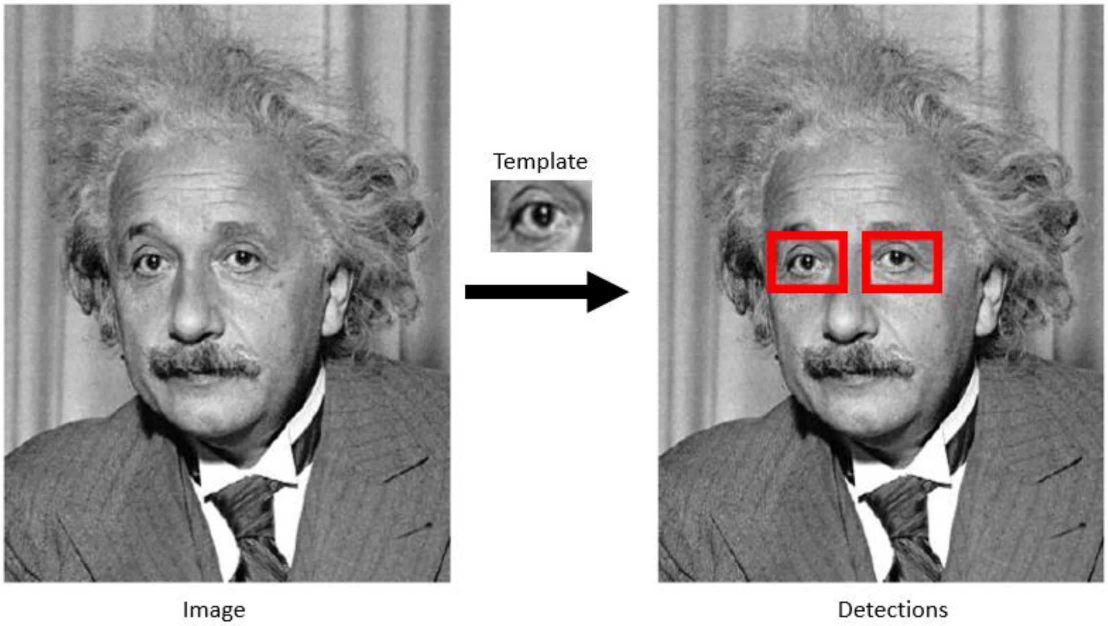


Figure 3-5: Sample image showing that searching for a specific image feature (an eye in this case) can identify multiple areas that are likely to contain that feature in specific circumstances where the object is similar to the template. Source adapted from: (44)

This application can sometimes work in highly structured images where we know the specific object being looked for. However, this simple methodology is often not sophisticated enough to reliably work. For example, if you apply this same methodology to identify and locate in a simple image you get good results, but in a more realistic image with noise and object variety the results are much worse. Figure 3-6 shows an example of successfully attempting to identify a chair in a simple image with a known object versus failing to identify a chair when the chair is slightly dissimilar in a busier image.

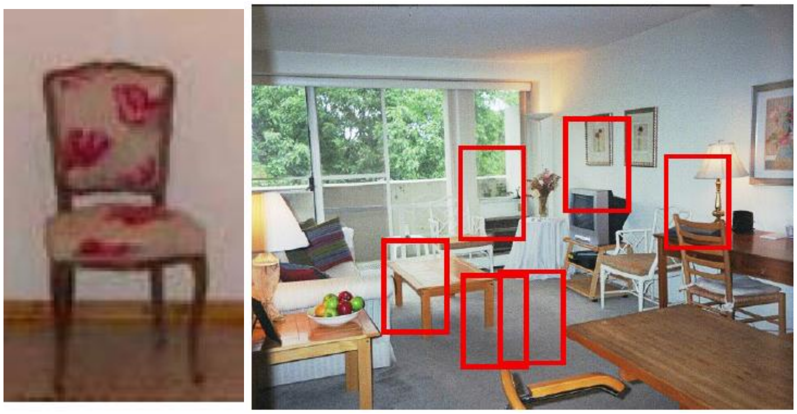


Figure 3-6: Image of a feature search of a chair (left) that fails when searching in a complex background and produces several false positives (right). Source: (44)

That is a fundamental problem of traditional computer vision: specific knowledge is required about the objects one is looking to identify, as well as the scale and perspective from which the image is taken. Fortunately, in medical imaging, this is often the case. Many medical imaging procedures are heavily standardized with images being taken at the same angle each time, and very specific objects (fractures, tumors, etc.) are looked for (11). This has enabled the development of model observers and detectability metrics to be created and applied to medical fields.

3.3: Model Observers

Model observers are a subfield of computer vision that work to quantify how an object will be detected in human observer tasks, such as many medical diagnostic tasks. Many of these model observers have been shown to correctly quantify the detection capabilities of various imaging tasks in various backgrounds (141). By understanding how these model observers work, we may be able to apply these techniques to determine the ability to detect an object in breast imaging tasks, in order to better quantify interval cancer risk. Over the years a variety of model observers have been developed for medical applications, including the design of CAD systems (51,141–147) . The model observers used in these studies all varied in the amount of information assumed about the signal of interest and in the background.

3.3.1 Non-Channelized Filters

There are two main classes of model observers: non-channelized and channelized observers. Non-channelized observers are observers that look for specific templates or objects and do not account for the fact that they eye responds to various frequencies at different amounts. Even with this limitation, it has been shown that these non-channelized observers correlate well with humans in various detection tasks (148,149). An advantage of these filters is that because they are generally more simple, they are easier and quicker to implement (141,150).

Non-prewhitening (NPW) filters have been used extensively in medical imaging (150). For its effectiveness, the NPW model observer is one of the simpler models to implement, as it has a filter that exactly matches the expected signal that is attempted to be detected. The NPW filter is unable to account for noise correlations in its model, but it assumes full knowledge of the signal of interest and. This model is usually successful in many predicting human performance in white noise but fails in certain anatomic backgrounds. The non-prewhitening eye (NPWE) filter adds an eye filter to the NPW filter, which helps the model observer take into account the fact that the eye responds differently to different spatial frequencies (141,148,150).

A more advanced non-channelized model observer is the Hotelling filter, which was proposed by Barrett et al (141). The Hotelling observer differs from the previous filters mainly that the template used accounts for the expected signal profile as well as the expected background statistics. This helps to better identify how the noise will affect the patterns and objects seen by a human observer in certain backgrounds. It has been shown that background noise statistics play a part in object and signal detectability (151,152), so it is often a good idea to consider which observer is best considering the background statistics.

3.3.2 Channelized Filters

Channelized model observers are fundamentally different from non-channelized filters in that they pre-process the image through a series of filters, or channels, that are tuned to specific spatial frequencies or orientations. These models then compute the correlation of each channel to obtain a response of each channel at a possible signal location. This allows the observer to combine information and responses of a variety of frequencies at once. The filters chosen can vary wildly, but two common sets of channels are the difference of gaussians and the Gabor channels (141,150).

In addition to a variety of filters, there are a variety of model observers that handle the channel responses differently. Simple models can take the sum of each channel response or just take the response of the most sensitive channel. Additionally, non-prewhitening channelized matched filters have been developed, which applied a NPW model to each image channel response (141). There are also channelized Hotelling filters that further weight the response of various channels based on the frequency and orientation. Each of these filters have a variety of applications in medical imaging (141).

3.4: Figures of Merit and detectability

Figures of merit are the measures that quantify the level of detectability after running an image through a model observer. There are a variety of ways to quantify the output of a model observer, each with its own benefits and drawbacks.

3.4.1 Forced Choice Tests

A method that is often used to determine detectability in both human observers and model observers is an alternative forced choice test (AFC). There are 2-alternative forced choice (2AFC) and 4-alternative forced choice (4AFC) tests (141,153,154). In both of these a set of either 2 or 4 images is shown with a similar background, one of which has the signal of interest. If the reader is able to correctly identify the image with the signal at a certain rate, the signal is deemed detectable. This process helps to specifically quantify whether a signal is detectable in a given background and can be automated by asking whether a model observer is able to correctly guess the signal-present image.

3.4.2 Metrics of Detectability

In addition to AFC tests, several indices of detectability have been created as measures of detectability strength. These indices measure the different between the channel outputs of the signal present and signal absent images divided by the overall variability, as seen in the following equation (141). These are often adjusted when accounting for image correlation between channels, certain types of background noise, or other factors. These metrics also can also include measures of intensity, signal to noise ratios, and other imaging metrics.

$$d' = \frac{\langle \lambda_s \rangle - \langle \lambda_n \rangle}{\sigma_\lambda}$$

4 | Model Observer Detectability in Mammography

Because of the various medical fields in which model observers have been applied to quantify detectability, it is a reasonable hypothesis that model observers could also be applied to quantify detectability in mammograms, especially in an image where the background is usually consistent. There has been much work to apply these computer vision methods to quantify detectability and detect potential lesions in mammography.

4.1: Detectability Limits in Mammography - CDMAM Phantom

The limits of detectability in mammography has been extensively studied and quantified for quality assurance (QA) tests by developing phantoms with human readers.

4.1.1 CDMAM Explained

The Contrast Detail Mammography (CDMAM) phantom is a well-established phantom that helps investigate detectability limits of small objects in mammograms, such as small tumors and calcifications (151,155). It is a phantom scanned on its own with several gold discs of varying diameters and thicknesses where the reader is tasked to identify the locations of the discs, making a reader perform a 4AFC test. In addition, this can be analyzed to help produce a Contrast Detail (CD) curve, a curve showing detectability thresholds for objects of various sizes (155,156). Figure 4-1 shows the CDMAM phantom as well as a characteristic CD curve that can be produced from it.

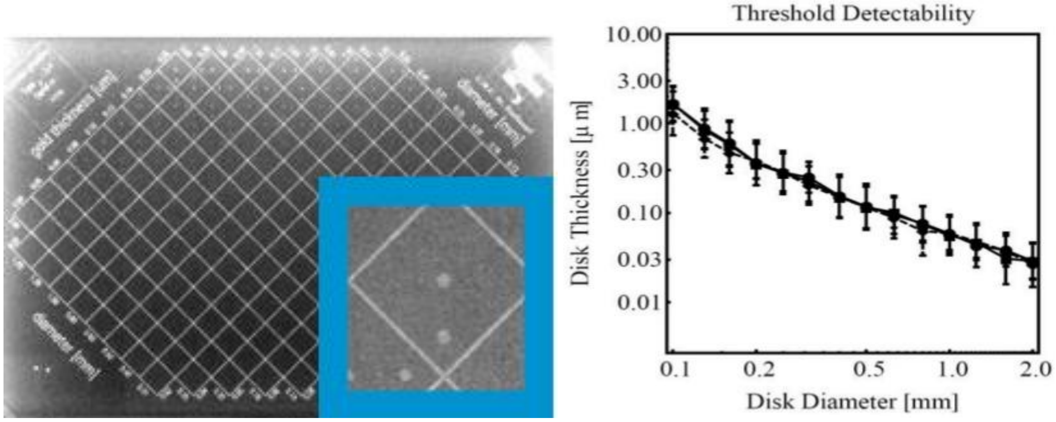


Figure 4-1: Image of a CDMAM phantom enhanced (left). At right, a CD curve is shown that identifies the minimum object thickness for detectability at each diameter. Source: (142,155)

The contrast detail curve is a curve summarizing the detectability limits of an image at various diameters. This helps radiologists identify the limits to detect calcifications or tumors if they exist at various sizes and is also used for QA tests (156). In general, the curve follows a logarithmic scale and the lower the curve, the better the detectability. This phantom helps quantify the smallest objects that can be detected in a mammogram with no background, often with objects roughly the size of microcalcifications. Because this phantom does not include an anatomically correct background, it cannot be directly applied to quantify risk of interval cancer or of missing a cancer directly in a mammogram, as the background noise plays a significant role in the detectability of an object.

4.1.2 IQF

In addition to contrast detail curves, the results of a CDMAM phantom test can be summarized as an Image Quality Factor, or IQF (149,155,157). The IQF summarizes the CD curve as a single number that is calculated as an inverse area under the curve. The larger the IQF, the better the performance of the mammography unit. While this IQF number does not differentiate detectability at different sizes like the CD curve, it can be split to identify detectability of only large, medium, or small diameter objects (155).

$$IQF = n / \sum_{i=1}^{n} T_{i,min} D_i$$

4.2: Simulating Lesions

Several groups have worked to determine methods to create, simulate, and model virtual pseudo lesions. The motivations behind these studies have been purely academic, to help identify differences between lesion shape and outcomes, or to better understand various obstacles in detecting tumors in mammography. Shaheen has developed methods that simulate the shape and appearance of microcalcification and lesions, both in 2D and 3D environments for use in virtual trials. Further, Pezeshk et all have shown the ability to simulate lesion insertions into mammograms, allowing for improved ability of virtual clinical trials (158–161). Most of these studies implemented their method by simulating the X-ray spectrum as well as the expected attenuation of breast tissue and lesions, based on known attenuation values of these tissues (66,162,163).

4.3: Study: Derived mammographic masking measures based on simulated lesions predict the risk of interval cancer after controlling for known risk factors: a case-case analysis (164)

If we are able to combine the work that has been done to simulate lesions as well as the work that has been done in computer vision to develop model observers that simulate human vision, it is likely that we could create a measure of lesion detectability throughout a mammogram that would be a strong risk factor for interval cancer.

4.3.1 Introduction

Digital mammography is the clinical standard in screening for breast cancer in average-risk women. However, radiologically dense and complex tissue can reduce the screening sensitivity of detection from 84% in low density mammograms to 68% in high density mammograms, leading to cancers missed by screening mammography (13,165). This effect is commonly called mammographic masking, and can lead to one type of interval cancer where lesions are missed by screening mammography due to dense tissue masking the presence of a lesion (15,20,21). Roughly 13 percent of the breast cancers diagnosed in the U.S. are interval cancers based on initial assessment (166).

Women with dense breast tissue have a higher rate of interval cancer, and legislation has been passed or is in process in 41 states to mandate the notification of women found to have breast tissue, including the recommendation to consider additional supplementary screening methods (13,167).While breast density measured using Breast Imaging-Reporting and Data System (BI-RADS) scores can serve as a rough proxy for masking in measuring interval cancer risk, the scores are subjective and do not account for the texture and distribution of dense tissue (168–170). As a result, the American College of Radiology has called for direct measures of mammographic masking as a way to predict risk of interval cancer (171).

We hypothesized that mammographic masking could be directly measured by developing software that inserts pseudo-lesions into clinical mammograms and these measures can predict the effectiveness of mammography to detect cancer at the time of screening mammogram acquisition. To test this hypothesis, we performed a case-case analysis of women who had developed breast cancer either by screening mammography (screen-detected cancer) or clinically during the interval after a normal screening (interval cancer) in a cohort of women with raw digital images. The purpose of this study was to create masking measures and apply them to a population of women who experienced either screen-detected or interval cancers after controlling for breast density.

4.3.2 Methods

The aim of the study was to examine if our generated masking measures could classify between screen-detected and interval cancers.

Subjects and Data

Raw digital screening mammogram images for both Craniocaudal (CC) and Mediolateral Oblique (MLO) views were collected from 2006 to 2015 from three radiology facilities. During this time period all interval cancers, defined as invasive cancers identified within 12 months of a negative screening examination, from these facilities were identified. An equal number of screen-detected cancers were matched by based on age, race, exam date (within 365 days) and time since last imaging examination. Screen-detected cancers were defined as invasive cancers identified within 12 months of a positive screening examination. All mammograms were interpreted prospectively by radiologists during the course of routine clinical care. Cancers were identified by annual linkage with the respective cancer registry. Ethics approval was obtained by the Institutional Review Board for this retrospective analysis of mammograms for masking properties, informed consent was given, and the study was Health Insurance Portability and Accountability Act (HIPAA) compliant. For each case identified, the mammograms prior to cancer detection were selected for analysis.

Derivation of X-Ray Spectra and attenuation curves

We extracted the relevant imaging technique factors from the Digital Imaging and Communications in Medicine (DICOM) image headers (kVp, mAs, breast thickness, anode and filtration materials and thicknesses) reproduced the characteristics of the incident X-ray spectra (67). We then used previously established methods to determine the additional attenuation from the breast tissue and the resulting X-ray spectrum at each pixel (172,173).

Creation of Virtual Lesions

Next, we created our virtual lesions by inserting a Gaussian profile of an additional attenuation into the raw data mammograms (67) as it corresponds to the most commonly clinically found tumor shapes (171) with full width half max (FWHM) to emulate clinically relevant tumor sizes (0.5 to 3 cm) (174). The magnitude of attenuation of our virtual lesions was calculated by combining the X-ray spectrum with the attenuations that would be observed in the Contrast Detail Phantom for Mammography (CDMAM) phantom, which contains gold discs with varying size and thickness and is used to quantify detectability thresholds of different imaging configurations by having readers attempt to identify the locations of these disks (151). We analyzed peak lesion thickness ranges of 0.03 – 4 um of gold attenuation, which correspond to 0.05 - 3 cm of lesion peak thickness (67). These virtual lesions are then blurred by the point spread function of the mammography unit (175).

Model Observer

Model observers are used often in medical literature to emulate the result of a human reader in a detection task.(176) We chose to implement the Non Pre-Whitening Matched Eye Filter (NPWE), as studies have shown the filter correlates with human observers in the case of detecting lesion like objects in a background similar to what is seen in mammography.(143) Equation 1 shows the NPWE filter, where E is an eye filter, g_s is the image template with the signal of interest (virtual lesion), and g_n is the image template without the signal of interest.

$$w = E^T[g_s - g_n]E \qquad (1)$$

When the result w is above a detectability threshold, the signal of interest is deemed detectable. We calculated IQF at 5 mm intervals patch by patch throughout the breast for the full range of diameters and visualized this through the image as an IQF map. We exclude non-breast regions from calculations by defining the skin edge and muscle region via thresholding and in-house software.

Determining Detectability Thresholds

In order to determine the threshold of detectability, we implemented a 2-Alternative Forced Choice (2-AFC) test.(143) In this test, two images are presented to the model observer. Only one contains the inserted lesion. The observer chooses the image with the highest response to the filter, and if this choice is correct more than 92% of the time(143) then the model observer is deemed able to accurately detect the lesion in that image.

To run a 2-AFC test, it is necessary to have lesion-free image patches similar to the image patch with the virtual lesion. For this, we generated simulated image patches with the same mean, standard deviation, and radially averaged power spectrum as each lesion-containing image region. We used these images as our patches without signal in order to perform the 2-AFC test.

Producing Masking Maps

Performing the 2-AFC test allowed us to calculate the threshold peak thickness value that was detectable for each FWHM virtual lesion size. From this, we produced a CD curve for each region, a curve that plots thresholds of detectability across multiple diameters and is important to image quality studies for mammography.(177,178) We then summarized this CD curve by calculating the IQF:

$$IQF = n / \sum_{i=1}^{n} T_{i,min} D_i \qquad (2)$$

Where n is the number of virtual lesion FWHM values and D_i and $T_{i,min}$ are FWHM values and threshold peak thicknesses of detectability for each virtual lesion FWHM. Larger IQF represents higher detectability in that patch, i.e., that masking is less likely. We calculated IQF patch by patch throughout the breast for the full range of diameters. This entire set of IQF values can then be visualized across the image as an IQF map, which highlights patches of high and low detectability. Next, we consider IQF map first order statistics, second order statistics, and other summary IQF measures that may classify between mammograms with low and high masking.

Statistical Analysis

Based on these IQF maps, we calculated various summary masking measures ranging from common statistics to measures of the grey level co-occurrence matrix (GLCM). We generated these IQF maps for CC and MLO views of the right and left breasts. We then used conditional logistic regression with interval vs. screen-detected cancer as the outcome and each of these masking measures as the predictor. We first fitted a conditional logistic regression using only BI-RADS density. We then fitted two conditional logistic regressions for each masking measure: one univariate regression using only the masking measures and one regression using BI-RADS density as well as the masking measures. We computed and compared the Receiver Operator Curve Area Under the Curve (ROC AUC) from the conditional logistic regressions (172). The best masking measure was selected as the measure with the largest improvement to the AUC compared to the model with only BI-RADS density, and also had a significant p-value of inclusion into the model with Bonferroni correction after controlling for BI-RADS density with a critical significance value of 2.85E-3.

Image analysis and calculation of masking parameters were carried out using MATLAB r2015a (Mathworks, Natick, MA). Conditional logistic regressions were carried out in R version 3.2.2 using the clogit function in the survival package. Analysis also used R.matlab, ROCR, dplyr, e1071, ICC, lattice, gdata, PredictABEL, and psych packages.

4.3.3 Results

Descriptive Statistics

Table 4-1 shows the demographic information of the women from each case-type. Within this dataset we had 182 women diagnosed with an interval breast cancer. These were matched by exam date (within 365 days) and time since last imaging examination to 173 women with screen-detected breast cancers. There were no screen-detected cancers that matched by age and race for 9 of the interval cancers, and these were excluded in the conditional logistic regressions. The descriptive statistics showed a difference in Body Mass Index (BMI) and BI-RADS density between groups, but the other demographic and risk information was not significantly different.

Table 4-1: Demographic information and statistics of the screen-detected and interval groups.

	Screen-Detected Group	Interval Group	P-Value
N (CPMC/MGH/UCSF)	173 (122/38/13)	184 (108/53/23)	
Age, years (Standard Deviation)	57.8 (10.9)	56.8 (11.8)	0.28
BMI, kg/m^2 (Standard Deviation)	24.9 (4.7)	23.5 (4.3)	<0.0001
Race:			0.88
White	127	129	
African American	3	4	
Chinese	25	27	
Filipina	3	3	
Hispanic	0	2	
Japanese	5	9	
Mixed	5	6	
Other Asian	2	1	
Other Non-Asian	3	3	
Menopausal status	119 (69%)	123 (67%)	0.69
Family history of breast cancer	47 (23%)	60 (33%)	0.25
Previous history of breast biopsy	55 (32%)	68 (37%)	0.33
BI-RADS Frequency:			0.008
A: Almost Entirely Fatty	11	3	
B: Scattered Fibroglandularities	50	33	
C: Heterogeneously Dense	61	78	
D: Extremely Dense	19	53	
Missing Data	19	7	
Unknown	13	10	

Figure 4-2 shows a sample region where a simulated Gaussian lesion was inserted into a region of a mammogram, with the peak thicknesses and FWHM of the virtual lesions shown in Table 4-2.

Table 4-2: Table showing all combinations of simulated lesions in terms of their FWHM and peak thickness.

Dimension	Sizes Used
Full Width Half Max (mm)	30, 28, 25, 22, 20, 18, 15, 12, 10, 8, 5, 4
Peak Thickness (um Au Equivalent)	4, 3.5, 3, 2.5, 2, 1.5, 1, 0.75, 0.5, 0.36, 0.25, 0.2, 0.16, 0.13, 0.1, 0.08, 0.05, 0.03

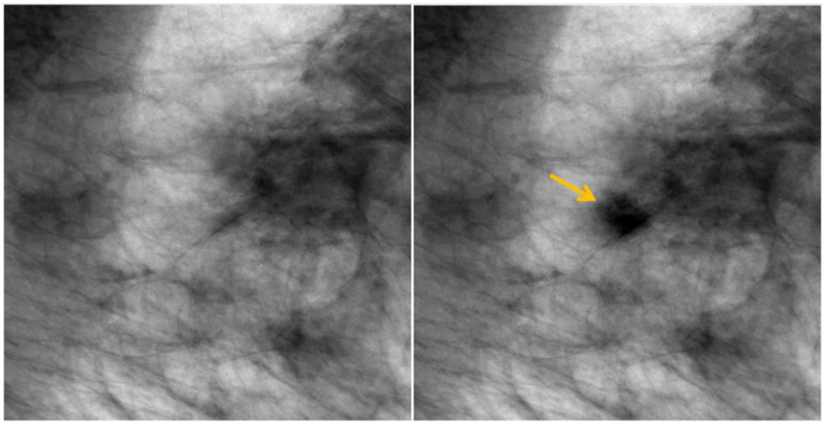


Figure 4-2: Sample image of a raw mammogram image (Left), and that same image region with a pseudo-lesion simulated and inserted into the image.

Figure 4-3 shows a masking map generated from a sample mammogram. Intuitively, we can see that the masking map has lower values in regions of high density.

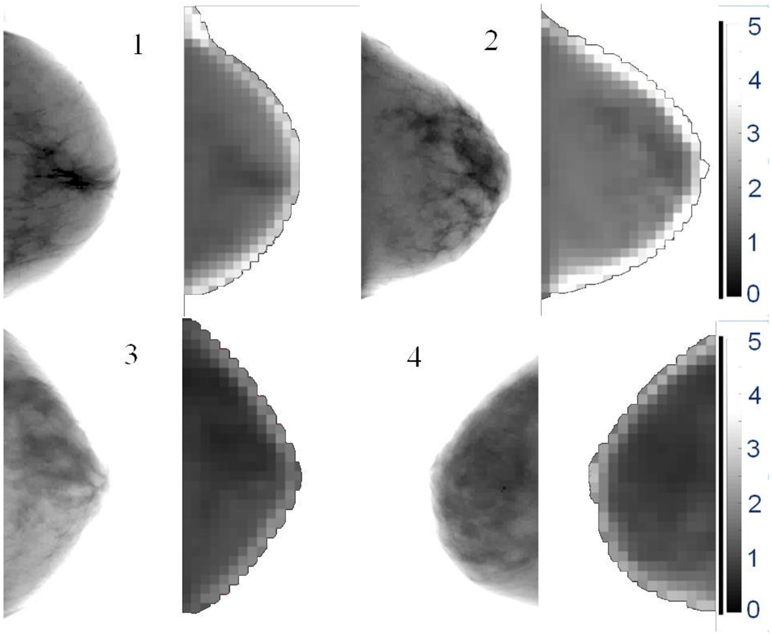


Figure 4-3: Raw data mammograms (CC View) and the respective generated IQF masking maps for sample images with BI-RADS density 1 (top-left), 2 (top-right), 3 (bottom-left), and 4 (bottom-right). Scale of IQF values are shown at right and are consistent across images. IQF values closer to zero are represented as darker pixels, indicate higher levels of masking, and are seen in the higher density images. Raw data mammograms have been contrast-enhanced to better see dense regions.

The full list of masking variables used in the conditional logistic regression between interval and screen-detected cancers and their univariate conditional logistic regression results are shown in Table 4-3.

Table 4-3: List of all masking measures analyzed and their respective AUC and P-value for the univariate classification and classification after controlling for BI-RADS density.

Masking measure	AUC for each masking measure	P-value in regression model	AUC after controlling for BI-RADS density	P-Value for inclusion of masking measure
IQF Mean	0.60	6.62E-07	0.68	1.19E-03
IQF Median	0.61	3.04E-06	0.68	02.18E-03
IQF Sum	0.63	8.58E-11	0.67	2.52E-06
IQF Entropy	0.59	6.64E-06	0.68	8.05E-04
IQF Kurtosis	0.56	0.015	0.67	0.214
IQF Skewness	0.59	1.78E-04	0.67	7.55E-02
IQF 10th Percentile	0.61	1.12E-06	0.69	1.72E-03
IQF 25th Percentile	0.61	7.02E-07	0.68	1.17E-03
IQF 75th Percentile	0.59	4.37E-06	0.69	2.91E-03
IQF 90th Percentile	0.60	1.15E-07	0.68	4.07E-04

Masking measure	**AUC for each masking measure**	**P-value in regression model**	**AUC after controlling for BI-RADS density**	**P-Value for inclusion of masking measure**
IQF Percent Area below 1	0.59	5.17E-06	0.68	6.72E-04
IQF Percent Area below 2	0.60	9.91E-07	0.68	2.43E-03
IQF Percent Area below 3	0.58	4.22E-05	0.68	5.58E-03
IQF Percent Area below 4	0.58	3.42E-04	0.68	1.22E-02
IQF GLCM Contrast	0.54	0.032	0.68	4.43E-02
IQF GLCM Correlation	0.58	1.72E-04	0.67	9.50E-04
IQF GLCM Energy	0.60	8.11E-07	0.68	4.58E-05
IQF GLCM Homogeneity	0.57	7.68E-03	0.67	3.91E-03

Many of the masking measures had similar AUC levels and similar ROC curves, as seen in Figure 4-4: This figure shows the ROC curves and associated AUC values of predicting interval and screen-detected cancer in the test set for the most significant masking measures. These measures improved upon the prediction after including other breast cancer risk factors.

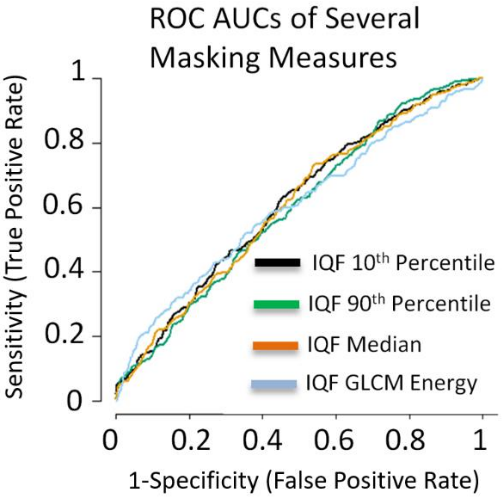


Figure 4-4: ROC curves for several of the masking measures in predicting interval vs screen-detected cancer. All masking measures had similar AUCs and ROC curves in the univariate analysis.

Figure 4-5 shows the ROC curves and associated AUC values of predicting interval and screen-detected cancer of the different models for the IQF 10th percentile, the masking measure that improved the AUC the most compared to the density only model.

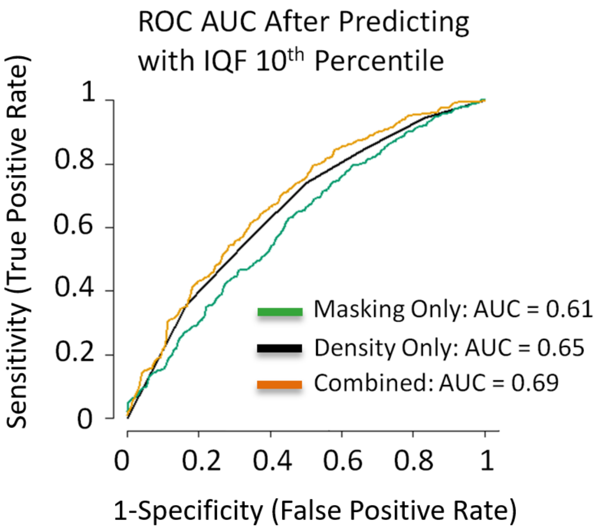


Figure 4-5: ROC curves of predicting interval vs screen-detected cancer for the best performing masking measure (IQF 10th percentile). After controlling for BI-RADS density, this masking measure improves the AUC from 0.65 to 0.69. P-value for inclusion of masking measure in combined model = 1.7 E-3.

Several IQF masking measures were statistically significant for inclusion in the conditional logistic regression, even after controlling for BI-RADS density. Table 4-4 contains the results of the proportion analysis. In Table 4-4 we see decreasing proportion of screen-detected cancers in the groupings with high levels of masking.

Table 4-4: Comparison of proportions of screen-detected cancers by BI-RADS density groupings and masking measure groupings, grouped by percentiles similar to BI-RADS density distribution. Left most quartiles of the masking measure correspond to the quartile with the lowest masking, and right most quartiles with the highest masking levels.
1st – 0-10th percentile of the value of the masking measure
2nd – 10-50th percentile of the value of the masking measure
3rd – 50-90th percentile of the value of the masking measure
4th – 90-100th percentile of the value of the masking measure
IQF: Image Quality Factor, BI-RADS: Breast Imaging-Reporting and Data System.

Masking measure	**Measure by quartile**					
Clinical BI-RADS	A	B	C	D	Unknow	Overall
Interval Cancers	3	32	77	53	17	182
Screen Detected	11	50	61	19	32	173
Proportion of screen-detected cancers	0.79	0.61	0.44	0.26	0.65	0.49
IQF 10th Percentile	4th	3rd	2nd	1st		Overall
Interval Cancers	11	62	84	25		182
Screen Detected	25	79	57	12		173
Proportion of screen-detected cancers	0.69	0.56	0.40	0.32		0.49

4.3.4 Discussion

We identified several measures of masking that are associated with interval compared to screen-detected cancers even after adjusting for BI-RADS density, and these masking measures may be useful to better identify groups at high risk of interval cancer. The IQF 10th percentile measure provided the largest gain in the AUC when added into the model, raising the AUC from 0.65 with density alone to 0.69. This indicates that these masking measures contain information about interval breast cancer risk that is not captured in the BI-RADS density classification alone.

The most significant measure, the 10th percentile of the IQF map, is an indicator of a region of the breast with low detectability. If such a region exists, it follows that a potential cancer would be less likely to be detected by the radiologist if in that region of the breast and that interval cancers are more likely. The fact that masking measures related to overall, local, and texture qualities were significant indicates that masking properties are complex and need to be further studied to identify all relevant factors at play. Analyzing the proportion of screen-detected cancers stratified by BI-RADS density and the IQF 10th percentile measure showed interesting interactions as well. In each case, the proportion of screen-detected cancers was highest in the low masking category and was the lowest in the high masking category. The difference between the lowest and the highest proportion was over 30%.

Little has been reported with regards to measuring mammographic masking. Mainprize et al. have performed a similar study in which they derive a masking measure by creating a regional detectability index of a Gaussian shaped simulated lesion with a 5 mm FWHM based on the signal to noise ratio, which can be derived from the normalized noise power spectrum, and several other imaging parameters of the mammogram (170). They found their masking measure correlated with breast density in several different ways and indicated it may be useful to identify risk of interval cancer. Our study expands upon this work by calculating detectability directly with a 2-AFC test of the model observer and by performing regressions to predict screen-detected and interval cancers. As the field investigating mammographic masking is growing, future developments and insights will be gained to best understand how to model and quantify masking.

This study has several strengths. First, it accounts for BI-RADS density, a known risk factor for interval cancer. This was important because as expected there was a significantly higher proportion of interval cancer cases in the high density categories compared to the screen detected category (13,167). Additionally, matching by age and race between the datasets helps control for confounding in our dataset.

There were several limitations that, if resolved, could improve upon the strength of the study. A more sophisticated model observer could potentially be used, such as a Channelized Hotelling filter (150). However, a NPWE filter may still be sufficient to properly quantify masking and predict risk of interval cancer in mammography. Additionally, this study was performed on a case-case dataset. In the future, comparing masking measures in interval cancers compared to women that don't have breast cancer would help better define the predictive value of the masking measure.

In conclusion, we've developed an automated method that better identifies mammograms with a high likelihood of masking invasive cancers compared to screen-detected cancers than BI-RADS density alone. Further, analysis of proportions of screen-detected cancers showed that these masking measures provide risk segmentation and may have the potential to identify low density groups at higher risk of interval cancer and high-density groups with lower risk of interval cancer. This method may be useful to objectively identify women that would benefit by either non-mammographic screening methods or supplemental breast cancer screening

4.4: IQF In Tomosynthesis - Feasibility

4.4.1 Tomosynthesis Imaging Chain

Tomosynthesis has been approved for use by the Food and Drug Administration (FDA) since 2011, with the goal to help resolve lesions at a higher specificity and sensitivity (179,180). Because of this and the uptake in screening using tomosynthesis, developing measures of interval risk and measures of detectability would be useful in this new technology as well.

Tomosynthesis is similar to mammography in that it helps with breast imaging using X-rays, but the main difference is that tomosynthesis machines take several images at slightly different angles, allowing for reconstruction of a pseudo 3D image to take place using reconstruction methods similar to CT (179,180). Figure 4-6 shows a schematic outlining the imaging process in tomosynthesis.

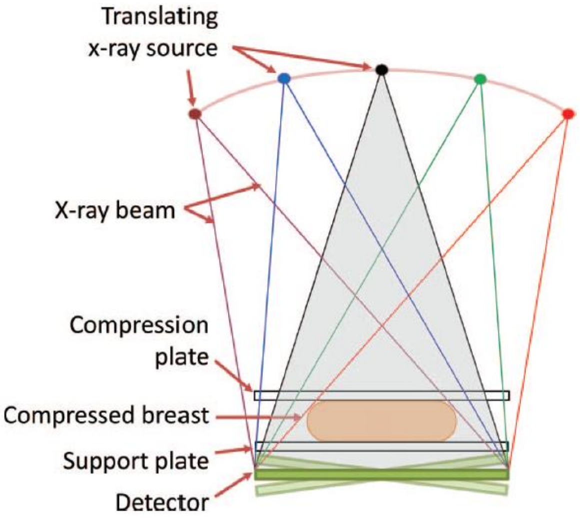


Figure 4-6: Schematic showing the imaging process of taking multiple projections for tomosynthesis imaging. Source: (179)

This technology was developed in the hope that tomosynthesis could help resolve tumors in fibroglandular regions or other hard to see areas of an image. The extra detection effect of tomosynthesis is still being studied (181), but it appears encouraging. While the extra projections add some extra calculations, it is still possible to simulate lesions create and calculate model observers in tomosynthesis (160). We performed a preliminary analysis to examine the feasibility of performing such a process to explore and compare interval risk between tomosynthesis and mammography.

4.4.2 Preliminary Results

To create and insert pseudo-lesions into a tomosynthesis image, a pseudo lesion must first be inserted at the correct location for all projections based on the specific geometry of the projection and the distance of tissue penetrated based on that geometry (160,180). After that, image reconstruction must be performed, which implements filtered back projection (similar to CT). This produces a pseudo 3D image which allows one to see the image in slices throughout the Z axis and can visualize the inserted lesion in the appropriate spot. From there, similar methods to a model observer could be performed in order to quantify detectability.

Figure 4-7 shows preliminary results of inserting a pseudo lesion into a tomosynthesis image. While this was primarily a feasibility and proof of concept study, it shows that it is certainly possible to create and insert pseudo lesions into tomosynthesis images to study the detectability of tomosynthesis images, the interval risk of those images, and compare them to detectability of mammogram images of women.

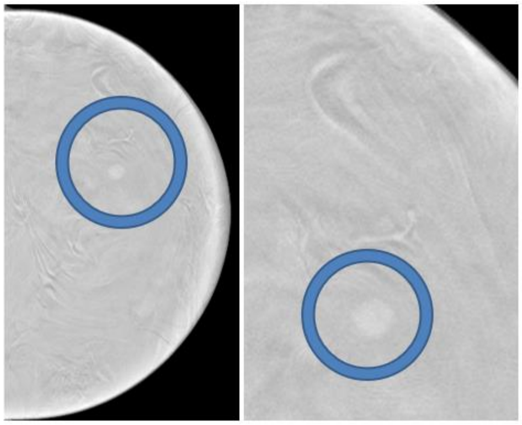


Figure 4-7: Full view (left) and zoomed in (right) tomosynthesis image showing a pseudo lesion created by inserting lesion-like attenuations into each projection and then reconstructing as an image.

In the future we plan to examine and compare the interval risk in mammograms and tomosynthesis images from the same women. This will not only allow to compare the detectability of lesions between tomosynthesis and mammography images, but also can examine the ability to determine interval risk between the two imaging modalities.

5 | Deep Learning

Deep learning is a relatively new branch of computer vision that has revolutionized the field, improving existing applications and enabling a wide new set of other applications. It has improved capabilities in a variety of fields, and introduced an outcome-based, rather than hypothesis-based, form of making decisions (71,182). In order to understand if deep learning could be used to create better measures and predictions of interval cancer risk, we need to better understand how deep learning works and what its advantages and drawbacks are.

5.1: History and development

5.1.1 Theory and Math

The theory behind deep learning has been theorized since the 70s and 80s, well before implementing the idea was actually possible (74,75). These papers theorized that it would be possible to have a series of interconnect variables, each with a series of weights. The variable and weights would predict a specific outcome and depending on the accuracy of prediction a back-propagation algorithm could be applied to change the weights and make the system and algorithm more accurate (44). An example of a fully connected set of layers and weights is shown in Figure 5-1.

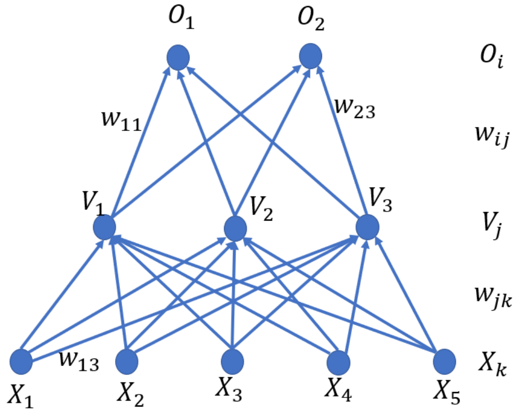


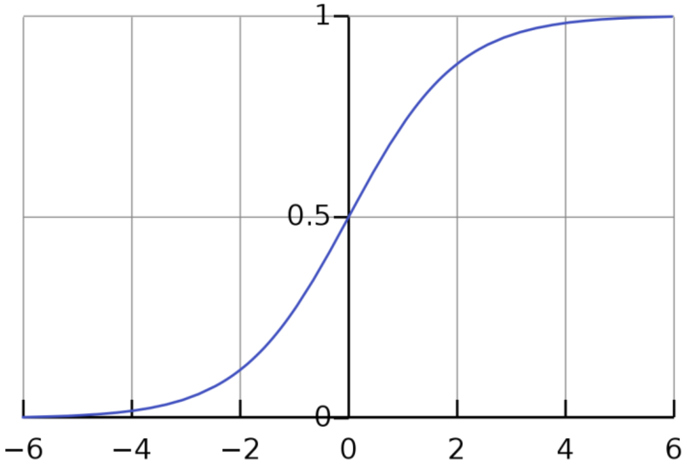
Figure 5-1: Schematic of a fully connected neural network. Each circle represents a node, and each arrow indicates a weight that the node is given from previous nodes or gives to future nodes. The summation of these operations produces values in each output node, leading to a decision. Source: (44)

In this figure, each arrow represents a multiplier and each circle represents a weight. In this system, there are five input variables (X) connected to three input weights, connected to two more output weights. This example network could be taking five blood marker levels to predict a likelihood of any two outcomes, such as diabetes or heart disease. The equation for this network would be set up its predictions through forward propagation, by multiplying the variables with the weights similar to matrix multiplication, by the following equations (44):

$$O_i = \sum w_{ij} V_j \qquad V_j = \sum w_{jk} x_k$$
$$O_i = \sum w_{ij} \sum w_{jk} x_k$$

After a prediction is made on some data, the network will adjust the (w) weights with each iteration in order to bring its predictions closer to the truth. In our diabetes example, if one of the inputs was blood sugar levels, the weights associated with that variable would get more substantial over time, as it is a known risk factor for diabetes. This can expand as far as the imagination can go, with as many variables per layer and as many layers as one wishes (44,182). This allows to include information from a wide variety of data types and makes this computation potentially robust.

Unfortunately setting up a network purely in this way has an inherent flaw: all of the matrix multiplication steps are linear at this point. This reduces the ability of the network to adjust, and drastically reduces the complexity and depth of the network by making it a series of linear combinations (44). To improve on this, researchers developed activation functions, which help add nonlinearities into each layer of these functions and prevent them from being simple linear combinations. A sample nonlinearity is shown in Figure 5-2.



$$g(z) = \frac{1}{1+e^{-z}}$$

Figure 5-2: The nonlinear rectified linear unit (reLU) function that is introduced to deep learning networks to add complexity to the model. Source (44)

This changes our forward propagation from linear combinations to a series of nonlinear steps, as seen in the following equations (44):

$$V_j = g\left(\sum w_{jk}x_k\right) \qquad O_i = g\left(\sum w_{ij}V_j\right)$$
$$O_i = g\left(\sum w_{ij}g\left(\sum w_{jk}x_k\right)\right)$$

Within these fully connected layers and nonlinearities, the forward propagation is now much more robust and complex than just a series of linear combinations. Once the predictions are made from forward propagation, a loss function is created that compares the predicted outcomes with the actual outcomes. There are a variety of optimization functions (182), but the vast majority of them compare the difference between the predictions (and confidence in those predictions) with the ground truth output and weight the loss based on how far from the truth the prediction was (44).

The loss function is then leveraged to determine the way the network should change its weights in order to produce better results in the future. This process involves taking the derivative of the loss function to determine where the largest changes in weights should occur and is called back propagation, and it is shown in Figure 5-3.

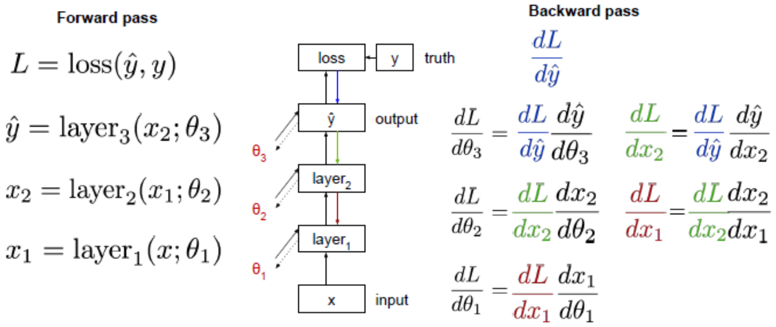


Figure 5-3: Schematic showing forward propagation to produce predictions (left equations) and back propagation to change the weighting factors (right equations). Source (44)

All of these mathematical steps help the network to make and adjust its predictions, not be limited to a linear combination, and adjust its weights to optimize a loss function. A lower output in the loss function indicates more confidence and more accuracy, and a decrease in this loss over time indicates the model is learning. Unfortunately, this also makes it difficult to compute and adjust these weights. In the 1980s, computing and updating the weights of these networks would take weeks and was just impractical considering the available computing power at the time (44). This made it impractical to apply these ideas for research in a meaningful way.

5.1.2 The Computing Revolution

It took decades between when deep learning was first theorized in the 70s and 80s to when it revolutionized computer vision in the 2010s. Improvements in the amount of data and its structure as well as improvements in computing ability helped to make implementation of deep learning more possible and practical (44,182,183). Figure 5-4 shows changes in both the amount of data and the computing power available over time.

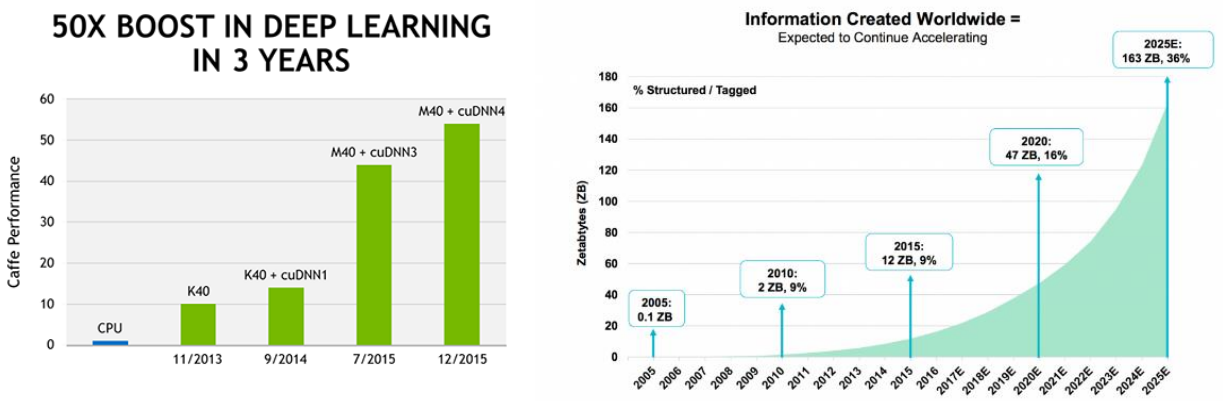


Figure 5-4: Graphs showing the growth in the amount of data existing and generated over time (left) and the increase in computing power over time (right). Both of these factors contributed to the explosive growth seen in deep learning over time. Source (183–185)

In order to actually apply deep learning methods, a large amount of computing ability is required. The recent increase in computing ability was hugely influential in the application of neural networks and is rapidly increasing from year to year. In addition to improving computational power over the years, researchers discovered that using graphics processing units (GPUs) as opposed to central processing units (CPUs) could improve the computational speed of deep learning algorithms by a factor of 10 or more (44,183). These advancements helped to drastically improve the ability of deep learning to be actually applied.

Another factor that served as a catalyst for deep learning applications was the explosion in both the amount and availability of data over time. Because of the large number of parameters inherent in neural networks, they will naturally overfit data unless they can train on a large dataset with a variety of data. In recent years the amount of data being produced and stored has doubled every two years, providing the large amount of data required. In addition, the advancement of the internet made it much easier to share and have access to that large amount of data, something that was incredibly difficult before. Further, the ability to share data through the internet made it much easier to collaborate with other groups, providing a larger variety of data into training.

This has led to very large public databases, like the ImageNet database. ImageNet is one of the premier deep learning competitions that occurs annually. In this competition, the goal is to use computer vision methods to correctly classify a set of over 1 million images in over 1,000 categories (44,75,186). Some sample ImageNet images are shown in Figure 5-5.

Figure 5-5: Sample images from a variety of the categories used in the ImageNet competition, showing the vast complexity and wide differences in types of images. Source: (44)

In the earliest years of this competition, traditional computer vision methods were used to categorize different classes. In 2010, the method had an accuracy of roughly 72% with these traditional computer vision methods, as seen in Figure 5-6. Once deep learning methods were introduced in 2012, it revolutionized the ImageNet competition and many other computer visions tasks. In 2012, the deep learning entry to the competition was over 10% more accurate than the best non-deep learning method, having an accuracy of 85% compared to 73%. Just two years later, every top competitor in the competition has used deep learning methods. These methods also took hold in a variety of fields including medical imaging, leading to a large spike in papers submitted over the years (77,78).

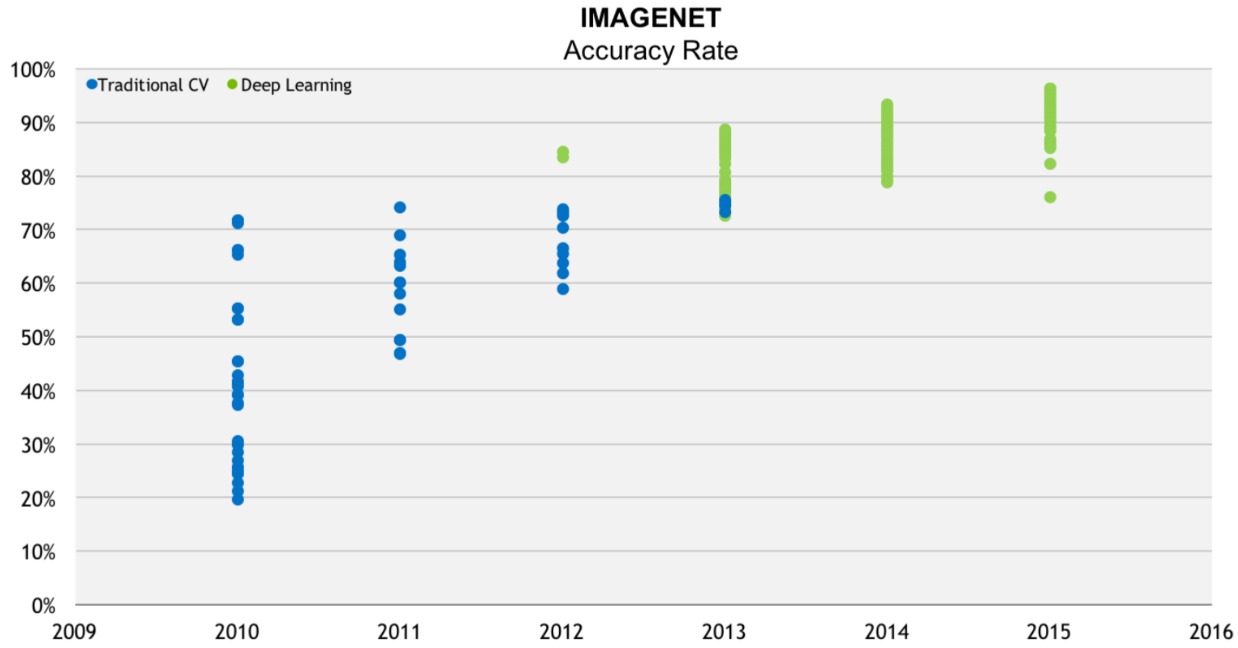


Figure 5-6: Methods and results over the past years of the leading ImageNet competition leaders. Once deep learning was introduced to this competition in 2012, it has outclassed every traditional computer vision method. Source (72).

5.2: Deep Learning Basics

5.2.1 Convolutional Neural Networks

In addition to standard deep learning methods, deep learning can be applied to make predictions based off of image data. For the most part these are created and applied through the use of convolutional neural networks (CNNs), an example of which is shown in Figure 5-7. In CNNs, instead of a set of variables and layers connecting to each other and changing weights, a series of filters of various sizes are connected to each other and change weights (44,187). These weights then are trained to look for specific features that are important for the discrimination task at hand.

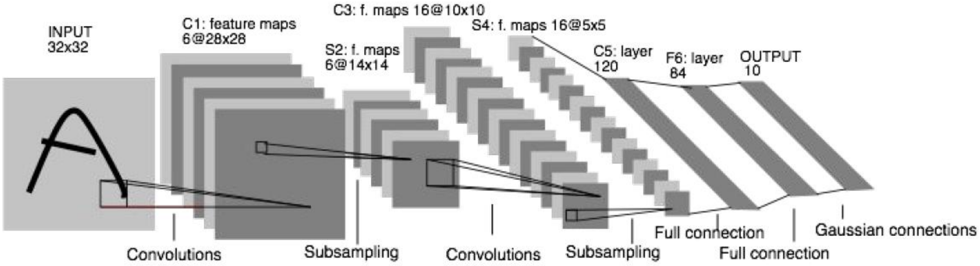


Figure 5-7: Schematic showing the layout of a convolutional neural network (CNN) and how convolutional layers and fully connected layers are stacked to produce full deep learning networks. In these networks instead of layers of nodes connected to each other through weights, we have layers of features that are convolved through the preceding image or feature map, with weights attached to those layers. These methods are used extensively in image processing competitions. Source: (44)

Once a CNN is trained, in general the first level features look for more general features like edges, curves, and intensity, and later features look for things specific to the classification task like eyes, wheels, mouths, and other high-level features. For example, in Figure 5-8, we can see the first filters look for edges and intensity measures and the later features are trained to look for hubcaps, windows, and other features specific to cars (44). In some cases, these filters look like some hand designed edge and feature filters created by humans, and they help to look for the same basic elements (edges, spheres) that those filters would in conventional computer vision.

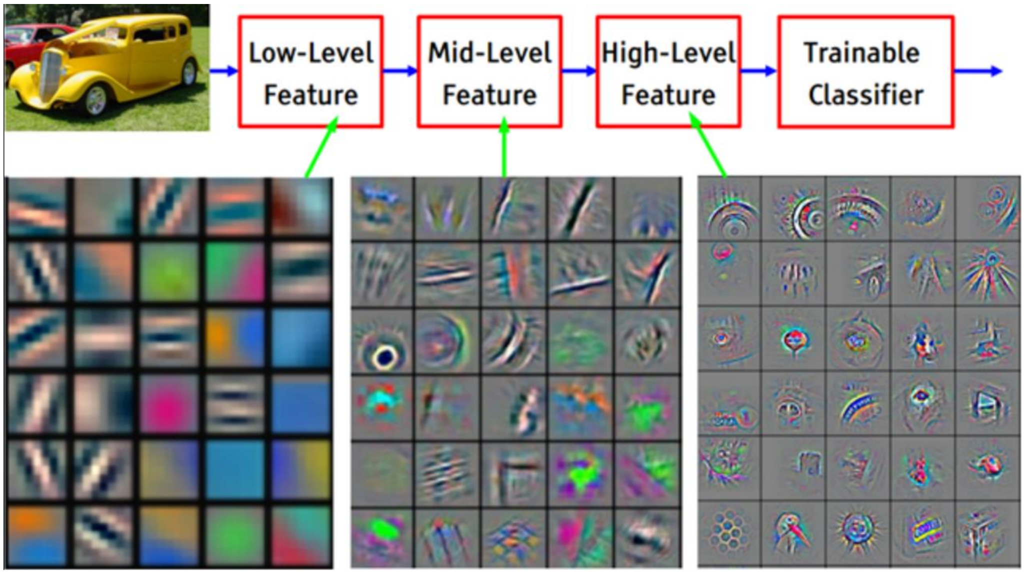


Figure 5-8: Schematic showing how low-level features in CNNs tend to identify base-level features like edges and intensity. Middle and higher-level features tend to show and look for more advanced features like curves and items specific to the task like wheels and windshields. Source: (44)

5.2.2 Outputs

Over time, deep learning networks have been adapted and trained to work on a variety of datasets and produce a variety of outcomes. Because of this, CNNs and deep learning have impacted nearly every facet of computer vision. They have been used to classify images into categories, to segment different regions of an image, to read and identify words, to identify the sentiment or style of a word or image, and to help fill in missing data in an image or even create new simulated images (71,72,76,77,187,188).

5.2.3 Drawbacks / Limitations

There is a key possible limitation and drawback in deep learning, and that is the great risk of deep learning networks to overfit and memorize data (78). Figure 5-9 shows a classic case of overfitting linear data to a polynomial equation.

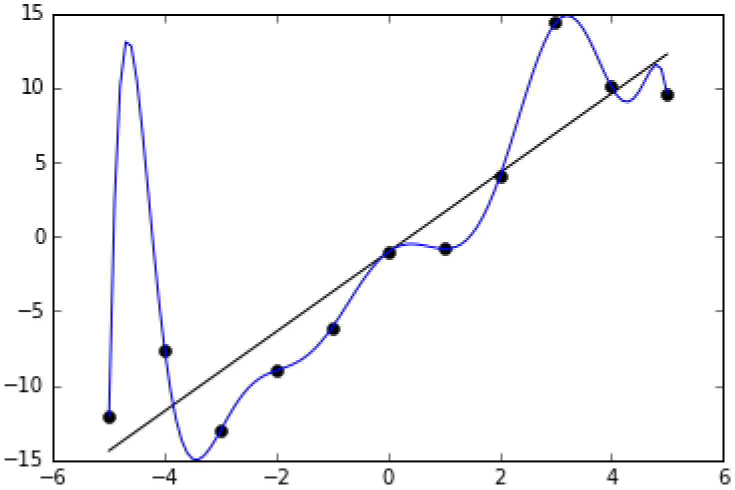


Figure 5-9: Classic example of overfitting a general linear dataset. In this case, linear data is being fit to a high-order polynomial. While it fits these specific data points well, extrapolating this polynomial to any point outside of those specific points produces wildly inaccurate results. Overfitting in deep learning applications

produces the same results, with great results on training data and poor accuracy on any outside data. Source: (189)

In deep learning, many networks have millions of weights and millions of interactions between those weights. This gives the networks a great risk to identify patterns in data for classification, but also a great ability for networks to memorize data and overfit (44).

Overfitting essentially happens in deep learning networks for two reasons. First, if the network is too complex for the type of analysis being done, for example if you choose an extremely complex network to classify digits, the network has more variables to potentially overfit and memorize the training data. Secondly, if you don't have enough training data, the neural network tends to memorize specific images in the dataset, meaning no learning is happening but the images are just being memorized. Lastly, if there is not enough variety in the dataset or the dataset is missing a key aspect of variation, the neural network will likely perform poorly on new data. For example, training a neural network to identify only printed words will not work effectively if the network tries to identify cursive words.

5.3: Training Heuristics

There are a number of methods to help mitigate overfitting from occurring in deep learning training or to help make your network more robust.

5.3.1 Training and test sets

It is important to have a metric that allows you to determine whether overfitting is occurring or not. To do this, a standard has evolved to split the data you have access to into a training dataset and test dataset (44). Often 80 percent of the data is placed into the training set and 20% of the data is placed into the test set.

The neural network then adjusts its weights and learns using the training dataset. It then performs predictions on the test dataset, and the results of the network on the training dataset and test dataset are compared. The benefit of this process is that during training, the network does not have access to the test dataset, so if the network starts to memorize the training dataset the results will get worse and worse in the test dataset. This lets one observe the training and test loss, and observing a decrease in test and train loss values indicates better learning over time. A sample curve showing results from the training and test datasets is shown in Figure 5-10

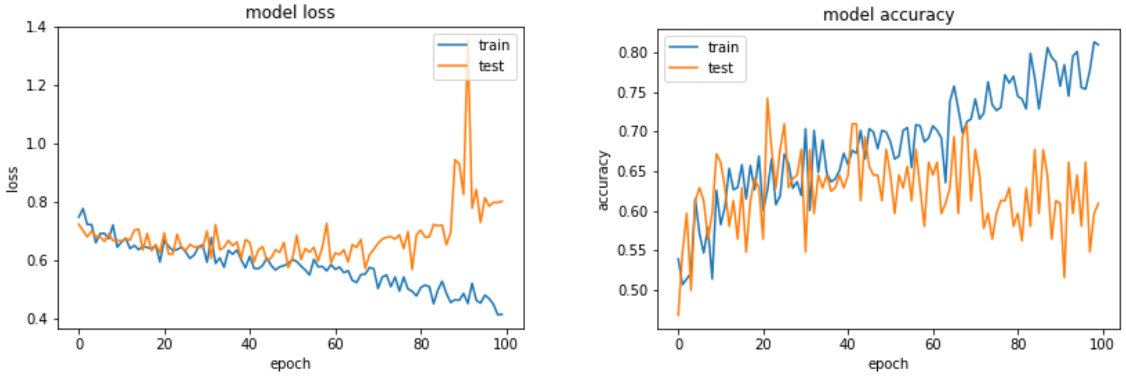


Figure 5-10: Typical loss (left) and accuracy (right) curves that are produced during deep learning training for both the test and train dataset. Ideally, loss curves decrease over time with test loss staying similar to training loss, and accuracy ideally rises over time. If test loss and accuracy deviate from the training results, like we see halfway through training, it indicates overfitting.

In this, the training and test accuracy and loss start out at similar levels. At this point in training, the network is doing true learning and not overfitting, as it is producing similar results in both data that it sees during training and the new data in the test set. After roughly 50 Epochs, however, the test loss starts to stabilize and then increase, and meanwhile the training accuracy continues to increase. This gap between the two curves is called the generalization gap, and it is an indicator of overfitting as the gap widens. At the point where the training accuracy is 85% and the test accuracy is 50%, a large amount of overfitting has occurred and the network has essentially just memorized each case in the training set. Keeping an eye on this generalization gap is key to identifying if overfitting is occurring, and there are several methods that exist to control this generalization gap and help prevent overfitting

5.3.2 Selecting the right network

There are a wide variety of different deep learning architectures, each with their own pros and cons (44,72,75,182). It has been shown that different types of data perform much differently using different architectures, so testing the different architectures on your dataset is very important (44). This can help to provide more consistent results, prevent overfitting, and generate more accurate predictions.

5.3.3 Learning Hyperparameters

Learning hyperparameters are user-defined parameters that tell a network how to treat data, how it should learn, and how aggressively it should learn (44). A sample of different learning hyperparameters are shown in Table 5-1.

Table 5-1: List of typical hyperparameters and their interpretation in deep learning networks.

Hyperparameter (Range)	***Interpretation***
Rotation	Range for a random rotation
Zoom	Range for a random zoom
Shear	Range for a random shear
Vertical/Horizontal Flip	Random chance of flip in respective direction
Momentum	Importance given to previous weight updates compared to current weight updates
Regularization	Penalty applied to large image weights
Decay	Learning rate decay over each update
Dropout	Percent of weights ignored between dense layers in the fully connected layer
Learning Rate	How big of a change is made with each weight update
Epochs (In	Number of times to run through and predict the whole dataset
Batch Size	Number of samples per weight update

Image Size	Input image size in pixels
nLayersRetrain	Number of layers allowed to have their weights altered.
Channel Shifts	Switches the RGB channels of the input image
Noise Addition	Addition of noise, removal of image regions.
Rescaling	Rescales image values
	Option to normalize or center the image statistics

Adjusting these hyperparameters will affect how aggressively the network learns and help change the generalization gap to prevent (or produce) overfitting. Many studies have been done to show baseline levels and typical ranges for many of these hyperparameters.

5.3.4 Hyperparameter Optimization

Although using standard levels of learning rate, momentum, and other hyperparameters can be useful, it is difficult to identify what to do if those sets of hyperparameters don't produce good results. While adjusting individual parameters can sometimes help, it is often useful, more practical, and more reproducible to standardize this process of identifying the best hyperparameter values to use.

This process is called hyperparameter optimization (44). There are several ways to perform hyperparameter optimization, but one of the most common and easy ways is to perform a grid search or hyperparameter sweep. This is where you take the full set of realizable hyperparameters for each image and train the network with each possible combination of hyperparameters. This allows you to see and identify different hyperparameters sets that would be useful and to better understand the interaction between these hyperparameters. This is important to do because while it is often the case that standard hyperparameters values work well during training, different types of data and different sizes of data often get better results in different hyperparameter spaces (44,76,182) .

5.3.5 Image Augmentation

Sometimes regardless of the hyperparameters chosen and the network chosen, it is difficult to get reasonable results that don't overfit because the dataset is just too small. This happens often in cases where the data available is limited, like in medical imaging and diagnostics (77). In these cases, it is often helpful to perform another process called image augmentation.

Image augmentation can help introduce variety to small datasets. The process involves performing a series of random image alterations before feeding the image into the deep learning network. This helps ensure that each time the network sees the same image, it is seeing the image at a slightly different angle or orientation, which helps prevent the network from memorizing the specific pixels or a specific image has a specific outcome (77,182,190). Figure 5-11 shows several options for image augmentation steps, which can include introducing random flips, zoom, shear, and rotation to the image before feeding the image into the network.

Image augmentation has been shown to make networks more robust and more resistant to overfitting (44). It has also been shown to produce features that re more significant to the underlying detection process, rather than features that help identify specific images.

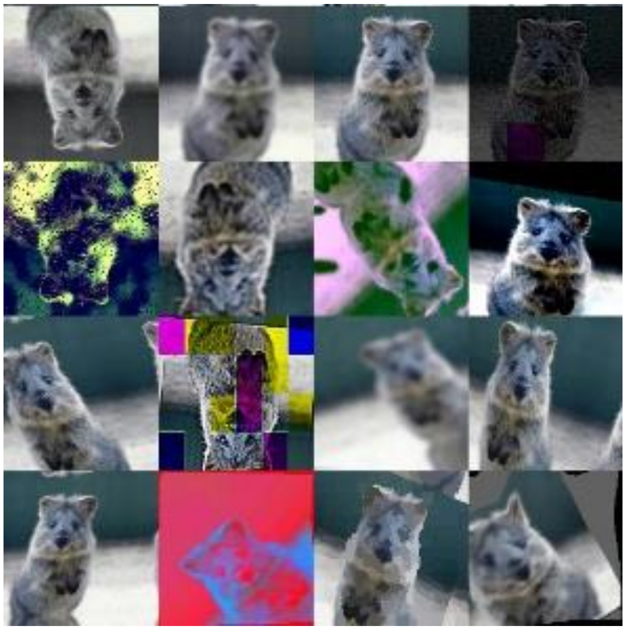


Figure 5-11: Example of the various image augmentation steps that can occurs before feeding into the network, allowing one to add variety to a limited dataset and reduce overfitting. Source: (190)

5.3.6 Transfer Learning

Even with all of these heuristics, in some settings the data is not available in the volume or variety required in order to prevent overfitting and despite best efforts, the network will lean towards memorizing the training set. In these cases, it is often useful to perform a process called transfer learning. In transfer learning, before training happens, you borrow network weights from a well-established training dataset that has already been performed (44,191). This helps to establish features from the outset that look for important general features in image classification such as edges, curves, common shapes, and common textures.

Then, you re-train those weights to adjust the weights and features to identify which features are most important for this specific problem, and adjusts the features to look for shapes, edges, and textures that are most important in this specific problem. A schematic showing the transfer learning process is shown in Figure 5-12

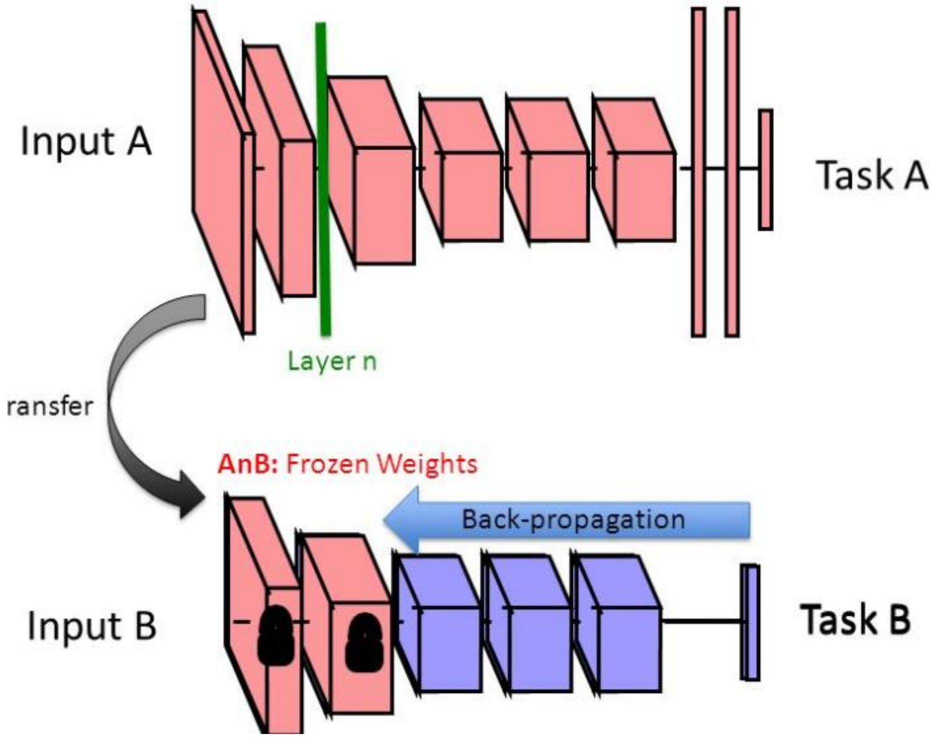


Figure 5-12: Schematic of how transfer learning works. A network is first trained on a large and robust dataset (such as ImageNet) for a specific task (task A). Those weights are then borrowed, a new end of the network is applied to classify a new task, and weights are retrained to best perform for that new task. Source: (192)

Often, this involves borrowing network weights from the ImageNet competition, as it is one of the most well-established competition with a huge amount and a huge variety of data (182). This ensures that in most cases it will have identified important features for all parts of image detection. During this process, you can unlock any level of layers to be retrained based on preference and results, allowing you do just refine top level features or refine all the weights and features as well.

This process of transfer learning has been shown to improve and help in many cases where data is limited (44,76,188,191,193). It helps to leverage the large datasets and things learned from that, and applies it to a specific problem

5.3.7 Tiered vs Categorical vs Binary.

While the learning of a deep learning network is done independent of the user, it is dependent on user-defined outcomes, whether it is classification, segmentation, or some other task. Because of this. When choosing between classes of objects, Binary classification is typically done when a decision needs to be made between two classes only, whereas categorical classification is used in multiclass problems (182).

In some cases, with a limited number of classes, some people hypothesized that a tiered classification system may outperform trying to classify all categories at once. For example, classifying images of animals first into cats versus dogs, and then having a network to differentiate dogs and a network to differentiate cats may outperform classifying all at once. However, it has been shown that most networks with enough depth are able to classify categorically quite efficiently even with a small number of categories (44).

5.4: Improvements to Medicine

Deep learning has made a plethora of improvements over the years. It has helped a variety of military applications, from automated flying of drones to machine learning. More relevant to us though, it has revolutionized the field of medicine (77,78).

Deep learning has consistently been applied to and improved on current diagnostic methods in many medical fields. CheXNet has been able to improve upon lung pathology diagnoses using chest X-ray images compared to an average of radiologist diagnoses, improving the F1 score from 0.387 to 0.435 (63). Deep learning was applied to diagnose diabetic retinopathy and was able to improve classification AUC from 0.937 using current detection techniques and human observers to 0.980 using deep learning methods. A negative of this network was a specificity level of 0.87, which would lead to a larger number of false positives (194). Several review papers have highlighted the various advancements, opportunities, and risks of deep learning in other medical fields (76,77,195).

Deep learning methods have already been applied to several areas of breast cancer research with promising results. An analysis of screening mammograms to perform lesion segmentation and classification with a deep learning network was able to achieve classification AUC of 0.95 and outperformed CAD systems (196). Deep learning networks were also able to categorize amount of dense tissue on par with experts using Cumulus-like thresholds (197). Lastly, a deep learning network lesion classification system achieved an AUC of 0.82 on digitized-screen film mammograms and 0.90 on breast tomosynthesis images (188). Our study, which has been able to differentiate between future interval and screen detected cancers more effectively than using current methods such as BI-RADS breast density, further contributes to the understanding that using deep learning can further our diagnostic capabilities and our understanding in several areas of breast cancer research.

6 | Deep Learning to Quantify Interval Risk

6.1: Preliminary Work: Optimizing deep learning for interval risk.

In the studies outlined in future sections, we attempted to implement deep learning methods to identify interval risk on a variety of datasets. In order to make this successful, we implemented several heuristics and methods to get the best possible results, as detailed below.

6.1.1 Transfer Learning

Medical imaging often has a limited number of images in each classification category, and that holds true in the studies that will be presented in Sections 6.2: and 6.3: as well. In initial work implementing deep learning without performing transfer learning, we observed a huge amount of overfitting. It was clear that the dataset we had was not large enough to create large and robust weights from scratch. To combat the effects that a limited dataset has on deep learning, we implemented transfer learning as seen in Section 5.3.6 : Transfer Learning.

6.1.2 Computational Limitations

Deep learning is always limited by the computational ability of the computer it is performed on. This work was performed on an NVIDIA (Santa Clara, CA, USA) Quadro K 2200 with 16 GB of RAM. While this is capable of running and training deep learning networks consistently, there were several compromises we undertook in order to allow for the technical specifications of the workstation.

The K2200 GPU limits the training speed of the deep learning networks. Because of this, image compression is forced in order to allow for reasonable training times so we can iterate on a variety of training techniques. Further, the 16GB of RAM in this system force limitations in both the image size (in pixels) and the batch size (number of images fed into the network at once). Because of this, it takes longer to train a network, and we need to compress the image size from their natural image size of 2000x3000 pixels per mammogram. We tested a number of image compression sizes to see the training speeds and maximum batch sizes for a variety of cases. As seen in Table 6-1, training time for a reasonable number of epochs becomes quite long, and compression down to roughly 300 pixels was necessary in many cases.

Table 6-1: Sample of maximum batch sizes possible and training speeds of various image input sizes on our workstation. Because of this in order to iterate on network on a practical time scale, we were limited to optimizing networks by compressing the images down to 331 pixels or below.

Image Size (Pixels)	Max Batch Size	Training Time per epoch (case-case analysis: Section 6.2:)	Train time 500 Epochs
224	16	18 sec	2.5 hrs
275	16	25 sec	3.5 hrs
331	8	35 sec	4.9 hrs
400	8	52 sec	7.2 hrs
450	4	65 sec	9.0 hrs
500	4	78 sec	10.8 hrs
600	2	123 sec	17.1 hrs

6.1.3 Image Input

A large proportion of the area of a mammogram is empty space without breast tissue. As a result, this area would not be useful in deep learning training. Because image compression is required from the computational limitations of our system, it was important to maximize the amount of useful information presented to the deep learning network. To do this, we implement a number of image preprocessing techniques that resulted in each mammogram being cropped into the rectangle that inscribes the mammogram area.

In this, we identify the skin edge through thresholding. We then crop the excess material out, and normalize the image on a 0 to 255 intensity scale as that is the expected intensity scale in these deep learning networks.

6.1.4 Image Collaging

An important consideration when designing a deep learning network is the method and way to input images into the deep learning network. Because mammograms are always taken in sets of four, we leveraged an idea called image collaging that combines all four mammograms of a single patient into a single image that is inputted into the network, as seen in Figure 6-1:

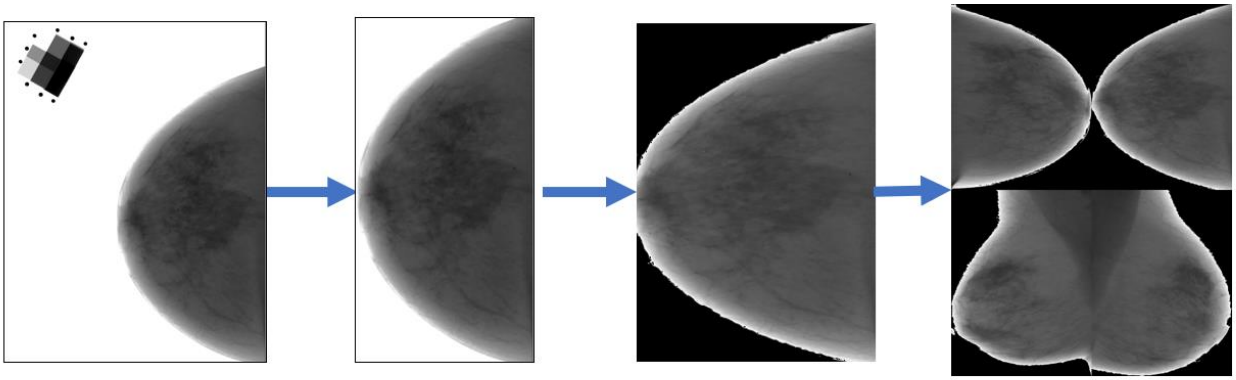


Figure 6-1: Schematic showing the process of collaging the preprocessed images together into a single image. At left is the raw 16-bit unprocessed mammogram. Middle left is the image after image cropping. Left and Middle left images were brightness and contrast enhanced in this figure for improved visibility. Middle right removes background values and convert the values to an 8-bit scale, and the rightmost image combines each of the four mammograms into a single image. This helped to maximize the amount of useful data introduced to the network at once.

This method has both benefits and drawbacks. A benefit of image collaging is that the network sees all imaging information about each case simultaneously, allowing it to make additional and deeper connections than if it sees each image individually. Further, this helps reduce image redundancy, as inputting multiple images from the same case as individual inputs implies image uniqueness when the images are in fact related. Unfortunately, this method also drastically reduces the amount of total data the deep learning network has access to, as it combines multiple images into a single image. Lastly, the type of image collaging performed here, where each image is forced into a different quadrant, does not preserve the aspect ratio of the image, meaning certain factors such as breast shape and total area may not be preserved by this pre-processing method.

In the end, several studies showed promising results from image collaging rather than using separate image inputs. We adopted this collaging methodology as we were able to get more accurate and more robust results from this method on our dataset, as seen in Figure 6-2.

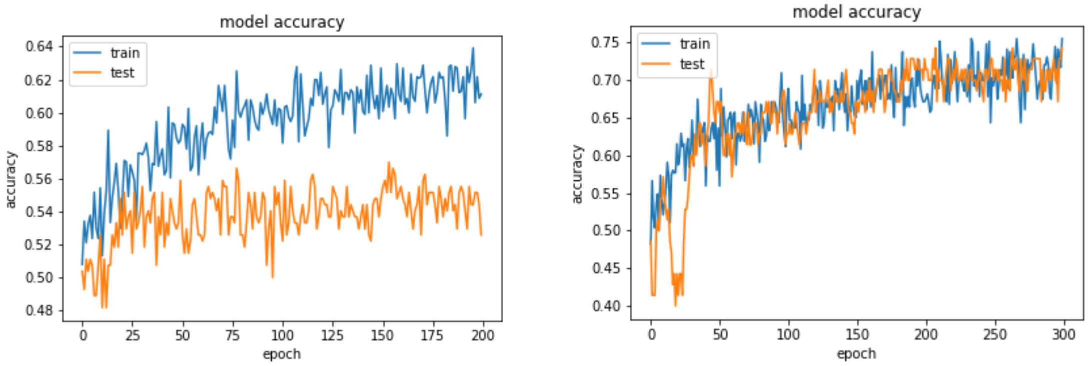
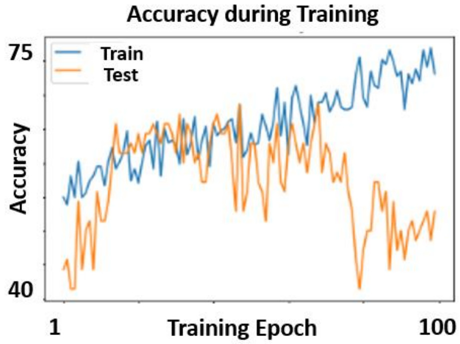
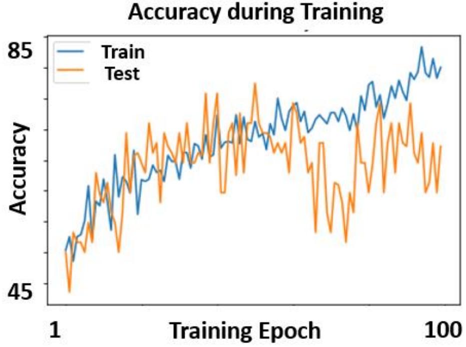
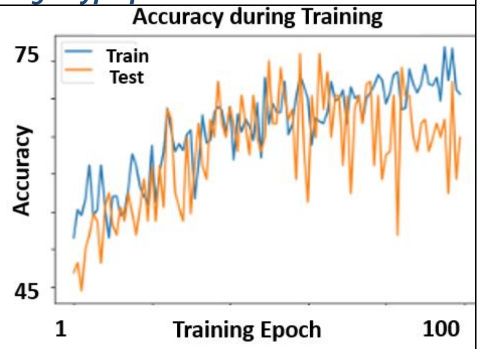
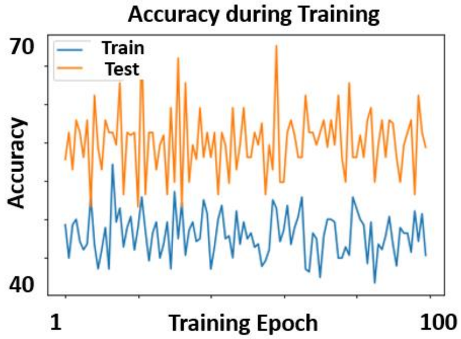
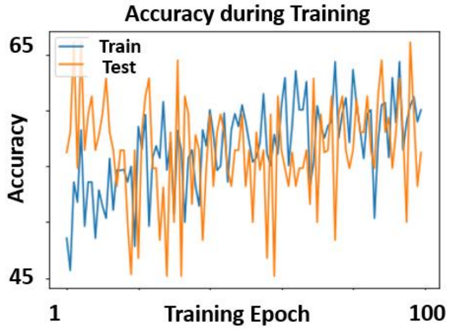
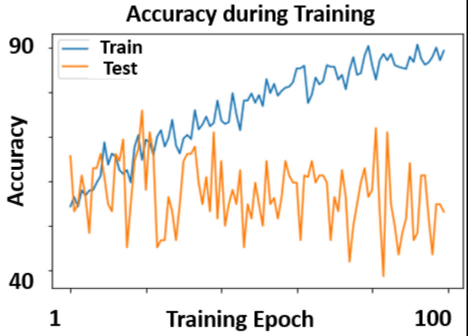
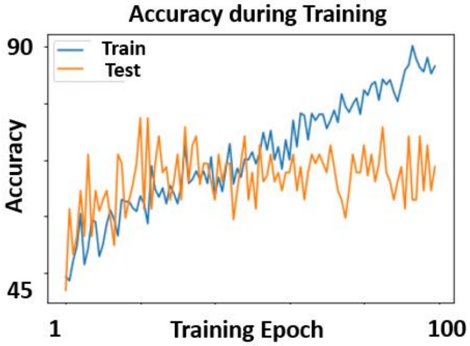
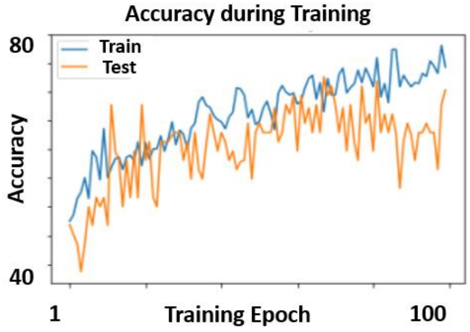
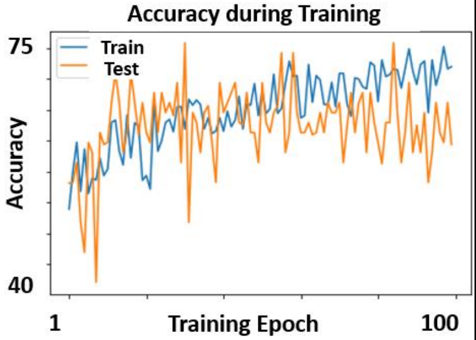


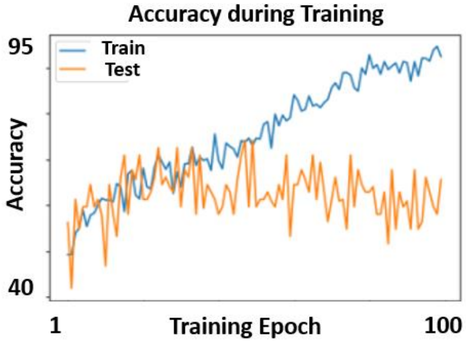
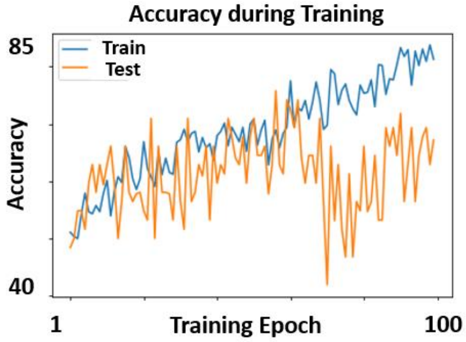
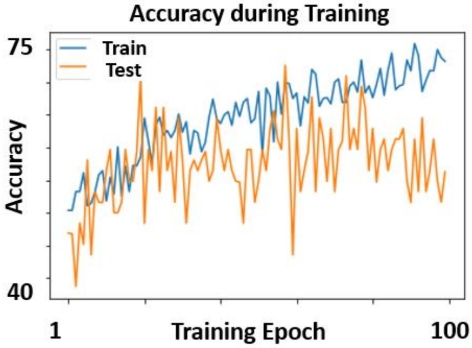
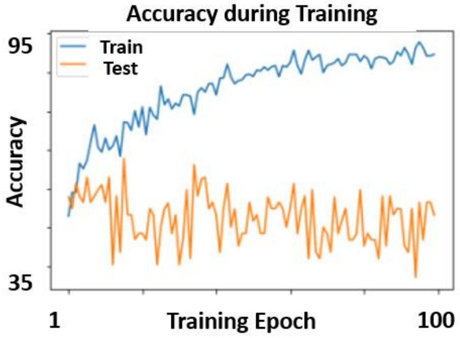
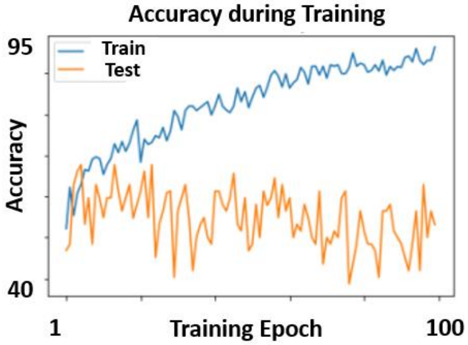
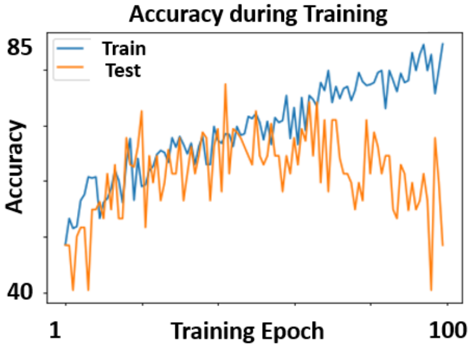
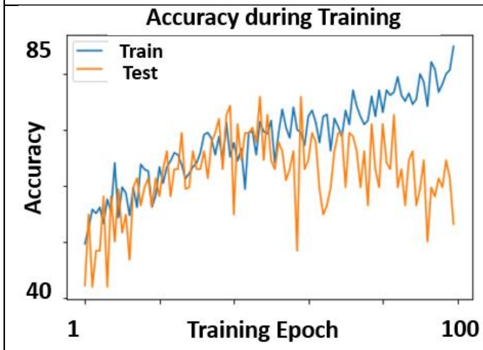
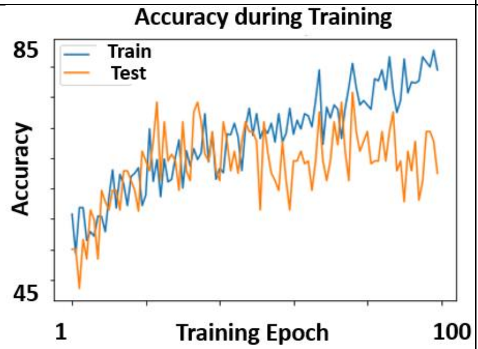
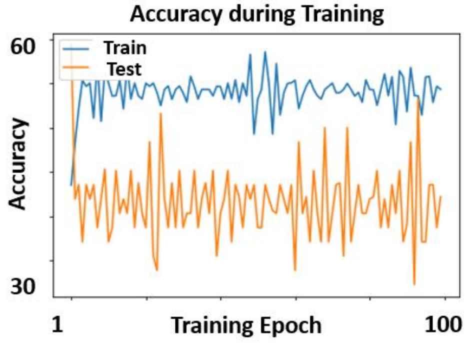
Figure 6-2: Comparison of test and training accuracy by inserting each mammogram individually (left) versus inserting the 4 mammograms as a collage (right). In this case each network was trained under the same set of hyperparameters, and shows how combining the images as a collage helped to reduce overfitting and improve accuracy.

6.1.5 Hyperparameter Sweeps

In deep learning networks, there are a variety of hyperparameters that are available to adjust in order to optimize and improve training, learning, and prevent overfitting. Different datasets require different hyperparameter sets for proper training, so in order to test which range of hyperparameters would produce good results with minimal overfitting, we performed a hyperparameter search.

In some cases, with a large amount of time and resources, a full grid search in implemented that tests every permutation of hyperparameters. Since our setup was limited by computational ability, we tested each hyperparameter across its full range independently, rather than doing a full grid search. Doing this, we were able to identify several trends and hyperparameter sets that would produce positive results, as seen in Figure 6-3.

Low hyperparameter	***Medium hyperparameter***	***High hyperparameter***
 Batch Size = 2	 Batch Size = 8	 Batch Size = 16
 Learning Rate = 1e-8	 Learning Rate = 1e-4	 Learning Rate = 1e-2
 Zoom = 0.1	 Zoom = 0.4	 Zoom = 0.5

Low hyperparameter	***Medium hyperparameter***	***High hyperparameter***
 Rotation = 10	 Rotation = 30	 Rotation = 60
 Dropout = 0.1	 Dropout = 0.5	 Dropout = 0.9
 Regularization = 1e-19	 Regularization = 1e-9	 Regularization = 1e-4

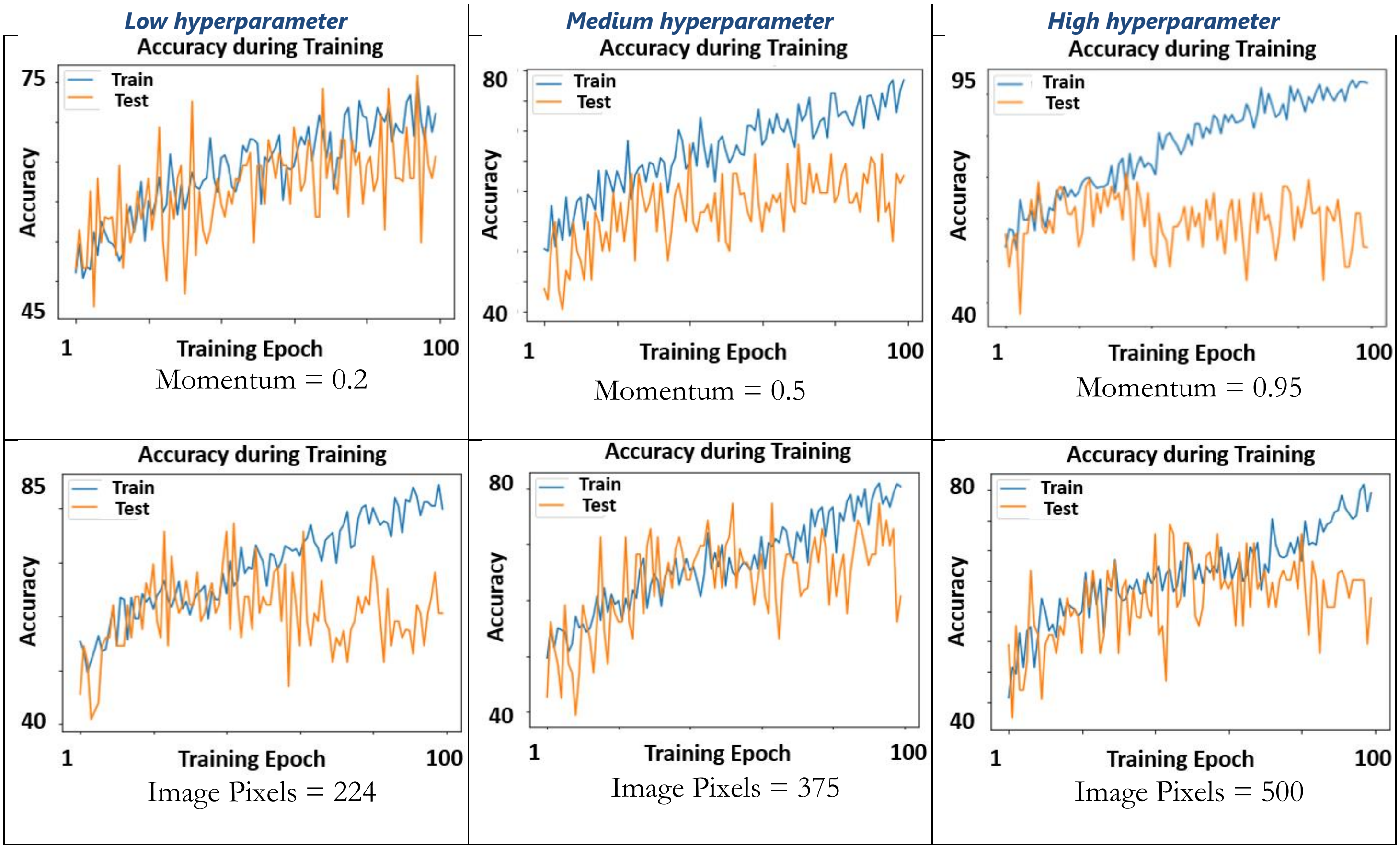


Figure 6-3: Schematic showing several graphs of the change in the accuracy of the training versus various hyperparameter changes. From this we see high overfitting with low amounts of rotation, zoom, batch size, and dropout, and increasing those hyperparameters helps to reduce overfitting. We see too high of a learning rate leads to overfitting and too low of a learning rate leads to no learning occurring.

6.1.6 Class Weights

When deep learning networks are trained, it assumes that each potential outcome has an equal chance of happening. During training, a deep learning network will change its weights in whatever way it can to get the best accuracy and minimize its loss function. If one image category has many more images or cases than the other categories, the network will quickly realize the best way to maximize accuracy is to always guess the most common category. While this maximizes accuracy, it does not actually produce true learning and the network will not perform well with new data.

In medical imaging, and in this case of identifying interval cancer risk, it is quite common to have one category (healthy cases) be much more common than another category (cancer cases). To avoid having the more common category dominate the decision making of the system, we implemented class weights in Section 6.3: . These class weights add an extra penalty to guessing the over-represented category and prevents deep learning networks from only guessing the most common category. Below you can see a case of training before and after class weights were implemented in a case where we had a healthy image category that outnumbered the other categories.

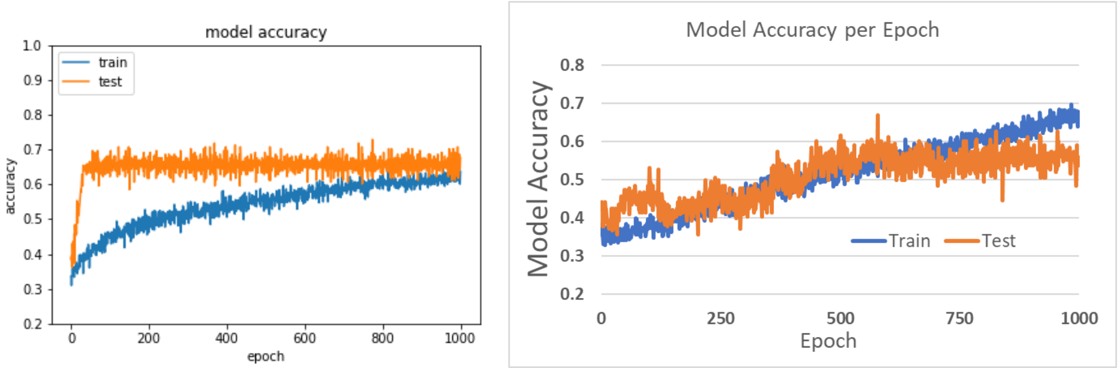


Figure 6-4: Training before (left) and after (right) the implementation of class weights into the loss metric. In this case the most common category contains 66% of the total images. One can see that before class weights were introduced, the network immediately skyrockets to 66% test accuracy, showing that the network was simply guessing the most common category.

6.2: Study: Deep learning networks find unique mammographic differences in previous negative mammograms between interval and screen-detected cancers (198)

In this study we aimed to use the advances in deep learning and apply them to our goal of identifying groups of women at higher risk of interval cancer.

6.2.1 Introduction

Mammography is the current gold standard in screening for breast cancer in average-risk women. However, radiologically dense and complex tissue can reduce screening detection sensitivity leading to cancers missed by screening mammography from the obscuring of the lesions (13,165).

These cancers discovered after normal screening mammograms are called interval cancers and the effect reducing sensitivity in mammograms is commonly called mammographic masking. Roughly 13% of breast cancers diagnosed in the U.S. are interval cancers (12).

Previous studies have shown BI-RADS density and other density measures are risk factors for interval cancer (13,16). While clinically measured BI-RADS breast density can serve as a rough proxy for masking in measuring interval cancer risk, the scores are subjective and do not account for texture of dense tissue (168–170). Because of this the American College of Radiology has asked for development of direct measures of masking and interval risk (10).

Previous studies have measured the ability of pre-defined kernels and model observers to quantify masking and interval cancer risk, indicating some promise in computer vision to identify interval risk (170). Advanced computer vision methods such as deep learning have shown promise in many computer vision tasks such as object identification and have performed extremely well in the ImageNet competition compared to traditional pre-defined kernel methods (73,76). Transfer learning (191,195) of these networks has been effective in medical applications, including breast cancer, where in many cases deep learning models were able to equal or improve current classification or diagnostic schemes performed (63,76,188,193,196,199,200). Another useful property of deep learning models is their ability to highlight pixels containing unique information relevant to that image's classification called saliency maps, which can be used in biological applications to develop hypothesis on the underlying biology or features associated with the classification of interest (63,187).

The purpose of this study was to implement a convolutional neural network to classify pre-cancer mammograms in a population of women who later experienced either screen-detected or interval cancers and compare this classification to a similar classification using BI-RADS density in order to see if neural networks can more effectively quantify risk of interval cancer. If successful, these methods could be expanded upon to improve risk models and identify high risk groups of women, develop automated methods or software that can aid radiologists, and further understand radiomic quantities and underlying biology that indicate high risk of interval cancer

6.2.2 Methods

Participants

Participants were selected from a screening population that had received full-field mammograms acquired from 2006 to 2015 from three medical centers University of California – San Francisco, California Pacific Medical Center, and Marin General Hospital. Ethics approval was obtained by the University of California – San Francisco Institutional Review Board for this retrospective analysis of mammograms for interval risk properties. Interval cancers were defined as invasive cancers identified within 12 months of a negative screening examination, from these centers were included. An equal number of screen-detected cancers were matched by age and race if such matching data existed, based on all screen-detected cancers diagnosed at the three centers. Screen-detected cancers were defined as invasive cancers identified within 12 months of a positive screening examination. All mammograms were interpreted prospectively by radiologists during the course of routine clinical care. Cancers were identified by annual linkage from the associated registry.

Mammography

The raw, “For Processing” representation of the standard four screening views (Mediolateral-oblique (MLO) and Cranio-Caudal (CC) images of both sides) were used for this study. All images were acquired on Hologic Selenia full-field digital mammography systems. These images were pre-processed in order to maximize the information provided to the network in the following ways. First, the skin edge of these images was identified via thresholding and excess background of the images was cropped out. The images were then normalized on a 0 to 255 scale and the four views were stacked as a 2x2 collage image for each case, with one view in each quadrant. This allowed all four views to be contained in a single image. These images were then separated randomly into a training and test set at an 80/20 split.

Deep Learning Models

An existing deep learning network architecture (ResNet50) was implemented with ImageNet transfer learning weights on all convolutional blocks (73). A fully connected layer was then added with 256 weights, a dropout layer, and a final weight with sigmoid activation to classify between screen-detected and interval cases. Figure 6-5 shows a diagram of the deep learning architecture and the fully-connected layer used in training (73). The weights of the fully-connected layer were randomly initialized and pre-trained on the training set images. During training, a binary cross entropy loss metric was optimized with a stochastic gradient descent optimizer. During model training, data augmentation was performed by introducing a random amount of shear, zoom, rotation, and horizontal and vertical reflection within specific ranges in order to increase the variety of data and reduce overfitting. For each epoch training and validation loss and accuracy was recorded and network weights were saved if it improved the validation loss. The final network weights used were the weights with the best validation loss throughout training.

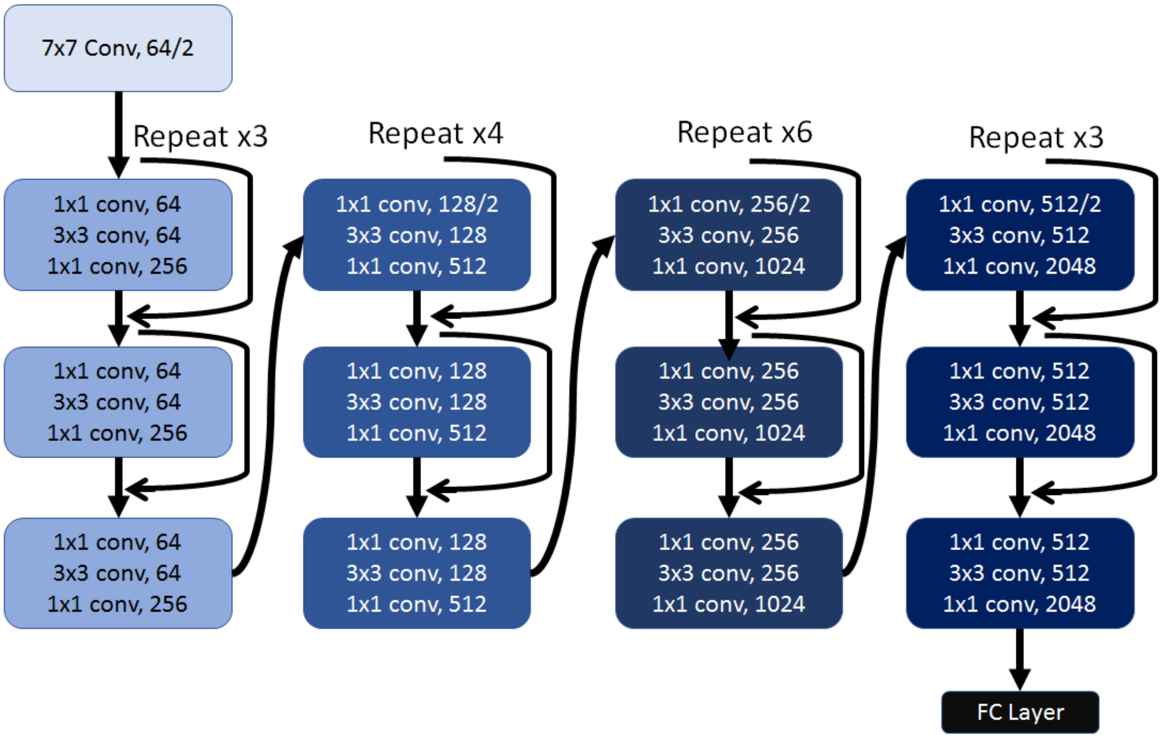


Figure 6-5: Schematic of the architecture of the deep learning network used in this study. YxY conv, M/N = M kernels of YxYx3 size and stride length of N (N=1 if only M is listed). Fully Connected (FC) Layer = Dense(256), Dropout, Dense(1)

Model hyperparameters for data augmentation, training parameters, and optimizing parameters were selected though hyperparameter sweeps of a variety of hyperparameters. The hyperparameters were swept through a full realizable range of values and examined to determine best values to improve training or reduce overfitting. Data augmentation hyperparameters were rotation, zoom, shear, horizontal reflection, and vertical reflection and applied to the training data. Training hyperparameters were learning rate, batch size, number of epochs, image input size, and number of convolutional layers to allow to re-train weights. Model optimizer hyperparameters were momentum, regularization, and decay. Loss and accuracy were computed in the train and validation set. Saliency maps were produced along with a contingency table enumerating the number of correct and incorrect predictions with some sample images in order to understand what factors contributed to incorrect and correct predictions. Image preprocessing was done in Matlab r2015a (Mathworks, Natick, MA), ResNet50 was implemented with Keras and Tensorflow (182) using Spyder 3.2.3 and Python 3.5.

Model Statistical Testing

After training was complete, conditional logistic regression was performed in three cases: one with BI-RADS density as a classifier, one with the deep learning network predictions as a classifier, and one with both. In all cases interval vs. screen-detected was the outcome. An ROC analysis and AUC curve was produced in all cases and compared. Statistical analysis and figure generation was performed via Spyder and R version 3.2.2.

6.2.3 Results

Demographic Information

Table 6-2 shows the demographic information of the women from each case-type. Within this dataset we had 182 women diagnosed with an interval breast cancer. These were matched by age and race to 173 women with screen-detected breast cancers. There were no screen-detected cancers that matched by age and race for 9 of the interval cancers, and these were excluded in the conditional logistic regressions. The descriptive statistics showed a difference in body mass index (BMI) and BI-RADS density between groups, but the other demographic and risk information was not significantly different.

Table 6-2: Descriptive statistics of the screen-detected and interval cancer groups

	Screen-Detected Group	Interval Group	P-Value
N	173	182	
Age, years (Standard Deviation)	57.8 (10.9)	56.8 (11.8)	0.28
BMI, kg/m^2 (Standard Deviation)	24.9 (4.7)	23.5 (4.3)	<0.0001
Race:			0.88
White	127	129	
African American	3	4	
Chinese	25	27	
Filipina	3	3	
Hispanic	0	2	
Japanese	5	8	
Mixed	5	5	
Other Asian	2	1	
Other Non-Asian	3	3	
Menopausal status	119 (69%)	123 (68%)	0.69
Family history of breast cancer	47 (23%)	60 (33%)	0.25
Previous history of breast biopsy	55 (32%)	68 (37%)	0.33
BI-RADS Frequency:			0.008
A: Almost Entirely Fatty	11	3	
B: Scattered Fibroglandularities	50	33	
C: Heterogeneously Dense	61	78	
D: Extremely Dense	19	53	
Missing Data	19	7	
Unknown	13	8	

Hyperparameter Training

Table 6-3 shows the end result of the hyperparameter sweep and optimal hyperparameters that were used in training our network. Of note we learned moderately aggressive image augmentation hyperparameters controlled overfitting while still allowing learning to take place. Additionally, a large batch size improved training by introducing the optimizer to more data and a learning rate in the range of $1e^{-3} – 1e^{-5}$ produced good learning results. High dropout in the final fully-connected layers helped to control overfitting as well. Optimal parameters were selected based on their ability to reduce overfitting and reduce validation loss.

Table 6-3: Chosen hyperparameters with brief description used in this analysis

Hyperparameter (Range)	Hyperparameter Type	Interpretation	Chosen Value
Rotation (0-90)	Data Augmentation	Range for a random rotation	20
Zoom (0-1)	Data Augmentation	Range for a random zoom	0.5
Shear (0-1)	Data Augmentation	Range for a random shear	0.3
Vertical/Horizontal Flip (Yes/No)	Data Augmentation	Random chance of flip in respective direction	Yes/Yes
Momentum (0-1)	Optimizer Parameter	Accelerates or dampens oscillations in given direction.	0.3
Regularization (0-1)	Optimizer Parameter	Penalty applied to large image weights	0
Decay (0-1)	Optimizer Parameter	Learning Rate decay over each update.	1e-5
Dropout (0-1)	Fully-connected Layer	Percent of weights dropped out between dense layers in the FC layer.	0.95
Learning Rate (0-1)	Training Parameter	Importance attributed to weight updates.	1e-3
Epochs (Integer)	Training Parameter	Number of epochs performed	1000
Batch Size (2^n any n)	Training Parameter	Number of samples per gradient update	16
Image Size (Minimum 224)	Training Parameter	Input image size in pixels	224
nLayersRetrain (Fully Connected only – All Layers)	Training Parameter	Number of layers allowed to have their weights altered.	All Layers (173)

Training Results

Figure 6-6 shows the loss and accuracy of the model over time. We can see that the generalization gap between train and test loss is small, indicating little overfitting occurred. Training loss continued to drop after 500 epochs. The best validation loss occurred in epoch 482, with a validation loss of 0.499 and validation accuracy of 75.2 percent.

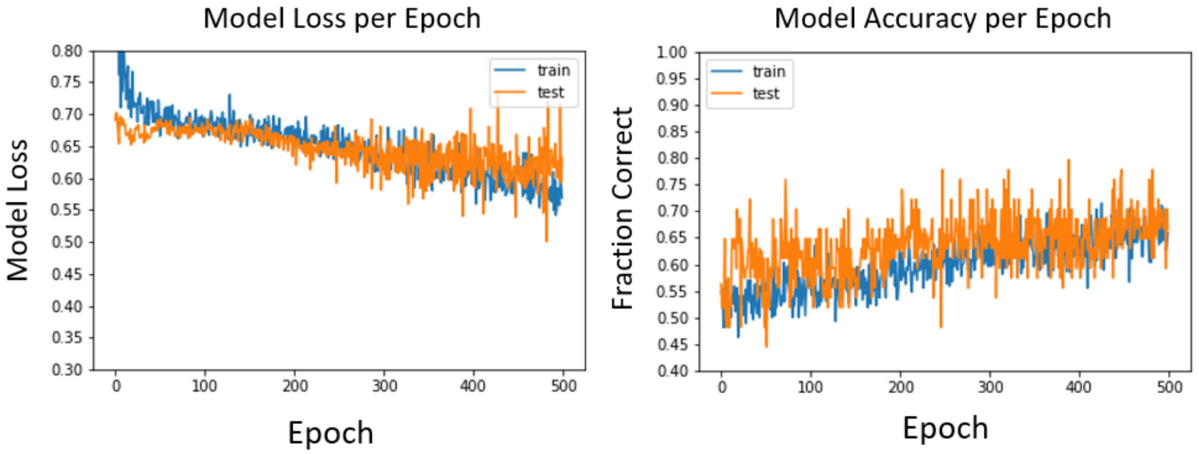


Figure 6-6: Loss and accuracy curves per epoch of the test and train set of the deep learning network. Best validation loss occurred on epoch 482. At that epoch training loss and accuracy were 0.58 and 67.4%, respectively, and test loss and accuracy were 0.499 and 75.2%, respectively.

Table 6-4 shows a contingency table quantifying the number and percent of correct and incorrect predictions in each category. Seventy-five percent of the images were correctly categorized.

Table 6-4: Contingency table of the number of correctly and incorrectly classified images from the dep learning network

Number (Percent)	Predicted Screened	Predicted Interval
Actual Screened	134 (37.6%)	38 (10.7%)
Actual Interval	50 (14.1%)	134 (37.6%)

ROC Analysis

Figure 6-7 compares the classification ROC analysis and AUC of the deep learning network versus using just BI-RADS density in a conditional logistic regression, and a final analysis combining both of these methods. We can see that the neural network outperforms using BI-RADS density alone in predicting interval versus screen-detected cancer.

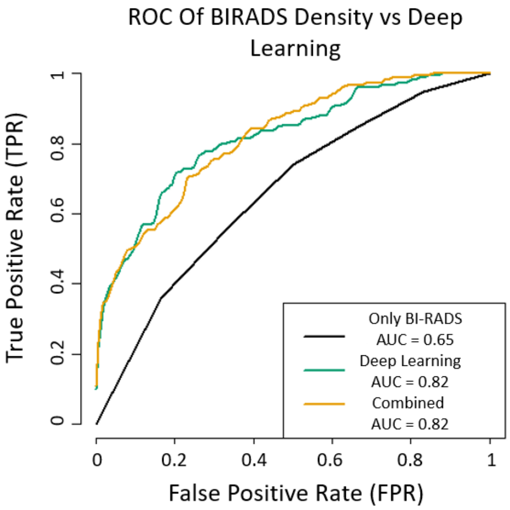


Figure 6-7: ROC Curves interval vs screen-detected cancer classification using BI-RADS density alone (Only BI-RADS) vs using the deep learning predictions (Deep Learning) vs using both as predictors (Combined). Prediction accuracy was 63% using BI-RADS density alone and 75% using deep learning alone.

Saliency Maps

Figure 6-8 shows the pseudo presentation mammograms (produced using methods described by Malkov, et al (201)), saliency maps, and then the superposition of the two for representative screen-detected and interval mammogram visits, both of which were correctly classified. The intensity of the saliency signal is shown from 0 to 255 color scale. A threshold was applied to highlight the regions above the 50 percent activation level in the network to improve image clarity. The side and quadrant (if available) where the cancer was found in subsequent mammograms is also shown. We observed that localized regions could highly influence the classification, but that broad regions of the breast could influence decision making as well.

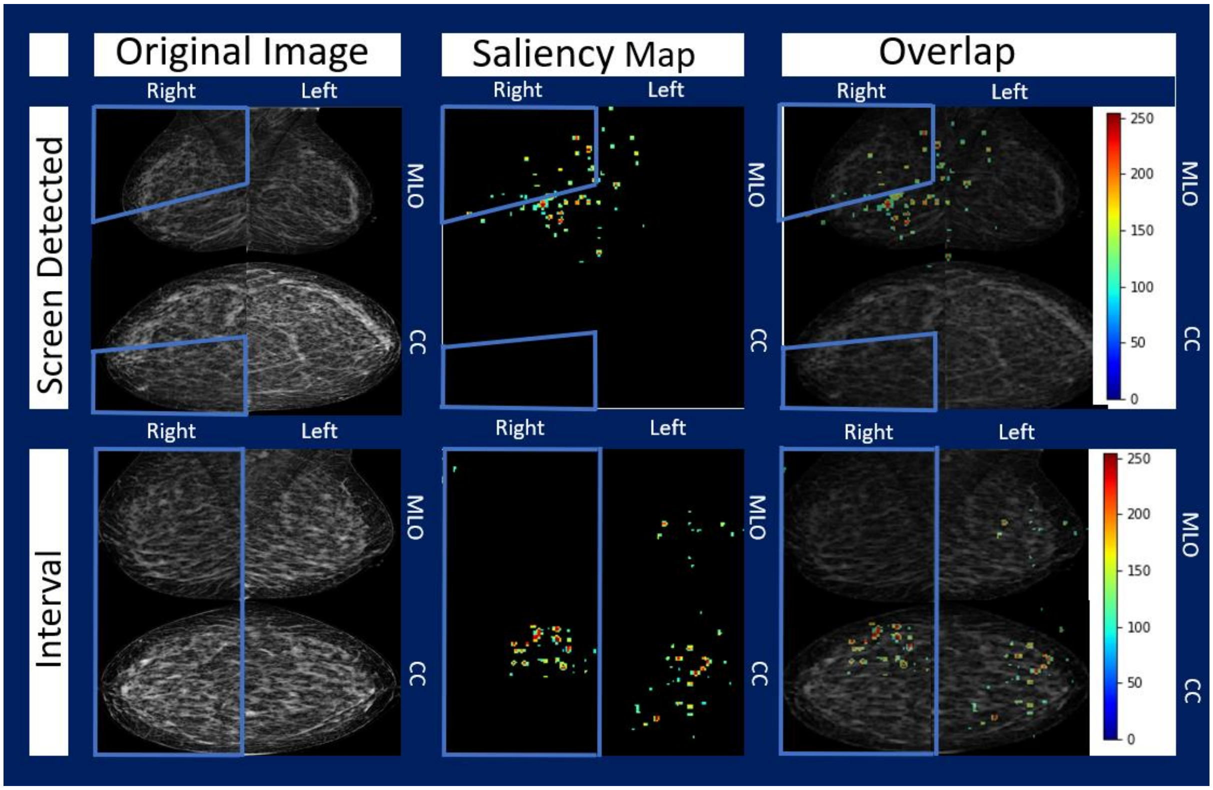


Figure 6-8: Saliency maps of sample screen-detected and interval images (both correctly classified). For each row, the pseudo-presentation images are shown (left) along with the saliency map (middle) that highlights the pixels that had above a 50 percent weight in classifying the image in its respective category (i.e. first row saliency map highlights weights that push towards decision of classifying as screen-detected decision). At right, the images are overlaid.

6.2.4 Discussion

This is the first application of neural networks to classify interval cancer risk known to the authors. The deep learning model provided an AUC of 0.82 with 75% classification accuracy on its own while using BI-RADS density alone produced an AUC of 0.65 and classification accuracy of 63%. This indicates that this deep learning network is able to pick up on information about interval breast cancer risk that is not captured in the BI-RADS density classification alone, such as mammographic masking.

Previous work by Kerlikowske et al. (13) showed that breast density was associated with increased prevalence of interval cancer in a screening population. Women in the highest breast density category (BIRADS D) had 5 to 10 times higher incidence of interval cancers than women with almost entirely fatty breast (BIRADS A). Furthermore, Kerlikowske showed that using a combination of density and age to identify women for discussion about supplemental results in the fewest women counseled per interval cancer occurrence. The most efficient strategy resulted in a ratio of 694 women targeted for supplemental screening per interval compared to 1124 using breast density alone. Recently, automated computerized methods to quantify breast density have been shown to produce similar levels of interval risk as subjective BI-RADS density scores (16). We are not able to directly compare our results to these studies since we did not include women that didn't develop breast cancer, but the implication of this study is that deep learning may be able to further stratify women for supplemental screening beyond age and density.

Other researchers have investigated radiomic features as a measure for interval risk. Strand et al identified several mammographic image features, eccentricity and skewness of intensity gradient, that differentiate between interval and screen detected breast cancer risk (169). Holm, et al identified some biological risk factors between interval and screen detected cancers after controlling for age and mammographic density, including family history of breast cancer, use of hormone replacement therapy (HRT), and BMI above 25 kg/m^2 (20). While no studies of interval risk have been performed using deep learning methods, Mainprize et al developed a direct measure of detectability that were significant for interval risk as well, indicating the promise of computer vision in understanding interval risk (170).

Deep learning models offer several advantages to conventional biomarker models for interval risk. First, deep learning models can continue to learn and improve as more data is added to the training set. This is not usually true with conventional biomarkers. Secondly, deep learning doesn't depend on an a priori hypothesis of what breast feature is important, allowing for identification of patterns or features that may be otherwise overlooked. Third, deep learning models are robust with regards to image artifacts and quality due to the inclusion of scans with a variety of imaging defects.

Deep learning has consistently been applied to and improved on current diagnostic methods in many medical fields. CheXNet has been able to improve upon lung pathology diagnoses using chest X-ray images compared to an average of radiologist diagnoses, improving the F1 score from 0.387 to 0.435 (63). Deep learning was applied to diagnose diabetic retinopathy and was able to improve classification AUC from 0.937 using current detection techniques to 0.980 using deep learning methods (194). While these results are exciting, it is important to note that more work needs to be done to fine-tune these methods, as the retinopathy deep learning network had a low specificity of 87%, leading to more unnecessary biopsies. This problem of false positives is also the case with several other deep learning networks in medical imaging. Several review papers highlight advancements, opportunities, and risks in deep learning in a variety of medical fields (76,77,195).

Deep learning methods have already been applied to several areas of breast cancer research with promising results. An analysis of screening mammograms to perform lesion segmentation and classification with a deep learning network was able to achieve classification AUC of 0.95 and outperformed CAD systems (196). Deep learning networks were also able to categorize amount of dense tissue on par with experts using Cumulus-like thresholds (197). Lastly, a deep learning network lesion classification system achieved an AUC of 0.82 on digitized-screen film mammograms and 0.90 on breast tomosynthesis images (188). Our study, which has been able to differentiate between future interval and screen detected cancers more effectively than using current methods such as BI-RADS breast density, further contributes to the understanding that using deep learning can further our diagnostic capabilities and our understanding in several areas of breast cancer research.

This study has several strengths. First, the dataset controls for age and race between interval and screening-detected cancers, helping to reduce possible bias. Being able to compare to predictions based on BI-RADS density directly allow for comparisons to current interval cancer risk factors (13,167). Additionally, the data pre-processing steps helped to combine data from multiple views and make connections between different views, and the data augmentation we performed in training helped to enhance the variability of our dataset, which helped to prevent or mitigate some of the threat of overfitting to the data. The hyperparameter sweeps performed enabled us to find hyperparameter sets that helped to improve training and test accuracy while minimizing the amount of overfitting.

The contingency table shows the vast majority of images analyzed were correctly classified, with slightly more actual interval images being misclassified compared to actual screen-detected images. This could very well be because the higher density of the interval images made it more difficult to distinguish and classify. The saliency maps provided interesting information about the images and which pixels and regions influenced the decisions to classify the image as an interval or a screen-detected image. While it appears the regions of interest in the interval image was related to density and that the screen detected images were influenced by the region that later developed cancer, further work must be done to examine how the pixels and regions of interest in the saliency maps relate to the underlying tissue and density.

There were several limitations to our study. First, we had a limited number of interval cases. When additional cases become available, an even more robust model may be possible. We attempted to mitigate this risk through the various data pre-processing and augmentation steps we performed, as well as careful selection of hyperparameters to prevent overfitting. Further, splitting the data into a test and train set helped to test against a dataset unseen to the network.

Additionally, this study did not include women that didn't develop cancer by design. Our hypothesis was that there were fundamental differences in the mammograms of women that develop interval versus screen-detected cancers. We found a strong signal to confirm this hypothesis, guiding the path to future studies that compare interval cases, screen detected cases, and non-cases. Lastly, we used a simple definition of interval cancer with the assumption that all of the interval cancers were due to masking. However, there are at least four causes of interval cancers: mammographically-occult cancers that are radiographically similar to normal breast tissue, missed cancers in dense tissue that would have been detected in fatty tissue (mammographic masking), missed cancers due to radiologist fatigue or error, and fast-growing cancers not present at the time of the previous mammogram. Identifying and separating these subgroups can be difficult. We did not separate these types of interval cancers, which introduced additional noise into the dataset. Had we trained on only truly masked cancers, there would have been less noise and the performance may have been even stronger.

This study expands upon current knowledge of the field by implementing cutting edge research in deep learning to improve upon current methods of quantifying interval risk. Previous models currently have implemented model observers in their predictions and produced promising results. Our study expands upon this work by improving the predictive ability and detecting factors important for classification not quantified in BI-RADS density alone. This work could be expanded upon further to develop automated methods or software that can aid radiologists, improve interval risk models, or identify groups of women or currently unknown textures or radiomic quantities that are at high risk of interval cancer.

We conclude that pre-cancerous mammograms contain imaging information beyond breast density that can be used to predict the probability of mammographic masking and interval cancer outcomes. Deep learning models can be trained to utilize this information and may be useful to triage women at high risk of interval cancer into alternative screening methods.

6.3: Study: Deep learning networks find unique mammographic interval risk and screen-detected risk measures (202)

Using deep learning to risk of interval versus screen detected cancer is valuable, but in order to make a robust method of interval risk, we need a deep learning network that is able to identify interval cancer risk in the presence of mammograms from and healthy patients as well as screen detected cases. In order to do this, we need a dataset of healthy controls as well as our matched screen detected and interval cohort.

6.3.1 Introduction

Mammography is the current gold standard in screening for breast cancer in average-risk women. However, radiologically dense and complex tissue can reduce screening detection sensitivity leading to obscuring breast lesions and cancers missed by screening mammography (13,165). These cancers discovered within 12 months after normal screening mammograms are called interval cancers, and the effect of breast density reducing mammography sensitivity is commonly called masking. Roughly 13% of breast cancers diagnosed in the U.S. are interval cancers (203).

Previous studies have shown Breast Imaging-Reporting and Data System (BI-RADS) breast density and other quantitative density measures are risk factors for interval cancer (13,204). While clinically measured BI-RADS breast density is a risk factor for interval cancer such that legislation has been passed in over 40 states to notify women if they have high BI-RADS density breast density (167), the classification is subjective and does not account for the texture of dense tissue (168–170). Because of this the American College of Radiology has asked for development of direct measures of masking and interval risk (10).

Previous studies have measured the ability of pre-defined kernels and model observers to quantify masking and interval cancer risk, indicating some promise in computer vision to identify interval cancer risk (170). Advanced computer vision methods such as deep learning have shown promise in many computer vision tasks and have performed extremely well in the ImageNet competition compared to traditional pre-defined kernel methods (73,76). Transfer learning (191,195) of these networks has been effective in medical applications, including breast cancer, where deep learning models were often able to equal or improve current classification or diagnostic schemes performed (63,76,188,193,196,199,200). Another useful property of deep learning networks is their ability to highlight pixels containing unique information relevant to that image's classification called saliency maps, which can be used in biological applications to develop hypothesis on the underlying biology or features associated with the classification of interest (63,187).

The purpose of this study was to implement a deep learning network to classify pre-cancer mammograms that result in either healthy cases, interval cancers, or screen-detected cancers within 12 months of the mammogram and to compare this classification to a similar classification using BI-RADS breast density. We hypothesized that because of the various advances deep learning has done in several fields, that these networks can more effectively quantify risk of interval or screen-detected cancer than BI-RADS breast density. If successful, these methods could be expanded to improve risk prediction models for interval cancer, develop automated methods or software that can aid radiologists in risk prediction, and further understand radiomic quantities as they relate to underlying cancer biology.

6.3.2 Methods

Participants

Participants were selected from a screening population that had received full-field digital mammograms acquired from 2006 to 2015 from four radiology facilities, University of California – San Francisco, California Pacific Medical Center, Marin General Hospital, and Novato Community Hospital. Ethics approval was obtained by the University of California – San Francisco Institutional Review Board for this retrospective analysis of mammograms. Interval cancers from these institutions were included, defined as invasive cancers identified within 12 months of a negative screening examination. An equal number of screen-detected cancers were matched by age and race if such matching data existed, based on all screen-detected cancers diagnosed at the four centers. Screen-detected cancers were defined as invasive cancers identified within 12 months of a positive screening examination. Two healthy controls were matched to each interval and screen-detected case. All mammograms were interpreted prospectively by radiologists during the course of routine clinical care. Cancers were identified by annual linkage to the state California Cancer Registry.

Mammography

The de-identified raw, "For Processing" representation of the standard four screening views (Mediolateral-oblique (MLO) and Cranio-Caudal (CC) images of both sides) were used for this study. All images were acquired on Hologic Selenia full-field digital mammography systems. These images were pre-processed in order to maximize the information provided to the network in the following ways. First, the skin edge of these images was identified via thresholding and excess background of the images was cropped out. The images were then normalized on a 0 to 255 scale and the four views were stacked as a 2x2 collage image for each case, with one view in each quadrant. This allowed all four views to be contained in a single image. These images were then separated into a training and test set at an 80/20 split.

Deep Learning Models

An existing deep learning network architecture (ResNet50) was implemented with ImageNet transfer learning weights on all convolutional blocks (73). A fully connected layer was then added with 256 weights, a dropout layer, and a final weight with sigmoid activation to classify between screen-detected and interval cases. Figure 1 shows a diagram of the deep learning architecture and the fully-connected layer (73). The weights of the fully-connected layer were randomly initialized and pre-trained on the training set images. During training, a categorical cross entropy loss metric was optimized with an adadelta optimizer (182). Class weights were introduced into the loss metric during training to account for class imbalance of the healthy control images (182). During model training, data augmentation was performed by introducing a random amount of shear, zoom, rotation, and horizontal and vertical reflection within specific ranges in order to increase the data variability and reduce overfitting. For each epoch, training and test loss and accuracy were recorded and network weights were saved if it improved the test loss. The final network weights used were the weights with the best test loss throughout training.

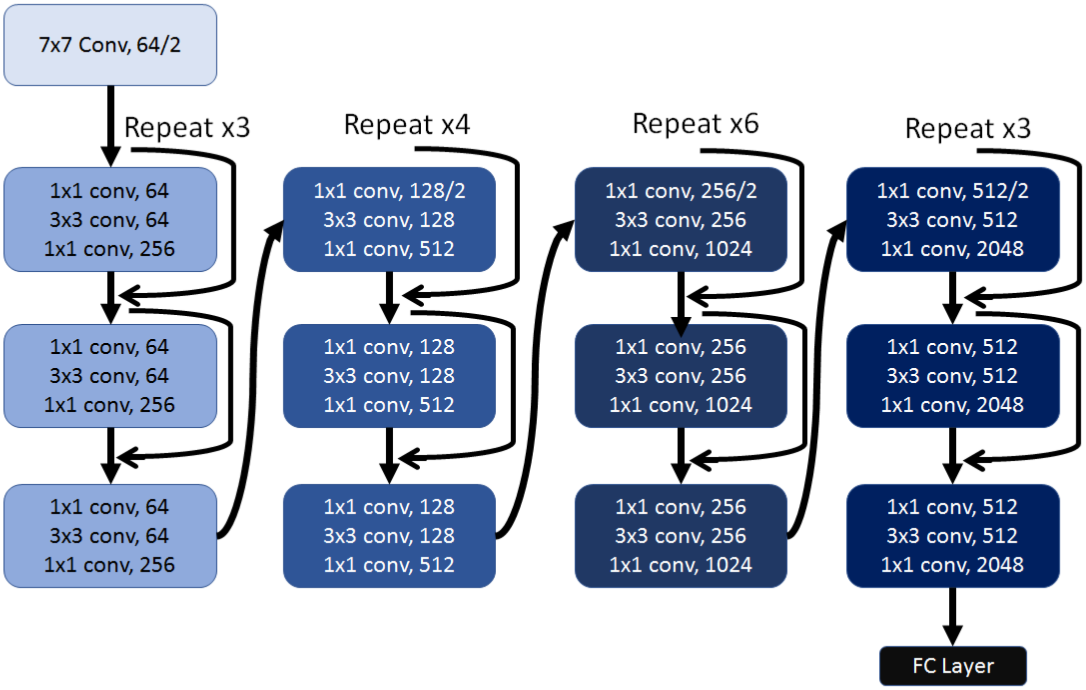


Figure 6-9: Schematic of the architecture of the deep learning network used in this study. YxY conv, M/N = M kernels of YxYx3 size and stride length of N (N=1 if only M is listed). Fully Connected (FC) Layer = Dense(256), Dropout, Dense(1)

Model hyperparameters for data augmentation, training parameters, and optimizing parameters were selected through hyperparameter sweeps of a variety of hyperparameters. The hyperparameters were swept through a full realizable range of values and examined to determine best values to improve training or reduce overfitting. Data augmentation hyperparameters were rotation, zoom, shear, horizontal reflection, and vertical reflection and applied to the training data. Training hyperparameters were learning rate, batch size, number of epochs, image input size, and number of convolutional layers to allow to re-train weights. Model optimizer hyperparameters were momentum, regularization, and decay. Loss and accuracy were computed in the training and test set. Saliency maps were produced along with a contingency table enumerating the number of correct and incorrect predictions with some sample images in order to understand what factors contributed to incorrect and correct predictions. Image preprocessing was done in Matlab r2015a (Mathworks, Natick, MA), ResNet50 was implemented with Keras and Tensorflow (182) using Spyder 3.2.3 and Python 3.5.

Statistical Analysis

After training was complete, accuracy and loss curves were generated. Further, contingency tables were generated to show the correct and incorrect classifications. Lastly, risk of interval cancer and screen-detected cancer were calculated using conditional logistic regression of each cancer case versus the respective matched controls, with BI-RADS density as a risk factor and using the deep learning predictions of each outcome as a risk measure, Statistical analysis and figure generation was performed via Spyder and R version 3.2.2.

6.3.3 Results

Table 1 shows the demographic information of the women from each category. Within this dataset we had 182 women diagnosed with an interval breast cancer. These were matched by age and race to 173 women with screen-detected breast cancers. There were then a total of 679 healthy controls. There were no screen-detected cancers that matched by age and race for 9 of the interval cancers. These were included in the deep learning training to maximize the dataset, but were excluded in the conditional logistic regressions to ensure matching.

Table 6-5 shows the end results of the hyperparameter sweep and optimal hyperparameters that were used in training our network. Of note we learned moderately aggressive image augmentation hyperparameters controlled overfitting while still allowing learning to take place. Additionally, a large batch size improved training by introducing the optimizer to more data and a learning rate in the range of $1e^{-3}$ – $1e^{-5}$ produced good learning results. High dropout in the final fully-connected layers helped to control overfitting as well. Optimal parameters were selected based on their ability to reduce overfitting and reduce training and test loss.

Table 6-5: Chosen hyperparameters with brief description. Hyperparameter sweep went through a realizable range for each hyperparameter and individual values were chosen to optimize training ability or to minimize overfitting, depending on the parameter.

Hyperparameter (Range)	**Hyperparameter Type**	**Interpretation**	**Chosen Value**
Rotation (0-90)	Data Augmentation	Range for a random rotation	20
Zoom (0-1)	Data Augmentation	Range for a random zoom	0.4
Shear (0-1)	Data Augmentation	Range for a random shear	0.3
Vertical/Horizontal Flip (Yes/No)	Data Augmentation	Random chance of flip in respective direction	Yes/Yes
Momentum (0-1)	Optimizer Parameter	Accelerates or dampens oscillations in given direction.	0.3
Regularization (0-1)	Optimizer Parameter	Penalty applied to large image weights	0
Decay (0-1)	Optimizer Parameter	Learning Rate decay over each update.	1e-6
Dropout (0-1)	Fully-connected Layer	Percent of weights dropped out between dense layers in the FC layer.	0.75
Learning Rate (0-1)	Training Parameter	Importance attributed to weight updates.	1e-3
Epochs (Integer)	Training Parameter	Number of epochs performed	1000
Batch Size (2^n any n)	Training Parameter	Number of samples per gradient update	16
Image Size (Minimum 224)	Training Parameter	Input image size in pixels	275
nLayersRetrain (Fully Connected only – All Layers)	Training Parameter	Number of layers allowed to have their weights altered.	All Layers (173)

Figure 6-10 shows the loss and accuracy of the model over time and compares the result of the training and test set. We can see that the generalization gap between train and test loss is small and that the curves in the test and train set are similar. This indicates little overfitting occurred and that classification results were similar in both training and test sets. The best test loss occurred in epoch 579, with a test accuracy of 67 percent.

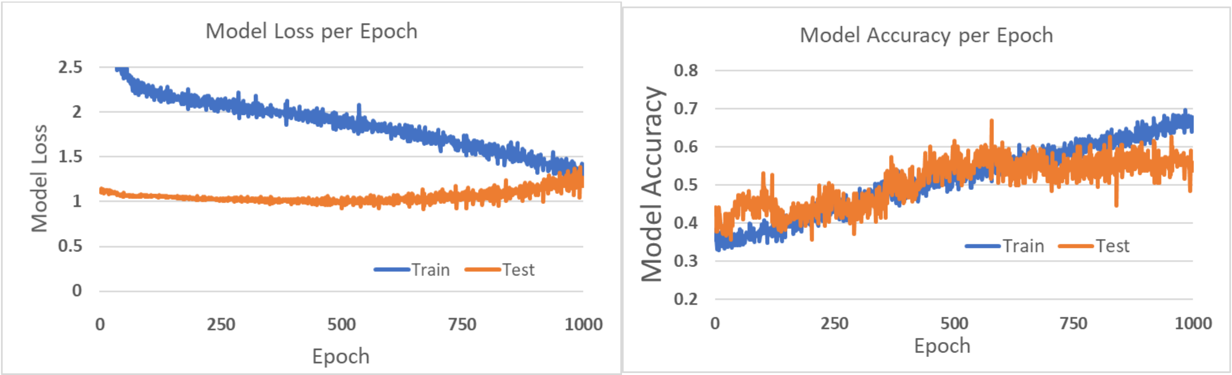


Figure 6-10: Loss and accuracy curves per epoch of the test and train set of the deep learning network. Best validation loss occurred on epoch 579. At that epoch training and test accuracy were 54% and 67%, respectively.

Table 6-6 shows a contingency table quantifying the number and percent of correct and incorrect predictions in each category. Sixty-seven percent of the images were correctly categorized.

Table 6-6: Contingency table of the number of correctly and incorrectly classified images from the deep learning network. A higher accuracy existed in predicting the healthy images, and there were few cases of one cancer type being misclassified as the other cancer type (i.e. an interval cancer being classified as screened, or vice versa).

Number (Percent)	Predicted Healthy	Predicted Interval	Predicted Screened	Total
Actual Healthy	500/679 (73.6%)	65/679 (9.5%)	114/679 (17.2%)	679
Actual Interval	64/184 (34.7%)	97/184 (52.7%)	23/184 (12.5%)	184
Actual Screened	68/172 (39.5%)	8/172 (4.7%)	96/172 (55.8%)	172
Total	632	170	223	1035

6.3.4 Discussion

We developed a deep learning algorithm based off of methods that were able to improve on risk of interval versus screen detected cancer, and included healthy controls into the training to get a true sense of interval risk. After training, the network was able to correctly classify between interval cancer, screen-detected cancer, and healthy images with a 67% classification accuracy. This was quite encouraging considering a baseline guessing accuracy would have been 33 percent accurate, and it indicates that a deep learning network is able to identify information in mammograms associated with interval breast cancer diagnosis that is not captured in the BI-RADS density classification alone.

Contingency tables showed the area where the deep learning network was failing. The network classified the future healthy controls at the highest accuracy, with 73.6% of controls being correctly identified. When healthy cases were misclassified, they were more often classified as future screen-detected cancers. This may be because the future interval cases may have had qualities that indicated high masking or low detectability, whereas the future screened and future healthy images may have had similar levels of detectability. The future interval and future screened images were correctly classified 52.7 and 55.8 percent of the time. In both cases, there was very little misclassification from one cancer category into the other cancer category, indicating the network was able to strongly distinguish between these two categories.

Future analyses will finalize this study by comparing classification accuracy using BI-RADS density. Further, the odds ratios for interval and screen detected cancer risk will be calculated using BI-RADS density as a predictor and compared with using the deep learning predictions as a predictor. We expect the results of this will show improvements in interval cancer risk using the deep learning predictions, but the data required for this analysis was not available at the time of publication.

Previous work by Kerlikowske et al. (13) showed that breast density was associated with increased prevalence of interval cancer in a screening population. Furthermore, Kerlikowske showed that using a combination of breast density and 5-year breast cancer risk to identify women for discussion about supplemental results in the fewest women counseled per interval cancer occurrence. Recently, automated methods to quantify breast density have been shown to produce similar levels of interval risk as subjective BI-RADS density scores (204).

Deep learning networks have the ability to introduce additional complexity, can become more robust with the addition of more data, and can identify patterns or features that may be otherwise overlooked with conventional biomarker models for interval risk. Deep learning has consistently been applied to and improved on current diagnostic methods in many medical fields (76,77,195), including lung pathology diagnosis (63).

Deep learning methods have also been applied to several areas of breast cancer research with promising results. Deep learning has been applied to improve lesion detection in computer-aided detection (196), to effectively categorize the amount of dense tissue in mammograms (197), and to improve lesion classification systems on breast tomosynthesis images (188). Our study further contributes to recent findings that deep learning can improve diagnostic capabilities and our understanding in breast cancer research.

This study has several strengths. First, the dataset controls for age and race, helping to reduce possible confounding. Further, comparing our network to predictions based on BI-RADS density provides comparisons against current interval cancer risk factors (13,167). Additionally, the transfer learning methods, the data preprocessing steps, data augmentation steps, class weights, and hyperparameter sweeps performed helped to maximize test accuracy while minimizing overfitting.

There were several limitations to our study. First, we had a limited number of interval cancers, which makes overfitting a concern to be considered. We mitigated this risk through transfer learning, data preprocessing and augmentation, and careful selection of hyperparameters. Additionally, comparing results of the test and train sets to ensure they had similar results helped to prevent overfitting. Lastly, we used a broad definition of interval cancer. While many interval cancers occur because the lesion is masked, some interval cancers occur due to radiologist fatigue or error and others from fast growing lesions that develop after the previous mammogram (17). Identifying and separating these subgroups can be difficult. We did not separate these types of interval cancers, which introduced additional noise into the dataset and may have weakened our results compared to using a dataset of only truly masked interval cancers.

We conclude that pre-cancerous mammograms, whether interval, screen detected, or healthy in the future, contain imaging information beyond breast density that can be used to predict the probability of breast cancer detection, and that deep learning models may be able to detect and identify that imaging information. This work could be expanded upon further to improve risk prediction models for interval cancer or screening detected cancer, develop automated methods or software that can aid radiologists in risk prediction, and understand if these deep learning predictions relate to underlying radiomic quantities or tissue biology.

7 | Future Directions

A limitation in training deep learning networks is the compute power of the system being used and the size of the dataset available. Being limited in compute power make training take longer, makes it less feasible to test a variety of hyperparameters, makes it less feasible to use large datasets, and forces image compression during training. In the future this lab will acquire methods to drastically increase compute power, which should improve the speed of iterating on models, allows for more robust training, and reduce the amount of necessary image compression. The NVIDIA DGX-1 has 170 teraflops of compute power while the NVIDIA Quadro K2200 used in these studies had roughly 2 teraflops of compute power, so with the new system with a 100x improved compute power will lead to drastic improvements in training time, required image compression, and ability to iterate on new networks (205,206).

In order to generate the strongest measures of interval risk, an analysis on a full prospective screening study will need to be performed and analyzed. To date, we have performed preliminary analysis on a subset of a prospective screening study with promising results. Our future analysis on the full prospective screening study will improve the performance of the network and make it more robust.

There are additional steps we will take in the future to produce more optimized training and additional understanding of deep learning predictions. The future implementation of the NVIDIA DGX-1 will allow for further optimization by training the deep learning network on higher resolution images. Further, the additional computation power of the DGX-1 will make it more practical to implement these deep learning networks on the much larger tomosynthesis images. Lastly, we plan to perform investigates that examine biological differences between the high and low interval risk images, to see why and how the tissue is different and how that contributes to interval risk.

The end goal of this work is to provide tools that will assist radiologists in identifying risk of interval cancer to suggest supplemental screening for women at high risk of interval cancer that will maximize interval cancers found early and minimize additional biopsies. In order to do this, significant work needs to occur to convert this from a black box prediction that a deep learning network outputs into a meaningful output that radiologists and clinicians can understand, identify with, and interact with. We will continue to work on saliency maps and GUIs that present the deep learning predictions in a digestible way. Further, tying these saliency maps to underlying biology and radiomic features or correlating known textural features or histology features with deep learning predictions will help clinicians better understand and grapple with these measures of interval risk.

8 | Conclusions

Throughout this dissertation, we developed and iterated on several imaging methods to improve upon previous measures of cancer risk, and we have identified when imaging methods fail to be useful for detecting cancer. These studies were built on both well-established methods to identify risk as well as novel technologies that are becoming more readily available. While the specific methods and techniques differ across these studies, the studies in this dissertation all aimed to produce stronger imaging-based risk factors and can be used to help personalize different screening and risk tools.

I began my work by studying risk of cancer as related to obesity and lean mass measures using DXA. Since the composition of specific limbs of the body can lead to more specific risk measures, we developed a means to quantify the expected composition for each limb in the body controlled for age, race, and gender through LMS normalization. Because of this work, reference values for limbs can be readily calculated for the US population and used in future studies on obesity and muscularity as they relate to cancer risk.

We next studied why mammography failed to detect breast cancer. We had a hypothesis that inherently high density and complex texture of some women's breast puts them at high risk of interval cancers compared to women with low density, low complexity breast and screen detected cancers. We used traditional computer vision methods in the form of model observers to emulate current detectability metrics used in mammography in the form of the CDMAM phantom. This led to the development of the IQF metric, which was able to improve on classifying interval versus screen detected cancers with an AUC of 0.69 compared to using 0.65 with breast density alone. While these results were not as strong as we had hoped, they confirmed that there were inherent differences in imaging metrics between interval cancers and screened cancers that is not detected in breast density alone. Although there are ways to extend traditional computer vision methods such as different detectability filters, deep learning had consistently outperformed traditional computer vision methods in a variety of other tasks (44,186), leading us to conclude that deep learning methods in future studies would outperform compared to expanding on traditional computer vision methods.

To further improve measures of interval risk, we leveraged the fact that developments in deep learning have improved classification and diagnosis in a variety of fields including the medical field. We developed deep learning networks to identify differences between future interval and screen detected images and classify between categories. This network utilized transfer learning, image collaging, image augmentation, and hyperparameter sweeps in order to mitigate issues that can come from smaller datasets seen in medical imaging such as overfitting, non-robust feature weights, and poor results on new data. These methods led to a classification ROC AUC of 0.82 with deep learning compared to 0.65 with breast density alone. This confirmed our hypothesis that there are imaging differences in breasts of women who later are diagnosed with interval versus screen detected cancer, and led to the next logical step, testing classification with healthy controls added in the dataset in order to generate true measures of interval risk.

By adding controls into the classification, we gain the ability to generate true measures of interval risk, but need to account for several factors. Because control images are more common, class weights needed to be introduced in order to prevent the deep learning network from over-guessing the most common category. By incorporating this as well as previous methods to prevent overfitting, we were able to correctly classify 67% of the images. 75% of control images were correctly classified, and there was very little misclassification occurred between cancer types (i.e. very few screen-detected cancers were classified as intervals, and vice versa). In the future odds ratios and other measures of risk are being developed. With controls added into the study, we've further shown how these deep learning methods can add to and improve on interval risk measures.

In summary, these works provide a detailed analysis of various methods to identify interval risk, and break ground on emergent methods to use deep learning to identify interval risk. These methods could be further developed to identify groups at high risk of interval cancer. By identifying these women and referring them to supplemental screening programs, we can help reduce the number of interval cancers and improve mortality rates for women.

References

1. DeSantis C, Ma J, Bryan L, Jemal A. Breast cancer statistics, 2013: Breast Cancer Statistics, 2013. CA Cancer J Clin. 2014 Jan;64(1):52–62.

2. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, et al. Cancer Statistics, 2008. CA Cancer J Clin. 2008 Jan 28;58(2):71–96.

3. Coughlin SS, Ekwueme DU. Breast cancer as a global health concern. Cancer Epidemiol. 2009 Nov;33(5):315–8.

4. Yabroff KR, Lund J, Kepka D, Mariotto A. Economic Burden of Cancer in the United States: Estimates, Projections, and Future Research. Cancer Epidemiol Biomarkers Prev. 2011 Oct 1;20(10):2006–14.

5. Max W, Sung H-Y, Stark B. The economic burden of breast cancer in California. Breast Cancer Res Treat. 2009 Jul;116(1):201–7.

6. McPherson K. ABC of breast diseases: Breast cancer---epidemiology, risk factors, and genetics. BMJ. 2000 Sep 9;321(7261):624–8.

7. Narod SA. Tumour size predicts long-term survival among women with lymph node-positive breast cancer. Curr Oncol [Internet]. 2012 Sep 27 [cited 2016 Sep 15];19(5). Available from: http://www.current-oncology.com/index.php/oncology/article/view/1043

8. Esserman L. Rethinking Screening for Breast Cancer and Prostate Cancer. JAMA. 2009 Oct 21;302(15):1685.

9. Esserman LJ, Thompson IM, Reid B, Nelson P, Ransohoff DF, Welch HG, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. Lancet Oncol. 2014 May;15(6):e234–42.

10. D'Orsi C, Sickles E, Mendelson E, Morris E, et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology; 2013.

11. Prince J, Links J. Medical Imaging Signals and Systems. Pearson Prentica Hall;

12. Sprague BL, Arao RF, Miglioretti DL, Henderson LM, Buist DS, Onega T, et al. National performance benchmarks for modern diagnostic digital mammography: update from the Breast Cancer Surveillance Consortium. Radiology. 2017;283(1):59–69.

13. Kerlikowske K, Zhu W, Tosteson A, Sprague B, Tice J, Lehman C, et al. Identifying Women With Dense Breasts at High Risk for Interval Cancer: A Cohort Study. Ann Intern Med. 2015 May 19;162(10):673–81.

14. Kerlikowske K. 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.

15. Gilliland FD, Joste N, Stauber PM, Hunt WC, Rosenberg R, Redlich G, et al. Biologic characteristics of interval and screen-detected breast cancers. J Natl Cancer Inst. 2000;92(9):743–749.

16. Kerlikowske K, Scott CG, Mahmoudzadeh AP, Ma L, Winham S, Matthew J, et al. Automated and Clinical BI-RADS Breast Density Measures Predict Risk of Screen-detected and Interval Cancers. Ann Intern Med. Awaiting Publication;

17. Strand F. Determinants of interval cancer and tumor size among breast cancer screening participants [Internet]. 2018 [cited 2018 Jun 20]. Available from: http://hdl.handle.net/10616/46330

18. Hofvind S, Geller B, Skaane P. Mammographic features and histopathological findings of interval breast cancers. Acta Radiol. 2008 Nov;49(9):975–81.

19. Siu AL, on behalf of the U.S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2016 Feb 16;164(4):279.

20. Holm J, Humphreys K, Li J, Ploner A, Cheddad A, Eriksson M, et al. Risk Factors and Tumor Characteristics of Interval Cancers by Mammographic Density. J Clin Oncol. 2015 Mar 20;33(9):1030–7.

21. Hofvind S, Yankaskas BC, Bulliard J-L, Klabunde CN, Fracheboud J. Comparing interval breast cancer rates in Norway and North Carolina: results and challenges. J Med Screen. 2009 Sep 1;16(3):131–9.

22. Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, hm-Velez MB, et al. Combined Screening With Ultrasound and Mammography vs Mammography Alone in Women at Elevated Risk of Breast Cancer. :13.

23. Corsetti V, Houssami N, Ghirardi M, Ferrari A, Speziani M, Bellarosa S, et al. Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: Interval breast cancers at 1year follow-up. Eur J Cancer. 2011 May;47(7):1021–6.

24. Are You Dense [Internet]. Available from: http://www.areyoudense.org

25. Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, et al. Breast density as a predictor of mammographic detection: comparison of interval-and screen-detected cancers. J Natl Cancer Inst. 2000;92(13):1081–1087.

26. Ciatto S, Visioli C, Paci E, Zappa M. Breast density as a determinant of interval cancer at mammographic screening. Br J Cancer. 2004 Jan 26;90(2):393–6.

27. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007;356(3):227–236.

28. Ciatto S, Houssami N, Apruzzese A, Bassetti E, Brancato B, Carozzi F, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. The Breast. 2005 Aug;14(4):269–75.

29. Redondo A, Comas M, Macià F, Ferrer F, Murta-Nascimento C, Maristany MT, et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. Br J Radiol. 2012 Nov;85(1019):1465–70.

30. Mayo Clinic [Internet]. Breast density - The four levels. Available from: https://www.mayoclinic.org/tests-procedures/mammogram/multimedia/breast-density-mdash-the-four-levels/img-20008862

31. Gail MH, Costantino JP. Validating and Improving Models for Projecting the Absolute Risk of Breast Cancer. JNCI J Natl Cancer Inst. 2001 Mar 7;93(5):334–5.

32. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. JNCI J Natl Cancer Inst. 1989 Dec 20;81(24):1879–86.

33. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, 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 Mar 4;148(5):337.

34. Mainprize JG, Alonzo-Proulx O, Jong RA, Yaffe MJ. Quantifying masking in clinical mammograms via local detectability of simulated lesions. Med Phys. 2016 Mar;43(3):1249–58.

35. Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, Miller AB, et al. Quantitative Classification of Mammographic Densities and Breast Cancer Risk: Results From the Canadian National Breast Screening Study. JNCI J Natl Cancer Inst. 1995 May 3;87(9):670–5.

36. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. Phys Med Biol. 1994 Oct 1;39(10):1629–38.

37. Hartman K, Highnam R, Warren R, Jackson V. Volumetric Assessment of Breast Tissue Composition from FFDM Images. In: Krupinski EA, editor. Digital Mammography [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008 [cited 2016 Oct 26]. p. 33–9. Available from: http://link.springer.com/10.1007/978-3-540-70538-3_5

38. Aitken Z, McCormack VA, Highnam RP, Martin L, Gunasekara A, Melnichouk O, et al. 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 Feb 1;19(2):418–28.

39. Ciatto S, Bernardi D, Calabrese M, Durando M, Gentilini MA, Mariscotti G, et al. A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. The Breast. 2012 Aug;21(4):503–6.

40. 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 Apr;4(2):173–82.

41. Wang J, Azziz A, Fan B, Malkov S, Klifa C, Newitt D, et al. Agreement of Mammographic Measures of Volumetric Breast Density to MRI. Matsuo K, editor. PLoS ONE. 2013 Dec 4;8(12):e81653.

42. Volpara [Internet]. Measuring Breast Density. Available from: https://volparasolutions.com/science-hub/breast-density/measuring-breast-density/

43. Gastounioti A, Conant EF, Kontos D. Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment. Breast Cancer Res [Internet]. 2016 Dec [cited 2018 Jun 25];18(1). Available from: http://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-016-0755-8

44. T D, A E. CS280: Computer Vision. Course Presentations presented at; 2016 Mar.

45. Strand F, Humphreys K, Cheddad A, Törnberg S, Azavedo E, Shepherd J, et al. Novel mammographic image features differentiate between interval and screen-detected breast cancer: a case-case study. Breast Cancer Res [Internet]. 2016 Dec [cited 2017 Sep 12];18(1). Available from: http://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-016-0761-x

46. Strand F, Humphreys K, Eriksson M, Li J, Andersson TML, Törnberg S, et al. Longitudinal fluctuation in mammographic percent density differentiates between interval and screen-detected breast cancer: Strand et al. Int J Cancer. 2017 Jan 1;140(1):34–40.

47. Byrne C, Schairer C, Wolfe J, Parekh N, Salane M, Brinton LA, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. J Natl Cancer Inst. 1995;87(21):1622–1629.

48. Tan M, Zheng B, Leader JK, Gur D. Association Between Changes in Mammographic Image Features and Risk for Near-Term Breast Cancer Development. IEEE Trans Med Imaging. 2016 Jul;35(7):1719–28.

49. Drukker K, Duewer F, Giger ML, Malkov S, Flowers CI, Joe B, et al. Mammographic quantitative image analysis and biologic image composition for breast lesion characterization and classification: Mammographic quantitative image analysis and biologic image composition. Med Phys. 2014 Feb 27;41(3):031915.

50. Malkov S, Shepherd JA, Scott CG, Tamimi RM, Ma L, Bertrand KA, et al. Mammographic texture and risk of breast cancer by tumor type and estrogen receptor status. Breast Cancer Res [Internet]. 2016 Dec [cited 2017 Jul 13];18(1). Available from: http://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-016-0778-1

51. Baydush AH, Catarious Jr DM, Floyd Jr CE. Computer-aided detection of masses in mammography using a Laguerre-Gauss channelized hotelling observer. In: Medical Imaging 2003 [Internet]. International Society for Optics and Photonics; 2003 [cited 2016 Oct 26]. p.

71–76. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=758480

52. Chan H-P, Wei J, Sahiner B, Rafferty EA, Wu T, Roubidoux MA, et al. Computer-aided Detection System for Breast Masses on Digital Tomosynthesis Mammograms: Preliminary Experience. Radiology. 2005 Dec;237(3):1075–80.

53. Davies DH, Dance DR. Automatic computer detection of clustered calcifications in digital mammograms. Phys Med Biol. 1990;35(8):1111.

54. Jalalian A, Mashohor SBT, Mahmud HR, Saripan MIB, Ramli ARB, Karasfi B. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clin Imaging. 2013 May;37(3):420–6.

55. Jinshan Tang, Rangayyan RM, Jun Xu, El Naqa I, Yongyi Yang. Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. IEEE Trans Inf Technol Biomed. 2009 Mar;13(2):236–51.

56. Wang Y, Aghaei F, Zarafshani A, Qiu Y, Qian W, Zheng B. Computer-aided classification of mammographic masses using visually sensitive image features. J X-Ray Sci Technol. 2017 Jan 28;25(1):171–86.

57. Baydush AH, Catarious DM, Lo JY, Floyd CE. Incorporation of a Laguerre–Gauss Channelized Hotelling Observer for False-Positive Reduction in a Mammographic Mass CAD System. J Digit Imaging. 2007 May 15;20(2):196–202.

58. History of Mammography [Internet]. GE Healthcare; Available from: http://newsroom.gehealthcare.com/wp-content/uploads/2016/03/mamography-final.pdf

59. Study Blue: X-Ray Tube [Internet]. Available from: https://www.studyblue.com/notes/note/n/tube-and-tubehead/deck/1900677?utm_source=homeworkhelp&utm_content=hhtest

60. Production of X-Rays [Internet]. Available from: https://www.radiologycafe.com/radiology-trainees/frcr-physics-notes/production-of-x-rays

61. Pisano ED, Cole EB, Hemminger BM. Image Processing Algorithms for Digital Mammography: A Pic- torial Essay. 20(5):13.

62. Karunanithi R, Ganesan S, Panicker TMR, Korath MP, Jagadeesan K. Assessment of bone mineral density by DXA and the trabecular microarchitecture of the calcaneum by texture analysis in pre- and postmenopausal women in the evaluation of osteoporosis. J Med Phys. 2007 Oct;32(4):161–8.

63. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. :7.

64. StackExchange [Internet]. Available from: https://physics.stackexchange.com/questions/150546/why-dont-x-rays-travel-through-water

65. Radiopaedia. Mammogram Systems [Internet]. Available from: https://radiopaedia.org/images/3165401

66. Boone JM, Lindfors KK, Cooper VN, Seibert JA. Scatter/primary in mammography: Comprehensive results. Med Phys. 2000;27(10):2408.

67. Boone JM, Cooper VN. Scatter/primary in mammography: Monte Carlo validation. Med Phys. 2000;27(8):1818.

68. Tan M, Pu J, Zheng B. Reduction of false-positive recalls using a computerized mammographic image feature analysis scheme. Phys Med Biol. 2014 Aug 7;59(15):4357–73.

69. Yaffe MJ. Mammographic density. Measurement of mammographic density. Breast Cancer Res [Internet]. 2008 Jun [cited 2018 Jun 25];10(3). Available from: http://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr2102

70. Dense Breast Advocacy [Internet]. Available from: densebreastadvocacy.org

71. NVIDIA. From the Frontline: How Deep Learning Plays Critical Role in Military Problem-Solving [Internet]. Available from: https://blogs.nvidia.com/blog/2016/06/29/deep-learning-6/

72. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. CoRR. 2015;

73. Fukushima K. Cognitron: A self-organizing multilayered neural network. Biol Cybern. 1975;20(3–4):121–36.

74. Miyake S, Fukushima K. A neural network model for the mechanism of feature-extraction: A self-organizing network with feedback inhibition. Biol Cybern. 1984 Aug;50(5):377–84.

75. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems [Internet]. 2012 [cited 2017 Oct 13]. p. 1097–1105. Available from: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks

76. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017 Dec;42:60–88.

77. Razzak MI, Naz S, Zaib A. Deep Learning for Medical Image Processing: Overview, Challenges and Future. :30.

78. Müller MJ. From BMI to functional body composition. Eur J Clin Nutr. 2013 Nov;67(11):1119–21.

79. Sood A, Qualls C, Li R, Schuyler M, Beckett WS, Smith LJ, et al. Lean mass predicts asthma better than fat mass among females. Eur Respir J. 2011 Jan 1;37(1):65–71.

80. Van Pelt RE, Evans EM, Schechtman KB, Ehsani AA, Kohrt WM. Contributions of total and regional fat mass to risk for cardiovascular disease in older women. Am J Physiol - Endocrinol Metab. 2002 May 1;282(5):E1023–8.

81. Jordan MJ, Aagaard P, Herzog W. Lower limb asymmetry in mechanical muscle function: A comparison between ski racers with and without ACL reconstruction: Bilateral asymmetry in ACL-R ski racers. Scand J Med Sci Sports. 2015 Jun;25(3):e301–9.

82. Breen L, Stokes KA, Churchward-Venne TA, Moore DR, Baker SK, Smith K, et al. Two Weeks of Reduced Activity Decreases Leg Lean Mass and Induces "Anabolic Resistance" of Myofibrillar Protein Synthesis in Healthy Elderly. J Clin Endocrinol Metab. 2013 Jun;98(6):2604–12.

83. Williams M, Hunter G, Kekes-Szabo T, Snyder S, Treuth M. Regional fat distribution in women and risk of cardiovascular disease. Am J Clin Nutr. 1996 Nov 7;65:855–60.

84. He Q. Trunk Fat and Blood Pressure in Children Through Puberty. Circulation. 2002 Mar 5;105(9):1093–8.

85. Sakai Y, Ito H, Egami Y, Ohoto N, Hijii C, Yanagawa M, et al. Favourable association of leg fat with cardiovascular risk factors. J Intern Med. 2005;257(2):194–200.

86. Villaça DS, Lerario MC, Corso S dal, Nápolis L, Albuquerque ALP de, Lazaretti-Castro M, et al. Clinical value of anthropometric estimates of leg lean volume in nutritionally depleted and non-depleted patients with chronic obstructive pulmonary disease. Br J Nutr [Internet]. 2008 Aug [cited 2016 Jul 27];100(02). Available from: http://www.journals.cambridge.org/abstract_S0007114507886399

87. Toth MJ, Tchernof A, Sites CK, Poehlman ET. Menopause-related changes in body fat distribution. Ann N Y Acad Sci. 2000;904(1):502–506.

88. Martin ML, Jensen MD. Effects of body fat distribution on regional lipolysis in obesity. J Clin Invest. 1991;88(2):609.

89. Caan BJ, Cespedes Feliciano EM, Prado CM, Alexeeff S, Kroenke CH, Bradshaw P, et al. Association of Muscle and Adiposity Measured by Computed Tomography With Survival in Patients With Nonmetastatic Breast Cancer. JAMA Oncol. 2018 Jun 1;4(6):798.

90. Braithwaite D, Satariano WA, Sternfeld B, Hiatt RA, Ganz PA, Kerlikowske K, et al. Long-term Prognostic Role of Functional Limitations Among Women With Breast Cancer. JNCI J Natl Cancer Inst. 2010 Oct 6;102(19):1468–77.

91. Cole TJ, Green PJ. Smoothing Reference Centile Curves: The LMS Method and Penalized Likelihood. Stat Med. 1992;11:1305–19.

92. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey [Internet]. National Health and Nutrition Examination Survey. 2015 [cited 2015 Oct 20]. Available from: http://www.cdc.gov/nchs/nhanes.htm

93. Kelly TL, Wilson KE, Heymsfield SB. Dual Energy X-Ray Absorptiometry Body Composition Reference Values from NHANES. Vella A, editor. PLoS ONE. 2009 Sep 15;4(9):e7038.

94. Kelly TL, Wilson KE, Heymsfield SB. Dual Energy X-Ray Absorptiometry Body Composition Reference Values from NHANES. Vella A, editor. PLoS ONE. 2009 Sep 15;4(9):e7038.

95. Hinton BJ, Fan B, Ng BK, Shepherd JA. Dual energy X-ray absorptiometry body composition reference values of limbs and trunk from NHANES 1999–2004 with additional visualization methods. Thomas DM, editor. PLOS ONE. 2017 Mar 27;12(3):e0174180.

96. Binkley N, Kiebzak GM, Lewiecki EM, Krueger D, Gangnon RE, Miller PD, et al. Recalculation of the NHANES Database SD Improves T-Score Agreement and Reduces Osteoporosis Prevalence. J Bone Miner Res. 2004 Nov 16;20(2):195–201.

97. Looker AC, Melton LJ, Harris TB, Borrud LG, Shepherd JA. Prevalence and trends in low femur bone density among older US adults: NHANES 2005–2006 compared with NHANES III. J Bone Miner Res. 2010 Jan;25(1):64–71.

98. Schoeller DA, Tylavsky FA, Baer DJ, Chumlea WC, Earthman CP, Fuerst T, et al. QDR 4500A dual-energy X-ray absorptiometer underestimates fat mass in comparison with criterion methods in adults. Am J Clin Nutr. 2005;81(5):1018–1025.

99. Looker AC, Melton LJ, Harris T, Borrud L, Shepherd J, McGowan J. Age, gender, and race/ethnic differences in total body and subregional bone density. Osteoporos Int. 2009 Jul;20(7):1141–9.

100. Flegal KM, Ogden CL, Yanovski JA, Freedman DS, Shepherd JA, Graubard BI, et al. High adiposity and high body mass index–for-age in US children and adolescents overall and by race-ethnic group. Am J Clin Nutr. 2010 Apr 1;91(4):1020–6.

101. Looker AC, Wright NC. Osteoporosis or Low Bone Mass at the Femur Neck or Lumbar Spine in Older Adults: United States, 2005-2008. :8.

102. American Diabetes Association. Economic Costs of Diabetes in the U.S. in 2012. Diabetes Care. 2013 Apr 1;36(4):1033–46.

103. Haffner SM, Lehto S, Rönnemaa T, Pyörälä K, Laakso M. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med. 1998;339(4):229–234.

104. Deurenberg P, Weststrate JA, Seidell JC. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. Br J Nutr. 1991;65(02):105–114.

105. Carnethon MR, De Chavez PJD, Biggs ML, Lewis CE, Pankow JS, Bertoni AG, et al. Association of weight status with mortality in adults with incident diabetes. Jama. 2012;308(6):581–590.

106. Burkhauser RV, Cawley J. Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. J Health Econ. 2008 Mar;27(2):519–29.

107. Rothman KJ. BMI-related errors in the measurement of obesity. Int J Obes. 2008 Aug;32:S56–9.

108. Weber DR, Moore RH, Leonard MB, Zemel BS. Fat and lean BMI reference curves in children and adolescents and their utility in identifying excess adiposity compared with BMI and percentage body fat. Am J Clin Nutr. 2013 Jul 1;98(1):49–56.

109. Müller MJ. From BMI to functional body composition. Eur J Clin Nutr. 2013 Nov;67(11):1119–21.

110. He Q. Trunk Fat and Blood Pressure in Children Through Puberty. Circulation. 2002 Mar 5;105(9):1093–8.

111. Sakai Y, Ito H, Egami Y, Ohoto N, Hijii C, Yanagawa M, et al. Favourable association of leg fat with cardiovascular risk factors. J Intern Med. 2005;257(2):194–200.

112. Villaça DS, Lerario MC, Corso S dal, Nápolis L, Albuquerque ALP de, Lazaretti-Castro M, et al. Clinical value of anthropometric estimates of leg lean volume in nutritionally depleted and non-depleted patients with chronic obstructive pulmonary disease. Br J Nutr [Internet]. 2008 Aug [cited 2016 Jul 27];100(02). Available from: http://www.journals.cambridge.org/abstract_S0007114507886399

113. Toth MJ, Tchernof A, Sites CK, Poehlman ET. Menopause-related changes in body fat distribution. Ann N Y Acad Sci. 2000;904(1):502–506.

114. Martin ML, Jensen MD. Effects of body fat distribution on regional lipolysis in obesity. J Clin Invest. 1991;88(2):609.

115. Wilson JP, Kanaya AM, Fan B, Shepherd JA. Ratio of Trunk to Leg Volume as a New Body Shape Metric for Diabetes and Mortality. Buzzetti R, editor. PLoS ONE. 2013 Jul 10;8(7):e68716.

116. Prado CM, Siervo M, Mire E, Heymsfield SB, Stephan BC, Broyles S, et al. A population-based approach to define body-composition phenotypes. Am J Clin Nutr. 2014 Jun 1;99(6):1369–77.

117. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, Boirie Y, Cederholm T, Landi F, et al. Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. Age Ageing. 2010 Jul 1;39(4):412–23.

118. Morley JE, Abbatecola AM, Argiles JM, Baracos V, Bauer J, Bhasin S, et al. Sarcopenia With Limited Mobility: An International Consensus. J Am Med Dir Assoc. 2011 Jul;12(6):403–9.

119. Visser M. Towards a definition of sarcopenia—results from epidemiologic studies. JNHA- J Nutr Health Aging. 2009;13(8):713–716.

120. Schutz Y, Kyle U, Pichard C. Fat-free mass index and fat mass index percentiles in Caucasians aged 18 – 98 y. Int J Obes. 2002;(26):953–60.

121. Albanese CV, Diessel E, Genant HK. Clinical Applications of Body Composition Measurements Using DXA. J Clin Densitom. 2003;6(2):75–85.

122. Deurenberg P, Deurenberg-Yap M. Validity of body composition methods across ethnic population groups. Acta Diabetol. 2003 Oct 1;40(0):s246–9.

123. Schoeller DA, Tylavsky FA, Baer DJ, Chumlea WC, Earthman CP, Fuerst T, et al. QDR 4500A dual-energy X-ray absorptiometer underestimates fat mass in comparison with criterion methods in adults. Am J Clin Nutr. 2005;81(5):1018–1025.

124. Wells JC, Williams JE, Chomtho S, Darch T, Grijalva-Eternod C, Kennedy K, et al. Body-composition reference data for simple and reference techniques and a 4-component model: a new UK reference child. Am J Clin Nutr. 2012 Dec 1;96(6):1316–26.

125. Pan H, Cole TJ. A comparison of goodness of fit tests for age-related reference ranges. Stat Med. 2004;23(11):1749–1765.

126. Oyhenart EE, Lomaglio DB, Dahinten SLV, Bejarano IF, Herráez Á, Cesani MF, et al. Weight and height percentiles calculated by the LMS method in Argentinean schoolchildren. A comparative references study. Ann Hum Biol. 2015 Sep 3;42(5):439–46.

127. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity: Extended international BMI cut-offs. Pediatr Obes. 2012 Aug;7(4):284–94.

128. Kranz S, Mahood LJ, Wagstaff DA. Diagnostic criteria patterns of U.S. children with Metabolic Syndrome: NHANES 1999–2002. Nutr J [Internet]. 2007 Dec [cited 2017 Jan 10];6(1). Available from: http://nutritionj.biomedcentral.com/articles/10.1186/1475-2891-6-38

129. Pang MYC, Eng JJ, McKay HA, Dawson AS. Reduced hip bone mineral density is related to physical fitness and leg lean mass in ambulatory individuals with chronic stroke. Osteoporos Int. 2005 Dec;16(12):1769–79.

130. Hassink SG, Sheslow DV, Lancey E de, Opentanova I, Considine RV, Caro JF. Serum Leptin in Children With Obesity: Relationship to Gender and Development. Pediatrics. 1996 Aug;98(2):201–3.

131. Hart NH, Nimphius S, Spiteri T, Newton RU. Leg strength and lean mass symmetry influences kicking performance in Australian Football. J Sports Sci Med. 2014;13(1):157.

132. Mooradian AD, Morley JE, Korenman SG. Biological actions of androgens. Endocr Rev. 1987;8(1):1–28.

133. Nussey S, Whitehead S. Endocrinology: An Integrated Approach [Internet]. Oxford: BIOS Scientific Publishers; 2001. Available from: https://www.ncbi.nlm.nih.gov/books/NBK22/?depth=10

134. Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. J Physiol. 1968 Mar 1;195(1):215–43.

135. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 1962 Jan 1;160(1):106–54.

136. Convolutional Neural Networks [Internet]. Convolutional Neural Networks. Available from: https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/convolutional_neural_networks.html

137. Shrivakshan GT. A Comparison of various Edge Detection Techniques used in Image Processing. 2012;9(5):8.

138. Mathworks. MATLAB [Internet]. Available from: https://www.mathworks.com/products/matlab.html

139. Glantz S. Primer of Biostatistics. 7th ed. McGraw-Hill Education;

140. Wikipedia [Internet]. Image Texture. Available from: https://en.wikipedia.org/wiki/Image_texture

141. Eckstein MP, Abbey CK, Bochud FO. A practical guide to model observers for visual detection in synthetic and natural noisy images. Handb Med Imaging. 2000;1:593–628.

142. Young KC, Cook JJH, Oduko JM, Bosmans H. Comparison of software and human observers in reading images of the CDMAM test object to assess digital mammography systems. In: Flynn MJ, Hsieh J, editors. 2006 [cited 2015 Dec 10]. p. 614206-614206–13. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1322588

143. Burgess AE, Jacobson FL, Judy PF. Human observer detection experiments with mammograms and power-law noise. Med Phys. 2001;28(4):419–437.

144. Yani Zhang, Pham BT, Eckstein MP. The effect of nonlinear human visual system components on performance of a channelized Hotelling observer in structured backgrounds. IEEE Trans Med Imaging. 2006 Oct;25(10):1348–62.

145. Platiša L, Goossens B, Vansteenkiste E, Park S, Gallas BD, Badano A, et al. Channelized Hotelling observers for the assessment of volumetric imaging data sets. JOSA A. 2011;28(6):1145–1163.

146. Young S, Bakic PR, Myers KJ, Jennings RJ, Park S. A virtual trial framework for quantifying the detectability of masses in breast tomosynthesis projection data. Med Phys [Internet]. 2013 [cited 2017 Jul 19];40(5). Available from: http://onlinelibrary.wiley.com/doi/10.1118/1.4800501/full

147. Burgess AE, Judy PF. Signal detection in power-law noise: effect of spectrum exponents. J Opt Soc Am A. 2007 Dec 1;24(12):B52.

148. Barrett HH, Yao J, Rolland JP, Myers KJ. Model observers for assessment of image quality. Proc Natl Acad Sci. 1993;90(21):9758–9765.

149. Barrett HH, Denny JL, Wagner RF, Myers KJ. Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance. JOSA A. 1995;12(5):834–852.

150. He X, Park S. Model Observers in Medical Imaging Research. Theranostics. 2013;3(10):774–86.

151. Grosjean B, Muller S. Impact of textured background on scoring of simulated CDMAM phantom. In: Digital Mammography [Internet]. Springer; 2006 [cited 2015 Nov 2]. p. 460–467. Available from: http://link.springer.com/chapter/10.1007/11783237_62

152. Eckstein MP, Abbey CK, Whiting JS. Human vs model observers in anatomic backgrounds. In: Medical Imaging'98 [Internet]. International Society for Optics and Photonics; 1998 [cited 2016 Aug 31]. p. 16–26. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=943219

153. Burgess AE. Comparison of receiver operating characteristic and forced choice observer performance measurement methods. Med Phys. 1995;22(5):643.

154. Elangovan P, Mackenzie A, Dance DR, Young KC, Cooke V, Wilkinson L, et al. Design and validation of realistic breast models for use in multiple alternative forced choice virtual clinical trials. Phys Med Biol. 2017 Apr 7;62(7):2778–94.

155. Rico R, Muller SL, Peter G, Noel A, Stines J. Automated scoring of CDMAM: a dose study. In: Medical Imaging 2003 [Internet]. International Society for Optics and Photonics; 2003 [cited 2015 Nov 2]. p. 164–173. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=758494

156. Karssemeijer N, Thijssen MAO. Determination of contrast-detail curves of mammography systems by automated image analysis. Digit Mammogr. 1996;96:155–160.

157. Young KC, Cook JJ, Oduko JM. Automated and human determination of threshold contrast for digital mammography systems. In: Digital Mammography [Internet]. Springer; 2006 [cited 2015 Dec 10]. p. 266–272. Available from: http://link.springer.com/chapter/10.1007/11783237_37

158. Pezeshk A, Petrick N, Sahiner B. Seamless lesion insertion in digital mammography: methodology and reader study. In: Tourassi GD, Armato SG, editors. 2016 [cited 2017 Aug 1]. p. 97850J. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2217056

159. Shaheen E, Van Ongeval C, Cockmartin L, Zanca F, Marshall N, Jacobs J, et al. Realistic Simulation of Microcalcifications in Breast Tomosynthesis. In: Martí J, Oliver A, Freixenet J,

Martí R, editors. Digital Mammography [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010 [cited 2017 Jul 19]. p. 235–42. Available from: http://link.springer.com/10.1007/978-3-642-13666-5_32

160. Shaheen E, De Keyzer F, Bosmans H, Dance DR, Young KC, Ongeval CV. The simulation of 3D mass models in 2D digital mammography and breast tomosynthesis: Simulation of 3D mass models. Med Phys. 2014 Jul 29;41(8Part1):081913.

161. Shaheen E, Van Ongeval C, Zanca F, Cockmartin L, Marshall N, Jacobs J, et al. The simulation of 3D microcalcification clusters in 2D digital mammography and breast tomosynthesis: Simulation of 3D microcalcification clusters. Med Phys. 2011 Nov 28;38(12):6659–71.

162. Chen RC, Longo R, Rigon L, Zanconati F, De Pellegrin A, Arfelli F, et al. Measurement of the linear attenuation coefficients of breast tissues by synchrotron radiation computed tomography. Phys Med Biol. 2010 Sep 7;55(17):4993–5005.

163. Tomal A, Mazarro I, Kakuno EM, Poletti ME. Experimental determination of linear attenuation coefficient of normal, benign and malignant breast tissues. Radiat Meas. 2010 Oct;45(9):1055–9.

164. Hinton BJ, Ma L, Mahmoudzadeh AP, Malkov S, Fan B, Greenwood H, et al. In Submission: Derived mammographic masking measures based on simulated lesions predict the risk of interval cancer after controlling for known risk factors: a case-case analysis. Med Phys. 2018 Aug 6;

165. Kerlikowske K. 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.

166. Sprague BL, Arao RF, Miglioretti DL, Henderson LM, Buist DS, Onega T, et al. National performance benchmarks for modern diagnostic digital mammography: update from the Breast Cancer Surveillance Consortium. Radiology. 2017;283(1):59–69.

167. Are You Dense [Internet]. Available from: http://www.areyoudense.org/

168. Malkov S, Shepherd JA, Scott CG, Tamimi RM, Ma L, Bertrand KA, et al. Mammographic texture and risk of breast cancer by tumor type and estrogen receptor status. Breast Cancer Res [Internet]. 2016 Dec [cited 2017 Jul 13];18(1). Available from: http://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-016-0778-1

169. Strand F, Humphreys K, Cheddad A, Törnberg S, Azavedo E, Shepherd J, et al. Novel mammographic image features differentiate between interval and screen-detected breast cancer: a case-case study. Breast Cancer Res [Internet]. 2016 Dec [cited 2017 Sep 12];18(1). Available from: http://breast-cancer-research.biomedcentral.com/articles/10.1186/s13058-016-0761-x

170. Mainprize JG, Alonzo-Proulx O, Jong RA, Yaffe MJ. Quantifying masking in clinical mammograms via local detectability of simulated lesions. Med Phys. 2016 Mar;43(3):1249–58.

171. ACR. ACR Breast Imaging Reporting and Data System Atlas. 2013;

172. 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.

173. van Engeland S, Snoeren PR, Huisman H, Boetes C, Karssemeijer N. Volumetric breast density estimation from full-field digital mammograms. IEEE Trans Med Imaging. 2006 Mar;25(3):273–82.

174. Narod SA. Tumour size predicts long-term survival among women with lymph node-positive breast cancer. Curr Oncol [Internet]. 2012 Sep 27 [cited 2016 Sep 15];19(5). Available from: http://www.current-oncology.com/index.php/oncology/article/view/1043

175. Samei E, Flynn MJ, Reimann DA. A method for measuring the presampled MTF of digital radiographic systems using an edge test device. Med Phys. 1998;25(1):102.

176. Eckstein MP, Abbey CK, Whiting JS. Human vs model observers in anatomic backgrounds. In: Medical Imaging'98 [Internet]. International Society for Optics and Photonics; 1998 [cited 2016 Aug 31]. p. 16–26. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=943219

177. Veldkamp WJH, Thijssen MAO, Karssemeijer N. The value of scatter removal by a grid in full field digital mammography. Med Phys. 2003;30(7):1712.

178. Ciatto S, Bernardi D, Calabrese M, Durando M, Gentilini MA, Mariscotti G, et al. A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. The Breast. 2012 Aug;21(4):503–6.

179. Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. Med Phys [Internet]. 2013 [cited 2017 Jul 19];40(1). Available from: http://onlinelibrary.wiley.com/doi/10.1118/1.4770279/full

180. Sechopoulos I. A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications. Med Phys [Internet]. 2013 [cited 2017 Jul 19];40(1). Available from: http://onlinelibrary.wiley.com/doi/10.1118/1.4770281/full

181. Helvie MA. Digital Mammography Imaging: Breast Tomosynthesis and Advanced Applications. Radiol Clin North Am. 2010 Sep;48(5):917–29.

182. Chollet, F. Keras: deep learning library for theano and tensorflow. 2015.

183. NVIDIA. Accelerating AI with GPUs: A New Computing Model [Internet]. Available from: https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-artificial-intelligence-gpus/

184. Reinsel D, Gantz J, Rydning J, Seagate. Data Age 2025: The Evolution of Data to Life-Critical Don't Focus on Big Data; Focus on Data That's Big. IDC. 2017 Apr;

185. fourquadrant. IDG CIO Tech Priorities – New Data Analytics Have the Edge [Internet]. Available from: https://www.fourquadrant.com/go-to-market-big-data-analytics-research/

186. Imagenet [Internet]. Available from: http://www.image-net.org/challenges/LSVRC/

187. Simonyan K, Vedaldi A, Zisserman A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. ArXiv13126034 Cs [Internet]. 2013 Dec 20 [cited 2018 May 22]; Available from: http://arxiv.org/abs/1312.6034

188. Samala RK, Chan H-P, Hadjiiski LM, Helvie MA, Richter C, Cha KH. Cross-domain and multi-task transfer learning of deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis. In: Mori K, Petrick N, editors. SPIE; 2018 [cited 2018 May 4]. p. 25. Available from: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10575/2293412/Cross-domain-and-multi-task-transfer-learning-of-deep-convolutional/10.1117/12.2293412.full

189. Overfitting [Internet]. Available from: https://en.wikipedia.org/wiki/Overfitting

190. Image Augmentation for Machine Learning [Internet]. Available from: https://github.com/aleju/imgaug

191. Long M, Wang J, Ding G, Pan SJ, Yu PS. Adaptation Regularization: A General Framework for Transfer Learning. IEEE Trans Knowl Data Eng. 2014 May;26(5):1076–89.

192. Transfer Learning [Internet]. Available from: http://slideplayer.com/slide/8370683/

193. Achanta HK, Misganaw B, Vidyasagar M. Integrating biological data across multiple platforms using importance-weighted transfer learning and applications to breast cancer data sets. In: Control Technology and Applications (CCTA), 2017 IEEE Conference on. IEEE; 2017. p. 955–960.

194. Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning. Investig Opthalmology Vis Sci. 2016 Oct 4;57(13):5200.

195. Greenspan H, van Ginneken B, Summers RM. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. IEEE Trans Med Imaging. 2016 May;35(5):1153–9.

196. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I. Detecting and classifying lesions in mammograms with Deep Learning. Sci Rep [Internet]. 2018 Dec [cited 2018 Apr 2];8(1). Available from: http://www.nature.com/articles/s41598-018-22437-z

197. Kallenberg M, Petersen K, Nielsen M, Ng AY, Diao P, Igel C, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. :10.

198. Hinton BJ, Ma L, Mahmoudzadeh AP, Malkov S, Fan B, Greenwood H, et al. In Submission: Deep learning networks find unique mammographic differences in previous negative mammograms between interval and screen-detected cancers. Med Image Anal. 2018 Aug 28;

199. Firoozbakht F, Rezaeian I, Porter L, Rueda L. Breast cancer subtype identification using machine learning techniques. In: Computational Advances in Bio and Medical Sciences (ICCABS), 2014 IEEE 4th International Conference on. IEEE; 2014. p. 1–2.

200. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015;13:8–17.

201. 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.

202. Hinton BJ, Ma L, Mahmoudzadeh AP, Malkov S, Fan B, Greenwood H, et al. In Submission: Deep learning networks find unique mammographic interval risk and screen-detected risk measures. Med Image Anal. 2018 Aug 28;

203. Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DSM, Kerlikowske K, et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. Radiology. 2017 Apr;283(1):49–58.

204. Kerlikowske K, Scott CG, Mahmoudzadeh AP, et al. Automated and clinical breast imaging reporting and data system density measures predict risk for screen-detected and interval cancers: A case–control study. Ann Intern Med. 2018 Jun 5;168(11):757–65.

205. NVIDIA V. NVIDIA Quadra Spec Sheet.

206. NVIDIA. DGX1 Launch Datasheet.

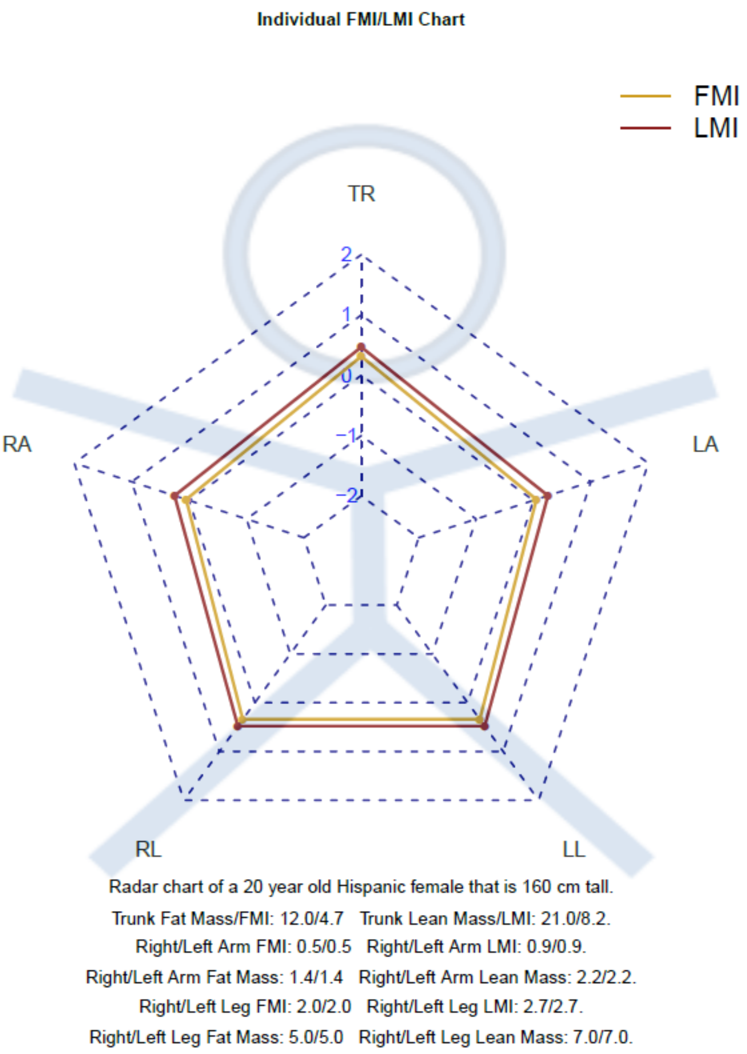
Appendix A: Code Availability

Several software tools and analysis scripts were developed over the course of this dissertation. These pieces of code are maintained in secure repositories on the Shepherd Lab GitHub (https://github.com/shepherd-lab). A listing of relevant repositories is provided below. Please direct inquiries to Ben Hinton (bhinton@berkeley.edu) or John Shepherd (johnshep@hawaii.edu).

- **deepLearningMasking (https://github.com/shepherd-lab/deepLearningMasking)**
  Python scripts to train deep learning networks for interval risk and generate predictions, visualization, and other tools.

Appendix B: Supplemental Tables

Supplemental tables and figures are provided below for the regional composition study: Section 2.3:



Radar chart of a 20 year old Hispanic female that is 160 cm tall.

Trunk Fat Mass/FMI: 12.0/4.7 Trunk Lean Mass/LMI: 21.0/8.2.

Right/Left Arm FMI: 0.5/0.5 Right/Left Arm LMI: 0.9/0.9.

Right/Left Arm Fat Mass: 1.4/1.4 Right/Left Arm Lean Mass: 2.2/2.2.

Right/Left Leg FMI: 2.0/2.0 Right/Left Leg LMI: 2.7/2.7.

Right/Left Leg Fat Mass: 5.0/5.0 Right/Left Leg Lean Mass: 7.0/7.0.

Figure 0-1: Sample radar chart output relating the FMI and LMI of each limb to the standards based on age, race, and sex.

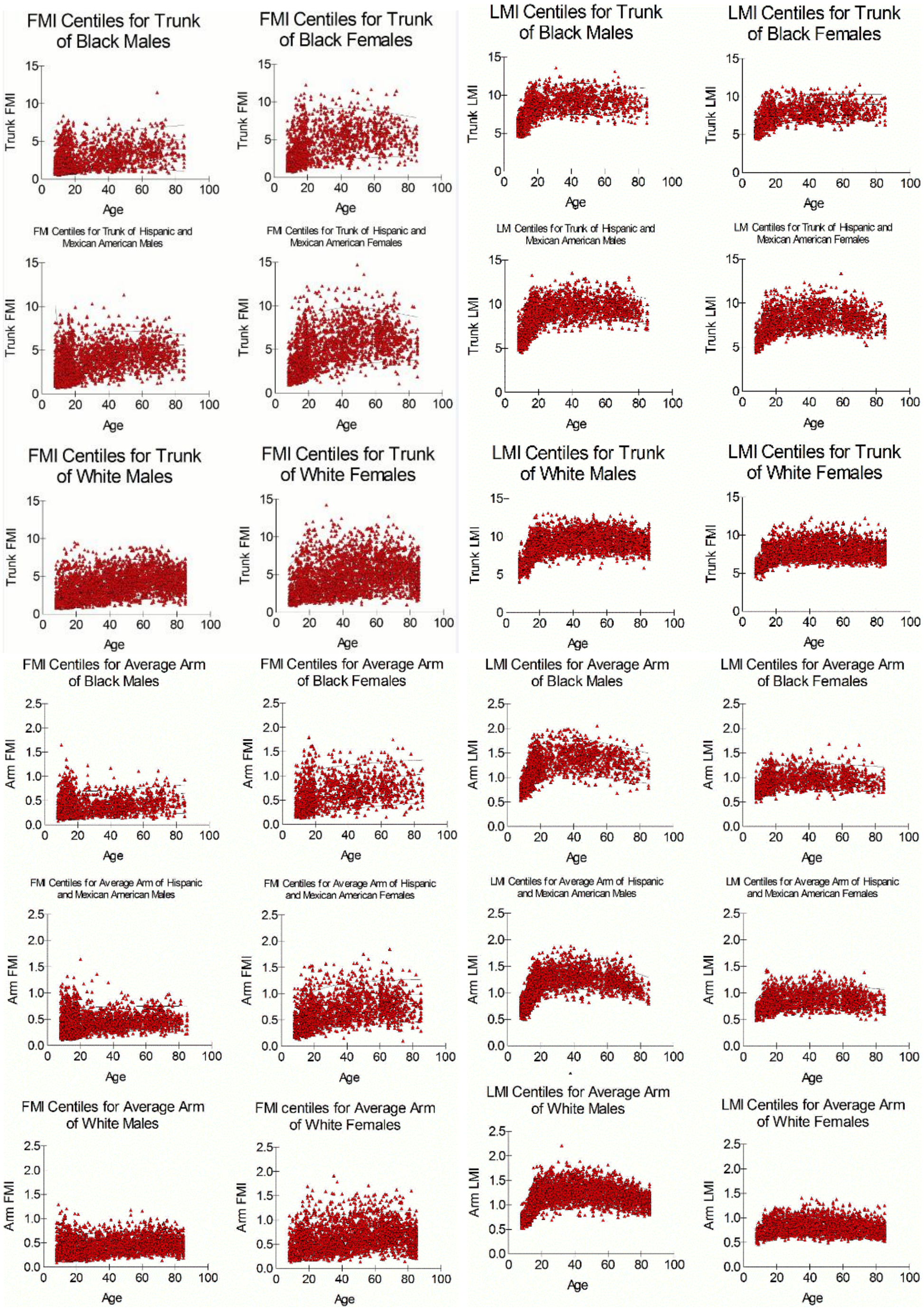


Figure 0-2: Distribution of FMI and LMI datapoints for the arms and trunk and centiles by age, race, and sex

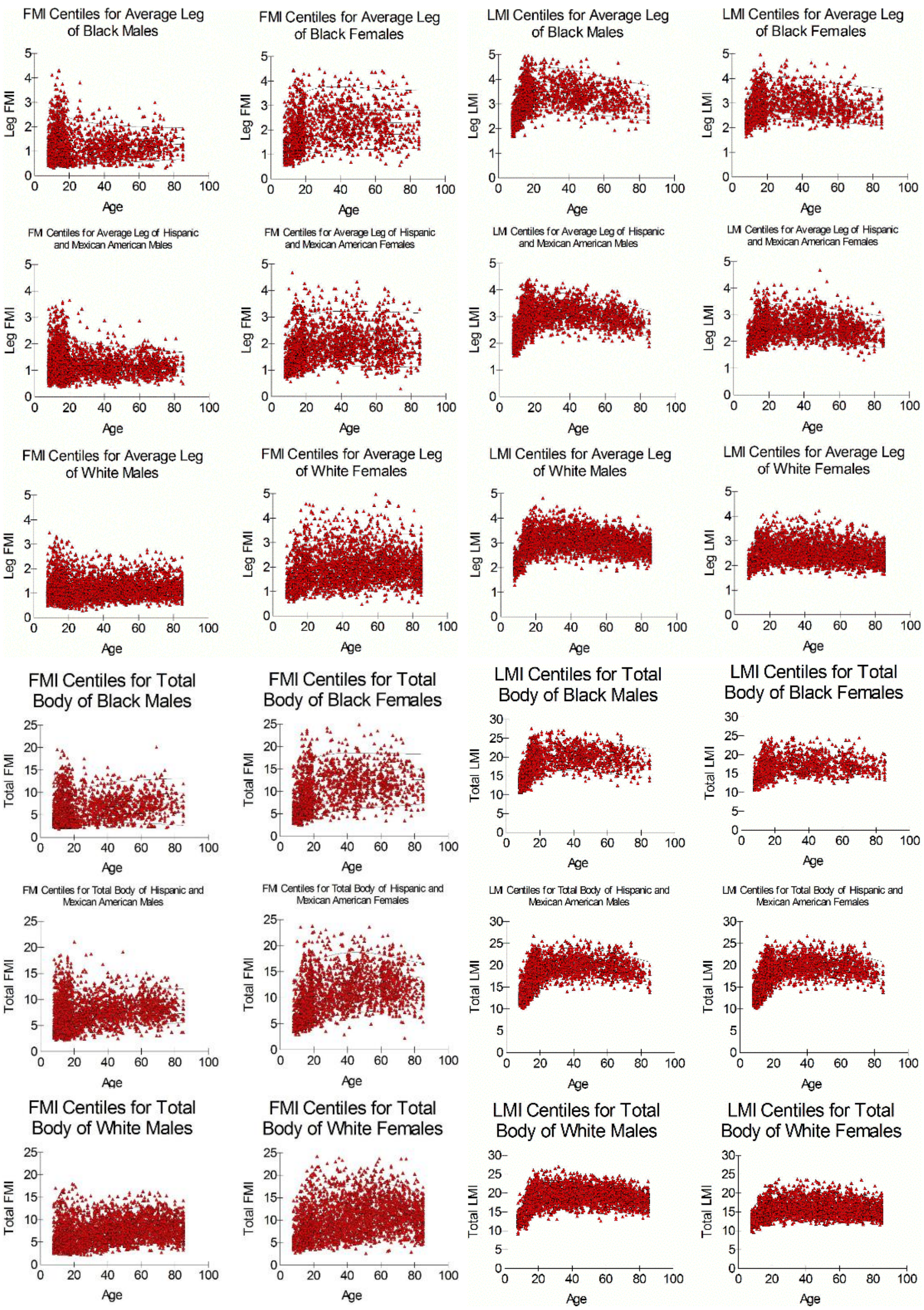


Figure 0-3: Distribution of FMI and LMI datapoints for the legs and total body and centiles by age, race, and sex

Table 0-1: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Females Ages 8-85 for Average Arm FMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.449	0.567	0.125	0.138	0.211	0.300	0.455	1.002	1.277
10	-0.346	0.544	0.146	0.161	0.248	0.351	0.519	1.023	1.239
12	-0.263	0.525	0.163	0.180	0.280	0.393	0.569	1.045	1.230
14	-0.192	0.509	0.178	0.197	0.307	0.428	0.610	1.065	1.232
16	-0.130	0.495	0.190	0.211	0.331	0.458	0.645	1.084	1.237
18	-0.076	0.483	0.202	0.224	0.352	0.485	0.675	1.102	1.244
20	-0.028	0.472	0.212	0.236	0.371	0.510	0.701	1.117	1.252
25	0.075	0.449	0.234	0.262	0.413	0.561	0.756	1.151	1.272
30	0.159	0.430	0.253	0.284	0.448	0.602	0.800	1.179	1.290
35	0.230	0.415	0.270	0.304	0.478	0.638	0.836	1.202	1.307
40	0.291	0.401	0.285	0.321	0.504	0.668	0.867	1.222	1.321
45	0.345	0.389	0.299	0.338	0.528	0.695	0.894	1.239	1.335
50	0.394	0.378	0.313	0.353	0.550	0.720	0.917	1.255	1.346
55	0.438	0.368	0.325	0.367	0.570	0.741	0.938	1.268	1.357
60	0.478	0.359	0.337	0.380	0.589	0.761	0.957	1.281	1.367
65	0.515	0.351	0.348	0.393	0.606	0.780	0.975	1.292	1.376
70	0.549	0.343	0.359	0.405	0.622	0.797	0.990	1.303	1.384
75	0.580	0.336	0.370	0.417	0.637	0.812	1.005	1.312	1.392
80	0.610	0.329	0.380	0.428	0.652	0.827	1.019	1.321	1.399
85	0.638	0.323	0.390	0.439	0.665	0.841	1.031	1.329	1.406

Table 0-2: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Males Ages 8-85 for Average Arm FMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.937	0.457	0.117	0.124	0.167	0.219	0.316	0.808	1.266
10	-0.880	0.454	0.121	0.129	0.174	0.229	0.327	0.771	1.111
12	-0.824	0.450	0.125	0.133	0.181	0.238	0.337	0.746	1.016
14	-0.769	0.447	0.129	0.138	0.188	0.247	0.348	0.729	0.953
16	-0.714	0.443	0.133	0.142	0.195	0.256	0.358	0.716	0.908
18	-0.660	0.440	0.137	0.146	0.202	0.264	0.368	0.707	0.876
20	-0.607	0.437	0.140	0.150	0.208	0.273	0.378	0.701	0.851
25	-0.477	0.428	0.149	0.160	0.225	0.294	0.401	0.694	0.813
30	-0.349	0.420	0.157	0.170	0.241	0.315	0.425	0.695	0.794
35	-0.223	0.412	0.164	0.179	0.256	0.336	0.447	0.700	0.786
40	-0.099	0.404	0.171	0.187	0.272	0.356	0.469	0.708	0.785
45	0.024	0.397	0.177	0.195	0.287	0.376	0.491	0.718	0.787
50	0.145	0.389	0.183	0.202	0.303	0.396	0.512	0.729	0.793
55	0.265	0.381	0.187	0.209	0.318	0.415	0.532	0.741	0.800
60	0.384	0.374	0.191	0.215	0.333	0.434	0.552	0.754	0.809
65	0.502	0.366	0.195	0.221	0.349	0.454	0.572	0.768	0.820
70	0.619	0.359	0.197	0.227	0.364	0.473	0.592	0.782	0.831
75	0.735	0.351	0.199	0.232	0.379	0.492	0.611	0.796	0.842
80	0.850	0.344	0.199	0.236	0.394	0.510	0.631	0.810	0.855
85	0.965	0.337	0.199	0.239	0.409	0.529	0.650	0.825	0.867

Table 0-3: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Females Ages 8-85 for Average Arm LMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.483	0.163	0.541	0.559	0.647	0.720	0.806	0.959	1.003
10	-0.483	0.163	0.589	0.609	0.704	0.784	0.878	1.045	1.092
12	-0.483	0.163	0.627	0.648	0.750	0.834	0.934	1.112	1.162
14	-0.483	0.163	0.655	0.678	0.784	0.872	0.976	1.162	1.215
16	-0.483	0.163	0.676	0.699	0.808	0.900	1.007	1.199	1.253
18	-0.483	0.163	0.691	0.715	0.827	0.920	1.030	1.226	1.282
20	-0.483	0.163	0.703	0.727	0.841	0.936	1.048	1.247	1.304
25	-0.483	0.163	0.722	0.747	0.863	0.961	1.076	1.280	1.339
30	-0.483	0.163	0.730	0.755	0.873	0.972	1.088	1.295	1.354
35	-0.483	0.163	0.732	0.757	0.876	0.974	1.091	1.298	1.358
40	-0.483	0.163	0.731	0.756	0.874	0.972	1.088	1.296	1.355
45	-0.483	0.163	0.727	0.752	0.869	0.968	1.083	1.289	1.348
50	-0.483	0.163	0.722	0.747	0.864	0.961	1.076	1.281	1.339
55	-0.483	0.163	0.717	0.742	0.858	0.954	1.068	1.272	1.330
60	-0.483	0.163	0.712	0.736	0.851	0.947	1.061	1.262	1.320
65	-0.483	0.163	0.706	0.731	0.845	0.940	1.052	1.253	1.310
70	-0.483	0.163	0.701	0.725	0.838	0.933	1.044	1.243	1.300
75	-0.483	0.163	0.695	0.719	0.832	0.926	1.036	1.233	1.290
80	-0.483	0.163	0.690	0.714	0.825	0.919	1.028	1.224	1.280
85	-0.483	0.163	0.685	0.709	0.819	0.912	1.021	1.215	1.271

Table 0-4: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Males Ages 8-85 for Average Arm LMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.429	0.156	0.552	0.571	0.657	0.728	0.810	0.955	0.996
10	-0.265	0.156	0.616	0.638	0.737	0.818	0.910	1.066	1.110
12	-0.131	0.156	0.699	0.724	0.839	0.932	1.036	1.210	1.257
14	-0.017	0.156	0.793	0.823	0.957	1.063	1.180	1.374	1.426
16	0.081	0.156	0.871	0.904	1.055	1.172	1.301	1.511	1.566
18	0.168	0.156	0.924	0.961	1.123	1.249	1.386	1.605	1.663
20	0.246	0.156	0.961	0.999	1.171	1.302	1.445	1.670	1.729
25	0.410	0.156	1.011	1.054	1.241	1.382	1.531	1.763	1.822
30	0.544	0.156	1.029	1.075	1.272	1.417	1.570	1.802	1.860
35	0.658	0.156	1.031	1.079	1.281	1.429	1.581	1.811	1.868
40	0.756	0.156	1.023	1.072	1.278	1.425	1.577	1.802	1.858
45	0.843	0.156	1.009	1.058	1.266	1.413	1.563	1.782	1.836
50	0.921	0.156	0.990	1.040	1.248	1.394	1.541	1.755	1.807
55	0.991	0.156	0.969	1.020	1.227	1.371	1.515	1.723	1.773
60	1.055	0.156	0.946	0.996	1.202	1.344	1.485	1.686	1.735
65	1.114	0.156	0.922	0.972	1.176	1.315	1.452	1.648	1.695
70	1.168	0.156	0.897	0.947	1.148	1.285	1.419	1.608	1.653
75	1.219	0.156	0.873	0.923	1.121	1.255	1.385	1.569	1.612
80	1.267	0.156	0.850	0.899	1.095	1.226	1.353	1.531	1.573
85	1.311	0.156	0.828	0.877	1.070	1.198	1.322	1.495	1.536

Table 0-5: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Females Ages 8-85 for Average Leg FMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.273	0.426	0.596	0.648	0.933	1.230	1.660	2.680	3.041
10	-0.190	0.416	0.679	0.740	1.071	1.407	1.876	2.927	3.280
12	-0.123	0.407	0.745	0.814	1.183	1.549	2.048	3.116	3.462
14	-0.066	0.400	0.801	0.876	1.277	1.668	2.189	3.266	3.606
16	-0.017	0.393	0.847	0.928	1.356	1.766	2.304	3.385	3.719
18	0.027	0.388	0.885	0.972	1.422	1.849	2.399	3.479	3.806
20	0.066	0.383	0.917	1.008	1.478	1.918	2.477	3.553	3.874
25	0.148	0.372	0.977	1.077	1.585	2.046	2.617	3.676	3.982
30	0.216	0.363	1.017	1.123	1.656	2.130	2.704	3.737	4.029
35	0.273	0.356	1.043	1.154	1.704	2.184	2.756	3.760	4.040
40	0.322	0.350	1.060	1.175	1.737	2.220	2.786	3.762	4.030
45	0.365	0.344	1.072	1.190	1.761	2.243	2.803	3.752	4.010
50	0.404	0.339	1.081	1.201	1.778	2.259	2.811	3.736	3.984
55	0.440	0.334	1.087	1.210	1.791	2.271	2.815	3.716	3.957
60	0.472	0.330	1.093	1.216	1.801	2.279	2.816	3.696	3.929
65	0.501	0.326	1.097	1.222	1.810	2.285	2.815	3.676	3.902
70	0.529	0.323	1.101	1.228	1.817	2.290	2.814	3.657	3.878
75	0.554	0.320	1.104	1.232	1.824	2.294	2.812	3.639	3.855
80	0.578	0.317	1.108	1.237	1.830	2.298	2.810	3.623	3.833
85	0.600	0.314	1.111	1.241	1.836	2.302	2.809	3.608	3.814

Table 0-6: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Males Ages 8-85 for Average Leg FMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.531	0.519	0.407	0.442	0.649	0.894	1.317	2.791	3.543
10	-0.487	0.513	0.411	0.447	0.659	0.907	1.324	2.686	3.335
12	-0.444	0.506	0.415	0.452	0.669	0.919	1.330	2.599	3.168
14	-0.402	0.500	0.419	0.457	0.678	0.931	1.337	2.525	3.032
16	-0.360	0.494	0.423	0.462	0.688	0.942	1.343	2.463	2.918
18	-0.319	0.487	0.427	0.467	0.698	0.954	1.350	2.409	2.823
20	-0.279	0.481	0.430	0.472	0.708	0.965	1.356	2.361	2.740
25	-0.179	0.466	0.440	0.484	0.732	0.993	1.373	2.267	2.579
30	-0.081	0.452	0.449	0.496	0.756	1.021	1.389	2.196	2.461
35	0.015	0.437	0.458	0.508	0.780	1.048	1.406	2.142	2.373
40	0.110	0.423	0.467	0.521	0.804	1.074	1.423	2.101	2.303
45	0.204	0.409	0.477	0.534	0.829	1.101	1.439	2.067	2.249
50	0.297	0.395	0.487	0.547	0.854	1.127	1.456	2.041	2.205
55	0.388	0.381	0.497	0.562	0.879	1.152	1.472	2.020	2.170
60	0.479	0.368	0.509	0.577	0.905	1.178	1.489	2.003	2.140
65	0.570	0.354	0.521	0.593	0.931	1.203	1.505	1.990	2.116
70	0.659	0.341	0.534	0.610	0.958	1.228	1.521	1.979	2.096
75	0.748	0.327	0.549	0.629	0.985	1.253	1.537	1.970	2.080
80	0.836	0.314	0.566	0.650	1.013	1.278	1.553	1.964	2.066
85	0.924	0.301	0.585	0.672	1.041	1.303	1.569	1.958	2.054

Table 0-7: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Females Ages 8-85 for Average Leg LMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.438	0.163	1.764	1.825	2.113	2.353	2.633	3.130	3.271
10	-0.438	0.163	1.915	1.981	2.293	2.554	2.858	3.398	3.551
12	-0.438	0.163	2.028	2.099	2.429	2.705	3.028	3.599	3.761
14	-0.438	0.163	2.109	2.182	2.526	2.813	3.148	3.742	3.911
16	-0.438	0.163	2.165	2.240	2.593	2.888	3.232	3.842	4.015
18	-0.438	0.163	2.205	2.281	2.641	2.940	3.291	3.912	4.088
20	-0.438	0.163	2.234	2.311	2.676	2.979	3.334	3.964	4.142
25	-0.438	0.163	2.275	2.354	2.725	3.034	3.396	4.037	4.219
30	-0.438	0.163	2.283	2.363	2.735	3.045	3.409	4.052	4.235
35	-0.438	0.163	2.271	2.350	2.720	3.028	3.390	4.029	4.211
40	-0.438	0.163	2.247	2.325	2.692	2.997	3.355	3.988	4.168
45	-0.438	0.163	2.219	2.296	2.658	2.960	3.313	3.938	4.116
50	-0.438	0.163	2.190	2.266	2.623	2.921	3.269	3.886	4.061
55	-0.438	0.163	2.161	2.236	2.589	2.882	3.226	3.835	4.008
60	-0.438	0.163	2.134	2.208	2.556	2.846	3.186	3.787	3.957
65	-0.438	0.163	2.108	2.182	2.526	2.812	3.147	3.741	3.910
70	-0.438	0.163	2.084	2.157	2.497	2.780	3.111	3.699	3.865
75	-0.438	0.163	2.061	2.133	2.469	2.749	3.077	3.658	3.823
80	-0.438	0.163	2.040	2.111	2.444	2.721	3.045	3.620	3.783
85	-0.438	0.163	2.019	2.090	2.419	2.694	3.015	3.584	3.745

Table 0-8: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Males Ages 8-85 for Average Leg LMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	0.118	0.147	1.668	1.729	2.001	2.211	2.440	2.807	2.903
10	0.118	0.147	1.937	2.008	2.323	2.567	2.833	3.259	3.371
12	0.118	0.147	2.169	2.248	2.601	2.874	3.172	3.649	3.774
14	0.118	0.147	2.357	2.443	2.826	3.123	3.447	3.965	4.101
16	0.118	0.147	2.485	2.576	2.980	3.293	3.634	4.181	4.325
18	0.118	0.147	2.562	2.656	3.073	3.396	3.748	4.311	4.459
20	0.118	0.147	2.609	2.704	3.129	3.457	3.816	4.390	4.540
25	0.118	0.147	2.661	2.758	3.191	3.526	3.892	4.477	4.631
30	0.118	0.147	2.664	2.761	3.195	3.530	3.896	4.482	4.636
35	0.118	0.147	2.643	2.739	3.169	3.502	3.865	4.447	4.599
40	0.118	0.147	2.609	2.704	3.129	3.457	3.815	4.389	4.540
45	0.118	0.147	2.568	2.662	3.080	3.403	3.756	4.321	4.469
50	0.118	0.147	2.526	2.618	3.029	3.347	3.694	4.250	4.395
55	0.118	0.147	2.482	2.573	2.977	3.289	3.630	4.176	4.320
60	0.118	0.147	2.438	2.527	2.924	3.231	3.566	4.103	4.244
65	0.118	0.147	2.395	2.482	2.872	3.174	3.503	4.030	4.168
70	0.118	0.147	2.352	2.438	2.821	3.117	3.440	3.958	4.093
75	0.118	0.147	2.310	2.395	2.771	3.062	3.379	3.888	4.021
80	0.118	0.147	2.271	2.354	2.723	3.009	3.321	3.821	3.952
85	0.118	0.147	2.233	2.315	2.678	2.959	3.266	3.758	3.887

Table 0-9: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Females Ages 8-85 for Trunk FMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.516	0.586	0.682	0.746	1.139	1.633	2.540	6.200	8.345
10	-0.419	0.568	0.833	0.916	1.413	2.015	3.059	6.594	8.308
12	-0.333	0.552	0.970	1.070	1.666	2.365	3.517	6.968	8.447
14	-0.255	0.537	1.097	1.215	1.907	2.696	3.941	7.334	8.668
16	-0.184	0.523	1.216	1.351	2.135	3.005	4.328	7.674	8.905
18	-0.117	0.511	1.324	1.477	2.348	3.291	4.678	7.975	9.126
20	-0.056	0.499	1.423	1.592	2.547	3.554	4.992	8.238	9.324
25	0.085	0.473	1.632	1.840	2.979	4.114	5.634	8.738	9.698
30	0.209	0.449	1.792	2.035	3.322	4.542	6.092	9.032	9.892
35	0.321	0.428	1.913	2.186	3.588	4.857	6.400	9.164	9.941
40	0.424	0.409	2.009	2.308	3.799	5.092	6.607	9.199	9.904
45	0.519	0.391	2.088	2.411	3.969	5.268	6.742	9.171	9.815
50	0.609	0.374	2.156	2.499	4.106	5.397	6.821	9.097	9.688
55	0.693	0.358	2.216	2.575	4.212	5.484	6.854	8.984	9.528
60	0.773	0.343	2.269	2.641	4.293	5.537	6.848	8.841	9.342
65	0.848	0.328	2.318	2.700	4.352	5.560	6.811	8.674	9.137
70	0.921	0.315	2.364	2.752	4.391	5.559	6.748	8.489	8.917
75	0.990	0.302	2.409	2.800	4.416	5.541	6.668	8.295	8.691
80	1.057	0.289	2.456	2.848	4.432	5.512	6.580	8.102	8.470
85	1.121	0.277	2.508	2.898	4.446	5.482	6.494	7.920	8.262

Table 0-10: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Males Ages 8-85 for Trunk FMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-1.120	0.458	0.680	0.720	0.954	1.244	1.817	6.541	25.065
10	-1.011	0.458	0.704	0.748	1.001	1.309	1.895	5.410	9.926
12	-0.912	0.458	0.728	0.774	1.047	1.374	1.975	4.914	7.439
14	-0.820	0.458	0.753	0.803	1.097	1.445	2.062	4.672	6.449
16	-0.736	0.458	0.782	0.837	1.154	1.524	2.163	4.571	5.971
18	-0.656	0.458	0.816	0.876	1.219	1.615	2.281	4.568	5.751
20	-0.580	0.458	0.856	0.921	1.295	1.721	2.419	4.636	5.684
25	-0.407	0.458	0.966	1.048	1.512	2.022	2.812	4.973	5.843
30	-0.250	0.458	1.065	1.165	1.726	2.323	3.204	5.356	6.133
35	-0.106	0.458	1.139	1.258	1.917	2.597	3.555	5.697	6.412
40	0.028	0.458	1.186	1.324	2.080	2.836	3.857	5.979	6.646
45	0.154	0.458	1.206	1.365	2.216	3.041	4.112	6.204	6.830
50	0.273	0.458	1.205	1.384	2.331	3.218	4.328	6.385	6.977
55	0.387	0.458	1.183	1.384	2.428	3.373	4.515	6.534	7.096
60	0.495	0.458	1.142	1.367	2.511	3.510	4.679	6.661	7.195
65	0.599	0.458	1.082	1.334	2.584	3.635	4.825	6.770	7.282
70	0.699	0.458	1.003	1.286	2.648	3.749	4.959	6.869	7.360
75	0.795	0.458	0.901	1.221	2.705	3.855	5.081	6.958	7.432
80	0.888	0.458	0.773	1.137	2.757	3.955	5.196	7.042	7.500
85	0.979	0.458	0.609	1.033	2.805	4.051	5.306	7.124	7.568

Table 0-11: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Females Ages 8-85 for Trunk LMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.472	0.131	4.429	4.554	5.127	5.589	6.115	7.012	7.259
10	-0.472	0.131	4.933	5.072	5.710	6.225	6.811	7.809	8.085
12	-0.472	0.131	5.325	5.475	6.163	6.719	7.352	8.429	8.726
14	-0.472	0.131	5.618	5.776	6.503	7.089	7.757	8.894	9.207
16	-0.472	0.131	5.832	5.996	6.751	7.359	8.052	9.232	9.558
18	-0.472	0.131	5.990	6.159	6.934	7.559	8.271	9.483	9.817
20	-0.472	0.131	6.110	6.283	7.073	7.711	8.437	9.674	10.014
25	-0.472	0.131	6.306	6.484	7.300	7.958	8.707	9.984	10.335
30	-0.472	0.131	6.410	6.591	7.420	8.089	8.850	10.148	10.505
35	-0.472	0.131	6.463	6.645	7.481	8.156	8.924	10.231	10.592
40	-0.472	0.131	6.490	6.673	7.513	8.190	8.961	10.275	10.636
45	-0.472	0.131	6.503	6.686	7.527	8.206	8.979	10.295	10.657
50	-0.472	0.131	6.507	6.691	7.532	8.211	8.985	10.301	10.664
55	-0.472	0.131	6.506	6.690	7.531	8.210	8.983	10.300	10.663
60	-0.472	0.131	6.503	6.686	7.527	8.206	8.979	10.295	10.657
65	-0.472	0.131	6.497	6.681	7.521	8.199	8.971	10.286	10.649
70	-0.472	0.131	6.491	6.674	7.514	8.191	8.963	10.276	10.638
75	-0.472	0.131	6.484	6.667	7.506	8.183	8.953	10.265	10.627
80	-0.472	0.131	6.477	6.660	7.498	8.174	8.944	10.254	10.616
85	-0.472	0.131	6.471	6.653	7.490	8.165	8.934	10.244	10.605

Table 0-12: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Males Ages 8-85 for Trunk LMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	0.110	0.132	4.317	4.458	5.080	5.557	6.072	6.890	7.103
10	0.110	0.132	4.729	4.883	5.564	6.086	6.651	7.547	7.780
12	0.110	0.132	5.218	5.388	6.141	6.717	7.340	8.328	8.586
14	0.110	0.132	5.779	5.967	6.800	7.437	8.128	9.222	9.507
16	0.110	0.132	6.237	6.440	7.339	8.027	8.772	9.953	10.261
18	0.110	0.132	6.538	6.751	7.694	8.415	9.196	10.435	10.757
20	0.110	0.132	6.737	6.956	7.927	8.671	9.476	10.752	11.084
25	0.110	0.132	7.029	7.258	8.271	9.046	9.886	11.217	11.564
30	0.110	0.132	7.157	7.390	8.422	9.212	10.067	11.422	11.776
35	0.110	0.132	7.213	7.448	8.488	9.284	10.145	11.512	11.868
40	0.110	0.132	7.239	7.475	8.518	9.317	10.182	11.553	11.910
45	0.110	0.132	7.244	7.480	8.524	9.323	10.189	11.561	11.918
50	0.110	0.132	7.234	7.469	8.512	9.310	10.174	11.544	11.901
55	0.110	0.132	7.207	7.441	8.480	9.275	10.136	11.501	11.857
60	0.110	0.132	7.164	7.397	8.430	9.221	10.077	11.434	11.787
65	0.110	0.132	7.108	7.340	8.364	9.149	9.998	11.344	11.695
70	0.110	0.132	7.041	7.270	8.286	9.062	9.903	11.237	11.585
75	0.110	0.132	6.969	7.196	8.201	8.970	9.802	11.122	11.466
80	0.110	0.132	6.897	7.122	8.117	8.878	9.701	11.008	11.348
85	0.110	0.132	6.829	7.051	8.036	8.789	9.605	10.899	11.236

Table 0-13: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Females Ages 8-85 for Total Body FMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.602	0.493	2.445	2.639	3.773	5.108	7.399	15.529	19.823
10	-0.458	0.472	2.862	3.102	4.478	6.025	8.495	15.710	18.805
12	-0.340	0.454	3.205	3.487	5.071	6.786	9.377	16.072	18.623
14	-0.241	0.440	3.501	3.821	5.590	7.442	10.121	16.472	18.712
16	-0.154	0.427	3.759	4.114	6.047	8.013	10.755	16.850	18.884
18	-0.078	0.416	3.984	4.371	6.449	8.509	11.295	17.183	19.068
20	-0.011	0.406	4.182	4.599	6.806	8.942	11.758	17.469	19.238
25	0.133	0.384	4.584	5.065	7.534	9.807	12.651	17.996	19.557
30	0.251	0.367	4.889	5.422	8.083	10.435	13.266	18.309	19.726
35	0.350	0.352	5.126	5.701	8.506	10.899	13.692	18.475	19.781
40	0.437	0.340	5.320	5.930	8.841	11.253	13.996	18.552	19.769
45	0.513	0.328	5.485	6.124	9.115	11.530	14.220	18.577	19.721
50	0.581	0.318	5.630	6.294	9.346	11.753	14.388	18.569	19.652
55	0.642	0.309	5.761	6.447	9.544	11.938	14.519	18.543	19.574
60	0.698	0.301	5.883	6.587	9.718	12.096	14.623	18.508	19.493
65	0.750	0.293	5.999	6.718	9.875	12.232	14.710	18.469	19.414
70	0.798	0.286	6.110	6.842	10.018	12.354	14.784	18.429	19.339
75	0.842	0.280	6.218	6.962	10.150	12.464	14.849	18.392	19.270
80	0.884	0.274	6.323	7.078	10.274	12.566	14.908	18.357	19.207
85	0.923	0.268	6.427	7.190	10.391	12.661	14.962	18.326	19.151

Table 0-14: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Males Ages 8-85 for Total Body FMI.

	Males									
			M							
Age	L	S	3	5	25	50	75	95	97
8	-1.125	0.410	2.302	2.428	3.153	4.009	5.581	14.167	24.111
10	-1.016	0.410	2.297	2.429	3.182	4.058	5.612	12.643	18.263
12	-0.918	0.410	2.292	2.430	3.210	4.106	5.646	11.733	15.656
14	-0.827	0.410	2.291	2.434	3.243	4.158	5.689	11.140	14.175
16	-0.743	0.410	2.299	2.449	3.290	4.229	5.761	10.769	13.272
18	-0.663	0.410	2.325	2.482	3.362	4.333	5.878	10.585	12.747
20	-0.588	0.410	2.369	2.536	3.463	4.473	6.046	10.554	12.489
25	-0.416	0.410	2.525	2.719	3.793	4.926	6.605	10.863	12.474
30	-0.260	0.410	2.674	2.901	4.135	5.398	7.189	11.325	12.757
35	-0.117	0.410	2.783	3.043	4.436	5.821	7.708	11.746	13.054
40	0.017	0.410	2.845	3.137	4.685	6.179	8.138	12.077	13.287
45	0.142	0.410	2.862	3.188	4.884	6.473	8.486	12.320	13.449
50	0.260	0.410	2.843	3.203	5.046	6.720	8.773	12.503	13.563
55	0.373	0.410	2.795	3.190	5.180	6.933	9.017	12.648	13.648
60	0.481	0.410	2.720	3.154	5.296	7.122	9.230	12.768	13.718
65	0.584	0.410	2.619	3.097	5.397	7.293	9.420	12.873	13.779
70	0.684	0.410	2.493	3.018	5.484	7.448	9.591	12.964	13.831
75	0.780	0.410	2.337	2.918	5.561	7.589	9.745	13.044	13.876
80	0.872	0.410	2.149	2.794	5.629	7.720	9.887	13.116	13.918
85	0.962	0.410	1.923	2.647	5.691	7.845	10.022	13.189	13.963

Table 0-15: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Females Ages 8-85 for Total Body LMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.898	0.126	10.556	10.821	12.062	13.095	14.309	16.485	17.112
10	-0.681	0.126	11.434	11.734	13.120	14.254	15.562	17.838	18.476
12	-0.503	0.126	12.453	12.792	14.342	15.590	17.009	19.421	20.085
14	-0.353	0.126	13.561	13.942	15.668	17.041	18.581	21.150	21.846
16	-0.223	0.126	14.431	14.848	16.722	18.196	19.831	22.514	23.233
18	-0.108	0.126	14.978	15.421	17.403	18.943	20.637	23.379	24.105
20	-0.005	0.126	15.322	15.786	17.846	19.434	21.163	23.930	24.656
25	0.212	0.126	15.785	16.288	18.488	20.149	21.925	24.697	25.411
30	0.390	0.126	15.916	16.445	18.732	20.429	22.215	24.952	25.646
35	0.540	0.126	15.881	16.430	18.774	20.486	22.266	24.950	25.623
40	0.670	0.126	15.771	16.334	18.719	20.436	22.202	24.829	25.482
45	0.785	0.126	15.620	16.195	18.608	20.324	22.072	24.643	25.277
50	0.888	0.126	15.450	16.035	18.470	20.181	21.909	24.426	25.042
55	0.981	0.126	15.257	15.851	18.300	20.003	21.710	24.172	24.771
60	1.065	0.126	15.043	15.644	18.100	19.791	21.473	23.882	24.465
65	1.143	0.126	14.811	15.416	17.872	19.549	21.205	23.560	24.127
70	1.216	0.126	14.562	15.171	17.623	19.282	20.911	23.211	23.763
75	1.283	0.126	14.310	14.921	17.366	19.006	20.607	22.854	23.391
80	1.346	0.126	14.066	14.679	17.114	18.735	20.310	22.507	23.029
85	1.405	0.126	13.833	14.447	16.874	18.477	20.026	22.176	22.686

Table 0-16: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Black Males Ages 8-85 for Total Body LMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.898	0.126	10.556	10.821	12.062	13.095	14.309	16.485	17.112
10	-0.681	0.126	11.434	11.734	13.120	14.254	15.562	17.838	18.476
12	-0.503	0.126	12.453	12.792	14.342	15.590	17.009	19.421	20.085
14	-0.353	0.126	13.561	13.942	15.668	17.041	18.581	21.150	21.846
16	-0.223	0.126	14.431	14.848	16.722	18.196	19.831	22.514	23.233
18	-0.108	0.126	14.978	15.421	17.403	18.943	20.637	23.379	24.105
20	-0.005	0.126	15.322	15.786	17.846	19.434	21.163	23.930	24.656
25	0.212	0.126	15.785	16.288	18.488	20.149	21.925	24.697	25.411
30	0.390	0.126	15.916	16.445	18.732	20.429	22.215	24.952	25.646
35	0.540	0.126	15.881	16.430	18.774	20.486	22.266	24.950	25.623
40	0.670	0.126	15.771	16.334	18.719	20.436	22.202	24.829	25.482
45	0.785	0.126	15.620	16.195	18.608	20.324	22.072	24.643	25.277
50	0.888	0.126	15.450	16.035	18.470	20.181	21.909	24.426	25.042
55	0.981	0.126	15.257	15.851	18.300	20.003	21.710	24.172	24.771
60	1.065	0.126	15.043	15.644	18.100	19.791	21.473	23.882	24.465
65	1.143	0.126	14.811	15.416	17.872	19.549	21.205	23.560	24.127
70	1.216	0.126	14.562	15.171	17.623	19.282	20.911	23.211	23.763
75	1.283	0.126	14.310	14.921	17.366	19.006	20.607	22.854	23.391
80	1.346	0.126	14.066	14.679	17.114	18.735	20.310	22.507	23.029
85	1.405	0.126	13.833	14.447	16.874	18.477	20.026	22.176	22.686

Table 0-17: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Females Ages 8-85 for Average Arm FMI.

	Females								
			M						
Age	M	S	3	5	25	50	75	95	97
8	0.314	0.454	0.151	0.164	0.235	0.314	0.435	0.769	0.904
10	0.366	0.440	0.176	0.192	0.276	0.366	0.500	0.837	0.963
12	0.410	0.428	0.197	0.215	0.310	0.410	0.553	0.892	1.012
14	0.447	0.418	0.216	0.235	0.340	0.447	0.598	0.937	1.053
16	0.480	0.409	0.232	0.253	0.367	0.480	0.636	0.977	1.089
18	0.510	0.401	0.247	0.270	0.391	0.510	0.671	1.011	1.120
20	0.536	0.394	0.260	0.284	0.412	0.536	0.701	1.040	1.147
25	0.590	0.380	0.288	0.315	0.457	0.590	0.762	1.099	1.200
30	0.633	0.368	0.310	0.340	0.493	0.633	0.809	1.141	1.238
35	0.667	0.358	0.328	0.360	0.522	0.667	0.845	1.172	1.266
40	0.694	0.349	0.344	0.378	0.546	0.694	0.874	1.196	1.287
45	0.717	0.341	0.357	0.392	0.566	0.717	0.897	1.214	1.302
50	0.736	0.334	0.369	0.405	0.584	0.736	0.916	1.228	1.313
55	0.753	0.328	0.379	0.416	0.598	0.753	0.932	1.238	1.322
60	0.767	0.322	0.388	0.426	0.612	0.767	0.945	1.246	1.328
65	0.779	0.317	0.396	0.435	0.623	0.779	0.957	1.253	1.333
70	0.790	0.312	0.403	0.444	0.634	0.790	0.967	1.259	1.337
75	0.800	0.308	0.411	0.452	0.644	0.800	0.976	1.264	1.341
80	0.810	0.303	0.417	0.459	0.653	0.810	0.985	1.269	1.344
85	0.819	0.299	0.424	0.466	0.662	0.819	0.993	1.273	1.347

Table 0-18: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Males Ages 8-85 for Average Arm FMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.174	0.579	0.119	0.134	0.221	0.323	0.484	0.915	1.081
10	-0.174	0.544	0.126	0.141	0.226	0.323	0.472	0.855	0.998
12	-0.174	0.515	0.132	0.146	0.230	0.322	0.461	0.806	0.931
14	-0.174	0.489	0.137	0.152	0.234	0.323	0.453	0.767	0.878
16	-0.174	0.465	0.144	0.159	0.240	0.326	0.450	0.741	0.843
18	-0.174	0.444	0.153	0.167	0.249	0.333	0.453	0.727	0.821
20	-0.174	0.425	0.162	0.177	0.259	0.342	0.459	0.722	0.810
25	-0.174	0.386	0.186	0.202	0.286	0.369	0.482	0.723	0.802
30	-0.174	0.357	0.208	0.224	0.310	0.393	0.502	0.730	0.802
35	-0.174	0.336	0.226	0.242	0.329	0.411	0.518	0.735	0.803
40	-0.174	0.321	0.240	0.257	0.344	0.425	0.530	0.739	0.804
45	-0.174	0.309	0.251	0.268	0.355	0.436	0.539	0.742	0.804
50	-0.174	0.299	0.260	0.277	0.365	0.445	0.546	0.744	0.804
55	-0.174	0.291	0.268	0.285	0.372	0.452	0.552	0.745	0.804
60	-0.174	0.285	0.274	0.291	0.379	0.457	0.556	0.746	0.803
65	-0.174	0.280	0.279	0.297	0.384	0.462	0.560	0.746	0.802
70	-0.174	0.275	0.284	0.302	0.388	0.466	0.562	0.746	0.801
75	-0.174	0.270	0.288	0.306	0.392	0.469	0.565	0.746	0.799
80	-0.174	0.266	0.292	0.310	0.396	0.473	0.567	0.745	0.798
85	-0.174	0.262	0.296	0.314	0.399	0.475	0.569	0.744	0.796

Table 0-19: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Females Ages 8-85 for Average Arm LMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-1.204	0.162	0.487	0.501	0.570	0.632	0.710	0.871	0.924
10	-1.073	0.162	0.525	0.541	0.617	0.684	0.768	0.936	0.989
12	-0.962	0.162	0.556	0.573	0.655	0.726	0.815	0.988	1.042
14	-0.865	0.162	0.580	0.598	0.684	0.760	0.852	1.028	1.082
16	-0.779	0.162	0.598	0.617	0.707	0.786	0.881	1.059	1.113
18	-0.701	0.162	0.612	0.632	0.726	0.807	0.904	1.084	1.137
20	-0.629	0.162	0.624	0.644	0.742	0.824	0.923	1.104	1.156
25	-0.473	0.162	0.645	0.667	0.771	0.858	0.959	1.140	1.192
30	-0.339	0.162	0.658	0.682	0.790	0.880	0.983	1.163	1.214
35	-0.223	0.162	0.666	0.690	0.802	0.894	0.998	1.177	1.226
40	-0.119	0.162	0.668	0.693	0.809	0.901	1.006	1.182	1.230
45	-0.025	0.162	0.667	0.693	0.810	0.903	1.008	1.181	1.227
50	0.061	0.162	0.662	0.689	0.807	0.901	1.004	1.173	1.218
55	0.140	0.162	0.655	0.682	0.801	0.894	0.997	1.162	1.206
60	0.214	0.162	0.646	0.673	0.793	0.886	0.987	1.148	1.190
65	0.283	0.162	0.636	0.663	0.783	0.875	0.974	1.132	1.172
70	0.348	0.162	0.626	0.653	0.773	0.864	0.961	1.114	1.154
75	0.409	0.162	0.615	0.642	0.762	0.852	0.948	1.097	1.135
80	0.467	0.162	0.605	0.632	0.751	0.840	0.935	1.080	1.117
85	0.522	0.162	0.595	0.622	0.741	0.829	0.922	1.064	1.100

Table 0-20: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Males Ages 8-85 for Average Arm LMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	0.175	0.162	0.498	0.519	0.610	0.681	0.759	0.883	0.916
10	0.175	0.158	0.553	0.575	0.674	0.750	0.834	0.967	1.002
12	0.175	0.155	0.637	0.662	0.773	0.858	0.952	1.101	1.140
14	0.175	0.152	0.740	0.768	0.894	0.991	1.097	1.266	1.310
16	0.175	0.150	0.826	0.857	0.996	1.102	1.218	1.403	1.451
18	0.175	0.147	0.883	0.916	1.062	1.174	1.296	1.489	1.539
20	0.175	0.146	0.921	0.955	1.105	1.220	1.344	1.542	1.594
25	0.175	0.142	0.978	1.013	1.167	1.285	1.413	1.615	1.667
30	0.175	0.139	1.007	1.042	1.197	1.315	1.443	1.644	1.697
35	0.175	0.136	1.023	1.058	1.212	1.329	1.455	1.654	1.706
40	0.175	0.133	1.030	1.064	1.216	1.331	1.455	1.651	1.702
45	0.175	0.131	1.029	1.063	1.211	1.325	1.446	1.638	1.687
50	0.175	0.130	1.021	1.054	1.200	1.310	1.429	1.615	1.663
55	0.175	0.128	1.008	1.040	1.181	1.288	1.403	1.584	1.631
60	0.175	0.126	0.989	1.020	1.157	1.261	1.372	1.547	1.592
65	0.175	0.125	0.966	0.997	1.129	1.229	1.336	1.503	1.547
70	0.175	0.124	0.938	0.967	1.094	1.190	1.293	1.453	1.495
75	0.175	0.122	0.907	0.935	1.056	1.148	1.246	1.399	1.438
80	0.175	0.121	0.875	0.901	1.017	1.104	1.198	1.344	1.381
85	0.175	0.120	0.844	0.869	0.979	1.063	1.152	1.291	1.326

Table 0-21: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Females Ages 8-85 for Average Leg FMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.710	0.309	0.734	0.774	0.983	1.195	1.497	2.247	2.532
10	-0.590	0.309	0.813	0.859	1.101	1.341	1.675	2.456	2.736
12	-0.492	0.309	0.875	0.926	1.196	1.459	1.818	2.621	2.897
14	-0.409	0.309	0.923	0.980	1.274	1.556	1.935	2.754	3.026
16	-0.338	0.309	0.962	1.023	1.337	1.636	2.031	2.860	3.128
18	-0.274	0.309	0.992	1.057	1.389	1.702	2.109	2.944	3.209
20	-0.218	0.309	1.015	1.083	1.432	1.756	2.173	3.011	3.271
25	-0.098	0.309	1.052	1.127	1.507	1.852	2.287	3.122	3.373
30	0.000	0.309	1.069	1.150	1.553	1.913	2.357	3.183	3.424
35	0.083	0.309	1.075	1.160	1.581	1.952	2.400	3.213	3.446
40	0.155	0.309	1.074	1.163	1.599	1.977	2.427	3.226	3.451
45	0.218	0.309	1.069	1.162	1.609	1.992	2.443	3.228	3.446
50	0.275	0.309	1.062	1.157	1.615	2.002	2.452	3.224	3.435
55	0.326	0.309	1.053	1.150	1.617	2.007	2.456	3.215	3.420
60	0.373	0.309	1.043	1.143	1.618	2.010	2.458	3.204	3.404
65	0.416	0.309	1.033	1.136	1.618	2.012	2.457	3.193	3.388
70	0.456	0.309	1.023	1.128	1.617	2.013	2.456	3.181	3.373
75	0.493	0.309	1.014	1.120	1.615	2.013	2.455	3.171	3.358
80	0.527	0.309	1.005	1.113	1.614	2.013	2.454	3.161	3.345
85	0.560	0.309	0.996	1.106	1.613	2.014	2.453	3.151	3.333

Table 0-22: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Males Ages 8-85 for Average Leg FMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.051	0.502	0.446	0.499	0.802	1.121	1.578	2.608	2.953
10	-0.051	0.478	0.465	0.519	0.815	1.121	1.552	2.502	2.815
12	-0.051	0.458	0.482	0.535	0.825	1.121	1.531	2.418	2.707
14	-0.051	0.441	0.497	0.550	0.835	1.121	1.513	2.350	2.619
16	-0.051	0.427	0.510	0.562	0.843	1.121	1.498	2.292	2.546
18	-0.051	0.414	0.522	0.574	0.850	1.121	1.485	2.243	2.482
20	-0.051	0.403	0.533	0.585	0.857	1.121	1.474	2.200	2.427
25	-0.051	0.378	0.557	0.608	0.870	1.121	1.450	2.111	2.315
30	-0.051	0.358	0.578	0.627	0.882	1.121	1.430	2.041	2.227
35	-0.051	0.342	0.596	0.644	0.892	1.121	1.414	1.984	2.156
40	-0.051	0.327	0.612	0.659	0.901	1.121	1.400	1.935	2.096
45	-0.051	0.314	0.626	0.673	0.908	1.121	1.388	1.894	2.044
50	-0.051	0.303	0.639	0.686	0.915	1.121	1.377	1.858	1.999
55	-0.051	0.293	0.652	0.697	0.922	1.121	1.367	1.825	1.960
60	-0.051	0.283	0.663	0.708	0.927	1.121	1.358	1.797	1.924
65	-0.051	0.274	0.674	0.718	0.933	1.121	1.350	1.770	1.892
70	-0.051	0.266	0.684	0.727	0.938	1.121	1.343	1.747	1.863
75	-0.051	0.259	0.693	0.736	0.943	1.121	1.336	1.725	1.836
80	-0.051	0.252	0.702	0.744	0.947	1.121	1.330	1.704	1.811
85	-0.051	0.245	0.711	0.752	0.951	1.121	1.324	1.686	1.788

Table 0-23: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Females Ages 8-85 for Average Leg LMI.

	Females								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-1.357	0.156	1.526	1.568	1.772	1.955	2.191	2.683	2.847
10	-1.197	0.156	1.661	1.708	1.936	2.138	2.393	2.908	3.073
12	-1.067	0.156	1.761	1.812	2.059	2.275	2.545	3.074	3.239
14	-0.956	0.156	1.831	1.885	2.146	2.373	2.652	3.189	3.353
16	-0.861	0.156	1.879	1.936	2.208	2.443	2.728	3.267	3.430
18	-0.776	0.156	1.913	1.972	2.253	2.493	2.783	3.323	3.483
20	-0.701	0.156	1.937	1.998	2.287	2.531	2.825	3.362	3.520
25	-0.541	0.156	1.970	2.035	2.337	2.589	2.886	3.416	3.568
30	-0.411	0.156	1.980	2.047	2.358	2.614	2.912	3.431	3.577
35	-0.301	0.156	1.975	2.043	2.360	2.618	2.915	3.423	3.564
40	-0.205	0.156	1.960	2.030	2.350	2.608	2.902	3.398	3.534
45	-0.121	0.156	1.937	2.008	2.330	2.587	2.877	3.360	3.491
50	-0.046	0.156	1.909	1.980	2.302	2.557	2.843	3.313	3.440
55	0.023	0.156	1.877	1.949	2.270	2.522	2.803	3.260	3.382
60	0.085	0.156	1.844	1.916	2.235	2.485	2.760	3.206	3.323
65	0.142	0.156	1.812	1.883	2.200	2.447	2.717	3.151	3.265
70	0.195	0.156	1.780	1.851	2.166	2.410	2.675	3.098	3.209
75	0.245	0.156	1.750	1.820	2.134	2.375	2.635	3.048	3.155
80	0.291	0.156	1.721	1.791	2.103	2.341	2.597	3.000	3.105
85	0.334	0.156	1.694	1.764	2.074	2.309	2.561	2.956	3.058

Table 0-24: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Males Ages 8-85 for Average Leg LMI.

	Males								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	0.120	0.155	1.502	1.560	1.819	2.020	2.241	2.596	2.689
10	0.120	0.150	1.752	1.817	2.109	2.334	2.581	2.977	3.081
12	0.120	0.146	1.976	2.048	2.368	2.615	2.884	3.315	3.427
14	0.120	0.143	2.163	2.240	2.581	2.844	3.130	3.587	3.706
16	0.120	0.140	2.288	2.367	2.720	2.992	3.287	3.757	3.879
18	0.120	0.138	2.358	2.438	2.794	3.068	3.365	3.837	3.961
20	0.120	0.136	2.397	2.478	2.834	3.107	3.402	3.872	3.994
25	0.120	0.131	2.453	2.532	2.882	3.150	3.439	3.897	4.016
30	0.120	0.127	2.487	2.565	2.909	3.171	3.453	3.899	4.015
35	0.120	0.124	2.510	2.587	2.924	3.181	3.457	3.891	4.004
40	0.120	0.121	2.520	2.595	2.925	3.176	3.445	3.868	3.978
45	0.120	0.119	2.516	2.589	2.912	3.156	3.418	3.828	3.934
50	0.120	0.117	2.500	2.571	2.885	3.122	3.376	3.774	3.877
55	0.120	0.115	2.473	2.543	2.848	3.077	3.324	3.708	3.808
60	0.120	0.113	2.439	2.507	2.802	3.025	3.263	3.634	3.730
65	0.120	0.111	2.400	2.465	2.751	2.966	3.196	3.554	3.646
70	0.120	0.110	2.355	2.418	2.694	2.902	3.123	3.469	3.557
75	0.120	0.108	2.307	2.368	2.635	2.835	3.049	3.382	3.467
80	0.120	0.107	2.260	2.319	2.576	2.770	2.976	3.296	3.378
85	0.120	0.106	2.214	2.271	2.520	2.707	2.906	3.215	3.294

Table 0-25: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Females Ages 8-85 for Trunk FMI.

	Females								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.432	0.518	0.841	0.919	1.371	1.899	2.773	5.498	6.729
10	-0.343	0.500	1.040	1.138	1.708	2.349	3.360	6.172	7.309
12	-0.266	0.484	1.218	1.337	2.013	2.752	3.871	6.732	7.804
14	-0.198	0.470	1.385	1.524	2.300	3.127	4.336	7.231	8.256
16	-0.136	0.457	1.542	1.700	2.571	3.476	4.760	7.677	8.665
18	-0.079	0.445	1.689	1.864	2.823	3.797	5.143	8.069	9.025
20	-0.027	0.434	1.823	2.015	3.054	4.087	5.481	8.402	9.329
25	0.090	0.410	2.107	2.337	3.541	4.683	6.151	9.009	9.866
30	0.191	0.389	2.339	2.600	3.929	5.140	6.636	9.394	10.188
35	0.281	0.370	2.536	2.821	4.246	5.498	6.995	9.640	10.378
40	0.362	0.353	2.705	3.012	4.506	5.778	7.259	9.785	10.473
45	0.437	0.337	2.850	3.174	4.715	5.991	7.441	9.846	10.488
50	0.506	0.323	2.971	3.307	4.874	6.138	7.546	9.826	10.425
55	0.570	0.309	3.070	3.414	4.988	6.228	7.585	9.739	10.297
60	0.630	0.297	3.152	3.500	5.066	6.274	7.576	9.605	10.125
65	0.687	0.285	3.221	3.570	5.116	6.287	7.531	9.440	9.924
70	0.741	0.274	3.281	3.628	5.147	6.277	7.462	9.258	9.709
75	0.792	0.263	3.335	3.680	5.164	6.252	7.380	9.071	9.492
80	0.841	0.253	3.388	3.728	5.175	6.220	7.295	8.887	9.281
85	0.888	0.243	3.441	3.775	5.183	6.188	7.211	8.713	9.083

Table 0-26: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Males Ages 8-85 for Trunk FMI.

	Males								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.672	0.617	0.781	0.850	1.276	1.842	2.999	10.132	17.520
10	-0.525	0.583	0.829	0.907	1.379	1.970	3.059	7.465	10.065
12	-0.404	0.555	0.858	0.943	1.447	2.050	3.074	6.397	7.943
14	-0.302	0.531	0.890	0.981	1.517	2.131	3.112	5.874	6.995
16	-0.214	0.511	0.945	1.045	1.623	2.263	3.235	5.710	6.628
18	-0.136	0.493	1.020	1.131	1.764	2.441	3.428	5.757	6.568
20	-0.066	0.476	1.107	1.232	1.925	2.645	3.659	5.913	6.660
25	0.082	0.442	1.336	1.494	2.339	3.162	4.245	6.409	7.070
30	0.203	0.414	1.534	1.720	2.686	3.581	4.698	6.780	7.382
35	0.305	0.391	1.694	1.903	2.957	3.891	5.012	6.997	7.551
40	0.393	0.370	1.827	2.055	3.170	4.122	5.229	7.115	7.627
45	0.471	0.352	1.942	2.185	3.340	4.296	5.380	7.170	7.646
50	0.541	0.336	2.045	2.298	3.480	4.431	5.487	7.187	7.632
55	0.604	0.321	2.137	2.399	3.595	4.534	5.558	7.174	7.591
60	0.661	0.308	2.221	2.488	3.688	4.611	5.602	7.139	7.531
65	0.714	0.296	2.296	2.567	3.764	4.667	5.623	7.086	7.455
70	0.764	0.284	2.366	2.638	3.826	4.707	5.629	7.022	7.371
75	0.809	0.274	2.433	2.705	3.879	4.737	5.626	6.955	7.286
80	0.852	0.264	2.498	2.769	3.927	4.762	5.620	6.891	7.205
85	0.892	0.255	2.562	2.832	3.972	4.785	5.614	6.831	7.130

Table 0-27: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Females Ages 8-85 for Trunk LMI.

	Females									
			M							
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-1.970	0.129	4.444	4.539	5.001	5.419	5.960	7.131	7.535
10	-1.769	0.129	4.952	5.061	5.589	6.060	6.659	7.905	8.318
12	-1.604	0.129	5.355	5.476	6.059	6.573	7.218	8.518	8.936
14	-1.465	0.129	5.666	5.797	6.425	6.972	7.651	8.988	9.408
16	-1.345	0.129	5.900	6.039	6.704	7.278	7.983	9.343	9.761
18	-1.239	0.129	6.080	6.226	6.921	7.517	8.240	9.613	10.029
20	-1.144	0.129	6.221	6.373	7.093	7.705	8.444	9.824	10.236
25	-0.942	0.129	6.456	6.619	7.388	8.031	8.794	10.175	10.577
30	-0.778	0.129	6.590	6.763	7.566	8.229	9.004	10.376	10.767
35	-0.639	0.129	6.666	6.846	7.675	8.352	9.133	10.490	10.870
40	-0.518	0.129	6.704	6.889	7.738	8.424	9.208	10.547	10.917
45	-0.412	0.129	6.712	6.902	7.766	8.458	9.241	10.560	10.920
50	-0.317	0.129	6.697	6.890	7.765	8.460	9.239	10.537	10.888
55	-0.231	0.129	6.665	6.861	7.743	8.439	9.214	10.489	10.832
60	-0.153	0.129	6.624	6.821	7.710	8.405	9.173	10.427	10.761
65	-0.081	0.129	6.577	6.777	7.669	8.363	9.125	10.357	10.683
70	-0.014	0.129	6.529	6.730	7.625	8.317	9.073	10.285	10.604
75	0.048	0.129	6.481	6.683	7.581	8.271	9.020	10.214	10.526
80	0.107	0.129	6.434	6.638	7.539	8.227	8.970	10.146	10.452
85	0.161	0.129	6.391	6.596	7.499	8.184	8.922	10.082	10.382

Table 0-28: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Males Ages 8-85 for Trunk LMI.

	Males								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.144	0.150	4.392	4.544	5.239	5.793	6.414	7.448	7.728
10	-0.144	0.145	4.803	4.964	5.696	6.277	6.926	8.001	8.291
12	-0.144	0.141	5.327	5.501	6.287	6.909	7.602	8.746	9.052
14	-0.144	0.137	5.945	6.135	6.989	7.662	8.411	9.641	9.971
16	-0.144	0.134	6.470	6.671	7.579	8.292	9.083	10.380	10.726
18	-0.144	0.132	6.814	7.022	7.957	8.691	9.503	10.830	11.183
20	-0.144	0.129	7.040	7.251	8.198	8.940	9.759	11.094	11.450
25	-0.144	0.124	7.394	7.607	8.560	9.303	10.120	11.446	11.798
30	-0.144	0.120	7.606	7.818	8.764	9.498	10.303	11.604	11.948
35	-0.144	0.116	7.762	7.972	8.907	9.631	10.422	11.697	12.033
40	-0.144	0.113	7.879	8.087	9.010	9.722	10.499	11.747	12.076
45	-0.144	0.111	7.963	8.168	9.078	9.778	10.540	11.761	12.081
50	-0.144	0.108	8.020	8.222	9.118	9.805	10.552	11.745	12.058
55	-0.144	0.106	8.047	8.246	9.125	9.799	10.530	11.695	12.001
60	-0.144	0.104	8.041	8.236	9.097	9.756	10.469	11.605	11.902
65	-0.144	0.102	7.996	8.186	9.027	9.669	10.363	11.466	11.754
70	-0.144	0.101	7.912	8.097	8.915	9.538	10.211	11.279	11.557
75	-0.144	0.099	7.801	7.981	8.774	9.377	10.028	11.060	11.328
80	-0.144	0.098	7.679	7.853	8.621	9.205	9.835	10.830	11.089
85	-0.144	0.096	7.556	7.725	8.470	9.035	9.644	10.605	10.855

Table 0-29: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Females Ages 8-85 for Total Body FMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.613	0.391	2.983	3.183	4.286	5.471	7.292	12.386	14.531
10	-0.523	0.381	3.394	3.628	4.903	6.240	8.222	13.342	15.329
12	-0.443	0.373	3.762	4.028	5.460	6.929	9.045	14.176	16.050
14	-0.371	0.365	4.099	4.395	5.974	7.561	9.788	14.922	16.712
16	-0.305	0.358	4.408	4.732	6.446	8.138	10.459	15.585	17.306
18	-0.243	0.352	4.688	5.039	6.877	8.661	11.058	16.165	17.826
20	-0.186	0.346	4.938	5.314	7.265	9.128	11.585	16.656	18.262
25	-0.055	0.332	5.449	5.879	8.063	10.074	12.621	17.551	19.032
30	0.060	0.320	5.837	6.312	8.676	10.781	13.360	18.111	19.485
35	0.164	0.309	6.141	6.654	9.157	11.321	13.897	18.457	19.736
40	0.260	0.299	6.385	6.929	9.540	11.736	14.287	18.655	19.851
45	0.349	0.290	6.575	7.146	9.836	12.043	14.551	18.727	19.847
50	0.431	0.281	6.711	7.303	10.042	12.240	14.688	18.667	19.717
55	0.510	0.273	6.798	7.405	10.168	12.337	14.712	18.492	19.475
60	0.584	0.266	6.849	7.466	10.231	12.360	14.654	18.238	19.158
65	0.654	0.258	6.872	7.495	10.247	12.326	14.534	17.930	18.792
70	0.721	0.251	6.876	7.502	10.228	12.251	14.373	17.588	18.396
75	0.786	0.244	6.870	7.496	10.187	12.153	14.189	17.235	17.994
80	0.847	0.238	6.861	7.486	10.137	12.044	13.998	16.887	17.601
85	0.907	0.232	6.853	7.476	10.084	11.933	13.810	16.555	17.229

Table 0-30: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Males Ages 8-85 for Total Body FMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.611	0.514	2.477	2.679	3.869	5.299	7.822	17.426	22.885
10	-0.487	0.487	2.524	2.738	3.972	5.388	7.706	14.886	18.142
12	-0.385	0.465	2.562	2.785	4.051	5.447	7.610	13.495	15.847
14	-0.300	0.447	2.606	2.839	4.134	5.516	7.563	12.652	14.525
16	-0.225	0.431	2.676	2.919	4.251	5.632	7.604	12.188	13.778
18	-0.160	0.417	2.771	3.026	4.405	5.797	7.726	11.979	13.390
20	-0.101	0.404	2.884	3.153	4.584	5.996	7.902	11.926	13.217
25	0.023	0.377	3.192	3.494	5.056	6.524	8.404	12.075	13.181
30	0.124	0.355	3.477	3.807	5.474	6.979	8.834	12.262	13.253
35	0.210	0.336	3.721	4.074	5.815	7.335	9.153	12.380	13.285
40	0.284	0.320	3.931	4.302	6.091	7.610	9.384	12.433	13.270
45	0.349	0.306	4.116	4.499	6.318	7.826	9.550	12.442	13.222
50	0.408	0.293	4.281	4.673	6.508	7.997	9.672	12.423	13.154
55	0.461	0.282	4.430	4.829	6.668	8.135	9.759	12.382	13.071
60	0.509	0.271	4.567	4.970	6.806	8.246	9.821	12.327	12.979
65	0.553	0.262	4.693	5.098	6.924	8.336	9.863	12.263	12.881
70	0.595	0.253	4.811	5.217	7.028	8.410	9.892	12.194	12.783
75	0.633	0.244	4.923	5.329	7.122	8.475	9.913	12.125	12.687
80	0.669	0.237	5.032	5.437	7.209	8.534	9.930	12.060	12.598
85	0.703	0.229	5.137	5.541	7.292	8.588	9.945	12.000	12.516

Table 0-31: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Females Ages 8-85 for Total Body LMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-2.075	0.126	9.866	10.069	11.062	11.962	13.131	15.678	16.566
10	-1.876	0.126	10.726	10.954	12.060	13.047	14.310	16.957	17.843
12	-1.707	0.126	11.431	11.680	12.884	13.946	15.284	18.004	18.886
14	-1.559	0.126	11.940	12.207	13.487	14.606	15.997	18.754	19.625
16	-1.427	0.126	12.289	12.569	13.910	15.070	16.496	19.261	20.117
18	-1.307	0.126	12.552	12.844	14.235	15.427	16.879	19.640	20.479
20	-1.198	0.126	12.759	13.062	14.495	15.715	17.186	19.937	20.760
25	-0.959	0.126	13.087	13.411	14.931	16.199	17.698	20.405	21.190
30	-0.756	0.126	13.272	13.614	15.199	16.502	18.014	20.669	21.420
35	-0.578	0.126	13.401	13.758	15.401	16.730	18.251	20.858	21.581
40	-0.419	0.126	13.488	13.860	15.552	16.904	18.429	20.992	21.691
45	-0.276	0.126	13.517	13.901	15.634	17.001	18.525	21.041	21.717
50	-0.145	0.126	13.449	13.842	15.601	16.974	18.486	20.945	21.598
55	-0.024	0.126	13.297	13.696	15.469	16.837	18.330	20.721	21.349
60	0.089	0.126	13.096	13.499	15.278	16.636	18.103	20.424	21.028
65	0.194	0.126	12.863	13.268	15.046	16.390	17.829	20.079	20.659
70	0.293	0.126	12.608	13.015	14.787	16.114	17.522	19.701	20.258
75	0.386	0.126	12.345	12.752	14.516	15.824	17.201	19.311	19.847
80	0.475	0.126	12.085	12.493	14.246	15.535	16.882	18.927	19.442
85	0.559	0.126	11.834	12.242	13.986	15.255	16.574	18.557	19.054

Table 0-32: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for Hispanic Males Ages 8-85 for Total Body LMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.093	0.138	9.955	10.276	11.722	12.855	14.109	16.158	16.703
10	-0.093	0.133	10.864	11.203	12.725	13.914	15.225	17.357	17.924
12	-0.093	0.129	11.953	12.316	13.941	15.206	16.597	18.853	19.451
14	-0.093	0.126	13.152	13.542	15.284	16.637	18.121	20.521	21.155
16	-0.093	0.124	14.115	14.524	16.351	17.766	19.316	21.815	22.475
18	-0.093	0.121	14.715	15.134	17.000	18.442	20.019	22.555	23.224
20	-0.093	0.119	15.100	15.522	17.400	18.850	20.432	22.973	23.641
25	-0.093	0.115	15.717	16.140	18.017	19.459	21.029	23.539	24.197
30	-0.093	0.111	16.092	16.512	18.368	19.790	21.333	23.792	24.436
35	-0.093	0.108	16.348	16.762	18.592	19.990	21.503	23.908	24.536
40	-0.093	0.105	16.507	16.915	18.714	20.085	21.566	23.915	24.527
45	-0.093	0.103	16.584	16.985	18.749	20.090	21.538	23.827	24.423
50	-0.093	0.101	16.597	16.990	18.717	20.028	21.440	23.669	24.249
55	-0.093	0.099	16.546	16.930	18.617	19.895	21.270	23.438	24.000
60	-0.093	0.097	16.434	16.809	18.453	19.697	21.033	23.136	23.682
65	-0.093	0.096	16.256	16.621	18.219	19.426	20.722	22.757	23.284
70	-0.093	0.094	16.008	16.362	17.909	19.077	20.329	22.293	22.801
75	-0.093	0.093	15.713	16.055	17.550	18.678	19.885	21.776	22.265
80	-0.093	0.091	15.399	15.729	17.173	18.261	19.424	21.244	21.714
85	-0.093	0.090	15.089	15.408	16.803	17.852	18.974	20.727	21.179

Table 0-33: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Females Ages 8-85 for Average Arm FMI.

	Females								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0.000	0.674	1.645	1.881
8	-0.468	0.491	0.154	0.167	0.243	0.331	0.474	0.913	1.109
10	-0.380	0.478	0.169	0.185	0.271	0.367	0.518	0.935	1.104
12	-0.308	0.468	0.182	0.199	0.294	0.397	0.553	0.956	1.108
14	-0.248	0.459	0.193	0.211	0.313	0.422	0.582	0.974	1.115
16	-0.195	0.452	0.202	0.222	0.330	0.443	0.607	0.990	1.123
18	-0.149	0.445	0.210	0.231	0.345	0.463	0.629	1.004	1.131
20	-0.108	0.439	0.217	0.239	0.358	0.480	0.648	1.017	1.139
25	-0.020	0.426	0.233	0.257	0.387	0.516	0.688	1.045	1.158
30	0.051	0.416	0.245	0.272	0.411	0.545	0.720	1.068	1.175
35	0.112	0.407	0.256	0.284	0.431	0.570	0.747	1.088	1.189
40	0.164	0.400	0.265	0.295	0.449	0.592	0.770	1.105	1.203
45	0.210	0.393	0.274	0.305	0.465	0.611	0.791	1.120	1.215
50	0.252	0.387	0.281	0.314	0.479	0.628	0.809	1.134	1.226
55	0.289	0.382	0.288	0.322	0.493	0.643	0.825	1.146	1.236
60	0.323	0.377	0.294	0.329	0.504	0.658	0.839	1.157	1.245
65	0.355	0.372	0.300	0.336	0.516	0.671	0.853	1.167	1.253
70	0.384	0.368	0.305	0.343	0.526	0.683	0.865	1.176	1.261
75	0.411	0.364	0.310	0.349	0.536	0.694	0.876	1.185	1.269
80	0.436	0.361	0.315	0.355	0.545	0.704	0.887	1.193	1.276
85	0.460	0.357	0.320	0.360	0.553	0.714	0.897	1.201	1.282

Table 0-34: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Males Ages 8-85 for Average Arm FMI.

	Males									
			M							
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0.000	0.674	1.645	1.881
8	-0.703	0.505	0.130	0.140	0.199	0.270	0.398	0.940	1.295
10	-0.595	0.484	0.133	0.144	0.205	0.276	0.397	0.812	1.025
12	-0.507	0.466	0.136	0.147	0.210	0.282	0.396	0.744	0.898
14	-0.432	0.452	0.139	0.151	0.216	0.287	0.398	0.704	0.828
16	-0.367	0.439	0.143	0.155	0.223	0.295	0.403	0.683	0.789
18	-0.310	0.428	0.149	0.161	0.231	0.305	0.412	0.674	0.769
20	-0.259	0.418	0.155	0.168	0.241	0.316	0.423	0.674	0.761
25	-0.152	0.396	0.170	0.185	0.266	0.345	0.453	0.686	0.761
30	-0.063	0.379	0.185	0.202	0.288	0.371	0.481	0.702	0.770
35	0.011	0.364	0.198	0.216	0.308	0.394	0.504	0.716	0.780
40	0.076	0.352	0.210	0.229	0.325	0.413	0.523	0.728	0.788
45	0.133	0.340	0.220	0.240	0.340	0.430	0.539	0.737	0.794
50	0.184	0.330	0.229	0.250	0.353	0.443	0.551	0.744	0.798
55	0.230	0.321	0.237	0.259	0.364	0.455	0.562	0.748	0.800
60	0.272	0.313	0.244	0.267	0.373	0.464	0.570	0.751	0.801
65	0.311	0.305	0.251	0.273	0.381	0.472	0.576	0.752	0.800
70	0.347	0.298	0.256	0.279	0.388	0.478	0.581	0.752	0.798
75	0.380	0.292	0.261	0.285	0.394	0.484	0.584	0.751	0.796
80	0.411	0.285	0.266	0.290	0.400	0.488	0.588	0.750	0.793
85	0.440	0.280	0.271	0.295	0.405	0.493	0.590	0.749	0.791

Table 0-35: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Females Ages 8-85 for Average Arm LMI.

	Females								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.383	0.157	0.498	0.515	0.594	0.659	0.734	0.865	0.902
10	-0.383	0.157	0.532	0.550	0.634	0.704	0.784	0.924	0.963
12	-0.383	0.157	0.558	0.577	0.665	0.738	0.823	0.970	1.011
14	-0.383	0.157	0.578	0.597	0.689	0.764	0.852	1.004	1.046
16	-0.383	0.157	0.592	0.613	0.707	0.784	0.874	1.029	1.073
18	-0.383	0.157	0.603	0.624	0.720	0.799	0.890	1.049	1.093
20	-0.383	0.157	0.612	0.632	0.730	0.809	0.902	1.063	1.108
25	-0.383	0.157	0.624	0.645	0.744	0.826	0.920	1.084	1.130
30	-0.383	0.157	0.629	0.651	0.751	0.833	0.928	1.094	1.140
35	-0.383	0.157	0.631	0.652	0.752	0.835	0.930	1.096	1.142
40	-0.383	0.157	0.629	0.651	0.751	0.833	0.928	1.094	1.140
45	-0.383	0.157	0.626	0.647	0.747	0.829	0.923	1.088	1.134
50	-0.383	0.157	0.621	0.642	0.741	0.822	0.916	1.079	1.125
55	-0.383	0.157	0.615	0.636	0.733	0.814	0.907	1.068	1.114
60	-0.383	0.157	0.608	0.628	0.725	0.804	0.896	1.056	1.101
65	-0.383	0.157	0.600	0.620	0.716	0.794	0.885	1.043	1.087
70	-0.383	0.157	0.592	0.612	0.706	0.784	0.873	1.029	1.073
75	-0.383	0.157	0.585	0.605	0.697	0.774	0.862	1.016	1.059
80	-0.383	0.157	0.577	0.597	0.688	0.764	0.851	1.003	1.046
85	-0.383	0.157	0.570	0.590	0.680	0.755	0.841	0.991	1.033

Table 0-36: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Males Ages 8-85 for Average Arm LMI.

	Males								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	0.512	0.141	0.514	0.534	0.621	0.684	0.751	0.852	0.878
10	0.512	0.141	0.554	0.575	0.668	0.737	0.808	0.918	0.945
12	0.512	0.141	0.634	0.659	0.766	0.844	0.927	1.052	1.083
14	0.512	0.141	0.740	0.769	0.894	0.985	1.081	1.227	1.264
16	0.512	0.141	0.830	0.862	1.001	1.104	1.212	1.375	1.416
18	0.512	0.141	0.889	0.924	1.073	1.183	1.299	1.474	1.518
20	0.512	0.141	0.925	0.961	1.116	1.230	1.350	1.533	1.579
25	0.512	0.141	0.960	0.998	1.159	1.278	1.403	1.592	1.640
30	0.512	0.141	0.973	1.011	1.174	1.295	1.421	1.613	1.661
35	0.512	0.141	0.982	1.020	1.185	1.307	1.434	1.628	1.677
40	0.512	0.141	0.985	1.024	1.189	1.311	1.439	1.633	1.682
45	0.512	0.141	0.981	1.019	1.184	1.306	1.433	1.626	1.675
50	0.512	0.141	0.970	1.008	1.170	1.291	1.416	1.608	1.656
55	0.512	0.141	0.953	0.991	1.151	1.269	1.392	1.580	1.628
60	0.512	0.141	0.933	0.969	1.126	1.242	1.363	1.547	1.593
65	0.512	0.141	0.908	0.943	1.096	1.208	1.326	1.505	1.550
70	0.512	0.141	0.879	0.913	1.061	1.170	1.284	1.457	1.501
75	0.512	0.141	0.847	0.880	1.023	1.127	1.237	1.404	1.447
80	0.512	0.141	0.815	0.847	0.983	1.084	1.190	1.351	1.391
85	0.512	0.141	0.783	0.814	0.946	1.043	1.144	1.299	1.338

Table 0-37: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Females Ages 8-85 for Average Leg FMI.

	Females								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.458	0.331	0.707	0.752	0.988	1.222	1.546	2.283	2.540
10	-0.386	0.331	0.769	0.820	1.085	1.343	1.695	2.472	2.734
12	-0.328	0.331	0.818	0.874	1.162	1.441	1.816	2.623	2.888
14	-0.279	0.331	0.858	0.918	1.227	1.523	1.916	2.745	3.013
16	-0.236	0.331	0.891	0.954	1.280	1.591	2.000	2.846	3.116
18	-0.198	0.331	0.917	0.983	1.325	1.648	2.069	2.929	3.199
20	-0.164	0.331	0.938	1.007	1.362	1.695	2.127	2.996	3.266
25	-0.093	0.331	0.974	1.049	1.430	1.782	2.232	3.114	3.381
30	-0.035	0.331	0.995	1.075	1.475	1.841	2.303	3.188	3.452
35	0.015	0.331	1.009	1.091	1.507	1.884	2.353	3.237	3.498
40	0.057	0.331	1.017	1.102	1.531	1.915	2.390	3.271	3.528
45	0.095	0.331	1.022	1.110	1.548	1.939	2.418	3.295	3.548
50	0.129	0.331	1.024	1.114	1.562	1.958	2.439	3.312	3.561
55	0.159	0.331	1.025	1.117	1.573	1.973	2.456	3.323	3.569
60	0.187	0.331	1.025	1.119	1.581	1.985	2.469	3.331	3.574
65	0.213	0.331	1.024	1.119	1.587	1.994	2.479	3.336	3.575
70	0.237	0.331	1.022	1.119	1.592	2.002	2.487	3.338	3.575
75	0.259	0.331	1.020	1.118	1.597	2.008	2.494	3.340	3.574
80	0.279	0.331	1.017	1.117	1.600	2.014	2.500	3.341	3.573
85	0.299	0.331	1.015	1.116	1.603	2.019	2.505	3.342	3.571

Table 0-38: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Females Ages 8-85 for Average Leg FMI.

	Males									
			M							
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.027	0.485	0.419	0.468	0.745	1.032	1.433	2.311	2.598
10	-0.027	0.466	0.439	0.489	0.764	1.045	1.433	2.270	2.540
12	-0.027	0.451	0.457	0.507	0.781	1.057	1.434	2.235	2.492
14	-0.027	0.437	0.473	0.523	0.795	1.067	1.434	2.206	2.451
16	-0.027	0.426	0.487	0.537	0.808	1.075	1.434	2.180	2.416
18	-0.027	0.415	0.500	0.551	0.820	1.083	1.434	2.158	2.384
20	-0.027	0.406	0.512	0.563	0.830	1.090	1.434	2.137	2.356
25	-0.027	0.385	0.539	0.590	0.853	1.105	1.434	2.094	2.297
30	-0.027	0.368	0.563	0.613	0.873	1.118	1.434	2.058	2.248
35	-0.027	0.353	0.584	0.634	0.890	1.129	1.433	2.027	2.207
40	-0.027	0.340	0.603	0.653	0.905	1.138	1.433	2.001	2.171
45	-0.027	0.329	0.621	0.670	0.919	1.147	1.432	1.977	2.140
50	-0.027	0.318	0.637	0.686	0.932	1.154	1.432	1.956	2.111
55	-0.027	0.309	0.652	0.701	0.944	1.161	1.431	1.937	2.086
60	-0.027	0.300	0.667	0.715	0.954	1.168	1.430	1.920	2.063
65	-0.027	0.292	0.680	0.728	0.965	1.174	1.430	1.903	2.041
70	-0.027	0.284	0.693	0.741	0.974	1.179	1.429	1.888	2.021
75	-0.027	0.277	0.706	0.753	0.983	1.185	1.429	1.874	2.003
80	-0.027	0.271	0.718	0.764	0.992	1.190	1.428	1.861	1.986
85	-0.027	0.264	0.729	0.775	1.000	1.194	1.428	1.849	1.969

Table 0-39: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Females Ages 8-85 for Average Leg LMI.

	Females								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.402	0.153	1.485	1.534	1.762	1.950	2.166	2.542	2.647
10	-0.402	0.153	1.661	1.716	1.971	2.181	2.423	2.843	2.961
12	-0.402	0.153	1.785	1.844	2.118	2.344	2.604	3.056	3.182
14	-0.402	0.153	1.866	1.927	2.214	2.450	2.722	3.194	3.326
16	-0.402	0.153	1.915	1.978	2.273	2.514	2.794	3.278	3.414
18	-0.402	0.153	1.946	2.011	2.310	2.555	2.839	3.332	3.470
20	-0.402	0.153	1.966	2.031	2.334	2.582	2.868	3.366	3.505
25	-0.402	0.153	1.983	2.049	2.354	2.604	2.893	3.395	3.536
30	-0.402	0.153	1.981	2.047	2.352	2.602	2.891	3.392	3.532
35	-0.402	0.153	1.975	2.041	2.345	2.594	2.882	3.382	3.522
40	-0.402	0.153	1.969	2.034	2.337	2.586	2.873	3.371	3.510
45	-0.402	0.153	1.959	2.024	2.325	2.572	2.858	3.353	3.492
50	-0.402	0.153	1.942	2.007	2.305	2.550	2.834	3.325	3.463
55	-0.402	0.153	1.921	1.985	2.280	2.522	2.802	3.289	3.424
60	-0.402	0.153	1.897	1.959	2.251	2.490	2.767	3.247	3.381
65	-0.402	0.153	1.870	1.932	2.220	2.456	2.729	3.202	3.334
70	-0.402	0.153	1.843	1.904	2.187	2.420	2.688	3.155	3.285
75	-0.402	0.153	1.815	1.875	2.154	2.383	2.647	3.107	3.235
80	-0.402	0.153	1.787	1.846	2.120	2.346	2.606	3.059	3.185
85	-0.402	0.153	1.759	1.818	2.088	2.310	2.567	3.012	3.137

Table 0-40: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Males Ages 8-85 for Average Leg LMI.

	Males								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.392	0.146	1.525	1.573	1.797	1.980	2.189	2.549	2.649
10	-0.243	0.143	1.731	1.788	2.043	2.248	2.479	2.866	2.971
12	-0.122	0.141	1.953	2.017	2.307	2.535	2.789	3.207	3.319
14	-0.020	0.139	2.160	2.232	2.553	2.803	3.079	3.525	3.643
16	0.069	0.137	2.309	2.386	2.730	2.995	3.284	3.746	3.867
18	0.148	0.136	2.395	2.476	2.833	3.106	3.401	3.868	3.989
20	0.218	0.134	2.440	2.523	2.887	3.163	3.460	3.924	4.044
25	0.366	0.131	2.470	2.555	2.924	3.199	3.490	3.938	4.052
30	0.488	0.129	2.470	2.556	2.926	3.197	3.481	3.911	4.020
35	0.590	0.127	2.474	2.561	2.931	3.199	3.477	3.894	3.999
40	0.679	0.125	2.475	2.563	2.932	3.198	3.471	3.877	3.978
45	0.758	0.123	2.467	2.555	2.923	3.185	3.453	3.847	3.945
50	0.828	0.122	2.448	2.536	2.901	3.159	3.420	3.803	3.897
55	0.891	0.121	2.422	2.509	2.869	3.122	3.377	3.749	3.840
60	0.949	0.120	2.389	2.475	2.830	3.078	3.326	3.686	3.774
65	1.002	0.119	2.349	2.434	2.782	3.024	3.266	3.614	3.699
70	1.052	0.118	2.303	2.386	2.727	2.962	3.197	3.533	3.614
75	1.098	0.117	2.254	2.336	2.669	2.897	3.125	3.449	3.527
80	1.140	0.116	2.205	2.285	2.610	2.833	3.053	3.366	3.441
85	1.181	0.115	2.158	2.236	2.554	2.770	2.984	3.286	3.359

Table 0-41: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Females Ages 8-85 for Trunk FMI.

	Females								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.418	0.537	0.913	1.000	1.513	2.120	3.139	6.388	7.880
10	-0.377	0.531	0.998	1.094	1.664	2.328	3.418	6.712	8.146
12	-0.337	0.524	1.080	1.187	1.812	2.530	3.687	7.014	8.397
14	-0.298	0.518	1.159	1.276	1.957	2.728	3.945	7.295	8.631
16	-0.259	0.512	1.235	1.362	2.098	2.920	4.192	7.554	8.846
18	-0.221	0.506	1.308	1.445	2.233	3.103	4.424	7.788	9.038
20	-0.183	0.500	1.376	1.523	2.363	3.278	4.642	7.997	9.206
25	-0.090	0.486	1.527	1.698	2.660	3.673	5.121	8.416	9.529
30	0.001	0.471	1.656	1.851	2.926	4.021	5.524	8.728	9.753
35	0.091	0.457	1.768	1.987	3.168	4.331	5.870	8.966	9.915
40	0.179	0.443	1.870	2.112	3.394	4.615	6.174	9.160	10.040
45	0.267	0.430	1.961	2.227	3.606	4.875	6.443	9.316	10.136
50	0.353	0.416	2.041	2.330	3.797	5.103	6.667	9.421	10.186
55	0.439	0.403	2.104	2.415	3.958	5.284	6.830	9.454	10.165
60	0.523	0.390	2.146	2.477	4.076	5.406	6.914	9.395	10.052
65	0.607	0.376	2.164	2.512	4.145	5.460	6.913	9.235	9.839
70	0.691	0.363	2.160	2.520	4.164	5.446	6.828	8.981	9.531
75	0.773	0.350	2.137	2.504	4.138	5.371	6.672	8.650	9.148
80	0.856	0.337	2.104	2.475	4.080	5.255	6.468	8.275	8.724
85	0.937	0.325	2.069	2.441	4.008	5.120	6.248	7.895	8.298

Table 0-42: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Males Ages 8-85 for Trunk FMI.

	Males								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.749	0.524	0.735	0.791	1.124	1.537	2.316	6.137	9.193
10	-0.663	0.513	0.778	0.839	1.202	1.640	2.429	5.649	7.653
12	-0.584	0.503	0.821	0.887	1.279	1.743	2.542	5.395	6.904
14	-0.511	0.493	0.868	0.941	1.364	1.855	2.670	5.289	6.523
16	-0.441	0.484	0.924	1.004	1.464	1.987	2.826	5.302	6.370
18	-0.375	0.476	0.988	1.076	1.579	2.138	3.009	5.400	6.362
20	-0.311	0.468	1.058	1.155	1.704	2.302	3.208	5.548	6.436
25	-0.163	0.449	1.233	1.354	2.024	2.719	3.708	5.971	6.743
30	-0.027	0.431	1.389	1.535	2.320	3.099	4.149	6.342	7.037
35	0.101	0.415	1.521	1.691	2.582	3.429	4.518	6.633	7.266
40	0.222	0.399	1.634	1.827	2.816	3.716	4.827	6.861	7.444
45	0.337	0.385	1.731	1.947	3.024	3.966	5.085	7.037	7.577
50	0.447	0.370	1.814	2.052	3.207	4.179	5.296	7.165	7.667
55	0.552	0.357	1.883	2.143	3.366	4.357	5.461	7.246	7.713
60	0.654	0.344	1.937	2.215	3.492	4.490	5.571	7.266	7.700
65	0.752	0.331	1.970	2.265	3.577	4.567	5.614	7.210	7.613
70	0.848	0.319	1.984	2.291	3.618	4.588	5.590	7.082	7.453
75	0.940	0.307	1.985	2.301	3.627	4.566	5.517	6.904	7.244
80	1.031	0.296	1.980	2.303	3.617	4.521	5.418	6.704	7.015
85	1.119	0.284	1.978	2.305	3.601	4.468	5.315	6.508	6.793

Table 0-43: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Females Ages 8-85 for Trunk LMI.

	Females								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.737	0.127	4.557	4.676	5.226	5.678	6.203	7.124	7.384
10	-0.737	0.127	5.003	5.133	5.738	6.234	6.810	7.821	8.107
12	-0.737	0.127	5.353	5.493	6.139	6.670	7.287	8.368	8.674
14	-0.737	0.127	5.626	5.773	6.452	7.010	7.658	8.795	9.116
16	-0.737	0.127	5.835	5.987	6.691	7.270	7.942	9.121	9.454
18	-0.737	0.127	5.993	6.149	6.873	7.467	8.157	9.368	9.710
20	-0.737	0.127	6.113	6.272	7.010	7.616	8.320	9.555	9.904
25	-0.737	0.127	6.300	6.464	7.225	7.850	8.575	9.848	10.208
30	-0.737	0.127	6.398	6.565	7.337	7.972	8.709	10.001	10.367
35	-0.737	0.127	6.450	6.618	7.397	8.037	8.779	10.083	10.451
40	-0.737	0.127	6.475	6.644	7.425	8.068	8.813	10.122	10.492
45	-0.737	0.127	6.481	6.650	7.432	8.075	8.822	10.131	10.501
50	-0.737	0.127	6.472	6.641	7.423	8.065	8.810	10.118	10.487
55	-0.737	0.127	6.453	6.621	7.401	8.041	8.784	10.088	10.457
60	-0.737	0.127	6.428	6.595	7.371	8.009	8.749	10.048	10.415
65	-0.737	0.127	6.399	6.565	7.338	7.973	8.709	10.002	10.368
70	-0.737	0.127	6.368	6.533	7.302	7.934	8.667	9.954	10.318
75	-0.737	0.127	6.336	6.501	7.266	7.895	8.624	9.905	10.266
80	-0.737	0.127	6.305	6.469	7.231	7.856	8.582	9.856	10.216
85	-0.737	0.127	6.275	6.439	7.196	7.819	8.542	9.809	10.168

Table 0-44: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Males Ages 8-85 for Trunk LMI.

	Males								
			M						
			3	5	25	50	75	95	97
Age	L	S	-1.881	-1.645	-0.674	0	0.674	1.645	1.881
8	-0.652	0.135	4.508	4.634	5.222	5.705	6.268	7.257	7.536
10	-0.494	0.133	4.857	4.996	5.631	6.146	6.734	7.742	8.021
12	-0.366	0.130	5.351	5.505	6.208	6.769	7.402	8.464	8.754
14	-0.257	0.129	5.958	6.132	6.915	7.534	8.225	9.365	9.672
16	-0.163	0.127	6.499	6.690	7.546	8.215	8.955	10.161	10.482
18	-0.080	0.126	6.878	7.081	7.988	8.691	9.462	10.704	11.032
20	-0.006	0.124	7.111	7.323	8.262	8.984	9.769	11.024	11.352
25	0.152	0.122	7.348	7.570	8.541	9.276	10.063	11.297	11.615
30	0.280	0.119	7.442	7.668	8.652	9.386	10.164	11.365	11.671
35	0.389	0.118	7.531	7.761	8.756	9.491	10.261	11.437	11.734
40	0.483	0.116	7.618	7.853	8.859	9.594	10.359	11.515	11.806
45	0.566	0.115	7.688	7.926	8.940	9.675	10.435	11.573	11.857
50	0.641	0.113	7.727	7.968	8.986	9.718	10.470	11.589	11.867
55	0.708	0.112	7.738	7.980	8.998	9.725	10.469	11.567	11.839
60	0.769	0.111	7.720	7.962	8.976	9.696	10.428	11.504	11.770
65	0.826	0.110	7.665	7.906	8.911	9.621	10.340	11.392	11.650
70	0.878	0.109	7.577	7.816	8.808	9.505	10.209	11.232	11.483
75	0.927	0.108	7.467	7.703	8.679	9.362	10.048	11.042	11.285
80	0.972	0.108	7.347	7.580	8.539	9.207	9.875	10.841	11.076
85	1.015	0.107	7.230	7.460	8.402	9.055	9.707	10.646	10.874

Table 0-45: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Females Ages 8-85 for Total Body FMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.471	0.422	3.048	3.281	4.581	5.981	8.116	13.856	16.150
10	-0.432	0.417	3.227	3.478	4.865	6.345	8.564	14.327	16.554
12	-0.394	0.413	3.402	3.670	5.145	6.700	8.998	14.776	16.941
14	-0.357	0.409	3.572	3.856	5.418	7.046	9.417	15.202	17.311
16	-0.320	0.405	3.735	4.037	5.682	7.380	9.817	15.600	17.656
18	-0.284	0.400	3.890	4.208	5.935	7.698	10.194	15.963	17.966
20	-0.248	0.396	4.035	4.370	6.174	7.997	10.543	16.285	18.236
25	-0.160	0.386	4.353	4.726	6.707	8.656	11.294	16.918	18.740
30	-0.073	0.376	4.619	5.027	7.164	9.212	11.900	17.360	19.059
35	0.012	0.367	4.848	5.289	7.566	9.691	12.404	17.678	19.263
40	0.096	0.357	5.054	5.528	7.933	10.121	12.840	17.925	19.408
45	0.179	0.348	5.243	5.750	8.274	10.513	13.226	18.124	19.515
50	0.261	0.338	5.413	5.950	8.582	10.857	13.550	18.259	19.565
55	0.342	0.329	5.552	6.117	8.838	11.132	13.786	18.294	19.520
60	0.422	0.320	5.648	6.237	9.023	11.312	13.906	18.196	19.341
65	0.502	0.311	5.695	6.302	9.120	11.381	13.890	17.943	19.006
70	0.581	0.302	5.690	6.308	9.127	11.334	13.738	17.534	18.515
75	0.660	0.293	5.638	6.262	9.051	11.184	13.464	16.995	17.894
80	0.738	0.284	5.556	6.179	8.916	10.960	13.109	16.374	17.196
85	0.815	0.275	5.463	6.083	8.753	10.704	12.722	15.736	16.486

Table 0-46: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Males Ages 8-85 for Total Body FMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.683	0.452	2.340	2.507	3.469	4.576	6.440	12.928	16.347
10	-0.602	0.442	2.411	2.589	3.597	4.729	6.566	12.299	14.970
12	-0.530	0.432	2.478	2.665	3.716	4.872	6.685	11.888	14.108
14	-0.463	0.423	2.546	2.742	3.835	5.014	6.810	11.625	13.544
16	-0.401	0.415	2.620	2.826	3.965	5.169	6.955	11.483	13.192
18	-0.343	0.408	2.703	2.920	4.107	5.339	7.125	11.438	12.995
20	-0.288	0.400	2.794	3.022	4.259	5.523	7.314	11.462	12.905
25	-0.161	0.384	3.032	3.290	4.659	6.002	7.816	11.676	12.924
30	-0.046	0.369	3.264	3.552	5.045	6.459	8.292	11.946	13.067
35	0.059	0.355	3.478	3.794	5.400	6.870	8.711	12.193	13.218
40	0.158	0.342	3.671	4.013	5.718	7.231	9.067	12.393	13.340
45	0.250	0.330	3.843	4.210	5.999	7.539	9.360	12.538	13.420
50	0.337	0.318	3.996	4.383	6.241	7.797	9.591	12.627	13.451
55	0.420	0.307	4.129	4.535	6.446	8.006	9.765	12.663	13.434
60	0.499	0.297	4.243	4.665	6.613	8.166	9.882	12.642	13.364
65	0.575	0.287	4.336	4.771	6.741	8.274	9.940	12.563	13.240
70	0.648	0.277	4.411	4.854	6.831	8.337	9.946	12.433	13.066
75	0.719	0.268	4.473	4.923	6.895	8.367	9.916	12.272	12.865
80	0.787	0.259	4.531	4.985	6.943	8.378	9.867	12.098	12.655
85	0.853	0.250	4.589	5.046	6.985	8.381	9.812	11.929	12.452

Table 0-47: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Females Ages 8-85 for Total Body LMI.

	Females								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.820	0.124	10.121	10.376	11.560	12.535	13.672	15.679	16.250
10	-0.820	0.124	10.838	11.111	12.378	13.422	14.639	16.789	17.400
12	-0.820	0.124	11.395	11.682	13.014	14.113	15.392	17.652	18.295
14	-0.820	0.124	11.823	12.121	13.503	14.643	15.971	18.316	18.983
16	-0.820	0.124	12.147	12.453	13.873	15.043	16.407	18.817	19.502
18	-0.820	0.124	12.390	12.702	14.150	15.344	16.736	19.193	19.892
20	-0.820	0.124	12.571	12.887	14.357	15.569	16.980	19.474	20.183
25	-0.820	0.124	12.842	13.166	14.667	15.905	17.347	19.894	20.619
30	-0.820	0.124	12.966	13.293	14.809	16.059	17.515	20.086	20.818
35	-0.820	0.124	13.014	13.341	14.863	16.117	17.579	20.160	20.894
40	-0.820	0.124	13.016	13.344	14.866	16.120	17.582	20.164	20.898
45	-0.820	0.124	12.986	13.313	14.831	16.083	17.541	20.117	20.850
50	-0.820	0.124	12.931	13.257	14.769	16.015	17.467	20.032	20.761
55	-0.820	0.124	12.859	13.182	14.686	15.925	17.369	19.919	20.645
60	-0.820	0.124	12.775	13.097	14.590	15.822	17.256	19.790	20.511
65	-0.820	0.124	12.686	13.006	14.489	15.712	17.136	19.652	20.368
70	-0.820	0.124	12.595	12.912	14.385	15.599	17.013	19.512	20.222
75	-0.820	0.124	12.505	12.820	14.282	15.488	16.892	19.372	20.078
80	-0.820	0.124	12.418	12.730	14.182	15.379	16.773	19.236	19.937
85	-0.820	0.124	12.334	12.645	14.087	15.276	16.661	19.107	19.803

Table 0-48: LMS Curve Fit Data providing L, M, and S values for 3rd through 97th percentiles for White Males Ages 8-85 for Total Body LMI.

	Males								
			M						
Age	L	S	3	5	25	50	75	95	97
8	-0.738	0.127	10.119	10.383	11.604	12.608	13.773	15.817	16.394
10	-0.540	0.124	10.904	11.194	12.523	13.593	14.811	16.886	17.458
12	-0.378	0.122	11.931	12.254	13.718	14.879	16.180	18.346	18.932
14	-0.241	0.121	13.105	13.465	15.082	16.347	17.747	20.036	20.647
16	-0.123	0.119	14.104	14.496	16.244	17.596	19.076	21.460	22.089
18	-0.018	0.118	14.783	15.199	17.038	18.446	19.973	22.402	23.037
20	0.075	0.117	15.193	15.625	17.521	18.959	20.507	22.942	23.573
25	0.274	0.114	15.593	16.044	18.002	19.460	21.001	23.375	23.980
30	0.435	0.112	15.727	16.190	18.173	19.626	21.143	23.441	24.019
35	0.572	0.111	15.842	16.314	18.319	19.768	21.264	23.501	24.059
40	0.691	0.109	15.936	16.416	18.437	19.882	21.359	23.545	24.086
45	0.795	0.108	15.978	16.463	18.493	19.929	21.387	23.524	24.049
50	0.889	0.107	15.953	16.442	18.471	19.894	21.329	23.415	23.925
55	0.974	0.106	15.872	16.362	18.383	19.789	21.197	23.231	23.726
60	1.051	0.105	15.738	16.227	18.232	19.617	20.997	22.977	23.457
65	1.122	0.104	15.541	16.027	18.008	19.367	20.714	22.637	23.102
70	1.188	0.103	15.289	15.769	17.718	19.047	20.359	22.222	22.670
75	1.249	0.102	15.004	15.477	17.390	18.687	19.962	21.765	22.197
80	1.306	0.101	14.709	15.175	17.050	18.315	19.554	21.297	21.714
85	1.360	0.101	14.423	14.882	16.720	17.954	19.158	20.847	21.250

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Benjamin Hinton

Author Signature

9/7/18

Date