

ABSTRACT

MANGUM, AMANDA JEAN. Self-Organizing Maps for Data Clustering in Acoustic Radiation Force Ultrasound Imaging of Cardiovascular Tissues. (Under the direction of Mansoor Haider.)

Atherosclerosis is a cardiovascular disease in which plaque accumulates along the wall of an artery, altering blood flow and increasing the risk for heart attack or stroke. Acoustic Radiation Force Impulse (ARFI) is an ultrasound imaging technique in which acoustic waves are focused at a point, causing displacement of the tissue that is then tracked over time to measure elastic and viscoelastic material properties from the imaging data. The emitted ultrasonic waves are sequenced in a manner that allows a pulse to be focused both spatially and temporally and the resulting tissue displacement to be recorded and analyzed. We investigate the application of data clustering algorithms, based on Self-Organizing Maps (SOMs), to ARFI imaging for early detection and characterization of atherosclerotic plaques. In this context, SOMs cluster images based on similar patterns in the data set that are identified via a projection of the data vector space onto a lower dimensional map. This map is obtained in a training phase that utilizes a neighborhood function to ensure that neighboring data clusters are more similar than clusters far away from each other (in terms of the map topology). Based on the dimension, size and scope of image patterns considered in this work, the SOM configuration used for all data sets considered was a 3x3 lattice of nine neurons. Once a SOM is trained using a particular data set, the number of used and unused neurons, the first metric, serves as an indicator of the scope of patterns identified in the data set. The second metric was a statistic calculated from the training data set that measured the extent to which these patterns were spread out in the data space. It was hypothesized that higher values of this measure would correlate with successful SOM performance.

The first set of data considered was comprised of ARFI peak displacement and relaxation time images for gels constructed to mimic arterial tissues with atherosclerotic plaques (phantoms) and was provided by the Gallippi Ultrasound Lab at UNC-Chapel Hill. When the data was used to train a SOM with vectorization, the trained SOM used two neurons to correctly identify the primary distinguishing feature as plaque stiffness (hard or soft), but failed to cluster on the other features reflected in the data set. However, subsequent training of SOMs on each data subgroup (soft and hard), resulted in a clustering that used three neurons and identified additional features such as plaque size and location. In an attempt to utilize more neurons in a trained SOM, two data sets were manufactured to mimic larger data sets with continuous variations in major features. The SOMs trained on these two manufactured data sets achieved better clusterings, using up to five of the nine available neurons. The final data set that was investigated came from an *ex vivo* ARFI imaging study on porcine iliac arteries performed at

the Gallippi Ultrasound Lab. An extensive analysis of SOM performance was carried out by training SOMs for possible combinations of expanded training sets comprised of either 6 or 7 original patterns. In each instance, an optimal sort resulting in only one unused neuron was identified. The second data expansion method was then applied to the original disease patterns by taking weighted averages of the original patterns, creating an expanded data set with more realistic and continuous variations in plaque features relative to the first expansion method. This expanded data set was partitioned into training and validation sets before training a SOM. On this more varied data set, optimally trained SOMs now contained mismatches but each pattern in the training or validation set mapped to no more than two neurons in the trained SOM. For clustering of data in both the training and validation sets, roughly 80% of the original disease patterns clustered to the same neuron. Results demonstrated a strong correlation between larger values of the statistic and the ability of the trained SOM to successfully cluster data in both the training and validation sets.

© Copyright 2014 by Amanda Jean Mangum

All Rights Reserved

Self-Organizing Maps for Data Clustering in Acoustic Radiation
Force Ultrasound Imaging of Cardiovascular Tissues

by
Amanda Jean Mangum

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Biomathematics

Raleigh, North Carolina

2014

APPROVED BY:

Alun Lloyd

Caterina Gallippi

Ralph Smith

Mansoor Haider
Chair of Advisory Committee

BIOGRAPHY

Mandy was born in Ohio but spent most of her time growing up in a small town in Tennessee. She graduated from Tullahoma High School in 2004 before heading to Centre College where she earned a B.S. in Mathematics in 2008. She then attended North Carolina State University where she received her M.S. in Mathematics and Ph.D. in Biomathematics under Dr. Mansoor Haider.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Mansoor Haider, for his help and guidance through my work and time at North Carolina State University. I would also like to thank my committee, Dr. Alun Lloyd, Dr. Caterina Gallippi, and Dr. Ralph Smith for their time, efforts, help, and support. A special thanks should also be given to Tomek Czernuszewicz for guiding me through the background and details of the data sets. My husband, Chad, also deserves many thanks for standing by me, helping me, and loving me. I would also like to thank my parents, family, and friends for all the support and encouragement they have given me through my life. Above all, thanks be to God for His unwavering faithfulness.

TABLE OF CONTENTS

| | |
|---|------------|
| LIST OF TABLES | vi |
| LIST OF FIGURES | vii |
| Chapter 1 Introduction | 1 |
| 1.1 Atherosclerosis | 1 |
| 1.2 Acoustic Radiation Force Impulse (ARFI) Imaging | 3 |
| 1.3 Machine Learning based on Self-Organizing Maps | 5 |
| Chapter 2 ANN Methods based on Self-Organizing Maps | 7 |
| 2.1 Introduction to Artificial Neural Networks | 7 |
| 2.2 Introduction to Self-Organizing Maps | 9 |
| 2.3 Weight Initializations | 14 |
| 2.4 Data Expansion | 15 |
| 2.5 Measuring the Success of a Clustering | 16 |
| 2.6 Training and Validation Data Sets | 17 |
| Chapter 3 SOM Clustering for Tissue Phantoms and Manufactured Data | 19 |
| 3.1 Gel-Based Tissue Phantoms | 19 |
| 3.1.1 ARFI images of tissue phantoms | 20 |
| 3.1.2 Application of SOMs | 22 |
| 3.1.3 Vectorization | 23 |
| 3.2 Trained SOMs for Analysis of Tissue Phantoms | 24 |
| 3.2.1 Initial Cropping | 24 |
| 3.2.2 Probing the Trained SOM | 26 |
| 3.2.3 Analysis of Data Subsets for Soft and Hard Inclusions | 27 |
| 3.2.4 Adding a Plaque-free Pattern | 30 |
| 3.2.5 Cropping for the Region of Interest | 31 |
| 3.2.6 Analysis of Data Subsets for Hard and Soft Inclusions | 32 |
| 3.3 Introduction to the Manufactured Data Sets | 37 |
| 3.3.1 Introduction to the Rectangular and Elliptical Data Sets | 37 |
| 3.3.2 Clustering Results for the Manufactured Data Sets | 41 |
| Chapter 4 SOM Clustering for Porcine Data | 44 |
| 4.1 Porcine Data Set | 44 |
| 4.1.1 Collection and Details of the Data Set | 44 |
| 4.1.2 Applying SOMs to the Porcine Data | 48 |
| 4.1.3 Parallelization and Cluster Computing | 49 |
| 4.2 Application of SOM based on the First Data Expansion | 49 |
| 4.2.1 Data Expansion and Clustering of the Diseased Data | 50 |
| 4.2.2 Crops of Diseased Data | 56 |
| 4.2.3 Partitioning the Data: Seven Training and Three Validation Patterns | 61 |
| 4.2.4 Partitioning the Data: Six Training and Four Validation Patterns | 66 |

| | | |
|-----------------------------|--|-----------|
| 4.3 | Second Data Expansion Method | 70 |
| 4.3.1 | Towards Optimizing the Choice of Training Sets | 75 |
| 4.4 | Summary and Conclusions | 76 |
| References | | 80 |

LIST OF TABLES

| | | |
|------------|---|----|
| Table 2.1 | General Algorithm for Training of 2-D Self-Organizing Maps | 12 |
| Table 2.2 | General method of expansion of data based on weighted averages of original disease patterns | 16 |
| Table 3.1 | The execution time for the training of a SOM and the effects of vectorization. Execution times with vectorization (a) and without vectorization (b) on a dual core processor (MacBook Pro with 1.83 GHz Intel Core Duo, 2GB RAM) and a dual quadcore processor (Mac Pro with dual 2.8GHz quadcore Intel Xeon, 4GB RAM) | 23 |
| Table 3.2 | This table contains the description of the plaque sizes and inverse stiffnesses in the nine basic patterns created for the manufactured data set. Three perturbations of each of these nine patterns were used to create the full manufactured rectangular data set. | 38 |
| Table 4.1 | Description of Plaque Composition Classifications | 45 |
| Table 4.2 | Numerical Ratings Based on Type | 46 |
| Table 4.3 | Description of the Data Based on the Gold Star Standards | 47 |
| Table 4.4 | This table shows validation for remaining three patterns the SOM was not trained on for Figure 4.9. Note that neuron 6 was unused in the training (because no data used for training was associated with it after training). The interpolation of final weights associated with each neuron provided a pattern that was closer to pattern II. than any of the used neurons. | 63 |
| Table 4.5 | This table shows the validation for the remaining three patterns the SOM was not trained on for Figure 4.10. Neuron 4 was unused after training, but pattern I. mapped to it during validation. | 64 |
| Table 4.6 | The validation results for remaining three diseased porcine patterns not used in the training of the SOM results presented in Figure 4.11 are shown. Neuron 2 had none of the training data clustered to it; however, one of the validation patterns (IX.) mapped to this neuron. | 65 |
| Table 4.7 | Validation results for the remaining four patterns. | 67 |
| Table 4.8 | The results of the validation on the trained SOM shown in Figure 4.20. The numbers in parenthesis after each pattern number indicates the number of images of that pattern type clustered to the neuron. | 73 |
| Table 4.9 | Results of the validation set of images for the trained SOM shown in Figure 4.21. The numbers in parenthesis after the pattern type give the number of images of that pattern mapped to that neuron. | 74 |
| Table 4.10 | The table displays three training sets predicted to train a SOM well (a), and three training sets predicted to not train a SOM well (b) based on their computed values of the statistic ρ for the training set. | 75 |

LIST OF FIGURES

| | | |
|------------|---|----|
| Figure 1.1 | Strokes are one possible effect of atherosclerosis. A healthy artery is shown in (a). The initial phase of atherosclerosis is shown in (b). The lesion progression is shown in (c), and thrombosis is shown in (d). (Reprinted from Nature, Peter Libby, Paul Ridker, and Goran Hansson, Progress and challenges in translating the biology of atherosclerosis, 317-325, Copyright (2011), with permission from Nature.) | 2 |
| Figure 1.2 | Illustration of ARFI imaging procedure in the Gallippi Ultrasound Lap at UNC [19]. This image depicts how the acoustic waves interact with the tissue and how the tracking pulse records tissue behavior. | 4 |
| Figure 2.1 | An illustration of how SOM training projects higher dimensional data (upper box) onto a lower dimensional neuron space comprised of five neurons (lower box). Note that if the data is projected down onto a one-dimensional neuron space, then the order of the clusters (color-coded) should be preserved. In this way, data mapped to neighboring clusters should be more similar than data mapped to neurons farther apart. | 10 |
| Figure 2.2 | An illustration of relationships between the data space, weight space and neuron topology in a two-dimensional Self-Organizing Map. Each neuron in the SOM (blue) has an associated weight vector that is the same dimension as the data. For instance, if the data has three dimensions (black), then each of the nine neurons in the SOM will have a 3-dimensional weight vector associated with it. | 10 |
| Figure 2.3 | Illustration of the training process for a 1-d SOM with five neurons arranged in a one-dimensional line. Here, neighbors are determined to be neurons one away from the winning neuron. The weights of the winning neuron and neighboring neurons move closer to the data point presented (open circle, top). After training, each neuron corresponds to a cluster of data points (below). | 11 |
| Figure 2.4 | Possibilities for defining the neighborhood function. For the two-dimensional neuron space as shown we could consider neighbors to be those neurons one grid step away or those one grid step away and “diagonals”. The relevant neuron is shown in black, and the two possible neighborhood definitions are shown graphically by the red neurons. | 13 |
| Figure 2.5 | Illustration of how the weights of each neuron in a 2-D SOM project into a 2-D space. If the circular space shown in (a) is full of ordered pairs used as data for an SOM, then the 4x4 neuron grid (b) may have weights (in the same 2-D space of the circle) that are positioned as in (c). | 14 |
| Figure 3.1 | Three images, B-mode ultrasound (a), peak displacement (b), and recovery time (c), given in ARFI data for a hard inclusion of size 2 mm. | 20 |

| | | |
|------------|---|----|
| Figure 3.2 | ARFI images of peak displacement for tissue phantoms with hard inclusions. The top row of images correspond to hard 2 mm inclusions, and the bottom row corresponds to 5 mm inclusions. | 21 |
| Figure 3.3 | ARFI images for peak displacement in tissue phantoms with soft inclusions. The top row of images corresponds to phantom tissues with 2 mm inclusions, and the bottom row corresponds to phantom tissues with 5 mm inclusions. | 22 |
| Figure 3.4 | A 3x3 SOM was trained on the full data set for gel tissue phantoms containing both hard and soft inclusions with cropped images in order to reduce the run time and noise. Shown: (a) index numbers used to refer to individual neurons, (b) the final weights corresponding to each neuron in the trained SOM, and (c) the average of all images clustered to each neuron. Since unused neurons appear as dark blue, it is clear that only two neurons had data associated with them after training. | 25 |
| Figure 3.5 | Each neuron in the SOM trained on data for tissue phantoms was probed to analyze the characteristics of images clustering to each of the two used neurons. (a) Images clustered to neuron 6. (b) Images clustered to neuron 8. Note that images with hard inclusions mapped to neuron 6 while images with soft inclusions mapped to neuron 8. | 26 |
| Figure 3.6 | The trained 3x3 SOMs for the cropped images separated into subsets containing only data for tissue phantoms with hard inclusions (a-b) and with only soft inclusions (c-d). The final weights in the trained SOM for each case are shown in (a) and (c), and the average of all images clustered to each neuron after training are shown in (b) and (d). Unused neurons appears as dark blue. | 28 |
| Figure 3.7 | Images clustered to each neuron that was used in the SOM trained on data for tissue phantoms containing only hard inclusions. The trained SOM mapped all 2 mm inclusions in this data subset to neuron 3 and all 5 mm inclusions to neuron 5. | 29 |
| Figure 3.8 | Images clustered to each neuron that was used in the SOM trained on data for tissue phantoms containing only soft inclusions. The trained SOM clustered the data based on both inclusion size and location. The data clustered to neuron 2 contains data with 2 mm inclusions located to the left and to the middle of the viewing region. Neuron 6 contains data with the inclusion located to the farthest right of the viewing rectangle, one with a 2 mm inclusion and one with a 5 mm inclusion. | 29 |
| Figure 3.9 | A SOM was trained on the data set for tissue phantoms with the three added images approximating tissue with no plaque. The final weights are shown in (a), and the averages of all the data mapped to each neuron after training is shown in (b). Unused neurons are displayed in dark blue. The trained SOM partitioned the data based on plaque stiffness (hard, soft, and no plaque). | 31 |

| | | |
|-------------|---|----|
| Figure 3.10 | A SOM was trained on the data set for tissue phantoms after each image was cropped to only show a region of interest near the plaque. The final weights associated with each neuron are shown in (a), and the averages of the data associated with each neuron in the trained SOM are shown in (b). The SOM sorted this data set based on the primary distinction of hard versus soft inclusion. | 32 |
| Figure 3.11 | A SOM was trained on the data set for tissue phantoms with hard inclusions cropped to only show the region of interest. The final weights associated with each neuron in the trained SOM are shown in (a), and the averages of all the data mapped to each neuron in the SOM after training is shown in (b). Three neurons (1, 3, and 4) had phantom data mapped to them after training. . . . | 33 |
| Figure 3.12 | Images clustered to each neuron that was used in the SOM trained on data for tissue phantoms cropped to the region of interest and containing only hard inclusions. All three images corresponding to a 2 mm inclusion mapped to neuron 1. The image containing the 5 mm inclusion located the most to the right (of the 5 mm inclusions) mapped to neuron 3, while the other two images corresponding to 5 mm inclusions mapped to neuron 4. | 34 |
| Figure 3.13 | A SOM was trained on the data set for tissue phantoms with soft inclusions cropped to only show the region of interest. The final weights associated with each neuron in the trained SOM are shown in (a), and the averages of the data mapped to each neuron in the SOM after training is shown in (b). Three neurons (2, 6, and 8) had phantom data mapped to them after training. . . . | 35 |
| Figure 3.14 | Images clustered to each neuron that was used in the SOM trained on data for tissue phantoms cropped to the region of interest and containing only soft inclusions. Two pieces of phantom data, both with 5 mm inclusions located to the left and right of the viewing window, were mapped to neuron 2. Two pieces of data with 2 mm inclusions (located in the middle and to the right of center in the viewing window) and a piece of data with a 5 mm inclusion located in the middle of the viewing window were mapped to neuron 6. The piece of phantom data with a 2 mm inclusion located to the left of the viewing region mapped to neuron 8. | 36 |
| Figure 3.15 | Images of each type in the rectangular manufactured data set (detailed in Table 3.2) to be trained in a SOM with a 3x3 lattice. | 39 |
| Figure 3.16 | Details of the plaque in the initialization scheme for each of the weights associated with each of the nine neurons to be trained in a SOM with a 3x3 lattice. | 40 |
| Figure 3.17 | The weights associated with each neuron in the 3x3 lattice were initialized to vary plaque size and stiffness in order to represent a range of possible plaque patterns that may be randomly generated for the elliptical manufactured data set. The size and stiffness of the plaques in the initialization increase from top left to bottom right. | 41 |

| | | |
|-------------|--|----|
| Figure 3.18 | A SOM was trained on the manufactured rectangular data set. The final weights associated with each neuron of the SOM are shown in (a), and the averages of the data mapped to each neuron after training the SOM are shown in (b). There were four neurons with manufactured data clustered to them after training. | 42 |
| Figure 3.19 | A SOM was trained on the 27 manufactured elliptical images with the weight initialization shown in Figure 3.15. The final weights associated with each neuron in the trained SOM are shown in (a) and the averages of the data clustered to each neuron after training the SOM are shown in (b). Five of the nine neurons (1, 3, 4, 7, and 8) had elliptical manufactured data clustered to them after training the SOM. | 43 |
| Figure 4.1 | Examples of B-mode images for the ex vivo data set are shown: (a) a control, (b) a false positive for a calcium deposit, (c) a clearly visible internal elastic lamina, (d) an artery with a lipid pool, a fibrous cap and a calcium deposit. . | 48 |
| Figure 4.2 | Original B-mode ultrasound images for diseased arteries in the Porcine data set. Each image is labelled with the pattern number I-X and distinguishes the different tissue samples. | 51 |
| Figure 4.3 | ARFI peak displacement images for the diseased arteries in the Porcine data set whose B-mode images are shown in Figure 4.2. Images are labeled with the pattern numbers I-X that distinguish tissue samples. | 52 |
| Figure 4.4 | The diseased porcine data was expanded with $\sigma = .8$ and then used to train a 3x3 SOM. The weight initialization for each neuron is shown in (a). The averages of the disease porcine data mapped to each neuron after training are shown in (b). All nine neurons were utilized, and there were no mismatches. The disease patterns that clustered to each neuron after training are shown in (c). | 53 |
| Figure 4.5 | The diseased porcine data was expanded with $\sigma = .8$ and then used to train a 3x3 SOM. The weight initialization for each neuron is shown in (a). The averages of the disease porcine data mapped to each neuron after training are shown in (b). All nine neurons were utilized, and there were no mismatches. The disease patterns that clustered to each neuron after training are shown in (c). | 54 |
| Figure 4.6 | The diseased porcine data was expanded with $\sigma = .8$ and then used to train a 3x3 SOM. The weight initialization for each neuron is shown in (a). The averages of the disease porcine data mapped to each neuron after training are shown in (b). All nine neurons were utilized, and there were no mismatches. The disease patterns that clustered to each neuron after training are shown in (c). | 55 |
| Figure 4.7 | The control patterns were added to the diseased porcine data that was expanded with $\sigma = 1$ and then used to train a 3x3 SOM. The weight initialization for each neuron is shown in (a). The averages of the disease porcine data mapped to each neuron after training are shown in (b). All nine neurons were utilized, but there were mismatches in the control data. The patterns that clustered to each neuron after training are shown in (c). | 56 |

| | | |
|-------------|---|----|
| Figure 4.8 | The porcine images were cropped in half to focus on the region of interest in the lower portion of the artery wall. The weight initializations are shown in (a). The averages for the data associated with each neuron in the trained 3x3 SOM is shown in (b). All nine neurons had porcine data associated with them. The patterns clustered to each neuron after training the SOM are shown in (c). | 57 |
| Figure 4.9 | New images after the automated cropping technique was applied to focus on the region of interest. Each of the patterns in Figure 4.3 was cropped so that the row containing the peak of the plaque in the lower arterial wall moved to the first row of the image. Pattern III contained considerable noise and therefore did not crop as well in the automated method. | 59 |
| Figure 4.10 | The porcine data was automatically cropped based on the peak of the plaque above the bottom wall to focus on the region of interest located below the artery wall and then used to train a SOM. The data expansion here used $\sigma = .2$. The final weights associated with each neuron (a) and averages of data associated with each neuron after training the 3x3 SOM (b) is shown. | 60 |
| Figure 4.11 | The SOM was trained on seven of the ten disease porcine patterns of data (I., III., IV., V., VI., VII., X.). The final weights (a) and averages of data (b) associated with each neuron in the trained 3x3 SOM is shown. | 62 |
| Figure 4.12 | The SOM was trained on seven of the ten patterns of data (II., IV., V., VI., VII., VIII., X.). The final weights (a) and data averages (b) for each neuron in the trained 3x3 SOM is shown. | 64 |
| Figure 4.13 | A SOM was trained on seven of the ten patterns corresponding to diseased porcine data (I., II., III., IV., VII., VIII., X.). The final weights corresponding to each neuron in the trained SOM (a) and the averages of the data clustered to each neuron after training (b) are shown. | 65 |
| Figure 4.14 | A SOM was trained on each possible combination of six of the ten diseased porcine patterns. The final weights associated with each neuron after training (a) and the averages of the training data clustered to each neuron after training (b) is shown for the following six training patterns (and their perturbed expansions): I., III., IV., V., VII., IX. | 66 |
| Figure 4.15 | The final weights associated with each neuron after training a SOM are shown in (a), and the final location of each of the six original disease patterns used during the training phase is shown in (b). Each of the six original patterns used in the training clustered to their own neuron. | 67 |
| Figure 4.16 | A comparison of final weights associated with each neuron in the trained 3x3 SOM shown in Figure 4.14 and the validation patterns mapping to the previously unused neuron 5. | 68 |
| Figure 4.17 | Comparison of final weights associated with each neuron in the trained SOM (a) and the validation pattern that mapped to neuron 6 (b). The absence of a noticeable vertical bar of low pixel-values does not lend the pattern to map to the first column (on the left) of final weights in (a). The peak shown in (b) is smooth and a relatively small arterial area is present, mapping the pattern to neuron 6. | 69 |

| | | |
|-------------|---|----|
| Figure 4.18 | A sorted bar graph of the uniformly distributed variable r , the rate at which the randomly selected original disease pattern is weighted in the process of creating a new image for the expanded data set. | 70 |
| Figure 4.19 | Bar graphs showing how many times each of the original disease patterns was used as the basis for a newly created image, (a), the closest data image D_{d_2} in the weighted average (b), the second closest data image D_{d_3} in the weighted average (c), and the third closest data image D_{d_4} in the weighted average described in Table 2.2. | 71 |
| Figure 4.20 | The table in (a) shows the final position of each of the original patterns chosen for the training set. The averages of the training data mapped to each neuron after the SOM was trained is shown in (b). | 72 |
| Figure 4.21 | The table in (a) shows the final position of each of the original patterns chosen for the training set. The number of images mapped to each neuron after training is shown in (b). The averages of the training data mapped to each neuron after the SOM was trained is shown in (c). | 74 |

Chapter 1

Introduction

1.1 Atherosclerosis

Heart attacks and strokes are among the leading causes of death in the United States for both men and women (WHO fact sheet no. 317, Feb. 2007) [40, 71]. One cause of these life-threatening events is atherosclerosis, a condition in which plaque composed of fatty substances, cholesterol, cellular waste products, calcium, and fibrin accumulates along the wall of an artery. Plaques are relatively stiff so that plaque build-up can lead to hardening and a loss of elasticity in the artery. Large lipid pools, thin or disrupted fibrous caps, and soft lipid cores are material characteristics that make plaques vulnerable [3, 25, 49, 67].

Atherosclerosis often progresses without notice because it may be asymptomatic for decades; however, proper testing can diagnose atherosclerosis as early as childhood. If detected early enough, lifestyle changes such as exercise, diet, and quitting smoking can slow or stop the progress of the disease before expensive or invasive medical action is required. Plaque can build up in arteries throughout the body and can remain a serious medical threat unless detected.

Figure 1.1 shows the progression of a healthy artery to an advanced stage of disease. A healthy muscular artery has three layers: the inner layer or tunica intima, the middle layer or tunica media, and the outer layer or adventitia. A monolayer of endothelial cells lines the tunica intima and is in contact with blood. The tunica media contains smooth muscle cells that are embedded in a complex extracellular matrix. Arteries damaged by atherosclerosis often contain areas of demarcated laminae in this layer of the artery, evidenced by layers of elastin between strata of smooth muscle cells [37, 73].

The initial stages of atherosclerosis involve a variety of changes to the structural composition of the arterial wall. Leukocytes in the blood adhere to the monolayer of cells lining the tunica intima, bound leukocytes begin to move into this first layer, monocytes begin to mature into macrophages, and the uptake of lipids begins creating foam cells.

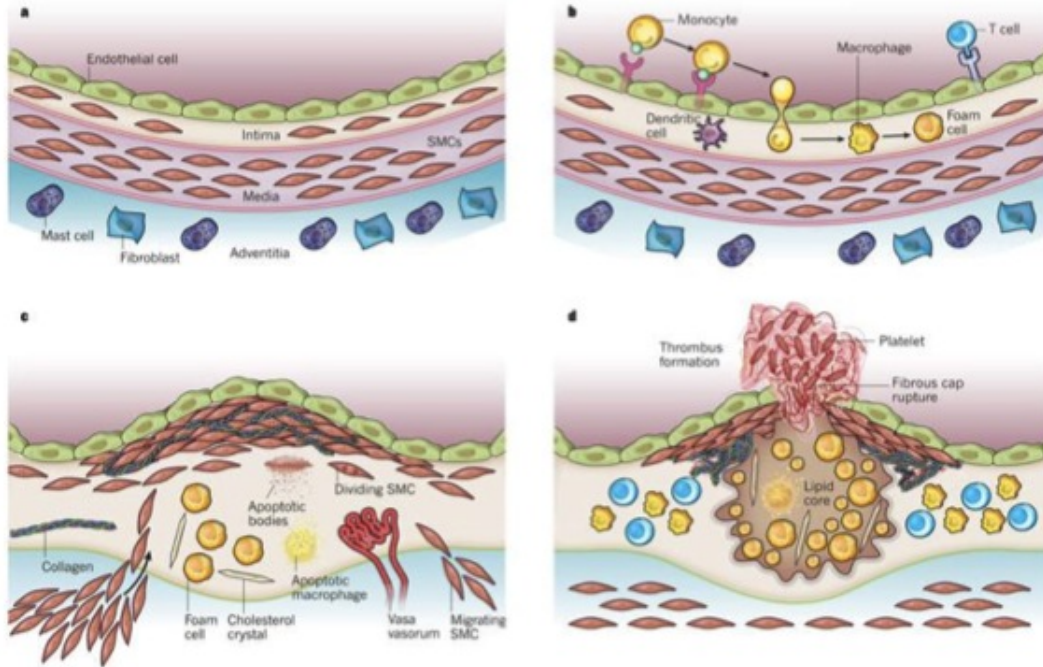


Figure 1.1: Strokes are one possible effect of atherosclerosis. A healthy artery is shown in (a). The initial phase of atherosclerosis is shown in (b). The lesion progression is shown in (c), and thrombosis is shown in (d). (Reprinted from Nature, Peter Libby, Paul Ridker, and Goran Hansson, Progress and challenges in translating the biology of atherosclerosis, 317-325, Copyright (2011), with permission from Nature.)

The disease continues to evolve as the smooth muscle cells from the tunica media move to the tunica intima. Smooth muscle cells in the intima begin to proliferate as well. Extracellular matrix macromolecules such as collagen, elastin, and proteoglycans also begin to be produced at heightened levels. As the lesion advances, some macrophages and smooth muscle cells can die, and these dead cells can accumulate in a central region of a plaque, forming a lipid, or necrotic core.

Thrombosis (more commonly referred to as a blood clot) is the most advanced stage of the disease. At this stage, a physical disruption of the atherosclerotic plaque may occur, causing the plaque to detach from the arterial wall and enter the blood stream. Plaques still attached to the wall at this stage may impede blood flow when the fibrous cap breaks and allows blood coagulation components to interact with the contents of the lipid core.

Classical imaging methods are proficient at detecting stenotic plaque that narrows the vessel lumen but fail to accurately detect nonstenotic plaques [3, 26, 38, 68]. Therefore, increased

knowledge of plaque composition and its risk for detaching from the arterial wall has shifted the focus of clinical methods away from traditional procedures and toward those able to report data on plaque composition and those capable of detecting stenotic and nonstenotic plaques alike. New imaging techniques being explored include X-ray computed tomography (CT) [68], cardiovascular magnetic resonance imaging (MRI) [82], integrated backscatter (IB), virtual histology (VH) methods, intravascular ultrasound (IVUS) [8, 10, 51], and acoustic radiation force impulse imaging (ARFI) [2, 3, 4, 16, 77]. These methods focus more on characterizing plaque composition than on lumen characteristics, such as flow resistance and severity of obstruction. Since thin fibrous caps, large lipid cores, and high macrophage contents are markers for at-risk plaques, these methods aim to identify plaque composition as well [84]. Some traditional methods, such as angiography, and new methods, such as IVUS elastography, carry inherent risk to the patient due to their invasive nature [10, 15, 66].

1.2 Acoustic Radiation Force Impulse (ARFI) Imaging

Acoustic Radiation Force Impulse (ARFI) imaging is one new technique being explored by several labs. ARFI relies on scattering and absorbing properties of acoustic waves that are propagating in tissue [56, 59]. Soft tissue generally absorbs acoustic waves; however, if acoustic frequencies are quickly varied (increased), the waves in the tissue become out of phase with the acoustic wave because the tissue cannot respond quickly enough to the change in positive and negative pressures [56, 59]. Relatively high intensity impulses excite the tissue for a short duration of time in the region to be imaged [3]. ARFI utilizes a clinical ultrasonic imaging scanner equipped with standard clinical imaging transducers and emits two pulses, an excitation pulse and then a tracking pulse, which are emitted over a short time [3]. These pulses compress the tissue in a specified region, and ARFI then measures the recovery of the tissue from the compression force by tracking tissue displacements. The raw ultrasound images are translated into ARFI images of peak displacement or a recovery time that is commonly defined as the time it takes the tissue to return to 67% of its pre-compressed height. It is hypothesized that ARFI technology can identify dangerous arterial plaque not detected by conventional B-mode ultrasound imaging because ARFI imaging can provide additional information about plaque stiffness, the presence of calcium deposits, and the relative composition of collagen and elastin [3, 69]. ARFI has also been shown to effectively delineate tissue structure in human as well as pig arteries [14, 76, 83]. This new technology is noninvasive and allows for small amounts of ultrasonic energy to be focused into very precise tissue locations.

The magnitude of this force can be modeled as:

$$|F| = \frac{2\alpha I}{c} \quad (1.1)$$

where F is a body force, α is a tissue absorption coefficient that depends on frequency, I is the acoustic intensity, and c is the sound speed (1540 m/s) in the tissue [56, 62].

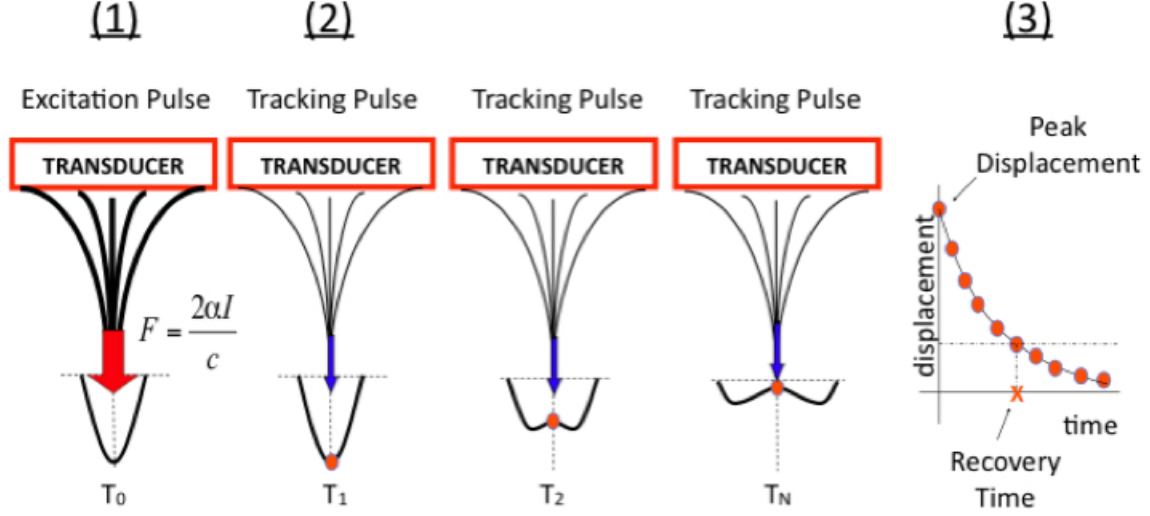


Figure 1.2: Illustration of ARFI imaging procedure in the Gallippi Ultrasound Lap at UNC [19]. This image depicts how the acoustic waves interact with the tissue and how the tracking pulse records tissue behavior.

ARFI research has been directed in three main arenas. These include using (zero-dimensional) ODE viscoelastic models, spatio-temporal models of wave propagation, and utilizing data mining techniques [9, 27, 39, 45, 46, 47, 56, 64]. Viscoelastic tissue properties have been estimated using ARFI technology, but challenges arise in large three-dimensional computations mostly due to tissue inhomogeneities [39]. While shear-wave propagation is used to estimate the Young's modulus, the relatively small displacements of the tissue pose difficulties for calculations [57]. Several labs are working on applying acoustic pulse technology to imaging tissues and organs such as the prostate, breast mass, tumors, the liver, atherosclerotic plaques, the gastrointestinal tract, lesions, abdominal organs and aortic aneurysms, thrombosis, myocardium, and muscle degeneration [1, 3, 7, 9, 11, 15, 17, 27, 29, 50, 58, 60, 61, 63, 79]. Research has also been conducted that uses ARFI to identify atherosclerosis in arteries which also leads to research focused on minimizing background noise and noise created by blood flow during imaging [3, 4, 5, 6, 16, 20, 21, 22, 23, 68, 74, 76, 80, 82]. While some of these studies have implemented K-Means clustering on temporal response patterns, they have not explored using Self-Organizing Maps for spatial pattern recognition of ARFI images to probe for specific plaque characteristics and predict characteristics of new images [46, 47].

1.3 Machine Learning based on Self-Organizing Maps

Artificial Neural Networks (ANNs) are a class of machine learning algorithms that are inspired by biological features of neuronal networks in the brain [43, 70]. Within the context of machine learning, there are two main types of learning: supervised learning and unsupervised learning. Supervised learning involves the presentation of input data and correct answers (output data) to an algorithm that adjusts weights to reproduce these relationship as accurately as possible. Supervised learning is well suited to applications where the ANN can be trained on a set of known input-output pairs of vectors. The trained ANN is then used to predict output values for new values of the inputs [70]. In contrast, unsupervised learning involves the sorting of data vectors into clusters that have similar features. Clusters are identified via weight vectors that have the same dimension as the data. Data contained in clusters that are far away from each other have different features when an unsupervised ANN is trained successfully [43, 70]. Self-Organizing Maps (SOMs) are a type of unsupervised ANN that project data onto a lower-dimensional map in order to cluster data according to the prominent features in a data set (detailed in Section 2.2). While SOMs are a clustering algorithm like K-Means, SOMs include a neighborhood function that ensures two neighboring clusters in a trained SOM will have more similarities than two non-neighboring clusters (detailed in Section 2.2). Several studies have been completed with ARFI imaging that implement K-Means clustering on temporal response data [46, 47], but SOMs have not been developed and applied to analysis of spatial images in this application area.

SOMs along with ARFI technology may greatly aid in early detection of atherosclerosis. Research has shown that SOMs are useful for clustering data relating to brain tumors, breast cancer, and pancreatic tumors [13, 42, 78, 81, 85]. ARFI technology allows for clear arterial images along with a measurement of arterial mechanical properties, and SOMs identify clusters based on patterns and similarities present in the data set. Therefore, the SOM may be used to sort the large images produced by ARFI imaging.

In this dissertation, specialized clustering algorithms were developed for analyzing ARFI ultrasound imaging data for gel-derived tissue phantoms, manufactured data, and for diseased tissue from porcine iliac arteries. In our application to ultrasound imaging data, SOMs achieve clustering by projecting high dimensional data onto a lower dimensional space of neurons arranged on a two dimensional lattice. In Chapter 2, the basic SOM algorithm is presented and factors relating to efficient implementation for high dimensional data, expansion methods for data sets of real images, and approaches for validation are developed. Two data expansion techniques are considered to increase the number of image samples as well as the variety of patterns contained in the images used to train the SOMs. Two metrics are also proposed in order to determine the quality of the data clustering. In Chapter 3, data sets of increasing complexity

are considered to develop and refine the SOM methods presented in Chapter 2. The first set of data considered is comprised of ARFI peak displacement and relaxation time images for gels constructed to mimic arterial tissues with atherosclerotic plaques (phantoms). In an attempt to improve the success of the trained SOM, two data sets were manufactured to mimic larger data sets with continuous variations in major features. In Chapter 4, the SOM techniques developed and evaluated in Chapters 2 and 3 are used to evaluate SOM-based data clustering for ARFI images of porcine tissues obtained from the Gallippi Ultrasound Lab. This data came from an *ex vivo* ARFI imaging study on porcine iliac arteries. The second data expansion method was then applied to the original disease patterns by taking weighted averages of the original patterns, creating an expanded data set with more realistic and continuous variations in plaque features relative to the first expansion method. In order to identify when a training set is correlated with optimal SOM clustering, the statistic developed in Chapter 2 is used to evaluate clustering capabilities of SOMs across all training sets.

Chapter 2

ANN Methods based on Self-Organizing Maps

2.1 Introduction to Artificial Neural Networks

Artificial Neural Networks (ANNs) are machine learning algorithms which are biologically motivated by the way the brain operates and learns. The neocortex, the most highly developed section of the brain, contains about 100,000 neurons (closely interconnected nerve cells) [70]. The dendritic tree of a specific neuron sums the output from the surrounding, connected neurons. If the sum (of the electric potential of each neighboring neuron) is above a certain threshold, a short electrical spike is produced and carried out along a pathway, affecting the synapses of other nearby neurons. Not only do nearby neurons have an electrical effect on their thousands of neighboring neurons, but the strength of the synaptic connections between neurons changes based on the firing of two connected neurons [43].

The first models based on neural networks date to the early 1940s. In 1943, McCulloch and Pitts introduced the idea that neurons could be treated in a binary sense—on or off [48]. Each neuron possessed a threshold, a pre-specified number of input channels, and an output channel. The neuron changes from the off state (0) to the on state (1) if the sum of the input channels (also in binary on or off states) reached the neuron’s threshold. This work was important for demonstrating that neurons may carry out logical operations but did not explain the formation of neuron connections during learning and depended on error-free functioning [70].

In 1949, Hebb expanded the theory presented by McCulloch and Pitts to include changes in neuron connections proportional to the activity of the connection receiving an input from a neighboring neuron and the output connection to other neighboring neurons [28, 70]. The Hebb Hypothesis is still implemented in learning models. One formulation of the hypothesis for the change in synaptic strength is as follows:

$$\Delta w_i = \epsilon \cdot y(\mathbf{x}) \cdot x_i \quad (2.1)$$

where the strength of the i^{th} synapsis is w_i , \mathbf{x} is a vector of the receiving inputs of the neuron where x_i is the input of the i^{th} synapse, $y(\mathbf{x})$ is the excitation of the neuron, and $\epsilon > 0$ is the size of a single learning step [70]. From there, the theory has expanded, and rapidly so since the advent of the modern computer.

Analysis of data using ANN models occurs in two phases. During the training phase, unknown weights in the network are iteratively changed until a set of pre-specified criteria are achieved. Once values of the weights are finalized, the ANN is “trained” and may be used in applications such as classification or predictive analysis. This stage is commonly referred to as validation. ANN training is also classified as either supervised or unsupervised, although hybrid approaches have also been used. In supervised learning, the data under consideration is partitioned into features or components that are inputs and those that are outputs. During the training phase, ANN weights are then varied by minimizing a cost function comparing predicted ANN output values to those in the training data set. Once trained, the supervised ANN is used to predict outputs for new input values. By contrast, unsupervised learning is more commonly used for sorting or clustering data. During unsupervised training, weights are adjusted based on inherent relationships between the data vectors and the weights themselves. Once trained, the unsupervised ANN sorts or classifies the data vectors into clusters. In validation, new data vectors presented to the ANN can then be classified into a representative cluster that was identified during the training phase. Due to these features, ANN models based on unsupervised learning also fall into the category of clustering algorithms, many of which have origins in field other than neural networks.

In general, clustering algorithms identify groups of data vectors that are “near” each other and, as a result, have similar features. One of the oldest and most commonly used clustering algorithms is K-Means clustering. In this algorithm, weight vectors have the same dimension as data vectors and one weight vector represents each cluster to be identified. The user sets the initial value of each of the weight vectors and, for each data vector, distances between that data vector and the weight vector associated with each cluster are computed. As each data vector is presented during training, the nearest weight vector (“the winner”) is moved closer to the piece of data. This process continues until a stop criterion is reached, at which point the weight vector for each cluster ends up “roughly” in the middle of the set of data vectors that are now associated with that cluster. When viewed in the context of ANN models, a prominent property of K-means clustering is that there is only one winner as each data vector is presented during the training phase. For example, the potential for weight vectors representing clusters near the (single) winner to be moved in the direction of the data vector is not considered.

The self-organizing map (SOM) is an algorithm that is similar to K-means, but also takes into account the aforementioned feature during training. While each weight vector is still associated with a data cluster, the weights are associated with neurons in a map (the SOM) in which the notion of neighbors is introduced. This additional feature is biologically motivated from the brain and the definition of neighboring neurons is based on the concept of nearest distance within an arrangement of these neurons into a SOM. Sense organs full of receptors such as the skin, eyes, and ears are connected to neuron maps in the cortex of the brain. These topographic maps preserve a (sometimes distorted) image of the receptors. Receptors close to each other on the skin, for example, cause the firing of neurons close to each other in the cortex. In this way, spatial proximity is preserved in the neuron map. Studies on the somatosensory cortex of monkeys, on the optical tectum of owls, and auditory cortex of bats have studied these maps in greater detail [32, 33, 75]. Relative to K-Means clustering, Self-Organizing Maps add the biologically inspired concept of neighborly influences to the process of iteratively adjusting weights during the training phase.

2.2 Introduction to Self-Organizing Maps

SOMs are a type of ANN based on Kohonen’s network model in which the algorithm maintains a spatial structure in the lower dimensional space comprised of neurons, incorporating neighboring relationships, while clustering data vectors that reside in a space of higher dimension [34, 35, 36]. Once a SOM has been trained, this means that clusters associated with neighboring neurons contain data that is more similar than for clusters far away from each other within the map topology. An illustration of how SOMs cluster by projecting the data onto a lower dimensional space of neurons is shown in Figure 2.1 for data that is two-dimensional and a SOM that is a one-dimensional arrangement of five neurons. The neurons are connected thus defining neighbor relationships and allowing topological relationships to be preserved throughout the training process. Each neuron of the SOM has an associated weight vector residing in a vector space of equivalent dimension to that of the data (Figure 2.2). This is called the weight space for the neurons (Figure 2.2). During training, the SOM algorithm cycles through the data vectors one at a time. For each chosen vector, the neuron (in the weight space) closest to that data vector is chosen as a winner. By analogy, the winning neuron is the one that would fire in the brain as a response to a sensory receptor being stimulated on the body. During SOM training, the winning neuron as well as its near neighbors then move their weight vectors in weight space towards the data vector that was presented to the SOM (Figure 2.3). Learning rate parameters, z and ζ (Table 2.1), specify how much the winning weight vector and its near neighbors move toward the presented data vector in the weight space. The full SOM algorithm is summarized in Table 2.1.

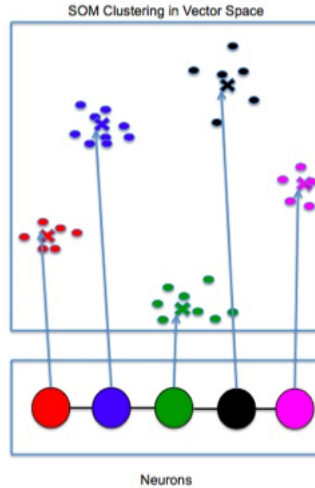


Figure 2.1: An illustration of how SOM training projects higher dimensional data (upper box) onto a lower dimensional neuron space comprised of five neurons (lower box). Note that if the data is projected down onto a one-dimensional neuron space, then the order of the clusters (color-coded) should be preserved. In this way, data mapped to neighboring clusters should be more similar than data mapped to neurons farther apart.

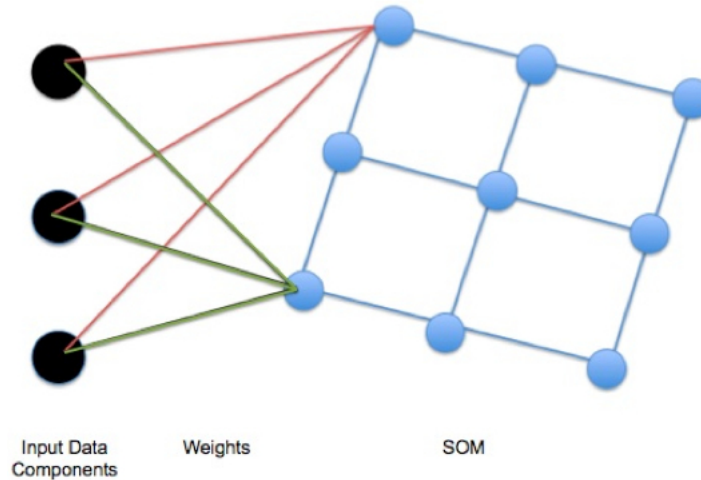


Figure 2.2: An illustration of relationships between the data space, weight space and neuron topology in a two-dimensional Self-Organizing Map. Each neuron in the SOM (blue) has an associated weight vector that is the same dimension as the data. For instance, if the data has three dimensions (black), then each of the nine neurons in the SOM will have a 3-dimensional weight vector associated with it.

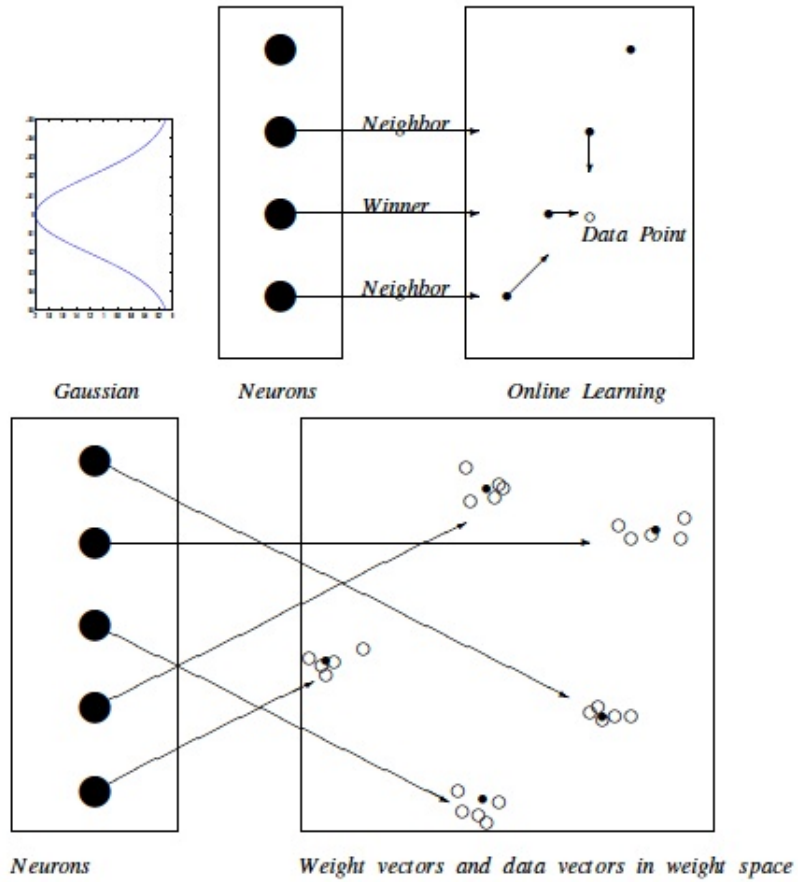


Figure 2.3: Illustration of the training process for a 1-d SOM with five neurons arranged in a one-dimensional line. Here, neighbors are determined to be neurons one away from the winning neuron. The weights of the winning neuron and neighboring neurons move closer to the data point presented (open circle, top). After training, each neuron corresponds to a cluster of data points (below).

Table 2.1: General Algorithm for Training of 2-D Self-Organizing Maps

General Algorithm:

1.) Chose the topological arrangement for the neurons ($n \times m$) and denote the total number of neurons as $N = n * m$.

2.) Initialize the weight vector $\mathbf{w}_j, j = 1, \dots, N$, associated with each neuron.

3.) For each data vector, \mathbf{d} , presented, determine the nearest weight vector, \mathbf{w}_{j^*} , by computing the Euclidean distance between the presented data vector and each weight vector \mathbf{w}_j . The neuron associated with the nearest weight vector is described as “winning”.

4.) For each data vector, \mathbf{d} , presented, move the winning weight vector in the direction of the presented data vector using the relation:

$$\mathbf{w}_{j^*} = \mathbf{w}_{j^*} + (z) * (\mathbf{d} - \mathbf{w}_{j^*}) \quad (2.2)$$

5.) Identify a set Q of indices for weight vectors of the neurons that are topological neighbors of the winning neuron. For $k \in Q$, move the weight vectors for these neighboring neurons in the direction of the presented data vector using the relation:

$$\mathbf{w}_k = \mathbf{w}_k + (\zeta) * (\mathbf{d} - \mathbf{w}_k) \quad (2.3)$$

6.) Update the two learning rate parameters z and ζ .

$$z = (\alpha) * z^{\frac{k}{k_{max}}} \quad (2.4)$$

$$\zeta = (\alpha) * \zeta^{\frac{k}{k_{max}}} \quad (2.5)$$

where k represents the current iteration count through steps 3–6, and α is chosen to be positive and less than one.

7.) Repeat steps 3–6 until the maximum number of iterations, k_{max} , is reached.

There are a number of parameters for the user to prescribe in the SOM algorithm. First, the dimension of the topological space of the neurons and the number of neurons are specified (Step 1 in Table 2.1). The neighborhood structure for the topological neuron space requires specification (for Step 5). For instance, in a two-dimensional neuron grid, near neighbors may either be defined as those neurons one away by the taxi cab metric in the neuron space or as the neurons one or two away by the taxi cab metric (Figure 2.4). (The taxi cab metric with neighbors

two away will include those neurons “diagonally” away from the winning neuron as neighbors on the grid.) A variation of SOM that also moves the neurons unidentified as a winner, or secondary neighbors to the winner, away from the data vector being presented may also be considered (between Steps 5 and 6). Specification of online versus batch learning is also required. In online learning, Steps 4 and 5 happen immediately after each data vector is presented; however, in batch learning, Steps 4 and 5 occur after the nearest weight vector has been determined for each data vector in the set. The stopping criteria for the SOM algorithm is the last parameter to be prescribed. Termination of the iterations of the algorithm occurs when the specified maximum iteration, k_{max} is achieved (Step 7). Alternatively, an SOM algorithm may be constructed such that termination occurs when the weight vectors change by less than a specified tolerance. For the the learning rate parameters, the user must specify the percentage, z , that winning weight vectors will move relative to the distance between the data and winning weight vector (Equation 2.2). A separate percentage, ζ , representing the shift of near neighboring weight vectors must also be set as a parameter (Equation 2.3). This neighborhood function causes SOM to differ from K-means and also leads to the preservation of nearness in corresponding regions of both the data/weight space and the geometrical arrangement of neurons in the SOM. As the algorithm progresses, both of these parameters will decrease at a chosen rate, α , at the end of each iteration (Equations 2.4 and 2.5). Since the data presented here comes from medical imaging, there are a large number of pixels recorded for each image, making the computations in Step 3 slow.

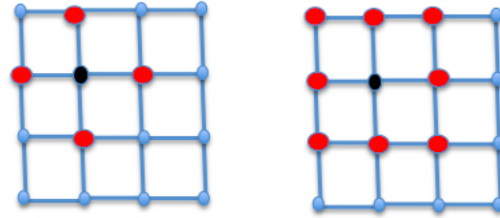


Figure 2.4: Possibilities for defining the neighborhood function. For the two-dimensional neuron space as shown we could consider neighbors to be those neurons one grid step away or those one grid step away and “diagonals”. The relevant neuron is shown in black, and the two possible neighborhood definitions are shown graphically by the red neurons.

It is clear that the consideration of neighbors in Step 5 of the SOM training algorithm is the main difference between K-Means and SOMs. While both algorithms locate clusters present in the data, SOMs rely on the structure of the neuron space during the training process. This neighbor structure ensures that the data associated with neighboring neurons will have

more in common than the data associated with non-neighboring neurons. When this concept is incorporated into SOMs in two (or higher) dimensions, the connection to K-Means is less clear. While both K-Means and SOM also leave room for interpolated patterns based on the data presented since it is not required that every neuron have an associated weight vector that is near a cluster of data, the neighborhood feature of the SOM ensures that the interpolated patterns have similarities with the nearby neurons. The difference between K-Means and SOMs is illustrated in Figure 2.5.

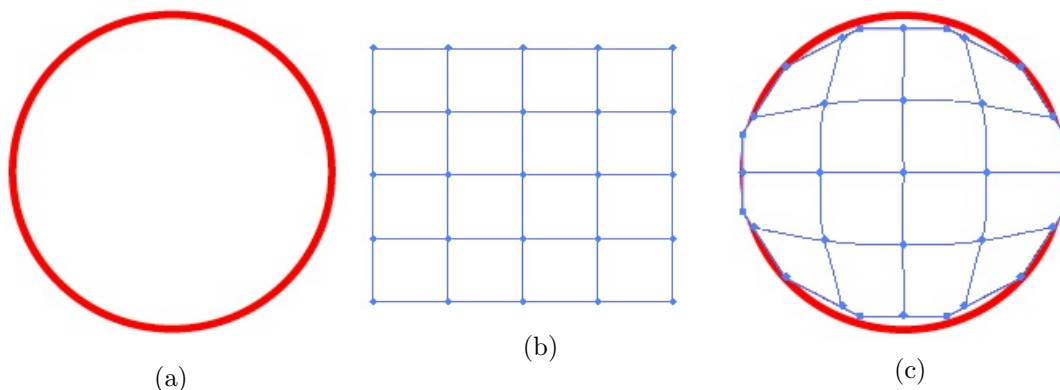


Figure 2.5: Illustration of how the weights of each neuron in a 2-D SOM project into a 2-D space. If the circular space shown in (a) is full of ordered pairs used as data for an SOM, then the 4x4 neuron grid (b) may have weights (in the same 2-D space of the circle) that are positioned as in (c).

2.3 Weight Initializations

The data clusterings resulting from SOM training are sensitive to the initialization of the weight vectors in Step 2 due to the SOM's dependence on the position of these weight vectors relative to the presented data vectors. Consequently, random initialization of weight vectors is not a wise strategy for most practical applications of SOMs to data clustering. Some effort is then needed to identify an initialization that can be generalized for the data set and that produces consistent and useful sorting within each application area.

The weight initialization should reflect the scope of the data since the training ends after a pre-determined amount of time. (Other termination criterion, for example, might signal the end of the loop when the weight vectors move less than a specified distance after one iteration through the data.) There are several ways to ensure this. One simple way to ensure that the weight vectors are spaced throughout the range of training data is to sort the data and assign

the initial weight vectors to evenly spaced (in terms of the sorted array) data vectors. Another way to achieve weight vectors appropriately spaced in the data space is to use the average values of sorted data vectors. This method is particularly helpful when there are less original data vectors than neurons. When the scope of the data is known, this is also fairly easy to implement because weights may be constructed in a manner that reflects the range of the major distinguishing features of the data.

In order to reduce the sensitivity of the results to the weight initializations, the majority of the results on one particular type of data set were computed using the same scheme for weight vector initialization. The weight initializations for certain data sets were changed, however, to specifically address the unique features in the data. For example, weight initializations were hand-selected and not automated for the manufactured data sets (Section 3.3) because the scope of the data was well-defined and known. The weight vector initializations for data obtained from ARFI scans, on the other hand, were created by sorting the given data set and selecting data vectors with increasing norms. Overall, in most applications a priori knowledge of the scope and features of the data set is sufficient to identify a weight initialization scheme that adequately distributes weight vectors within the data space prior to the commencement of SOM training.

2.4 Data Expansion

The ARFI image data considered in this study comes from either gels created in the lab that are designed to mimic tissues (phantoms), or from ultrasound imaging of arteries in sacrificed pigs. Typically, the number of neurons used in a SOM is on the order of the total number of unique patterns associated with the specific application. In all data sets considered in this work, only a small number (often one) image was available for each unique pattern. Consequently, data sets were expanded prior to training to enhance the robustness and predictive capability of the trained SOMs. The first approach we used perturbed each pixel of the original data pattern using a normal distribution where the original value of the pixel, which ranged from 0 to 25, was the mean and σ was generally equal to .2. Sensitivity to the choice of σ was evaluated by training on data sets using increasing values of σ in the expansion. Its value was found to have relatively little impact until σ was about 1.5, so large that expanded patterns lacked resemblance to a biologically plausible pattern. After running multiple experiments with different croppings and subsets with this data, we determined that a richer expansion was needed to demonstrate the full power of using a SOM with this data, and, therefore, a second method of data expansion was considered.

The second method for expanding the diseased porcine data set created more varied patterns that are still medically plausible. In order to achieve this, new images were added to the data set by taking weighted averages of existing disease images rather than perturbing individual

pixels of a single image. We first created a random array \vec{a} of 90 integers between 1 and 10, where 10 was chosen since it was the number of unique disease patterns in the data set. Then, an array \vec{r} of 90 random real numbers chosen from a uniform distribution between 0.8 and 1.0 was constructed to provide values for construction of new images via a weighted average. For each integer in the first array a_j , a new piece of disease data is created in the following manner:

Table 2.2: General method of expansion of data based on weighted averages of original disease patterns

Let the matrix D be comprised of rows that are the ten original disease patterns. Let D_k denote the k^{th} row of D .

For each integer a_i and weight r_i :

- 1.) Choose D_{a_i} as the base pattern to be used as the basis for creating a new pattern.
- 2.) Compute the Euclidean distance between D_{a_i} and all other original patterns $D_{j,:}$, keeping track of all the computed distances.
- 3.) Sort the computed distances, and save the indices of the sorted distances in the array \vec{d} .
- 4.) Create one new data pattern by using a weighted average of disease images as follows:

$$NewPattern = r_i \cdot D_{d_1} + \frac{1-r_i}{2} \cdot D_{d_2} + \frac{1-r_i}{3} \cdot D_{d_3} + \frac{1-r_i}{6} \cdot D_{d_4}$$

An expanded data set created on the basis of the technique in Table 2.2 has several advantages. The primary advantage is that new patterns constructed as weighted averages of real images create new images that are more realistic than those obtained via random perturbations. This approach is also consistent with the progression of atherosclerotic disease in a spatially and temporally continuous manner. Furthermore, by choosing values of the weights r closer to 1, we have an idea of which original pattern the new piece of data will most closely resemble, which can then be used to evaluate the success of a given clustering.

2.5 Measuring the Success of a Clustering

After completion of the SOM training algorithm, clusters are computed by mapping each data vector to the nearest weight vector (same as Step 3). Each vector in the data set is then associated with the neuron associated with that nearest weight vector. One way to grade the success of the clustering is to examine the data associated with each neuron at the end of

this final clustering based on the fully trained SOM. A well-trained SOM should give rise to clusters of data with clear common properties or features. Furthermore, if the data set has been expanded, then the majority of the data created through expansion to resemble a particular pattern should cluster to the same neuron.

The number of neurons that have data associated with them after the final iteration of the algorithm is yet another way to determine the success of the SOM training. When training the SOM, the user may have an idea of the number of distinct patterns in the training set based on prior knowledge of that particular set of data. If the trained SOM has data mapping to roughly the same number of neurons as unique patterns in the training set, then the clustering may be counted as a success. For example, a data set may contain six pieces of data but have only two types of unique patterns. If the SOM has data associated with only two of its neurons after training and if the data associated with each neuron all have the same common type, then the clustering may be called successful because the SOM appropriately sorted the data based on the main distinguishing feature. Neurons with data associated with them after training will be referred to as “used” while neurons with no data associated with them after training are referred to as “unused.” It is important to note that there may be more neurons in the map than distinct features or patterns in the data used to train the SOM.

Unused neurons that do not have any data associated with them after training still have associated weight vectors in the same space as the data. These final weights associated with unused neurons are interpolated patterns. Successful clusterings should produce final weights for unused neurons that are biologically plausible due to the topological properties of the SOM algorithm. Specifically, the weight vectors for the unused neurons potentially represent realistic patterns that were not present in the data set used for training the SOM. For example, a SOM may be trained on images representing ten types and varying degrees of atherosclerosis; however, these ten disease patterns should not be expected to exactly represent every possible combination of disease type and severity physically feasible across all subjects. The new patterns associated with unused neurons in the trained SOM may then prove to be closer to images for new subjects relative to the patterns from the original subject population used to train the SOM. One method for quantitatively assessing the SOM’s ability to interpolate useful patterns via unused neurons is to partition data into non-overlapping subsets that can be used for training and validation.

2.6 Training and Validation Data Sets

Large enough data sets are often partitioned into two distinct, non-overlapping subsets: a training set and a validation set. The training set should contain data that fairly represents the scope of the data set and is used to train and thus arrive at final values of the weights in the SOM.

The validation set should contain data corresponding to patterns that are not included in the training set. The trained SOM is then used to cluster data in the validation set and hence determine the winning neuron for each piece of data. Validation data mapping to a neuron classified as used after the training phase should have features that are similar to the training data mapped to the same neuron. However, since the SOM also interpolates new patterns based on weights of unused neurons, some of the newly presented patterns should also cluster to neurons that were unused after SOM training. When data is partitioned into training and validation subsets, the success of the clustering may then be measured by analyzing the total number of unused neurons subsequent to clustering of data in both the training and validation subsets.

In developing SOM algorithms and assessing their capabilities in a specific application area, the choice of an ideal or optimal training set is not well understood. Also, the data available to train a SOM may be limited or constrained due to considerations such as cost or privacy. As a result, it is beneficial to analyze the effectiveness of SOMs as the size and composition of training sets are varied. In order to determine the most effective way to split the data into training and validation sets, all possible combinations of choosing six of the original ten data patterns (and any expanded data belonging to those chosen patterns) were used to train and then validate a SOM. For each run c of a possible combination of $\binom{10}{6}$, the number of used neurons after the training phase of the SOM was recorded in the array u , and the number of previously unused neurons that had validation data cluster to them was recorded in the array n . A basic rule of thumb to find optimal partitions of the data defined by most used neurons after both training and validation becomes:

$$\max_{c \in \binom{10}{6}} u_c + n_c \quad (2.6)$$

This formula may be refined, though, to include the condition that the number of used neurons after training on any set of data for six original patterns, u_c , should not exceed six, since this would suggest that the final clustering of the training data set would map data based on one type of original pattern to more than one neuron after training. This training mistake is more likely to happen in the case where the data expansion was performed using the second method described in Section 2.4 of weighted averages.

Chapter 3

SOM Clustering for Tissue Phantoms and Manufactured Data

3.1 Gel-Based Tissue Phantoms

Gel-based materials with mechanical properties that are tunable in a controlled manner can serve as “phantoms” that mimic tissues for the purposes of evaluating imaging techniques. Agar/gelatin formulations containing type-A gelatin (Acro Organics, Geel, Belgium), agar (Fisher Scientific, Fair Lawn, NJ, USA), Photo-Flo (Eastman Kodak Co., Rochester, NY, USA), *n*-propanol, powdered graphite, 1:10 buffered formalin, and de-ionized water, ingredients from published recipes, were created in the Gallippi Ultrasound Lab at UNC [3, 41]. Specific elastic modulus values in the agar/gelatins were created by varying the amount of agar while keeping other ingredients constant. A ~ 3 mm layer (in the axial dimension) was included in each gel to represent an arterial wall based on the range of human carotid plaque thicknesses [18, 72]. The background tissue was created to be a ~ 20 mm layer beneath the region representing the arterial wall. Two inclusion sizes with width 2.5 mm or 5 mm in the lateral dimension were created with two different stiffnesses per inclusion size. Soft inclusions were manufactured to have an inclusion stiffness of ~ 110 kPa embedded in an arterial wall with stiffness ~ 190 kPa while hard inclusions were manufactured with a stiffness of ~ 190 kPa and embedded into arterial walls with stiffness ~ 110 kPa.

In this chapter, we demonstrate that the SOM methods outlined in Chapter 2 can effectively sort patterns in the data set of images of tissue phantoms manufactured as outlined above. This set of images was produced by applying ARFI imaging techniques, using a Siemens SONOLINE AntaresTM (Siemens Medical Solutions USA, Inc., Ultrasound Division) on gel tissue phantoms that mimic atherosclerotic tissue. Twelve 2055x40 data images were created by imaging four phantom gels (one for each combination of inclusion stiffness and size) with the inclusion lat-

erally offset 0, -3, and -6 mm from the center of the VF7-3 transducer mounted on an optical air table (Newport Corp., Irvine, CA, USA). The data were analyzed using a SOM with nine neurons arranged on a 3x3 lattice. This data set contained only two basic patterns—soft or hard inclusion. The variation in plaque stiffness caused the most easily detectable difference among the phantom images because it alters the values of more pixels per image than the size variation.

3.1.1 ARFI images of tissue phantoms

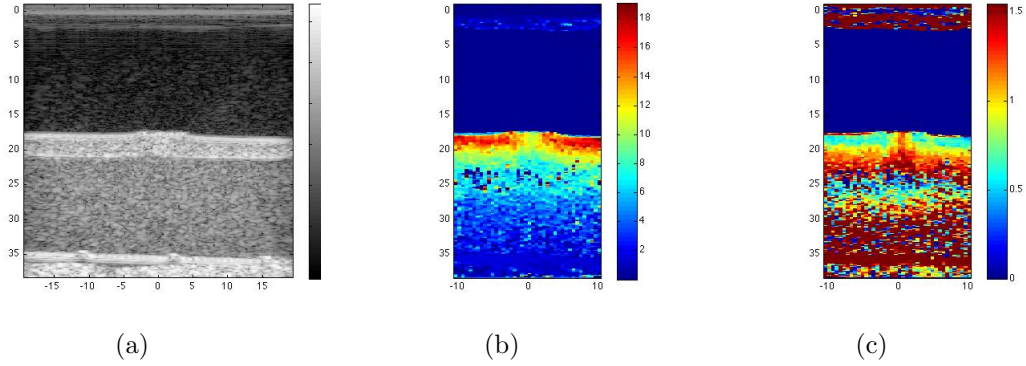


Figure 3.1: Three images, B-mode ultrasound (a), peak displacement (b), and recovery time (c), given in ARFI data for a hard inclusion of size 2 mm.

Figure 3.1 shows the three types of images that are collected during ARFI imaging. Figure 3.1(a) depicts a B-mode ultrasound scan of a phantom artery with a 2mm hard inclusion in the middle of the viewing area. The depth and lateral position of the inclusion are typically reported in millimeters. The image in Figure 3.1(b) shows the same region in terms of peak displacement, where every pixel represents the maximum displacement over time, in microns, from the ARFI push registered at that point in space. Since a higher maximum displacement would mean that the tissue is relatively soft, soft regions show up in red, denoting a higher registered value of tissue displacement. Alternatively, regions of stiff tissue are colored in light green and yellow. The inclusion can be seen as the yellow and (light) green region breaking the horizontal red region in the middle of the viewing area in Fig. 3.1b. It can be determined that this is a hard inclusion because the peak displacement values in and around the plaque are very low compared to the other arterial wall regions. It should be noted that since the artery is full of blood, its interior does not compress much (if any) during the push and appears to have a value of zero (shown in navy blue). The third ARFI image shown in Figure 3.1(c) depicts the time it takes (in milliseconds) for the tissue to recover to 33 % of the maximum displacement. This recovery

time which measures how long it takes to return 67 % of the tissue's natural position pre-ARFI push can capture both elastic and viscoelastic effects. Faster recovery times are typically associated with high tissue elastic shear modulus, low tissue density, whereas slower recovery times correlate with higher levels of internal viscoelastic damping. In Fig. 3.1c, the red region where the plaque is located suggests that possible properties of the plaque may include a low shear modulus, higher density or increased viscoelastic properties relative to the surrounding tissue (light blue). Taken together, data for both the peak displacement and recovery time suggest that the plaque has a high stiffness, high density and increased viscoelasticity relative to the surrounding tissue.

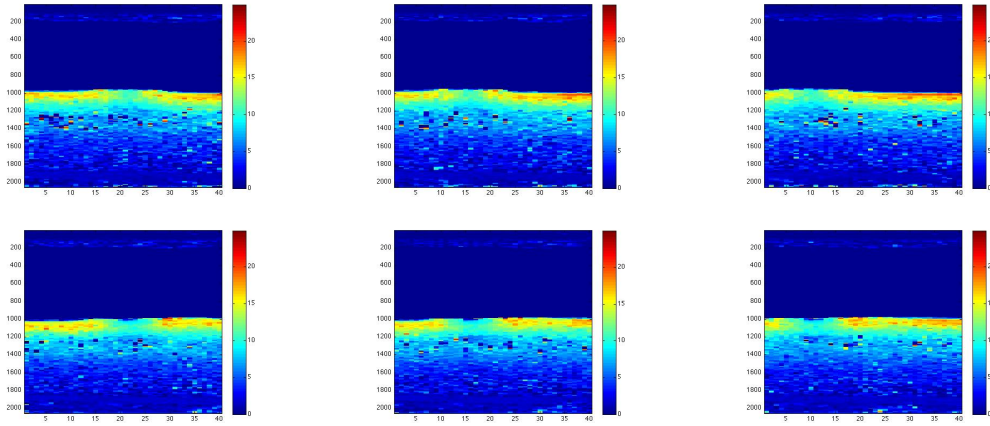


Figure 3.2: ARFI images of peak displacement for tissue phantoms with hard inclusions. The top row of images correspond to hard 2 mm inclusions, and the bottom row corresponds to 5 mm inclusions.

Figure 3.2 shows the ARFI images for the phantom data set that include hard inclusions of size 2 mm and 5 mm. The position of the inclusion is also slightly varied in the lateral direction. The plaque is located on the bottom surface of the phantom artery in each image (roughly in the middle of each image vertically). In the set of hard inclusions, the lower arterial wall has a relatively high displacement everywhere except at the site of the inclusion.

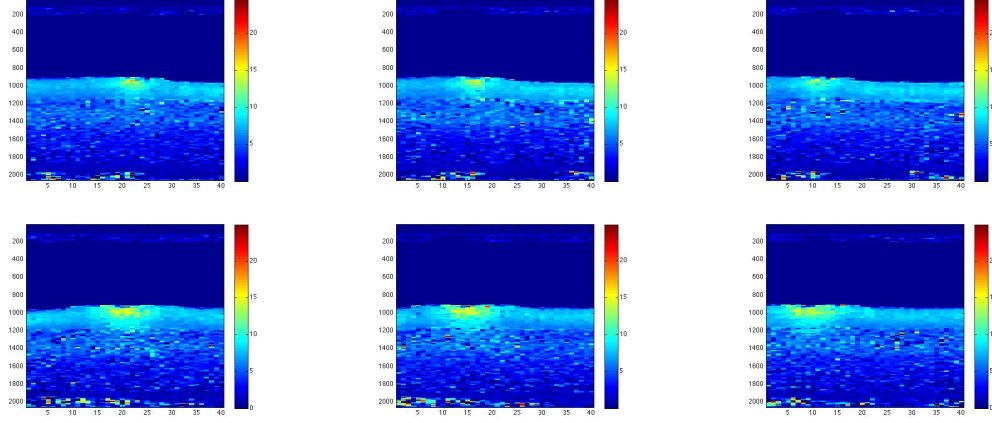


Figure 3.3: ARFI images for peak displacement in tissue phantoms with soft inclusions. The top row of images corresponds to phantom tissues with 2 mm inclusions, and the bottom row corresponds to phantom tissues with 5 mm inclusions.

The ARFI soft inclusion peak displacement data for tissue phantoms is shown in Figure 3.3. Again, there are two inclusion sizes, 2 mm and 5 mm, with laterally varying locations. These inclusions are also along the lower arterial wall (roughly in the center of the image vertically). The coloration shows that the inclusions are softer than the surrounding arterial wall. In both Figures 3.2 and 3.3 some noise is present in the lower half of the image.

3.1.2 Application of SOMs

The ANN modeling techniques described in Section 2.2 were implemented for the case of a 3x3 SOM comprised of nine neurons arranged on a grid. Close neighbors were defined as those neurons one grid interval away in the horizontal or vertical direction, with diagonal neighbors being those located across the diagonal within the same grid cell. During training, close neighbors of winning neurons were moved closer to the presented data vector than the diagonal neighbors in a ratio that reflects the difference in distance along the 3x3 grid between near neighbors and diagonal neighbors. Since the size of the phantom data set is relatively small (12 images), the weights were initialized by creating a single marker of one pixel per weight initialization where plaques were roughly located (on the axial axis) in the data images. This initialization method does not bias the data significantly because the size or depth of the plaque is not estimated and since a lab technician would commonly identify an anchor point for the region of interest in the viewing window. Similarly, the weight initializations for the manufactured data sets (to be discussed in Sec. 3.3) were constructed by hand, utilizing knowledge that all plaques were centered at the same point in this manufactured data set and were designed with a fixed range

of stiffness in relation to the background tissue as well as a fixed range of plaque size. The weights were initialized in such a way that the full spectrum of possible features of the data in the manufactured data set were represented.

3.1.3 Vectorization

All code was written for and implemented in MATLAB®(7.10 R2010a). To ensure maximum flexibility and the ability to tailor algorithms, the Neural Network Toolbox was not used but, instead, custom code was developed for all SOMs used in this work. In order to take full advantage of MATLAB®’s capabilities on multicore processors, vectorization was implemented throughout the code. With vectorization, the SOM algorithms easily scaled up to the vectors of relatively large dimension associated with ultrasound imaging, taking in data images with sizes roughly 500x40 to 2055x40 pixels (equivalent to vectors of dimension 20,000 and 82,200, respectively). Note that the images in Figure 3.1 are scaled and do not reflect the true size of the data images seen in Figures 3.2 and 3.3. By feeding in each image as one vector and by taking advantage of MATLAB®’s built-in vector operations, Steps 3-5 in the SOM algorithm (Table 2.1) were accelerated, allowing the code to run in an acceptable amount of time. In MATLAB® vectorized pieces of the code distribute operations to the available cores of the multicore processor. Table 3.4 demonstrates typical differences in computation time for the SOM algorithm on two different multicore processors with and without vectorization.

Table 3.1: The execution time for the training of a SOM and the effects of vectorization. Execution times with vectorization (a) and without vectorization (b) on a dual core processor (MacBook Pro with 1.83 GHz Intel Core Duo, 2GB RAM) and a dual quadcore processor (Mac Pro with dual 2.8GHz quadcore Intel Xeon, 4GB RAM)

| Dual Core | Dual Quadcore | Dual Core | Dual Quadcore |
|-----------|---------------|------------|---------------|
| 45.62 min | 3.46 min | 214.33 min | 60.76 min |
| (a) | | (b) | |

Code was also initially designed to make Step 3 of Table 2.1, the winning neuron search, occur in a parfor loop. However, the overhead setup for the parfor loop of a rather quick computation (computing a norm) resulted in slower run times. Therefore, only the vectorized version of the code described above was implemented for all results presented in this dissertation.

3.2 Trained SOMs for Analysis of Tissue Phantoms

In order to fully explore the data and the capabilities of the SOM to cluster proficiently, based on the methods outlined in Section 2.6, multiple SOMs were trained on different manipulations of the data set for tissue phantoms. Since the data is of very large dimension (roughly 80,000 pixels per image), computation can be further accelerated if unnecessary pixels in peripheral regions of each image are cropped out. This cropping results in there being less work involved in computing the nearest neighbor and in moving the weights associated with neighboring neurons towards the winning weight vector (Steps 3-5 in Table 2.1). Cropping was carried out in two stages. First, all tissue phantom images were cropped to 1205x40 pixels to eliminate the majority of the artery below the region of interest and resulted in cropping out roughly the top half of each image. Later, the data set was cropped further to 500x40 pixels to focus even more closely on the region of interest near the plaque. Code was not constructed to automate the process of identifying the region of interest since, in practice, a lab technician would use their discretion to recognize any regions of interest that should be explored in more depth and via manual section using a rectangular box.

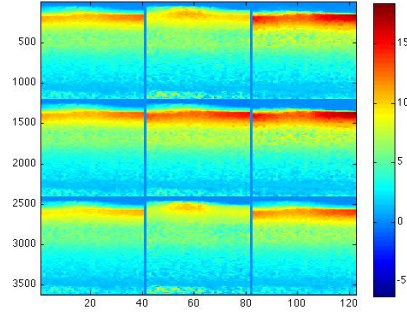
3.2.1 Initial Cropping

SOM clustering results for the data set of tissue phantom images are shown in Figure 3.4. Figure 3.4a shows the numbering system used for the nine neurons in the 3x3 SOM. The images for the final weights and data averages are arranged in this 3x3 pattern. Figure 3.4b shows the final weights of the SOM trained on all 12 phantom ARFI images for peak displacement. During the training phase, the evolving weights corresponding to each neuron serve as representative data vectors for all real data clustered to that neuron, or as interpolated patterns in the case that the neuron is unused. The final weights shown have interpolated patterns corresponding to both soft inclusions and hard inclusions. Each final weight is a possible pattern that could appear in clinical data even if none of the 12 phantom tissue patterns actually map to it.

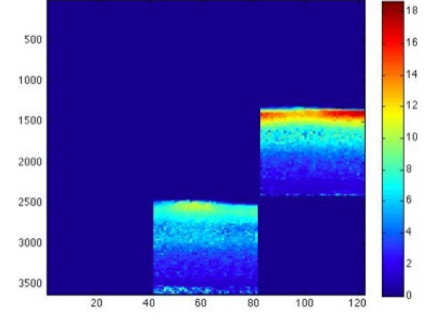
Subsequent to SOM training, the neurons that were used are shown in Figure 3.4c. For each of the two neurons that were used, the pattern representing the average of all images clustered to that neuron is shown. The neurons represented in dark blue were given an average of zero in each pixel to readily identify unused neurons in relating the actual data to the trained SOM. There were only two neurons (neuron 6 and neuron 8) that had training data mapped to them. All of the hard inclusions were mapped to neuron 6 while all the soft inclusions were mapped to neuron 8. Hence, the SOM clustered the data according to the difference in inclusion stiffness. Even though the plaque size and position varied for each inclusion type, the type (hard versus soft) was the primary feature that the SOM used to differentiate the data.

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 4 | 5 | 6 |
| 7 | 8 | 9 |

(a)



(b)



(c)

Figure 3.4: A 3x3 SOM was trained on the full data set for gel tissue phantoms containing both hard and soft inclusions with cropped images in order to reduce the run time and noise. Shown: (a) index numbers used to refer to individual neurons, (b) the final weights corresponding to each neuron in the trained SOM, and (c) the average of all images clustered to each neuron. Since unused neurons appear as dark blue, it is clear that only two neurons had data associated with them after training.

3.2.2 Probing the Trained SOM

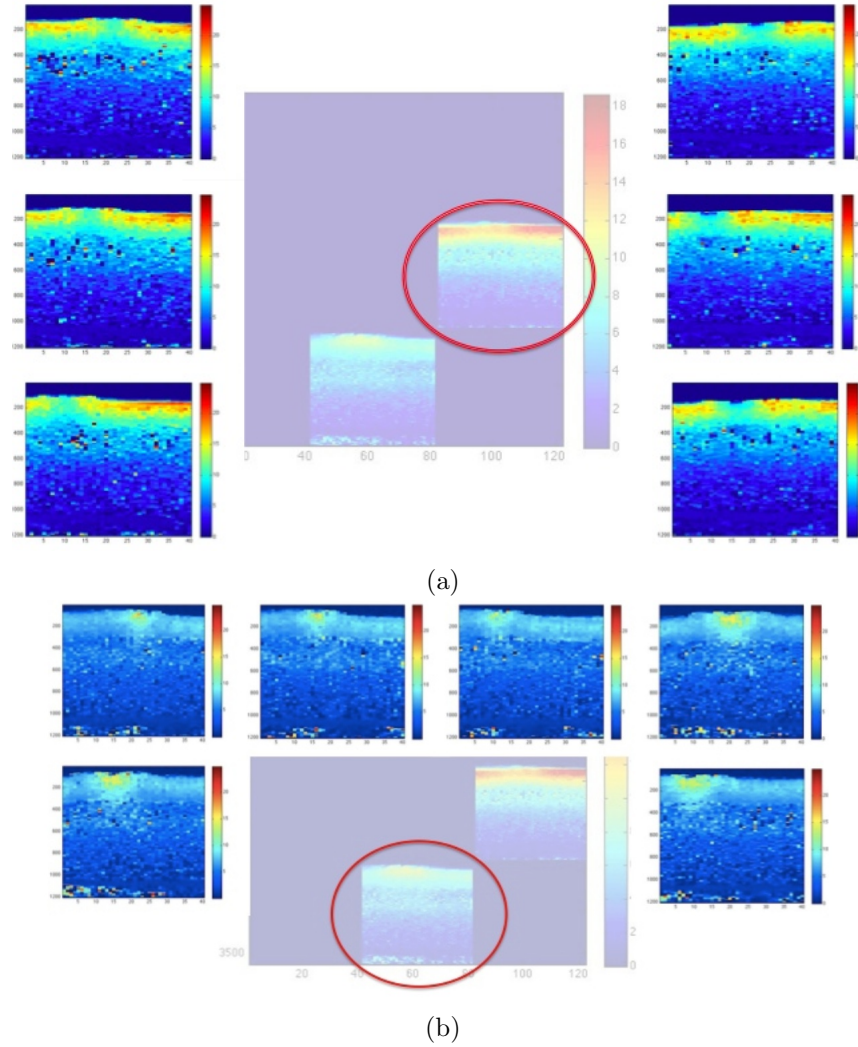


Figure 3.5: Each neuron in the SOM trained on data for tissue phantoms was probed to analyze the characteristics of images clustering to each of the two used neurons. (a) Images clustered to neuron 6. (b) Images clustered to neuron 8. Note that images with hard inclusions mapped to neuron 6 while images with soft inclusions mapped to neuron 8.

By probing each neuron to determine the images clustering to that neuron at the end of the SOM training, it was observed that all data corresponding to hard inclusions mapped to neuron 6, while all data corresponding to soft inclusions mapped to neuron 8. Figure 3.5 also shows each of the six images that mapped to neuron 6 (a) and each of the six images that mapped to

neuron 8 (b). It is observed that the SOM did not make distinctions based on the size of the inclusion (2 mm versus 5 mm) or the lateral location. Instead, the SOM separated the data in a manner that identified the primary feature as being the hardness or softness of the inclusion. This illustrates an attractive feature of unsupervised learning algorithms in that the primary feature for sorting the data was not specified a priori, but is identified during the training process. As mentioned in Section 2.5, we can conclude that the clustering was successful in that the SOM correctly identified two main groups of data based on inclusion stiffness; however, the SOM had seven of nine neurons unused at the end of the training phase, indicating that the feature identified (inclusion stiffness) dominated other features that may have been present in the data set.

3.2.3 Analysis of Data Subsets for Soft and Hard Inclusions

The trained SOM for the full set of ARFI imaging data for tissue phantoms only used two neurons after training, clustering all the soft inclusions to one neuron and all the hard inclusions to the other. To further analyze SOM clustering capabilities, new SOMs were trained on subsets of the data that included tissue phantoms with only hard inclusions and, then a SOM considering tissue phantoms with only soft inclusions. The aim of this retraining was to assess the ability of the SOM to detect the more subtle differences of inclusion size and location within each subset. Results for the two trained SOMs are shown in Figure 3.6. The final weights for the SOM trained on the hard inclusion data subset show only image patterns that are reflective of hard inclusion data (Fig. 3.6a). The data averages for each of the two neurons that were used subsequent to training show that the SOM clustered the hard inclusion data by inclusion size (Fig. 3.6b). All the data in this subset with 2 mm inclusions mapped to neuron 3 while all the data corresponding to 5 mm inclusions in this subset mapped to neuron 5 (Figure 3.7). Similarly, the SOM trained on the subset containing only images for tissue phantoms with soft inclusions interpolated only patterns reflective of soft inclusion images in the final weights associated with each neuron, shown in Figure 3.6c. The data averages for the data mapping to each neuron after training show that the data was clustered on both inclusion size and position (Figure 3.6d). Data corresponding to 2 mm inclusions in the left and middle of the viewing window mapped to neuron 2 while the data corresponding to 5 mm inclusions located to the left and middle of the viewing region mapped to neuron 8 (Figure 3.8). The 2 mm and 5 mm inclusions located the farthest right in the viewing window each mapped to neuron 6 (Figure 3.8). Therefore, the trained SOM for the soft inclusions clustered the data based on both inclusion size and location. Overall, the sequential use of SOMs for the tissue phantom data set demonstrated the capability of delineating image patterns based on primary features including inclusion stiffness, size and location within the image.

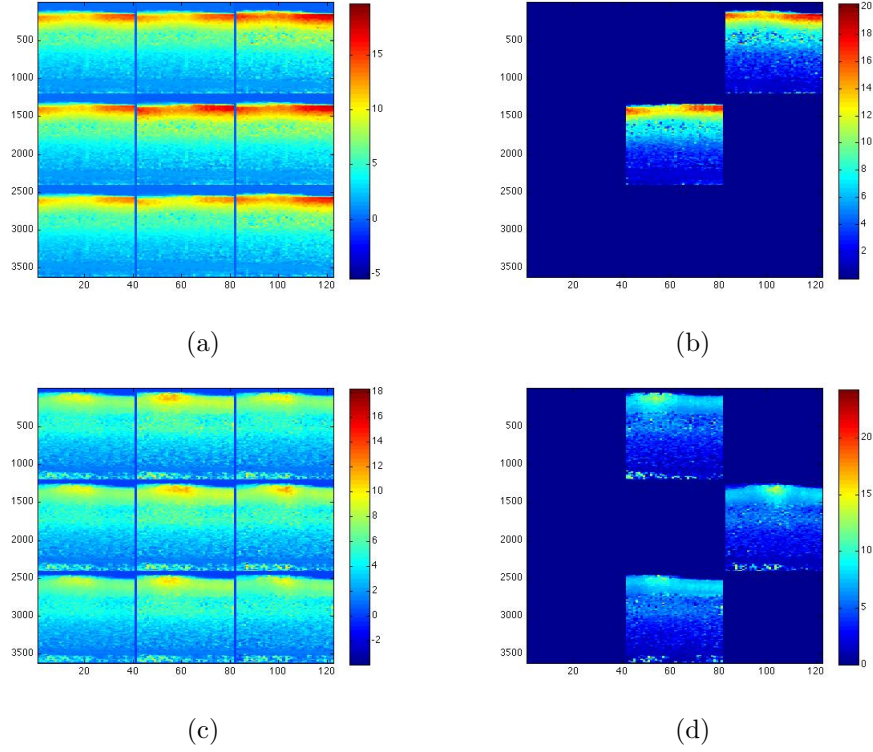


Figure 3.6: The trained 3x3 SOMs for the cropped images separated into subsets containing only data for tissue phantoms with hard inclusions (a-b) and with only soft inclusions (c-d). The final weights in the trained SOM for each case are shown in (a) and (c), and the average of all images clustered to each neuron after training are shown in (b) and (d). Unused neurons appears as dark blue.

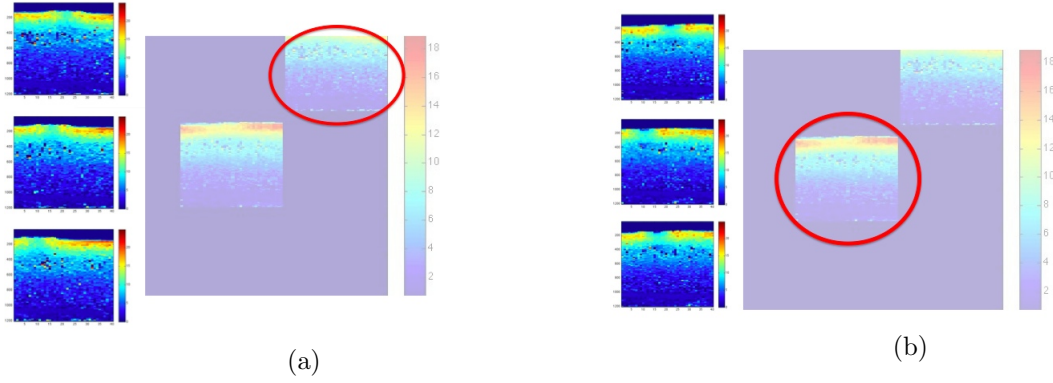


Figure 3.7: Images clustered to each neuron that was used in the SOM trained on data for tissue phantoms containing only hard inclusions. The trained SOM mapped all 2 mm inclusions in this data subset to neuron 3 and all 5 mm inclusions to neuron 5.

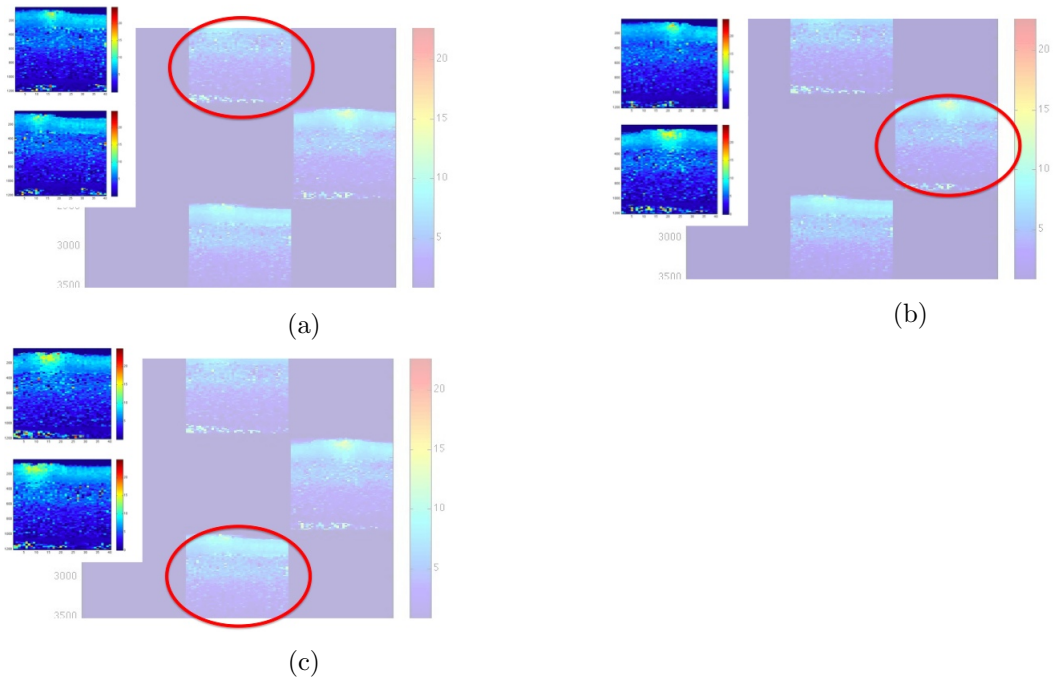


Figure 3.8: Images clustered to each neuron that was used in the SOM trained on data for tissue phantoms containing only soft inclusions. The trained SOM clustered the data based on both inclusion size and location. The data clustered to neuron 2 contains data with 2 mm inclusions located to the left and to the middle of the viewing region. Neuron 6 contains data with the inclusion located to the farthest right of the viewing rectangle, one with a 2 mm inclusion and one with a 5 mm inclusion.

3.2.4 Adding a Plaque-free Pattern

Since the SOM successfully clustered the full data set for tissue phantoms by the main distinguishing features, the data set was augmented with an additional pattern to make it more reflective of patterns for a patient population. Specifically, the new image was a manufactured pattern that serves as a rough approximation of a sample with no plaque. It should be noted that this pattern was not created in the lab and that the nature of its composition was not based on known AFRI images of arterial tissue regions that are free of plaque. Three manufactured pieces of data were used to represent tissue with no plaque and were added to the data set prior to training. The extent to which these new patterns might change the interpolated weights associated with each neuron in the final SOM was evaluated.

Results for clustering using the SOM trained on the expanded data set are shown in Figure 3.9. Even though the two existing patterns for the tissue phantoms were more similar to each other than the new manufactured data, the trained SOM still identified three distinct patterns in the data: hard inclusions, soft inclusions, and the new no plaque pattern (Figure 3.9). Probing used neurons for the clustered data after training, it was observed that all images corresponding to hard inclusions mapped to neuron 6, all data corresponding to soft inclusions mapped to neuron 2, and all of the manufactured data approximating images with no plaque mapped to neuron 8. The interpolated final weights associated with each neuron are quite different than those in Figure 3.4 where the SOM was trained on the data not containing a no plaque image. The interpolated weight vectors (Fig. 3.9a) are observed to exhibit patterns reflecting a range of possible images for patients with little or no plaque present in the left and lower left portions of the trained SOM.

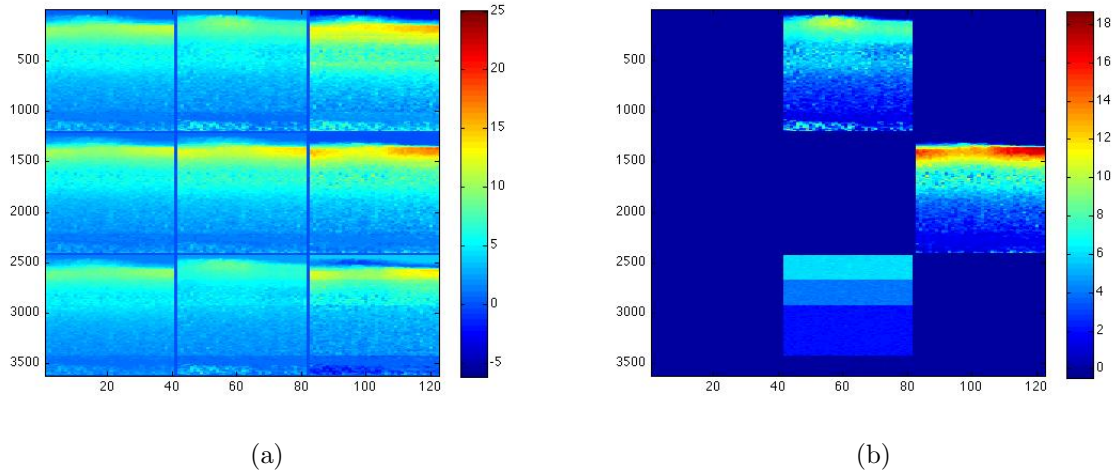


Figure 3.9: A SOM was trained on the data set for tissue phantoms with the three added images approximating tissue with no plaque. The final weights are shown in (a), and the averages of all the data mapped to each neuron after training is shown in (b). Unused neurons are displayed in dark blue. The trained SOM partitioned the data based on plaque stiffness (hard, soft, and no plaque).

3.2.5 Cropping for the Region of Interest

To further reduce computational cost and mimic imaging procedures in a clinical setting, images were cropped down to a region of interest. This manipulation of the data set also reduced the number and variation in values of pixels due to noisy regions of the image located far away from the plaque. Note that the manufactured data approximating images with no plaque were not considered for this training.

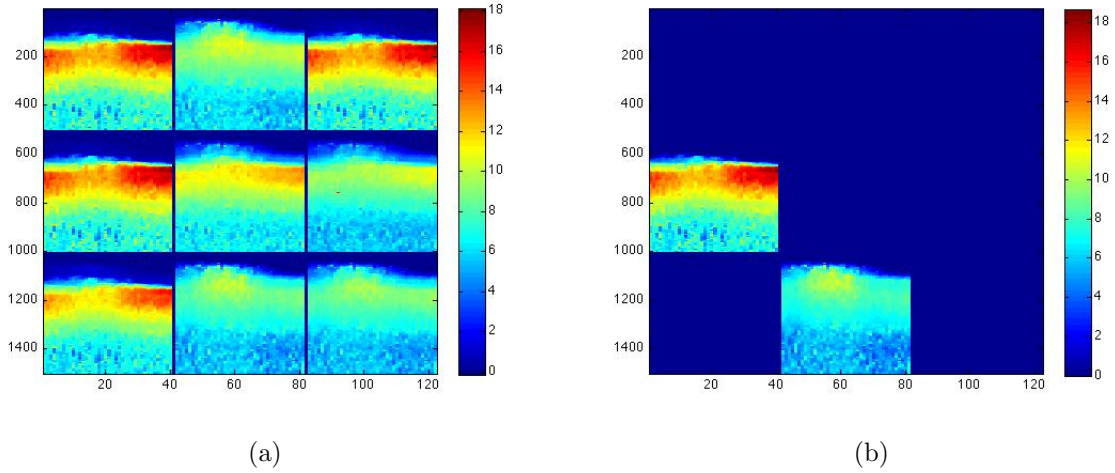


Figure 3.10: A SOM was trained on the data set for tissue phantoms after each image was cropped to only show a region of interest near the plaque. The final weights associated with each neuron are shown in (a), and the averages of the data associated with each neuron in the trained SOM are shown in (b). The SOM sorted this data set based on the primary distinction of hard versus soft inclusion.

The final weights associated with each neuron in the trained SOM and the average values of the data associated with each neuron are shown in Figure 3.10(a) and (b), respectively. It is observed that images corresponding to the interpolated weights reflect a richer set of patterns due to containment to the region of interest in the data set for tissue phantoms. The interpolated patterns exhibit more detailed characteristics of a plaque than the interpolated patterns shown for images covering a wider field of view (Figure 3.4). However, again, as in Figure 3.4, the interpolated final weights associated with each neuron in the trained SOM only show patterns that include a plaque. The interpolated patterns in the final weights move from patterns representing hard inclusions in the top and left (with the exception of neuron 2) to patterns corresponding to soft inclusions in the bottom right. While seven of the nine neurons in the trained SOM are unused, the interpolated patterns shown in the final weights for the trained SOM are reflective of real patterns that may occur in patients. As a result, newly collected data may map to one of these unused neurons when clustered using this trained SOM.

3.2.6 Analysis of Data Subsets for Hard and Soft Inclusions

The data for tissue phantoms cropped to a region of interest was again partitioned into two subgroups: hard inclusions and soft inclusions. The SOM was separately trained on each subgroup. Figure 3.11(a) shows the final weights clustered to each neuron in the trained SOM,

and Figure 3.11(b) shows the averages of the phantom data associated with each neuron after training. Unlike the trained SOM from Figure 3.6 (a-b), the trained SOM used three (instead of two) neurons. The data was clustered in a manner that delineated both inclusion size and location. All three patterns corresponding to 2 mm hard inclusions mapped to neuron 1. Neuron 3 contained the one piece of data corresponding to a 5 mm hard inclusion located to the right, and neuron 7 contained the two pieces of data corresponding to a 5 mm hard inclusion located to the left and middle of the viewing region (Figure 3.12).

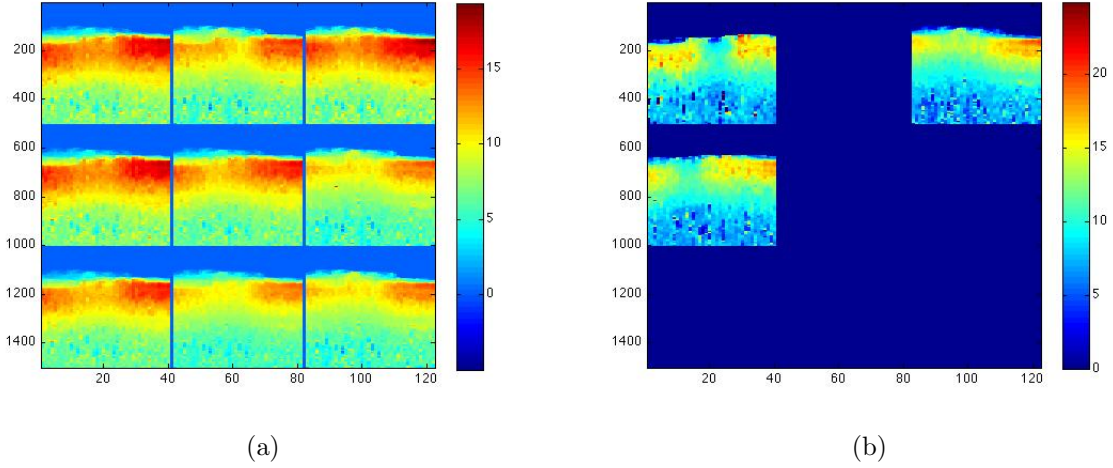


Figure 3.11: A SOM was trained on the data set for tissue phantoms with hard inclusions cropped to only show the region of interest. The final weights associated with each neuron in the trained SOM are shown in (a), and the averages of all the data mapped to each neuron in the SOM after training is shown in (b). Three neurons (1, 3, and 4) had phantom data mapped to them after training.

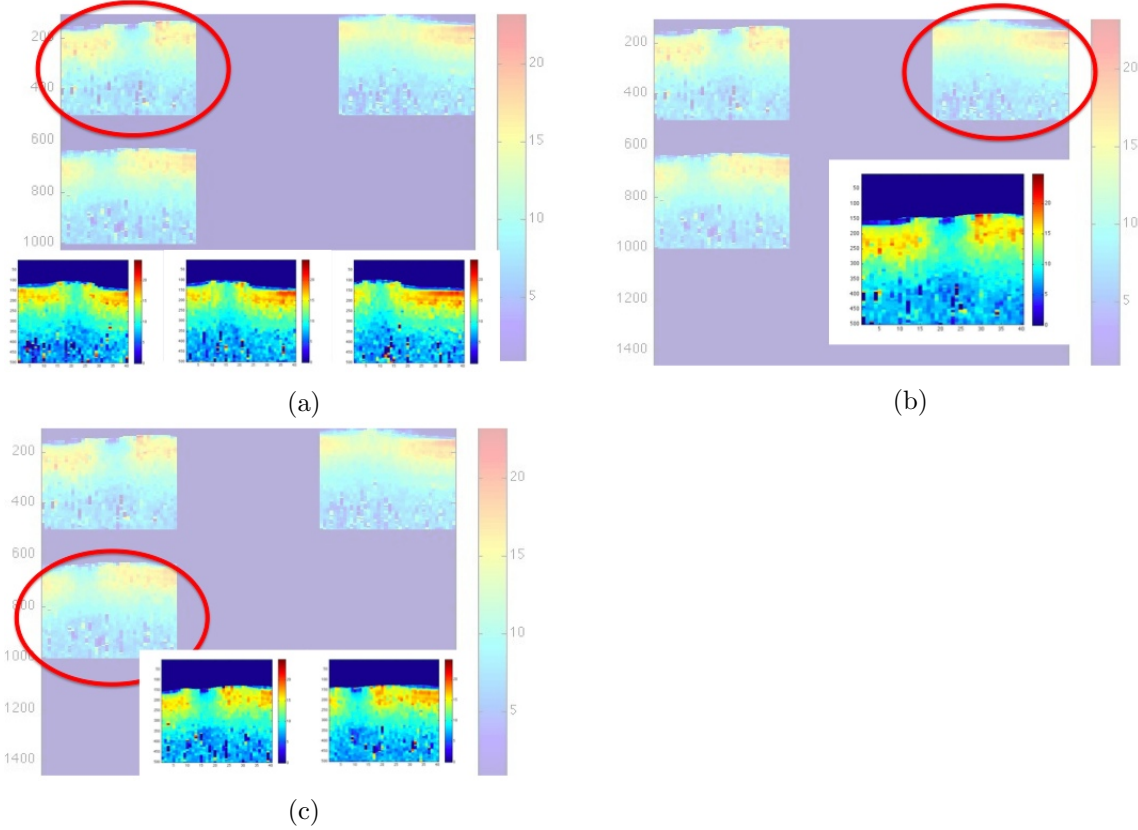


Figure 3.12: Images clustered to each neuron that was used in the SOM trained on data for tissue phantoms cropped to the region of interest and containing only hard inclusions. All three images corresponding to a 2 mm inclusion mapped to neuron 1. The image containing the 5 mm inclusion located the most to the right (of the 5 mm inclusions) mapped to neuron 3, while the other two images corresponding to 5 mm inclusions mapped to neuron 4.

Figure 3.13(a) shows the final weights associated with each neuron in the SOM trained on the soft inclusions cropped to the region of interest, and (b) shows the averages of the phantom data associated with each neuron after training. Figure 3.13(a) shows that the trained SOM only interpolated patterns with soft inclusions (as opposed to hard inclusions or no inclusions or a combination). The size and stiffness of the soft inclusions varies among the interpolated patterns shown in the weights associated with the unused neurons. Similar to Figures 3.6(d) and 3.8, three neurons (specifically neurons 2, 6, and 8) had data associated with them after the SOM training was completed. However, the SOM trained on data for images for soft inclusions cropped to a region of interest produced a different sort. Figure 3.14 shows the images clustered to each neuron after the training was completed. The data associated with neuron 2 were both 5 mm inclusions that were located to the middle and to the left of middle in the viewing region.

Neuron 6 contained two 2 mm inclusions and one 5 mm inclusion that were all centrally located in the viewing region. (The two 2 mm inclusions were located in the middle and slightly to the right of center in the viewing regions.) Neuron 8 had only one piece of phantom data associated with it, a 2 mm inclusion located to the left of the viewing rectangle.

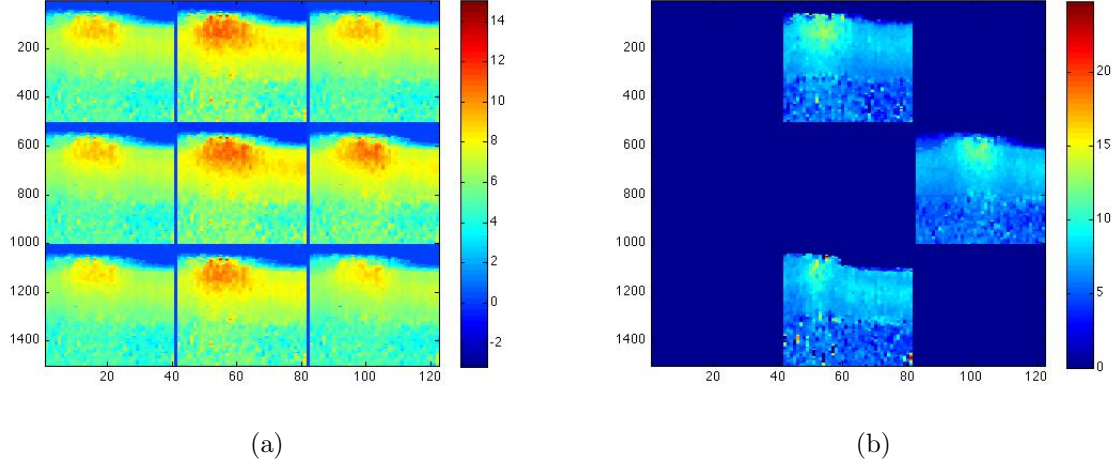


Figure 3.13: A SOM was trained on the data set for tissue phantoms with soft inclusions cropped to only show the region of interest. The final weights associated with each neuron in the trained SOM are shown in (a), and the averages of the data mapped to each neuron in the SOM after training is shown in (b). Three neurons (2, 6, and 8) had phantom data mapped to them after training.

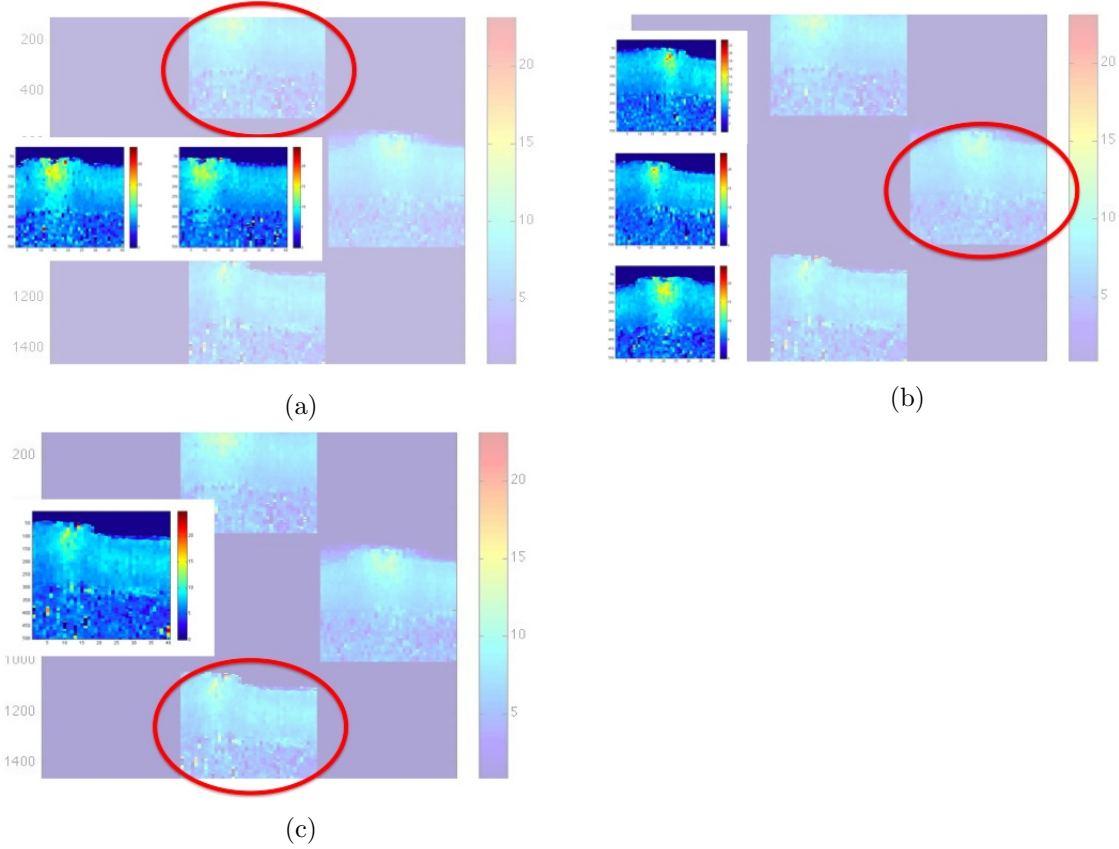


Figure 3.14: Images clustered to each neuron that was used in the SOM trained on data for tissue phantoms cropped to the region of interest and containing only soft inclusions. Two pieces of phantom data, both with 5 mm inclusions located to the left and right of the viewing window, were mapped to neuron 2. Two pieces of data with 2 mm inclusions (located in the middle and to the right of center in the viewing window) and a piece of data with a 5 mm inclusion located in the middle of the viewing window were mapped to neuron 6. The piece of phantom data with a 2 mm inclusion located to the left of the viewing region mapped to neuron 8.

While the SOM clearly delineates the major, distinct features in the phantom data set, a larger and more diverse data set is needed for further model development. Such a data set could contain more realistic patterns that could occur in patients and therefore create more diverse interpolations that correspond to the diverse patterns seen in patients at varying stages in the progression of the disease.

3.3 Introduction to the Manufactured Data Sets

The most neurons a trained SOM used when clustering the data for tissue phantoms was 3, and this occurred only when we trained a SOM on either the subset of all hard inclusions or all soft inclusions (otherwise only two neurons were used). The SOM sorted on the feature of plaque stiffness, as opposed to location, because plaque stiffness affects more pixels in the image than location. The distinct difference between soft and hard inclusions creates much greater distance between these images in the data space than between images of the same plaque stiffness but where the plaque location varies. Therefore, we considered manufacturing data to simulate more continuous transitions in image features and, hence, increase the usage of more neurons in a trained SOM.

3.3.1 Introduction to the Rectangular and Elliptical Data Sets

In order to display the robust nature of the clustering SOMs can offer, two data sets containing 500x40 pixel images with more features were manufactured. First, as a rough proof of concept, a manufactured data set was created to incorporate a small arterial region, tissue, and a rectangular plaque. The size and stiffness of the plaque was varied, but the location of the rectangular plaque was always centered laterally. Since a technician could center the plaque in the viewing region during ARFI imaging, the location of the center of a plaque within the viewing window can be largely controlled. The inverse tissue stiffness was set at 15 (relatively elastic), while the inverse plaque stiffness ranged from 2 to 15, where a “plaque” inverse stiffness of 15 represents no plaque. Nine basic patterns were created, and each of the nine basic patterns was perturbed at each pixel by using the original pixel value as the mean and drawing the perturbation from a normal distribution with $\sigma = .7$ to create three random perturbations of each of the nine basic patterns. The plaque size and stiffness for the rectangular manufactured data can be seen in Table 3.2 and Figure 3.15 while the SOM initialization can be seen in Figure 3.16. The initialization was constructed such that plaque size increased from left to right and from top to bottom and such that the plaque stiffness decreased from top to bottom and was constant along rows with the exception of the pattern corresponding to no plaque.

Table 3.2: This table contains the description of the plaque sizes and inverse stiffnesses in the nine basic patterns created for the manufactured data set. Three perturbations of each of these nine patterns were used to create the full manufactured rectangular data set.

| Plaque Size | Inverse Stiffness of Plaque |
|---------------|-----------------------------|
| 0x0 No Plaque | 15 |
| 15x3 | 12 |
| 10x4 | 12 |
| 30x6 | 9 |
| 100x20 | 2 |
| 150x4 | 5 |
| 25x24 | 11 |
| 5x15 | 13 |
| 50x10 | 9 |

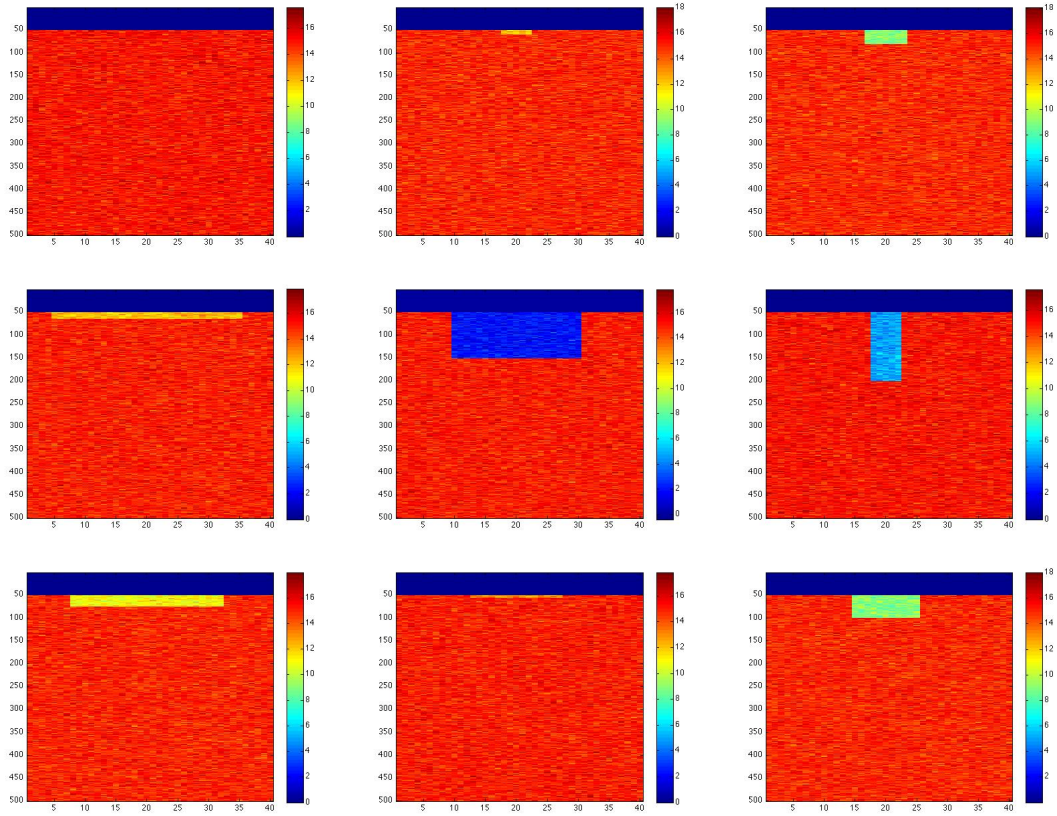


Figure 3.15: Images of each type in the rectangular manufactured data set (detailed in Table 3.2) to be trained in a SOM with a 3x3 lattice.

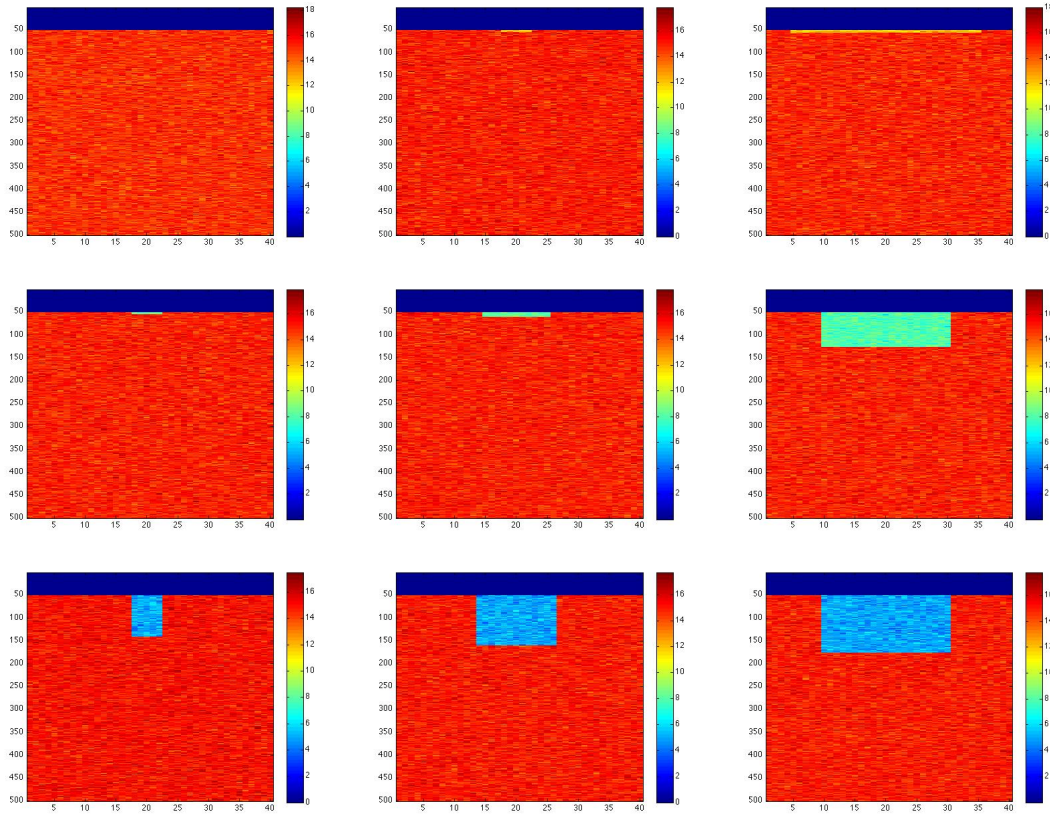


Figure 3.16: Details of the plaque in the initialization scheme for each of the weights associated with each of the nine neurons to be trained in a SOM with a 3x3 lattice.

Next, a manufactured elliptical data set was constructed in order to allow for more flexibility in the 500x40 data created. In this data set, a small region of pixels with value zero in the top of each image corresponds to the small arterial portion of the tissue phantom data cropped to highlight the region of interest. Plaques were represented by randomly generated ellipses and given a variety of “stiffnesses” in relation to the background tissue. The ellipses in the 27 new images were generated by randomly selecting the length of the major and minor axes. Similar to the rectangular manufactured data, each ellipse was centered to control for the position of the plaque laterally. Figure 3.17 shows the initial weights associated with each neuron before the SOM was trained. The initial weights were chosen to represent a number of plaque sizes and stiffnesses possibly presented in the elliptical manufactured data. Plaque size and stiffness increased from top left to bottom right in the 3x3 lattice.

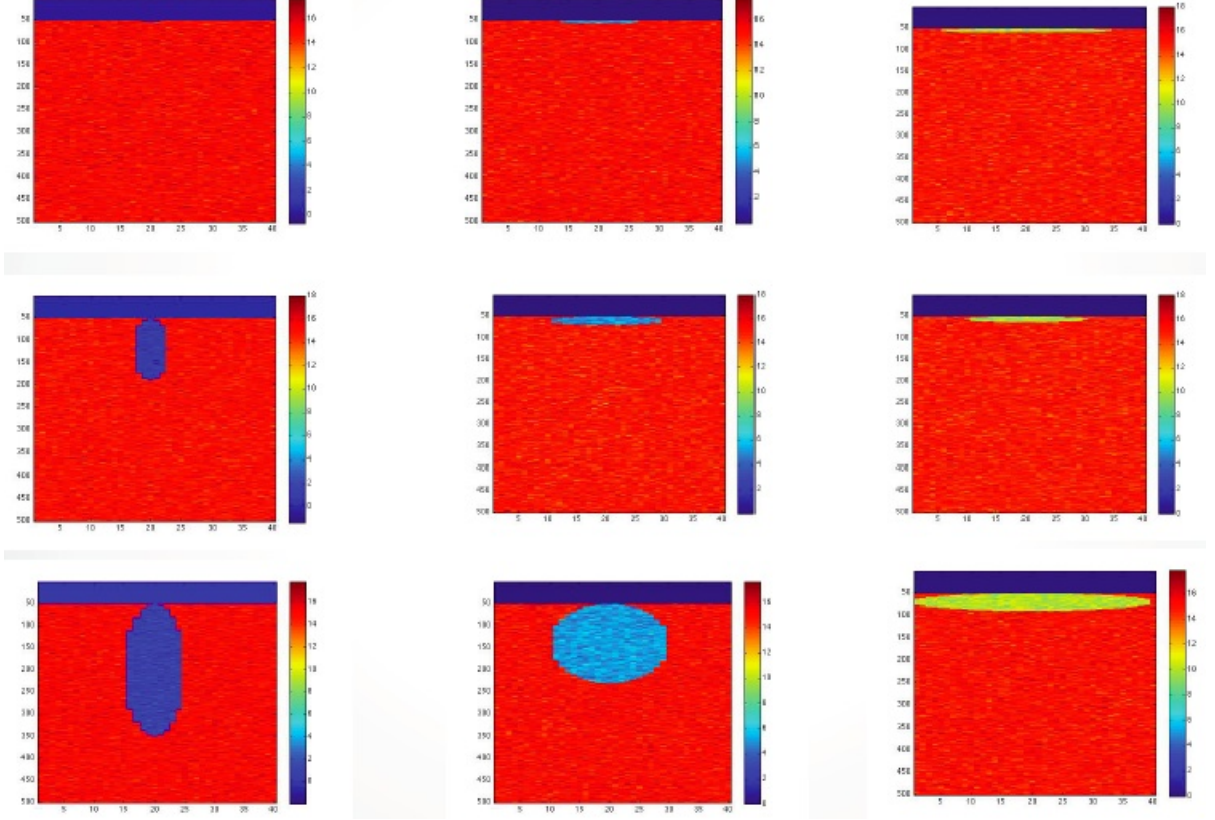


Figure 3.17: The weights associated with each neuron in the 3x3 lattice were initialized to vary plaque size and stiffness in order to represent a range of possible plaque patterns that may be randomly generated for the elliptical manufactured data set. The size and stiffness of the plaques in the initialization increase from top left to bottom right.

3.3.2 Clustering Results for the Manufactured Data Sets

There is some flexibility related to calibration of SOM algorithm parameters, including the weight initializations, and we will only present our best results for each type of manufactured data set to evaluate the ability to increase the number of used neurons in a data set comprised of images with more continuously varying features.

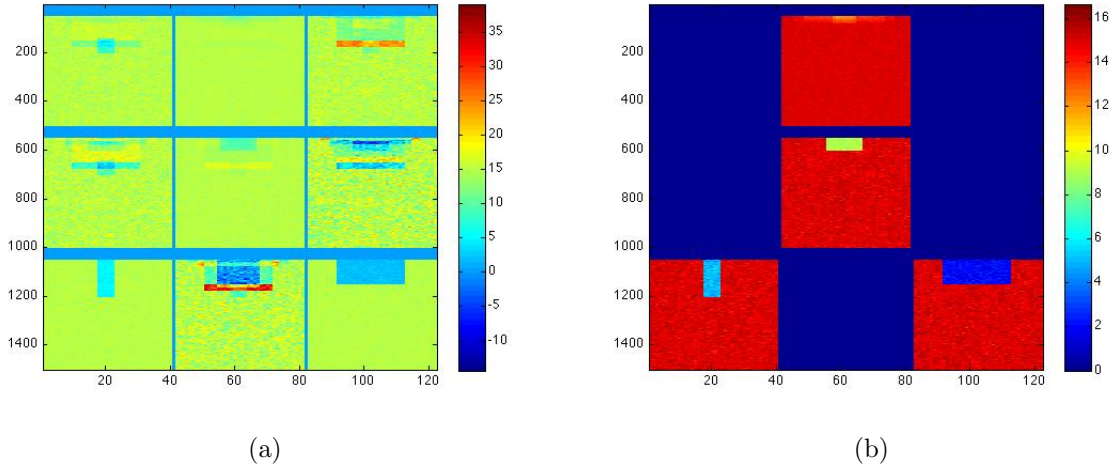


Figure 3.18: A SOM was trained on the manufactured rectangular data set. The final weights associated with each neuron of the SOM are shown in (a), and the averages of the data mapped to each neuron after training the SOM are shown in (b). There were four neurons with manufactured data clustered to them after training.

A SOM was trained on this new, rectangular manufactured data set. The final weights associated with each neuron in the trained SOM (Figure 3.18(a)) show interpolated patterns with sharp (90°) corners of varying sizes and stiffnesses where each interpolated pattern still maintains the centered position of the plaque. The area of the plaque region increases from left to right, and the plaque stiffness increases from top to bottom in the trained SOM (Figure 3.18(a)). Figure 3.18(b) shows the trained SOM has data clustered to four of the nine neurons (2, 5, 7, and 9). The final weights of neurons 2, 7, and 9 strongly represent the patterns of the manufactured data clustered to them. The resemblance between the average of the data clustered to neuron 5 and the final weight for neuron 5 is not as strong, but neuron 5 also has the most neighbors that affect the weight of the neuron during training of any neuron in the 3x3 lattice.

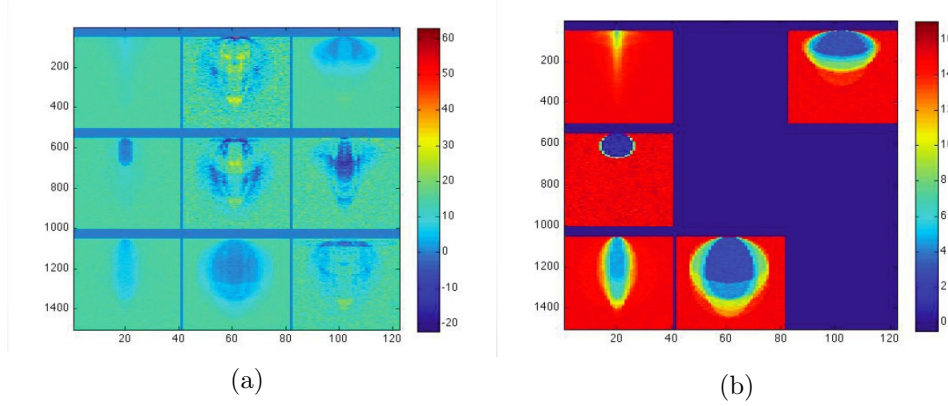


Figure 3.19: A SOM was trained on the 27 manufactured elliptical images with the weight initialization shown in Figure 3.15. The final weights associated with each neuron in the trained SOM are shown in (a) and the averages of the data clustered to each neuron after training the SOM are shown in (b). Five of the nine neurons (1, 3, 4, 7, and 8) had elliptical manufactured data clustered to them after training the SOM.

The final weights after training the SOM on the 27 images generated for the manufactured elliptical data set are shown in Figure 3.19(a). Even though the weights were initialized containing elliptical plaques and the final weight patterns were created from training data containing only elliptical plaques, non-elliptical patterns emerged as two dissimilar ellipses were averaged together in Steps 5 and 6 of the SOM training process (Table 2.1). While these non-elliptical patterns are not represented in the manufactured training data, these interpolated patterns represent other, more complex plaque shapes that may occur. The randomly generated elliptical data clustered to five of the nine neurons (1, 3, 4, 7, and 8) after the SOM was trained. The data averages shown in Figure 3.19(b) reveal that the SOM clustered images containing small plaques to the upper left while images with larger plaques were sorted to the lower right. The plaque stiffness also increases from upper left to lower right in the 3x3 lattice.

These two manufactured data sets demonstrate that the SOM can utilize more neurons if major features in the training data set have more continuous variations. Therefore, we would like to investigate a realistic and a more diverse data set corresponding to porcine atherosclerosis. We investigate such a data set in Chapter 4.

Chapter 4

SOM Clustering for Porcine Data

4.1 Porcine Data Set

4.1.1 Collection and Details of the Data Set

We have further developed and optimized our SOM algorithms within the context of data for an *ex vivo* study of porcine arteries from the Gallippi lab in which ARFI imaging was combined with histological analysis of plaque composition. Two types of pigs were represented in this study: pigs on a normal, low cholesterol diet that do not develop atherosclerosis and pigs receiving a high fat diet leading to hypercholesterolemia and atherosclerosis [3, 12, 52, 53, 54, 55]. The two pigs imaged for this data set were a 3 year 6 month old normocholesterolemic female used as a control and a 3 year 1 month old dietary hypercholesterolemic female [3]. Porcine samples were used because the atherosclerotic lesions developed in the dietary hypercholesterolemic pigs resembles the histopathology observed in humans [3, 12, 24, 52, 65]. This data set includes ARFI images from 22 distinct locations in the porcine iliac arteries of two pigs. Image locations were varied along the left and right femoral arteries. Three images were obtained at each location, one for each of three different beam sequence parameters. Twelve of the images are controls, coming from arteries with no plaque, and the remaining ten images are of diseased arteries that contain varying degrees of plaque. Subsequent to ARFI imaging, the arteries were harvested, cut longitudinally, and prepared for histological scoring based on light microscopy [30, 31]. The microscopy images were divided into four equal subsections, based on dividing the ARFI image into four equal subsections. An experienced pathologist then graded the light microscopy images according to the criteria established by the American Heart Association Committee on Vascular Lesions. If plaque was detected, structural and compositional features of the plaque such as collagen deposition, calcium deposition, degradation of the internal elastic lamina (IEL), the presence of lipid pools, and fibrous caps were also graded (Tables 4.1–4.3).

Identifiable landmarks were used to spatially correlate histology to ARFI images [3].

The availability of quantitative histological scores for each image in the data set allows us to evaluate capabilities of the SOM in identifying underlying structural features of atherosclerotic plaques. For example, images clustered to the same neuron can be probed for common histological features or quantitative histological measures can be plotted along the trained SOM to identify correlations and trends among tissue mechanical responses (measured using ARFI) as well as tissue composition. Furthermore, expansion of the data set to enable partitioning into training and validation subsets can accelerate the path to realization of optimal SOM modeling approaches.

Table 4.1: Description of Plaque Composition Classifications

| Type | Plaque Composition |
|------|---|
| I | Isolated macrophage foam cells |
| II | Multiple foam cell layers formed, fatty streak |
| III | Pre-atheroma with isolated extracellular lipid pools |
| IV | Atheroma with confluent extracellular lipid core |
| Va | Fibroatheroma surrounded by fibromuscular tissue layers with lipid or necrotic core (sometimes classified as simply Type V) |
| Vb | Calcification predominates (sometimes classified as Type VII) |
| Vc | Fibrous tissue changed predominate, absent/minimal lipid core (sometimes classified as Type VIII) |
| VI | Fissure, ulcerated, hemorrhagic, thrombotic lesion |

Table 4.2: Numerical Ratings Based on Type

| Plaque Detection | Plaque Not Present | Type I | Type II | Type III | Type IV | Type V | Type VI |
|----------------------------|------------------------|----------------------|-------------------|----------------------|------------------------|--------|---------|
| Collagen | Severe Decrease (>50%) | Mild Decrease (<50%) | No Change | Mild Increase (<50%) | Severe Increase (>50%) | — | — |
| Internal Elastic Lamina | Intact | Somewhat Disrupted | Heavily Disrupted | — | — | — | — |
| Calcium | None | Low | High | — | — | — | — |
| Lipid Pool/ Fibrous Cap | None | Small | Large | — | — | — | — |
| Ordinal Value | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

Table 4.3: Description of the Data Based on the Gold Star Standards

| Data Label | Histo Label | NC or DH | Plaque Thickness | C or D | Sect 1 | Sect 2 | Sect 3 | Sect 4 |
|------------|-------------|----------|------------------|--------|--------|--------|--------|--------|
| D1 | 1 | DH | — | D | 1 | 1 | 1 | 1 |
| C3 | 2 | NC | — | C | 0 | 0 | 0 | 0 |
| C4 | 3 | DH | — | C | 1 | 1 | 0 | 0 |
| D2 | 4 | DH | 1.008 | D | 1 | 4 | 5 | 5 |
| C5 | 5 | DH | — | C | 1 | 1 | 1 | 1 |
| D3 | 6 | DH | 1.908 | D | 1 | 5 | 1 | 1 |
| D4 | 7 | DH | 1.601 | D | 1 | 2 | 5 | 5 |
| D5 | 8 | DH | 1.293 | D | 5 | 5 | 1 | 0 |
| D6 | 9 | DH | .959 | D | 0 | 2 | 5 | 5 |
| D7 | 10 | DH | 1.599 | D | 5 | 5 | 2 | 1 |
| C6 | 11 | DH | — | C | 1 | 1 | 1 | 1 |
| D8 | 12 | DH | .971 | D | 0 | 1 | 3 | 2 |
| C7 | 13 | DH | — | C | 1 | 1 | 1 | 1 |
| D9 | 14 | DH | 1.031 | D | 2 | 5 | 2 | 2 |
| D10 | 15 | DH | 2.129 | D | 1 | 5 | 5 | 5 |
| C8 | 16 | NC | — | C | 1 | 1 | 1 | 1 |
| C9 | 17 | NC | — | C | 0 | 0 | 0 | 0 |
| C10 | 18 | NC | — | C | 0 | 0 | 1 | 0 |
| C11 | 19 | NC | — | C | 0 | 1 | 1 | 0 |
| C12 | 20 | NC | — | C | 0 | 0 | 1 | 1 |
| C1 | 21 | NC | — | C | 0 | 0 | 1 | 0 |
| C2 | 22 | NC | — | C | 1 | 0 | 0 | 0 |

| | | |
|---------|-------------------------|---|
| Legend: | Table Abbrevia- tion | Explanation |
| | C or D | Control or Disease |
| | NC or DH | Normo-Cholesterolemic Diet or Dietary-Hypercholesterolemic Diet |

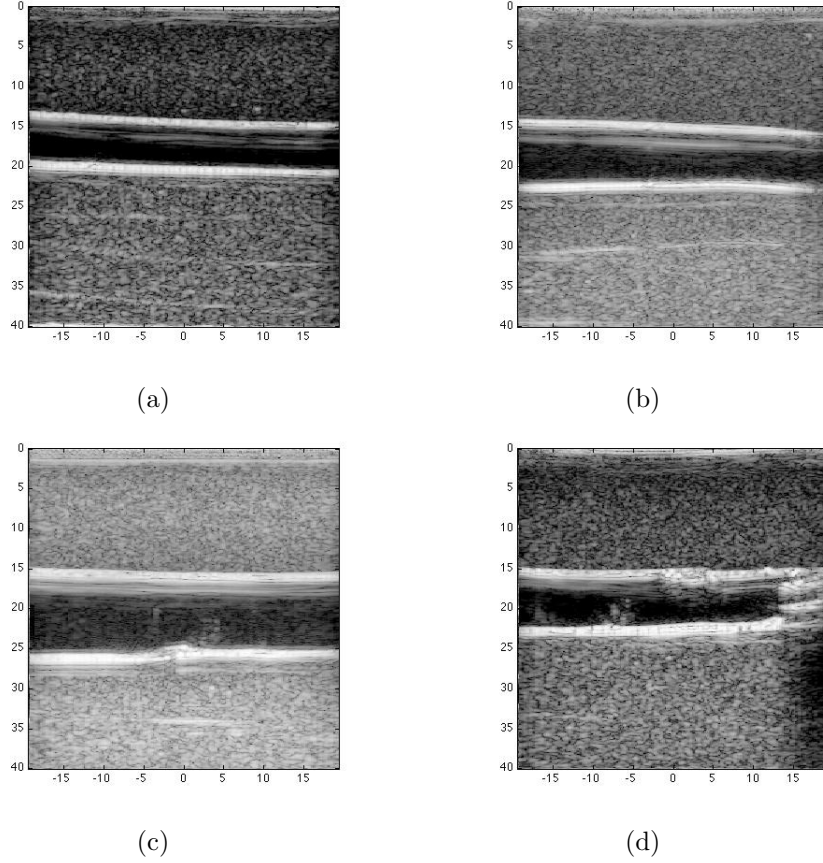


Figure 4.1: Examples of B-mode images for the ex vivo data set are shown: (a) a control, (b) a false positive for a calcium deposit, (c) a clearly visible internal elastic lamina, (d) an artery with a lipid pool, a fibrous cap and a calcium deposit.

ARFI ultrasound images for all porcine data in this data set were obtained from the Gallippi Lab at the University of North Carolina at Chapel Hill. Some sample B-mode ultrasound images are shown in Figure 4.1. There are 12 control images and 10 diseased images in the data set. The artery sizes vary slightly in this data set and are not all “centered” axially in the image. ARFI peak displacement and recovery time data was also collected for each piece of data.

4.1.2 Applying SOMs to the Porcine Data

As in Chapter 3, SOMs with a 3x3 lattice of nine neurons with two types of neighbors (close neighbors and diagonal neighbors) were implemented for all the results presented in this chapter. For details on the implementation please refer to Section 2.2. To mimic imaging data for a large population sample size, two approaches were used to expand the data set. In the first approach,

the original set of porcine images was expanded by perturbing each pixel of the original images using the technique outlined in Section 2.4 to create an expanded data set containing 80 images. Weight initializations needed to be constructed to accommodate the new porcine data. Several initializations (described with their results in Section 4.2.1) were implemented first. In order to construct the initializations for the weights associated with each neuron in cropped data sets, various forms of sorting the data vectors based on their means were implemented. In the most complex case, averages of every possible pair of original data were taken. The original data and their averages were then placed in a vector, and that vector was sorted according to size of each data entry. The number of entries in this sorted vector was divided by nine, the number of neurons: temporarily call the floor of this result n . The weights were initialized to be every n^{th} entry from the sorted vector of original data and their averages. A second approach was also considered and the original set of images was expanded through the method of weighted averages of original images that is outlined in Section 2.4.

4.1.3 Parallelization and Cluster Computing

The computing cluster in the Mathematics Department at North Carolina State University can be accessed remotely and is comprised of a Dell PowerEdge R620 with dual Intel Xeon E5-2690 2.90GHz processors and 128GB of RAM. Code was run by remotely logging into the cluster and by submitting batches of jobs. As the complexity of our computational studies increased, the code was run in parallel with three workers (as described in Sections 4.2.4, 4.2.5, and 4.3). Additional workers would further speed up computation time, but three enables the code to run in reasonable time without too much overhead required to distribute the computations. Typical run times for the most computationally intensive cases considered in this chapter were about three hours, noting that jobs from other users logged into the cluster were often running simultaneously.

4.2 Application of SOM based on the First Data Expansion

As mentioned earlier, the porcine data set consisted of 10 unique images and two methods, as outlined in Section 2.4, were used to expand the data set to a total of 80 images. The first method, in which the original images are perturbed with noise applied at each pixel of the the image, is considered in this section. As with the tissue phantom data, the expanded data sets were also cropped prior to training to focus in on a region of interest near the plaque and to accelerate the SOM algorithm. The expanded porcine data set was then partitioned into training and validation sets as outlined in Section 2.6. Lastly, the SOM was used to analyze sensitivity to training and determine the training sets that yielded the most used neurons in

the trained SOM after both training and validation. Such an assessment of SOM performance evaluates both the ability of SOM training to sort data into meaningful clusters as well as capabilities to interpolate weight vectors resembling new patterns in the validation set.

4.2.1 Data Expansion and Clustering of the Diseased Data

B-Mode ultrasound images of the porcine iliac arteries are shown in Figure 4.2, and ARFI peak displacement images for the same positions in the porcine iliac arteries are shown in Figure 4.3. The ten diseased porcine peak displacement data images were used as the basis patterns for the expanded data set. It should be noted that the images in Figure 4.3 include a substantial amount of noise away from the plaque region along the bottom wall of the artery and that the size of the image (2040x220) is large. It should also be noted that the peak displacement image for pattern III. is significantly different than the patterns presented in the remaining nine images. As with the data for tissue phantoms, the computation time for the SOM algorithm can be accelerated significantly if the images are cropped.

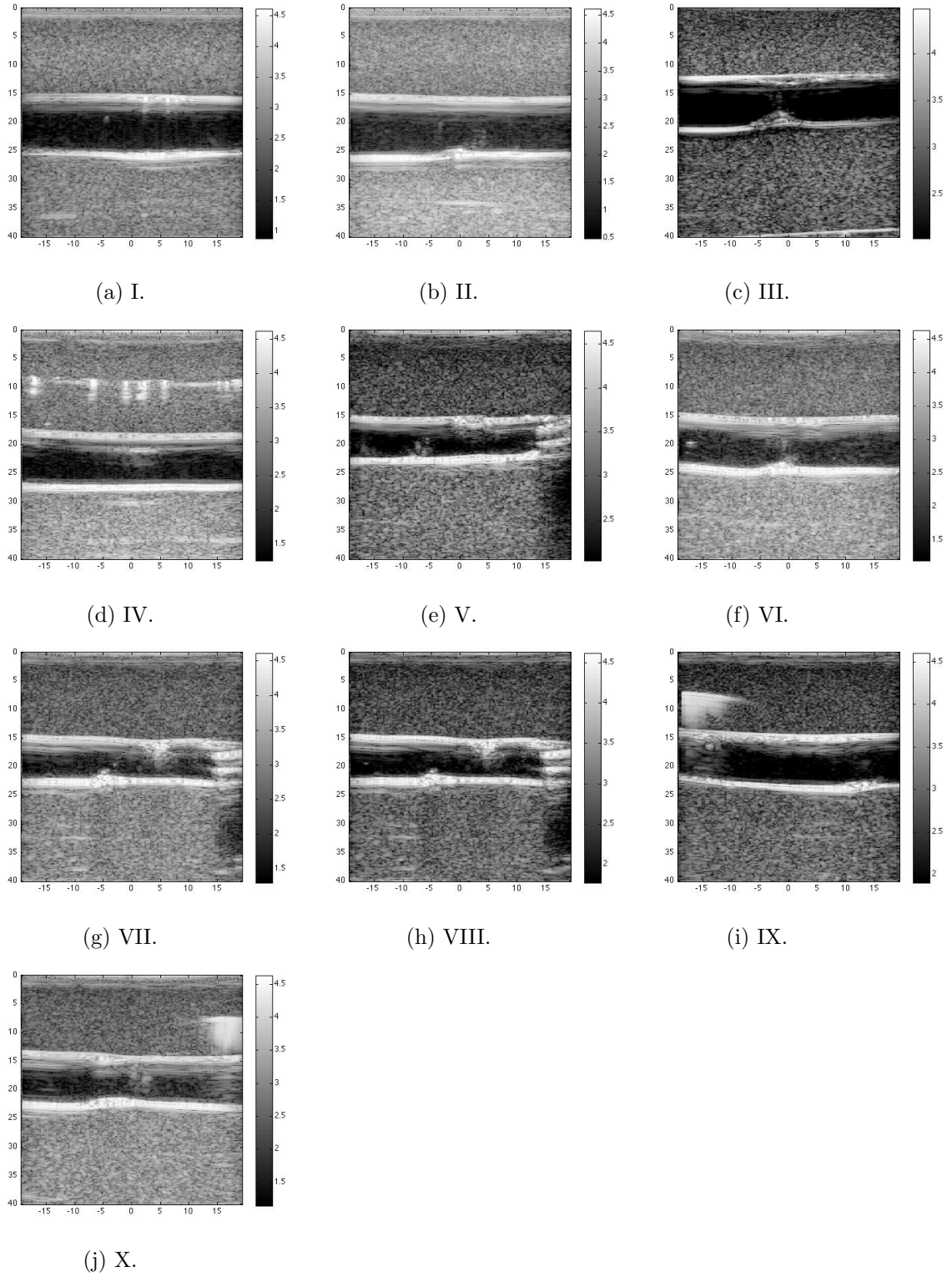


Figure 4.2: Original B-mode ultrasound images for diseased arteries in the Porcine data set. Each image is labelled with the pattern number I-X and distinguishes the different tissue samples.

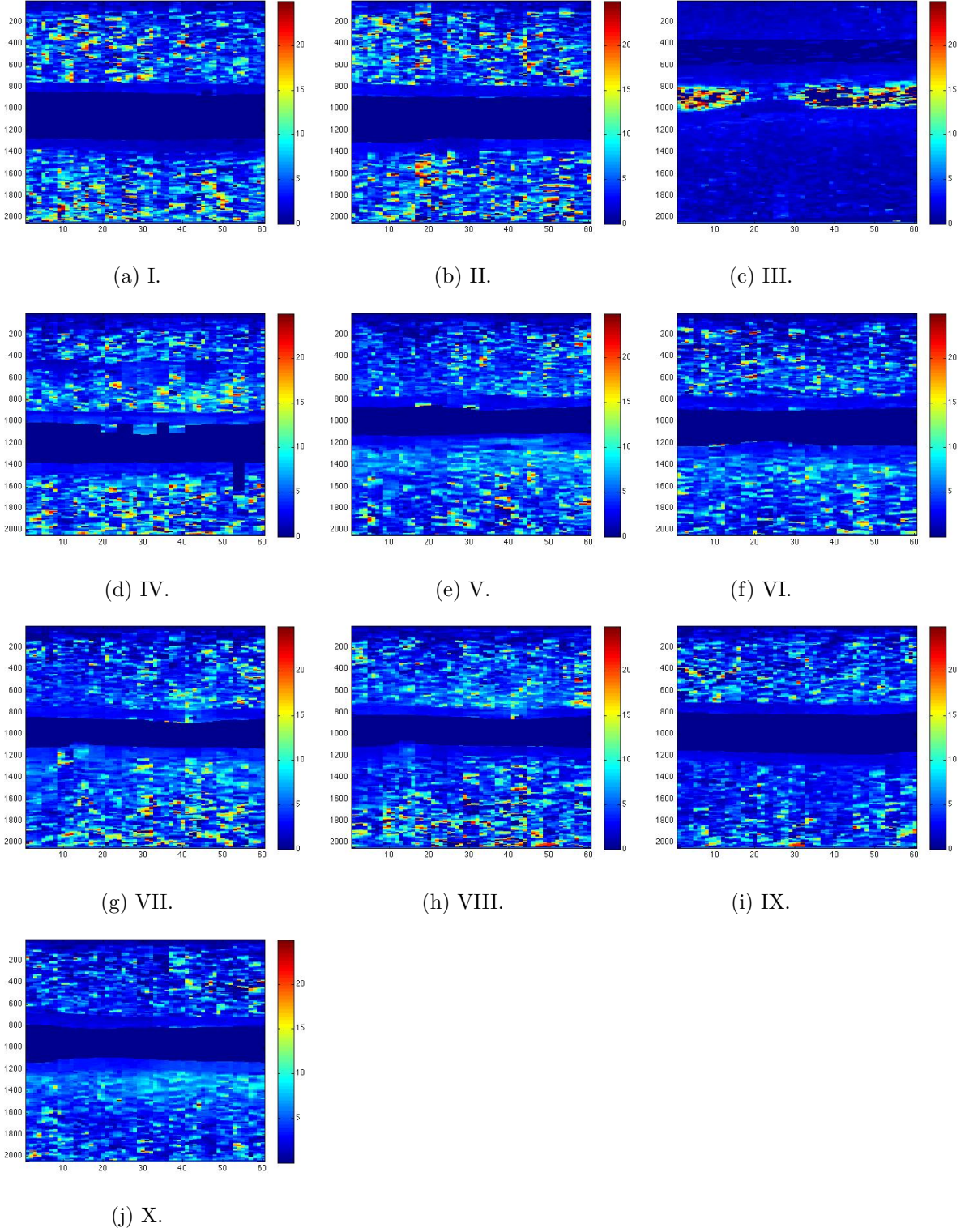


Figure 4.3: ARFI peak displacement images for the diseased arteries in the Porcine data set whose B-mode images are shown in Figure 4.2. Images are labeled with the pattern numbers I-X that distinguish tissue samples.

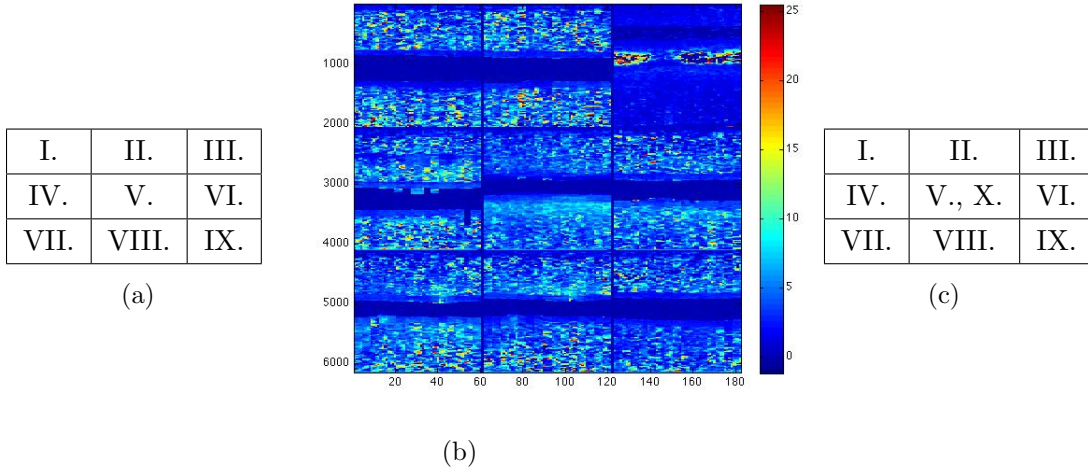


Figure 4.4: The diseased porcine data was expanded with $\sigma = .8$ and then used to train a 3x3 SOM. The weight initialization for each neuron is shown in (a). The averages of the disease porcine data mapped to each neuron after training are shown in (b). All nine neurons were utilized, and there were no mismatches. The disease patterns that clustered to each neuron after training are shown in (c).

A SOM with a 3x3 lattice of neurons was trained on an expansion of the diseased porcine data set using $\sigma = .8$. The average of the data clustered to each neuron is shown in Figure 4.4. All of the perturbations of each pattern clustered to the same neuron as the original disease pattern itself, and hence there were no mismatches in the trained SOM. Pattern X was the only original disease pattern that did not have an initial weight vector assigned to it before training. We observe from Figure 4.3 that pattern X is similar to pattern V (both clustered to neuron 5) and that the final weight for neuron 5 will be the most influenced by other weight patterns since it has the most neighbors of any neuron in the 3x3 lattice. It is unclear, however, if the SOM clustered the data on information relevant to the plaques present in each image or on background information. In order to test this, we evaluated the effects of changing the weight initialization and also analyzed the effects on the trained SOM of cropping the data.

The same expanded diseased porcine data set was then used to train a SOM where the weights were initialized by sorting the average values of the ten original disease images (Figure 4.5). Again, there were no mismatches in the trained SOM because all of the perturbed patterns clustered to the same neuron as the original disease pattern from which they were created, and all nine neurons were used. Pattern VII. was the only original disease pattern that did not have a corresponding initial weight, and all of the data corresponding to pattern VII. clustered to neuron 3. We can also conclude from Figure 4.3 that this is a reasonable placement for pattern VII. As with the previously trained SOM in Figure 4.4, it is not clear yet if the SOM sorted

on a feature specific to the plaques or if it trained primarily on background information within the image.

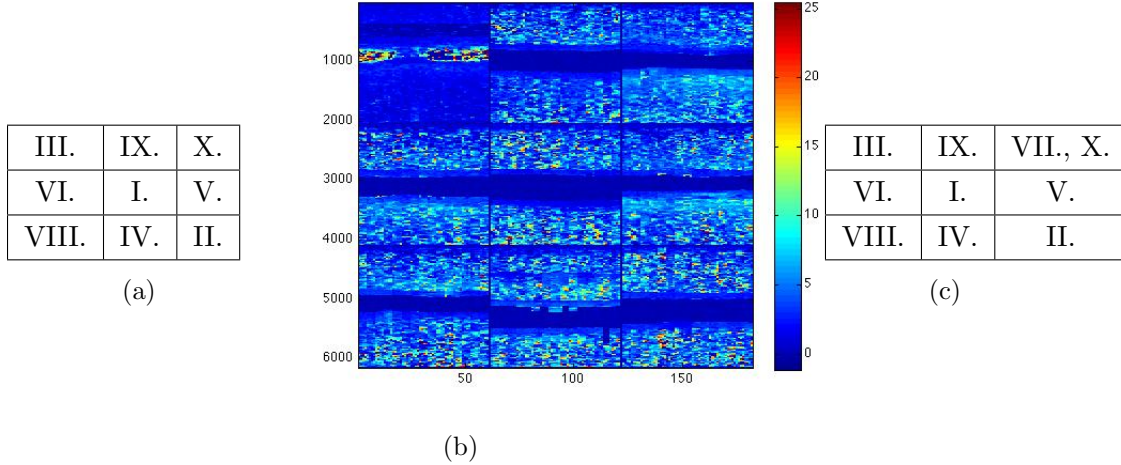
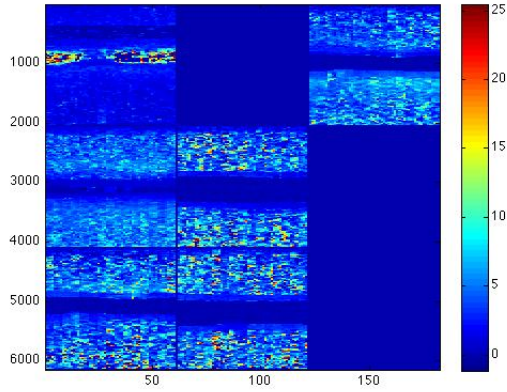


Figure 4.5: The diseased porcine data was expanded with $\sigma = .8$ and then used to train a 3x3 SOM. The weight initialization for each neuron is shown in (a). The averages of the disease porcine data mapped to each neuron after training are shown in (b). All nine neurons were utilized, and there were no mismatches. The disease patterns that clustered to each neuron after training are shown in (c).

The diseased porcine data expanded with $\sigma = .8$ was used to train another SOM with different weight initializations. This time the averages of each image in the expanded data set were sorted into a vector. Since there were 80 data images, the image corresponding to every 8th entry in the sorted vector of averages was used for a weight initialization. The weight initializations are shown in Figure 4.6(a). The letters a - g are used here to denote the seven perturbed images (in order) that were created based off of an original disease pattern. Even though more than one image corresponding to pattern V. was used in initializing the weights of the neurons in the SOM, there were still no mismatches after training. The SOM effectively gathered all images corresponding to pattern V. in neuron 4 and, across the entire trained SOM, all perturbed patterns clustered to the same neuron as the original disease pattern. There are, however, three unused neurons seen in Figure 4.6. We observe that the weights shifted substantially during the training because neuron 9 was initialized to have a perturbation of pattern X, and yet, all patterns corresponding to pattern X. clustered to neuron 3.

| | | |
|-------|--------|-----|
| V.d | I.c | V.f |
| VII.e | VIII.a | V.e |
| VI.b | II.f | X.f |

(a)



(b)

| | | |
|-------------------|-----|----------|
| III. | | VII., X. |
| IV., V., VI., IX. | I. | |
| VIII. | II. | |

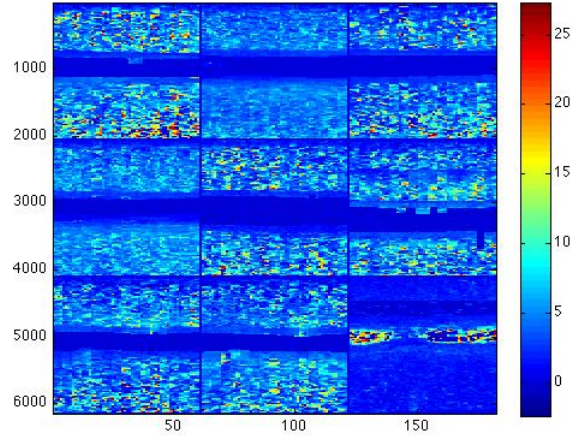
(c)

Figure 4.6: The diseased porcine data was expanded with $\sigma = .8$ and then used to train a 3x3 SOM. The weight initialization for each neuron is shown in (a). The averages of the disease porcine data mapped to each neuron after training are shown in (b). All nine neurons were utilized, and there were no mismatches. The disease patterns that clustered to each neuron after training are shown in (c).

The weight initialization was changed again, but this time the control patterns were included in the expanded porcine disease data set with $\sigma = 1$. The first three numbers in a random permutation of integers one to twelve, corresponding to the number of original control patterns, and the first six numbers in a random permutation of integers one to ten, corresponding to the number of original disease patterns, was used to determine the weight initialization (Figure 4.7(a)). The averages of all the data clustered to each neuron in the trained SOM is shown in Figure 4.7(b). We observe that there were no unused neurons; however, there were many mismatches in the control image group. The control patterns clustered to four of the nine neurons, and neurons 1 and 3 contained only one one control pattern each after training the SOM. The SOM did, however, correctly cluster each perturbed image with its corresponding original disease pattern. Neurons 2 and 4 contained both control and diseased data after the training of the SOM.

| Control 10 | Control 12 | Control 8 |
|------------|------------|-----------|
| VI. | I. | IV. |
| VIII. | VII. | III. |

(a)



(b)

| Control 10 | Controls 1, 2, 3, 5, 9, 11, 12 IX., X. | Control 8 |
|-------------------------------|---|-----------|
| Controls 4, 6, 7 II., V., VI. | I. | IV. |
| VIII. | VII. | III. |

(c)

Figure 4.7: The control patterns were added to the diseased porcine data that was expanded with $\sigma = 1$ and then used to train a 3x3 SOM. The weight initialization for each neuron is shown in (a). The averages of the disease porcine data mapped to each neuron after training are shown in (b). All nine neurons were utilized, but there were mismatches in the control data. The patterns that clustered to each neuron after training are shown in (c).

4.2.2 Crops of Diseased Data

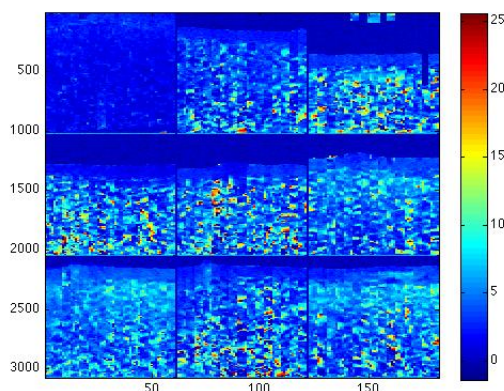
To accelerate the computation time for SOM training and focus in on the region of interest for atherosclerotic plaques, several approaches to cropping the diseased data were considered. In the first approach, the images were cropped to retain the lower half of the image, resulting in cropped images with dimensions of 1020x220 pixels. As shown in Figure 4.8, the trained SOM clustered each subset of the data corresponding to one particular disease pattern to one neuron. However, with this cropping method, it is still unclear whether the main feature the SOM clustered on was related to features of the plaque such as its stiffness or its size. Data corresponding to pattern III. clustered to neuron 1, but this pattern is significantly different from the other disease patterns. The remaining disease patterns appear to cluster within the 3x3 lattice based on the area of the arterial flow region in the image as well as the level of noise

present in the tissue. For example, the size of the arterial flow region present in the clustered images decreased from top to bottom in the trained SOM.

Results of the initial cropping method demonstrate that a new cropping technique that is individualized or automated may be required. Specifically, at the stage of atherosclerosis in the porcine data set, the size and characteristics of the plaque are the primary feature of interest. In contrast, for more advanced stages of the disease, narrowing of the arterial flow region may be of greater importance. For example, there is a much larger arterial region in the data clustered to neuron 3 than neuron 9 and it is not clear whether this is due to significant narrowing or, more likely, due to the need to register images to a common anatomical point across images taken from different pigs. Ultimately, in order to omit extra pre-processing of data, customized standards will need to be developed within a clinic to direct clinicians to center the ARFI image on the relevant region of interest.

| | | |
|------|-------|-----|
| III. | IX. | IV. |
| I. | II. | VI. |
| X. | VIII. | V. |

(a)



(b)

| | | |
|----------|-------|-----|
| III. | IX. | IV. |
| I. | II. | VI. |
| VII., X. | VIII. | V. |

(c)

Figure 4.8: The porcine images were cropped in half to focus on the region of interest in the lower portion of the artery wall. The weight initializations are shown in (a). The averages for the data associated with each neuron in the trained 3x3 SOM is shown in (b). All nine neurons had porcine data associated with them. The patterns clustered to each neuron after training the SOM are shown in (c).

To improve cropping, an automated cropping was developed for each of the ten porcine disease images and aimed to locate the top of the plaque on the lower arterial wall. Once this point was identified, the image was cropped so that the peak of the lower arterial wall plaque was at the top of the cropped image. Pixels were retained in a manner such that each cropped image was comprised of the row of the image containing the peak of the arterial wall and the next 400 consecutive rows beneath the peak. Since the blood in the artery should be imaged as navy blue (having a value of zero for peak displacement), cropping was implemented by determining the first nonzero pixel below the beginning of the artery, defined as the first row of consecutive zeros. This automated cropping procedure reduces some of the superfluous differences in the ten disease patterns that are unrelated to the primary disease marker in the porcine data set. Figure 4.9 shows the results of the automated cropping of each of the ten original porcine disease patterns.

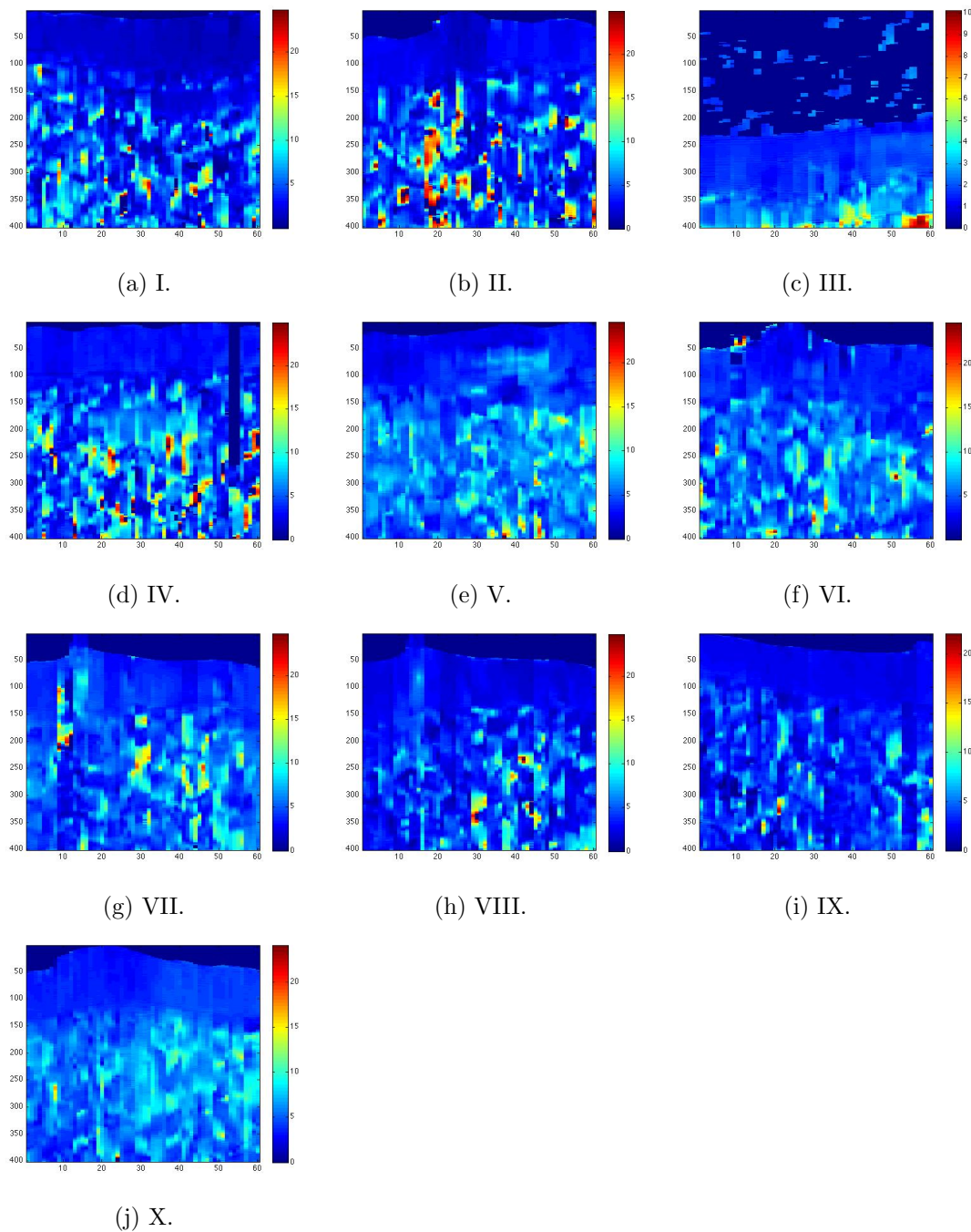


Figure 4.9: New images after the automated cropping technique was applied to focus on the region of interest. Each of the patterns in Figure 4.3 was cropped so that the row containing the peak of the plaque in the lower arterial wall moved to the first row of the image. Pattern III contained considerable noise and therefore did not crop as well in the automated method.

Once these images are cropped using the automated method, features other than artery size and background noise become the prominent differences between each pattern (Figure 4.9). It is much easier to identify features governed by the presence of plaque in the artery once images have been cropped to show the regions of interest. It should be noted that patterns VII. and VIII. are quite similar to each other, within the range of the ten patterns in the original data set, in that both images exhibit a sharp peak in the lower arterial wall in the left-hand portion of the cropped image (Figure 4.9). These patterns are also very similar to each other in that their peaks are very narrow. Pattern VI. also has one distinct peak, but it is much wider, like the peak in pattern X. It should also be noted that pattern III. is unique in that it is very different from all other patterns present. It is the only pattern that contains extra noise in the fluid region of the artery, making the automated cropping of the image more difficult. After the automated cropping, differences in the area of the arterial flow region now correspond to information relating to the plaque content of the image rather than an arbitrary or standard choice for the viewing window across all images. The greater likelihood that the SOM will now cluster on the size and shape of the plaque enables the image analysis to focus on the primary feature relating to the atherosclerosis present on the arterial wall.

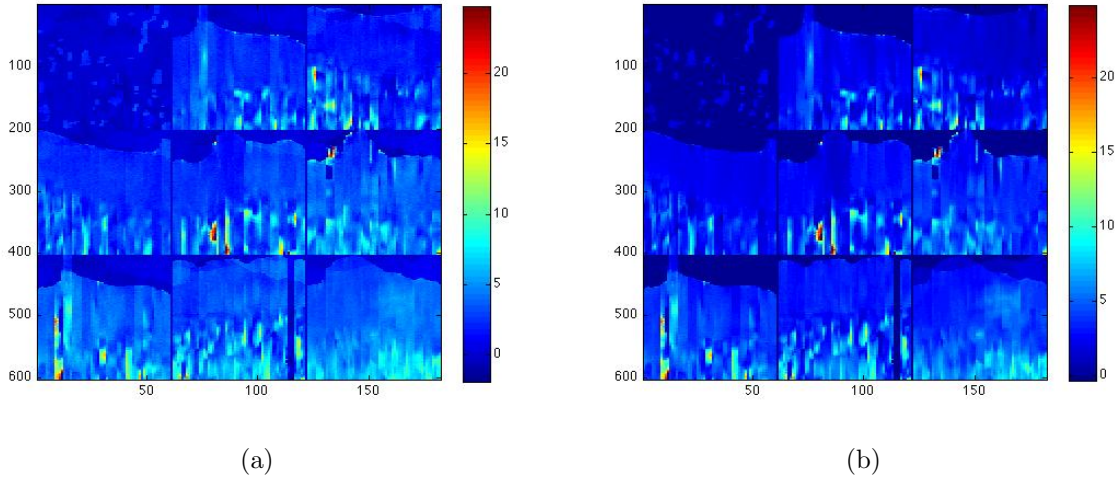


Figure 4.10: The porcine data was automatically cropped based on the peak of the plaque above the bottom wall to focus on the region of interest located below the artery wall and then used to train a SOM. The data expansion here used $\sigma = .2$. The final weights associated with each neuron (a) and averages of data associated with each neuron after training the 3x3 SOM (b) is shown.

A SOM was trained on the expanded porcine disease data set with the automated cropping

technique and the results are shown in Figure 4.10. The SOM mapped all images associated with the same disease pattern to the same neuron, resulting in no mismatches. Pattern III. mapped to neuron 1, and this pattern has an obvious effect on the final weight associated with neuron 1 in Figure 4.10(a), making it very different than the other final weight patterns. We observe that the SOM does still use some information from background noise to sort these images. For example, the final weight vectors associated with the top row (neurons 1-3) have less noise on the bottom half of the image than the bottom row of neurons (7-9) (Figure 4.10(b)). While the shape of the plaque is similar in the final weights of neurons 2 and 7, these two patterns did not cluster to neighboring neurons due to their differences in background noise throughout the bottom half of the image. Overall, the sequence of SOM methods and cropping approaches resulted in successful clustering of images by neuron number in the trained SOM map. To further analyze the capabilities of the SOM for the porcine data set, a partition of the original data pattern into training and validation sets was considered. Training using a data set with reduced patterns enables evaluation of capabilities of the trained SOM to interpolate new, physiologically realistic patterns that may not be present in the training set.

4.2.3 Partitioning the Data: Seven Training and Three Validation Patterns

The ability of a SOM to both accurately sort and interpolate or infer patterns within a larger set of images can be evaluated by partitioning data into training and validation subsets. Since there are numerous ways to partition, criteria can be specified to identify the best ways to partition a given data set. The diseased porcine data set was partitioned into two groups: a training set containing all of the images corresponding to seven of the ten original disease images and a validation set containing all the images corresponding to the remaining three patterns. The SOM was trained on the first set of data and then the final weights were used to determine the closest associated weight (winning neuron) for each of the patterns in the validation set. The code circulated through all combinations of choosing seven (of the original ten) data patterns for the training data set, and a parfor loop was implemented to run several realizations of training data sets at the same time. The images associated with the remaining three patterns were then used for validation after the SOM training was complete. Since the training set only contains seven patterns (and their perturbations), a successfully trained SOM will have at least two unused neurons at the end of the training process, implying that all of the images associated with one particular pattern clustered to the same neuron.

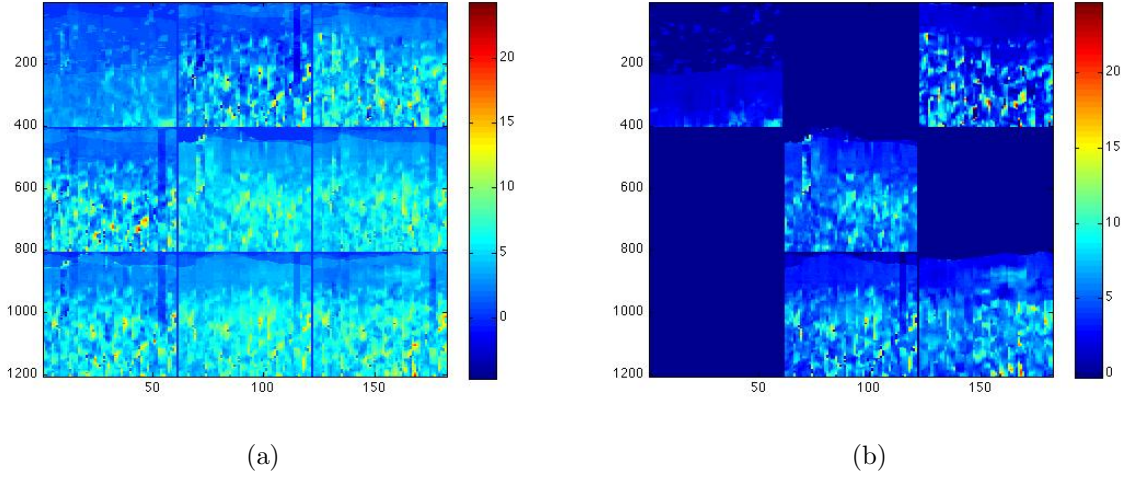


Figure 4.11: The SOM was trained on seven of the ten disease porcine patterns of data (I., III., IV., V., VI., VII., X.). The final weights (a) and averages of data (b) associated with each neuron in the trained 3x3 SOM is shown.

The trained SOM in Figure 4.11 has four unused neurons, so multiple patterns mapped to the same neuron in two instances. Patterns VI. and VII. both mapped to neuron 5 (Figure 4.11(b)). These two patterns both display a peak to the left of the image. Note that these two images are similar to each other considering that pattern VIII. was not included in the validation set. Even though patterns VII. and VIII. are quite similar, pattern VIII. did not map to neuron 5 in the validation phase. This occurred because the final weight for neuron 5 was influenced by both patterns VI. and VII. during training. Therefore, the final weight pattern associated with neuron 1 was closest to pattern VIII. during the validation phase. Pattern II. mapped to an “unused” neuron (6) during the validation phase. Both the final weight associated with neuron 6 and pattern II. have a gradual peak in the left half of the viewing region. Clustering of images in the validation set to a neuron with no data previously associated with it shows that the interpolation of patterns fine-tuned during the SOM training is useful for detecting new, possible patterns that may not be present in the training set.

Table 4.4: This table shows validation for remaining three patterns the SOM was not trained on for Figure 4.9. Note that neuron 6 was unused in the training (because no data used for training was associated with it after training). The interpolation of final weights associated with each neuron provided a pattern that was closer to pattern II. than any of the used neurons.

| Pattern: | Mapped to Neuron Number: |
|----------|--------------------------|
| II. | 6 |
| VIII. | 1 |
| IX. | 1 |

By cycling through all possible combinations of choosing seven of the ten disease patterns to formulate the training set, optimally trained SOMs could be identified by the number of unused neurons at the end of training and validation.

The trained SOM in Figure 4.11 has four unused neurons after training and three after validation since neuron 6 was unused after the SOM was trained but the images corresponding to pattern II. are closest to the final weight associated with neuron 6 (Figure 4.11 and Table 4.4). The SOM was trained on the seven original patterns I., III., IV., V., VI., VII., and X. Images with a peak of plaque on the lower arterial wall mapped to neurons 5 and 8. The level of noise in the tissue under the artery decreases from left to right and top to bottom in the trained SOM. An optimally trained SOM could have less unused neurons, however, since the seven distinct training patterns only clustered to five distinct neurons during training.

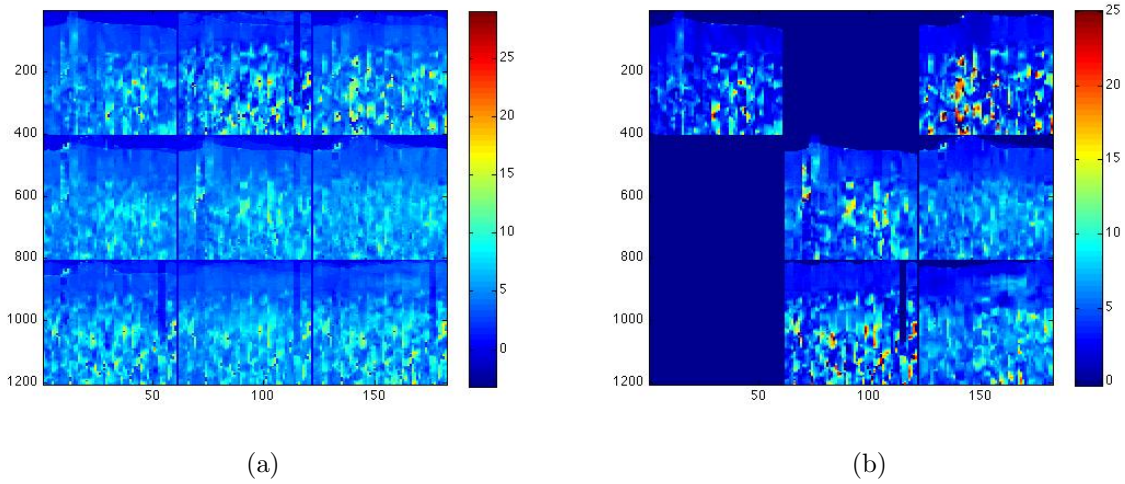


Figure 4.12: The SOM was trained on seven of the ten patterns of data (II., IV., V., VI., VII., VIII., X.). The final weights (a) and data averages (b) for each neuron in the trained 3x3 SOM is shown.

Table 4.5: This table shows the validation for the remaining three patterns the SOM was not trained on for Figure 4.10. Neuron 4 was unused after training, but pattern I. mapped to it during validation.

| Pattern: | Mapped to Neuron Number: |
|----------|--------------------------|
| I. | 4 |
| III. | 1 |
| IX. | 1 |

Based on a different partition into training and validation subsets, the trained SOM in Figure 4.12 now has only three unused neurons after the training phase. The plaque peaks in the final weights decrease from top to bottom and from left to right along the trained SOM. The vertical bar of zero values appearing in pattern IV. (mapped to neuron 8) also affects the final weights for its neighbors (neurons 5, 7, and 9). Patterns VI. and X. both mapped to neuron 6 during training. Both of these patterns have little arterial fluid imaged above the lower wall and have a solid portion under the wall before any noise begins below. Pattern I. mapped to previously unused neuron 4 during the validation phase (Table 4.5 and Figure 4.12). Both the final weight associated with neuron 4 and pattern I. display a small amount of arterial fluid,

a well-imaged region without much noise under the arterial wall, and lack the vertical bar of zero-values.

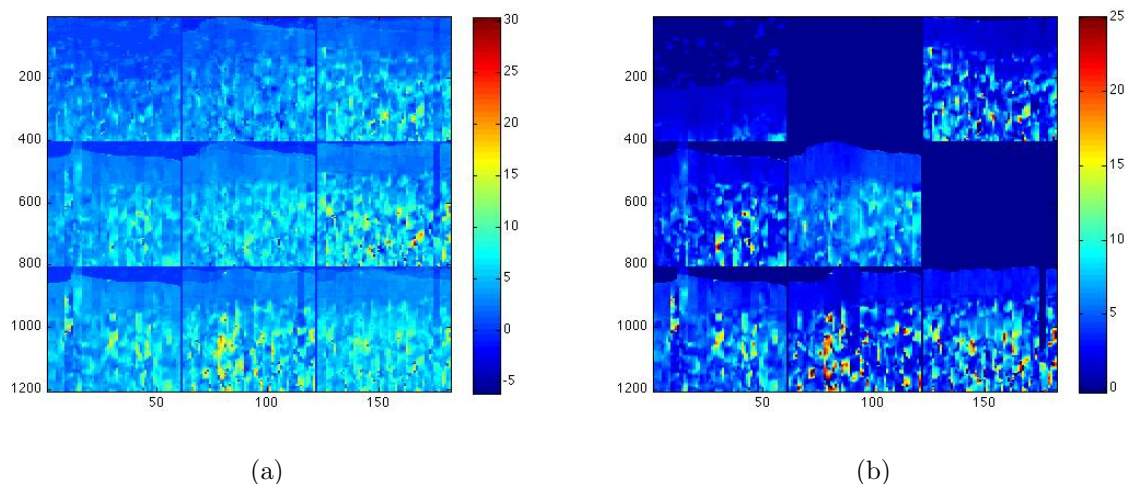


Figure 4.13: A SOM was trained on seven of the ten patterns corresponding to diseased porcine data (I., II., III., IV., VII., VIII., X.). The final weights corresponding to each neuron in the trained SOM (a) and the averages of the data clustered to each neuron after training (b) are shown.

Table 4.6: The validation results for remaining three diseased porcine patterns not used in the training of the SOM results presented in Figure 4.11 are shown. Neuron 2 had none of the training data clustered to it; however, one of the validation patterns (IX.) mapped to this neuron.

| Validation Pattern: | Mapped to Neuron: |
|---------------------|-------------------|
| V. | 5 |
| VI. | 5 |
| IX. | 2 |

The SOM shown in Figure 4.13 optimized the number of unused neurons after both training the SOM on seven of the ten diseased porcine patterns and then determining which final weights were closest to the remaining three diseased porcine patterns. There were only two neurons that had no training data clustered to them after the training phase which is optimal since

there were seven patterns and nine neurons used. During validation, one of the two previously unused neurons was used as well (Table 4.6). This leaves only one neuron that does not have data associated with it after the training and validation phases are complete. All realizations of possible combinations of seven of the ten disease patterns were considered (along with their perturbed expansions) to train a SOM, and this case minimized the number of neurons that have no data associated with them after both the training and validation phases. The final weights for the trained SOM show that the weights associated with the neurons have a decreasing area of arterial fluid from left to right and have decreasing peaks from left to right. Pattern IX. in the validation set mapped to previously unused neuron 2 because the final weight associated with neuron 2 was closest in that there is little noise, no large peak, and no vertical rectangle of zero-values.

4.2.4 Partitioning the Data: Six Training and Four Validation Patterns

This process of cycling through patterns to train and validate was repeated for all possible combinations of training on six of the diseased porcine patterns and validating on the remaining four patterns.

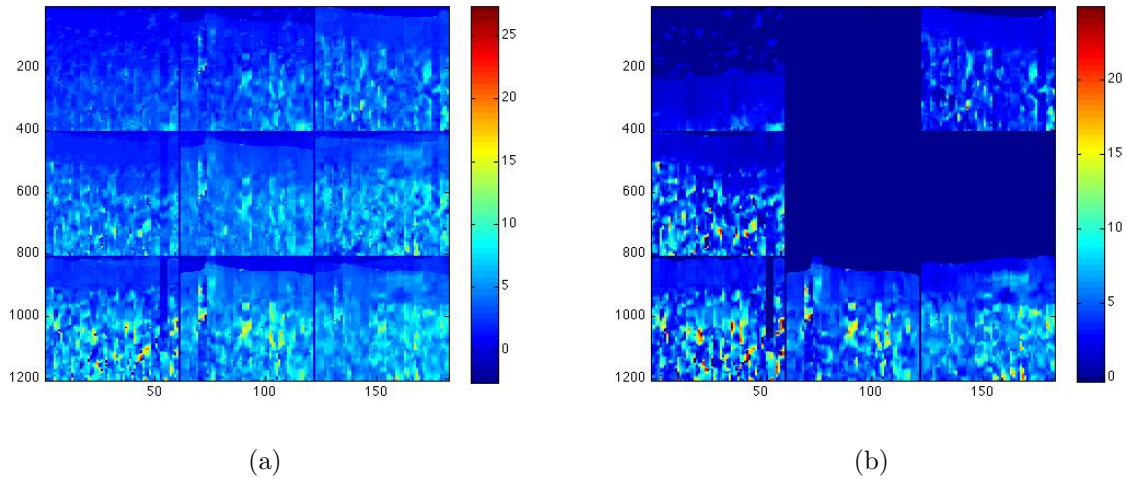


Figure 4.14: A SOM was trained on each possible combination of six of the ten diseased porcine patterns. The final weights associated with each neuron after training (a) and the averages of the training data clustered to each neuron after training (b) is shown for the following six training patterns (and their perturbed expansions): I., III., IV., V., VII., IX.

Table 4.7: Validation results for the remaining four patterns.

| Validation Pattern: | Mapped to Neuron: |
|---------------------|-------------------|
| II. | 6 |
| VI. | 5 |
| VIII. | 5 |
| X. | 9 |

The trained SOM shown in Figure 4.14 has diseased porcine data mapped to the maximum number of neurons after the training phase since there were only six distinct patterns (and their perturbed expansions). This realization of a SOM trained on six of the ten diseased porcine data is also optimal because it has the maximum number (two) of new neurons with data mapped to them during validation of all the SOMs that were trained and evaluated for validation. Neurons 5 and 6 do not have any of the six training patterns clustered to them, but patterns VI. and VII. from the validation set cluster to neuron 5 and pattern II. clusters to neuron 6 after training has been completed (Table 4.7 and Figure 4.14(b)).

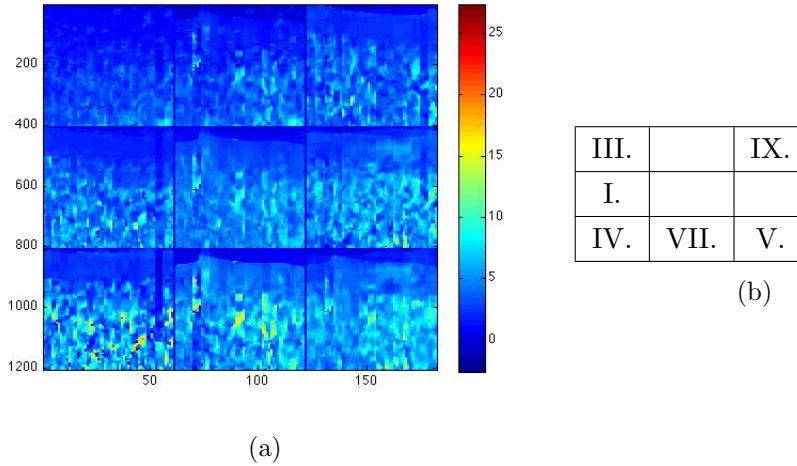


Figure 4.15: The final weights associated with each neuron after training a SOM are shown in (a), and the final location of each of the six original disease patterns used during the training phase is shown in (b). Each of the six original patterns used in the training clustered to their own neuron.

Comparing the original six disease patterns to the final weights associated with each neuron

in the trained SOM, it is again evident that pattern IV. (mapped to neuron 7) affected the final weights of its neighbors, creating a trace of the vertical rectangle of zero-values in each of the final weights in the first column (Figure 4.15). The noise along the right-hand side of the image for pattern I. also exhibits a trace of this column of zero-values, making this pattern neighbors with pattern IV. instead of pattern IX., which has minimal arterial fluid, and a well-imaged (low-noise) region below the lower arterial wall. There is also a distinct column of vertical noise in the left of the image for pattern VII. that affects the final weights in the second column of the 3x3 grid (Figure 4.15).

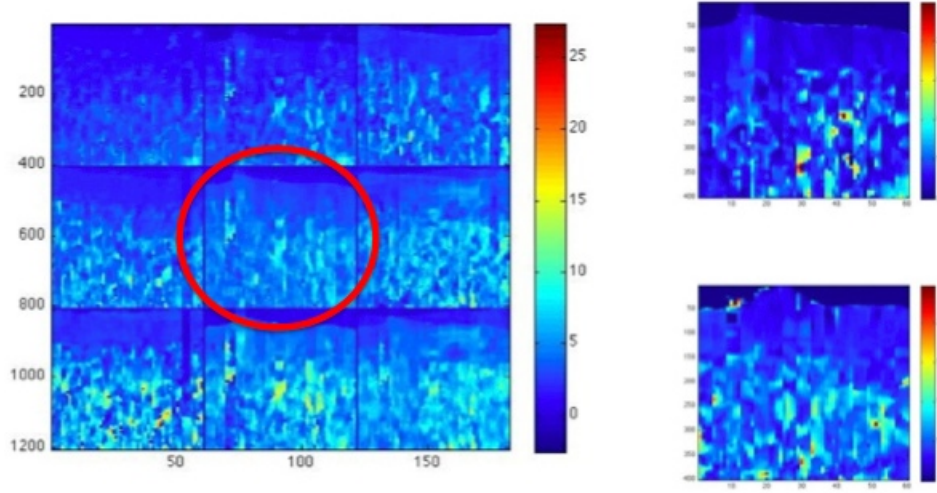


Figure 4.16: A comparison of final weights associated with each neuron in the trained 3x3 SOM shown in Figure 4.14 and the validation patterns mapping to the previously unused neuron 5.

The noise level increases from top to bottom in the final weights associated with each neuron in the trained SOM (Figure 4.16). Both patterns VI. and VIII. have moderate noise and a peak toward the left of the viewing area (Figure 4.16). Both of these patterns mapped to neuron 5 in the validation phase (Table 4.7). Since the final weights are set (and no longer move) before the validation data is presented, the fact that pattern VI. mapped to neuron 5 does not affect the trained SOM or the fact that pattern VIII. mapped to neuron 5 (and vice versa).

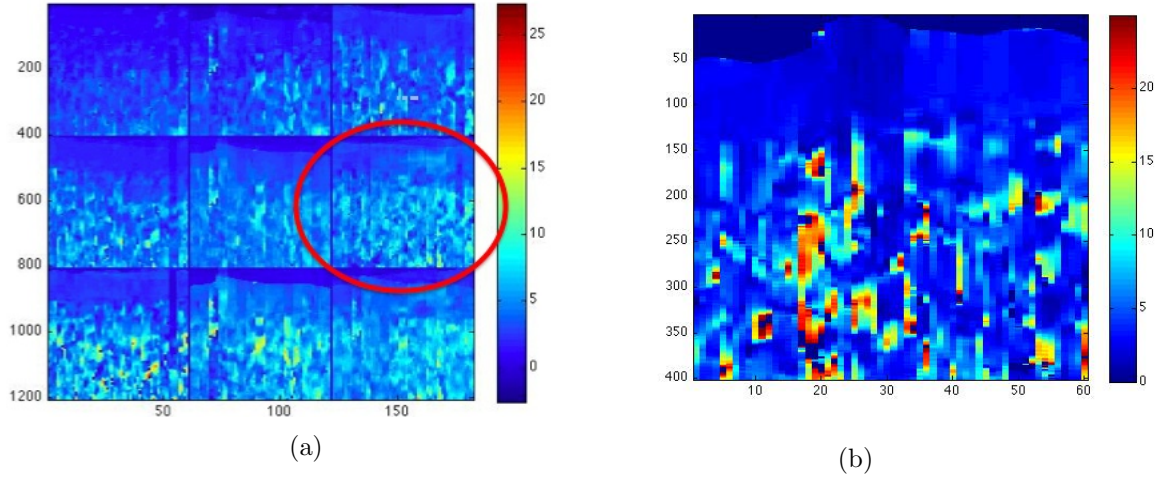


Figure 4.17: Comparison of final weights associated with each neuron in the trained SOM (a) and the validation pattern that mapped to neuron 6 (b). The absence of a noticeable vertical bar of low pixel-values does not lend the pattern to map to the first column (on the left) of final weights in (a). The peak shown in (b) is smooth and a relatively small arterial area is present, mapping the pattern to neuron 6.

Pattern II. also displays moderate noise, placing it in the same range as the noise levels of the second row of the final weight vectors associated with the neurons in the second row of the SOM (Figure 4.17). Pattern II. follows the trend of the lower arterial wall having a dip and then two peaks from left to right (Figure 4.17(b)). This is most similar to the pattern seen in the interpolated pattern of the weight vector for neuron 6, where pattern II. mapped in the validation process.

No data mapped to the second neuron because the interpolated pattern associated with neuron 2 has a pattern unlike the ten given disease patterns (Figure 4.9). The final weight associated with neuron 2 has more noise in the arterial fluid than the other final weights associated with the other neurons. This noise is due to the fact that it neighbors the neuron (1) that the data for pattern I. mapped to during training. Neuron 1 does not have as much arterial noise, though, because it is also affected by its neighbors, such as pattern I that is mapped to neuron 4 in training.

Throughout this chapter, the majority of the trained SOMs exhibited the property that perturbed data originating from the same image disease pattern always cluster to the same neuron. In other words, the trained SOMs rarely exhibited mismatches during the data clustering. While mismatches would be generated but increasing σ , image patterns at this level of noise looked unrealistic. This finding suggests that the method used to the expand the data set may need to be extended to create new images that mimic more realistic variations across

patient populations. Therefore, in the next section we explore a new method of expanding the diseased data that gives a richer set of patterns for the SOM to cluster.

4.3 Second Data Expansion Method

In this section, we will explore a second method for expanding the diseased porcine data set as an alternative to randomly perturbing each individual pixel. The second method also creates more variation in patterns yet retains medically plausible images. In order to achieve this, we constructed new images by taking weighted averages of the original disease images, as outlined in Table 2.2. One of the ten original disease patterns was assigned the heaviest weighting in the average (Step 4 in Table 2.2) and also serves as the pattern used to evaluate success of the clustering in both the training and validation sets. The set of original disease images was used to create 90 new patterns. First, a vector of 90 weight values r (Step 1 in Table 2.2) were drawn from a uniform distribution between 0.5 and 1. The sorted weight values generated in this manner are shown in Figure 4.18. The number of times that each of the original disease patterns was chosen as a base pattern for expansion is shown in Figure 4.19(a). The number of times that each of the original disease patterns was used as the second, third, and fourth closest patterns in the weighted average in Step 4 of Table 2.2 is shown in Figure 4.19(b)-(c).

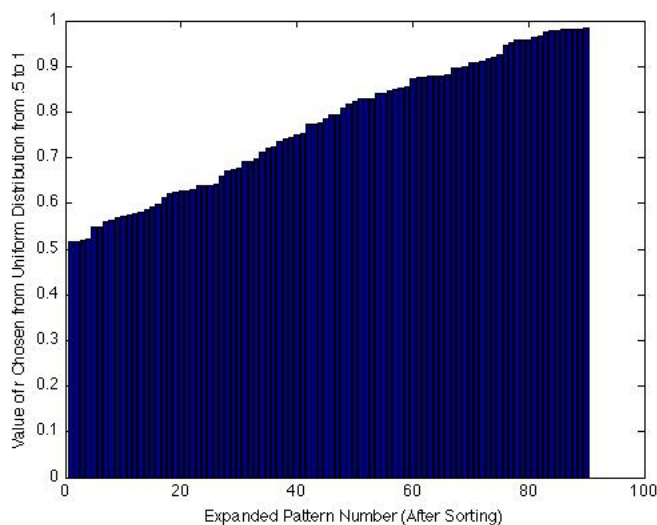
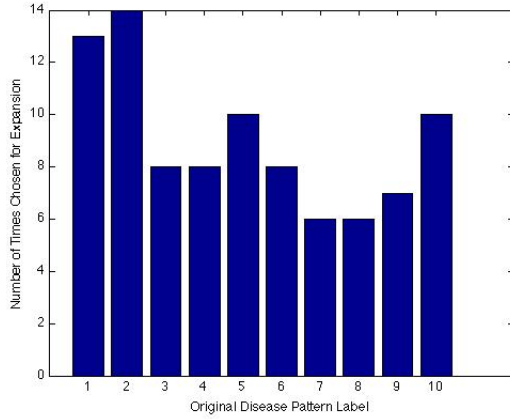
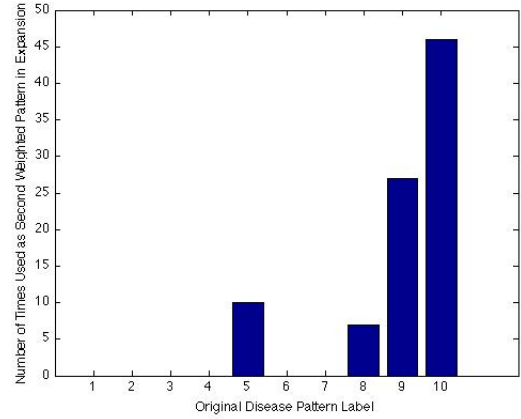


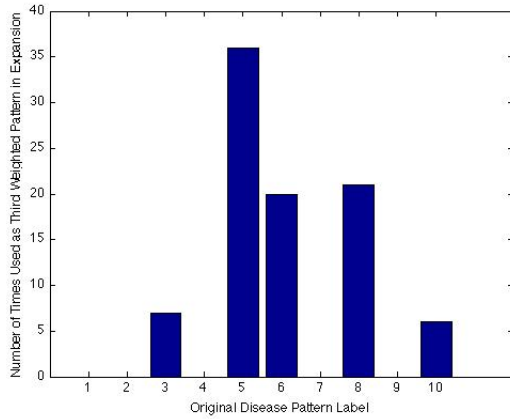
Figure 4.18: A sorted bar graph of the uniformly distributed variable r , the rate at which the randomly selected original disease pattern is weighted in the process of creating a new image for the expanded data set.



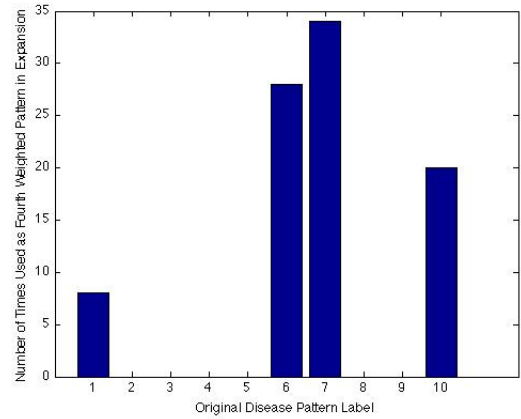
(a)



(b)



(c)



(d)

Figure 4.19: Bar graphs showing how many times each of the original disease patterns was used as the basis for a newly created image, (a), the closest data image D_{d_2} in the weighted average (b), the second closest data image D_{d_3} in the weighted average (c), and the third closest data image D_{d_4} in the weighted average described in Table 2.2.

The newly expanded data set was used to train multiple SOMs. After expansion, the 100 image patterns (10 original and 90 new) were partitioned into training and validation sets as in Section 4.2.5. To create the training set, all combinations of six of the ten original disease patterns were used in the training phase. In contrast to Section 4.2.5, the patterns identified for the training set each contain different numbers of images (Figure 4.19). Consequently, four images were chosen for each pattern in the training set and any remaining patterns were assigned to the validation set. Therefore, there were always 24 images in the training set. In contrast to the partitioning approach used for the first data expansion, the new method reserves images

associated with patterns in the training set that can be evaluated to determine to which neuron they cluster in the validation phase. The weight vectors were initialized by sorting the original six patterns chosen for the training set and all of their pair-wise averages as described in Section 4.1.2.

SOMs were trained for all possible combinations of the training and validation sets. Trained SOMs were evaluated and compared based on the number of unused neurons after training and the extent to which these unused neurons were used in clustering images from the validation set. Figure 4.20 shows SOM results for an example of an optimal combination of training and validation sets based on training with patterns I, II, III, IV, V and VI. After training, there were three unused neurons (neurons 2,3, and 4) which is optimal since the training set contained six patterns. There was only one mismatch after training: an image created from patterns VI. (dominant), X., V. and VII. was mapped to neuron 8 instead of neuron 6. It should be noted, though, that pattern V. is the second closest (in the 2-norm sense) to pattern VI. after X., which was not included in this training set. This mismatch was probably also related to a low r -value being assigned to the original pattern during the image's creation.

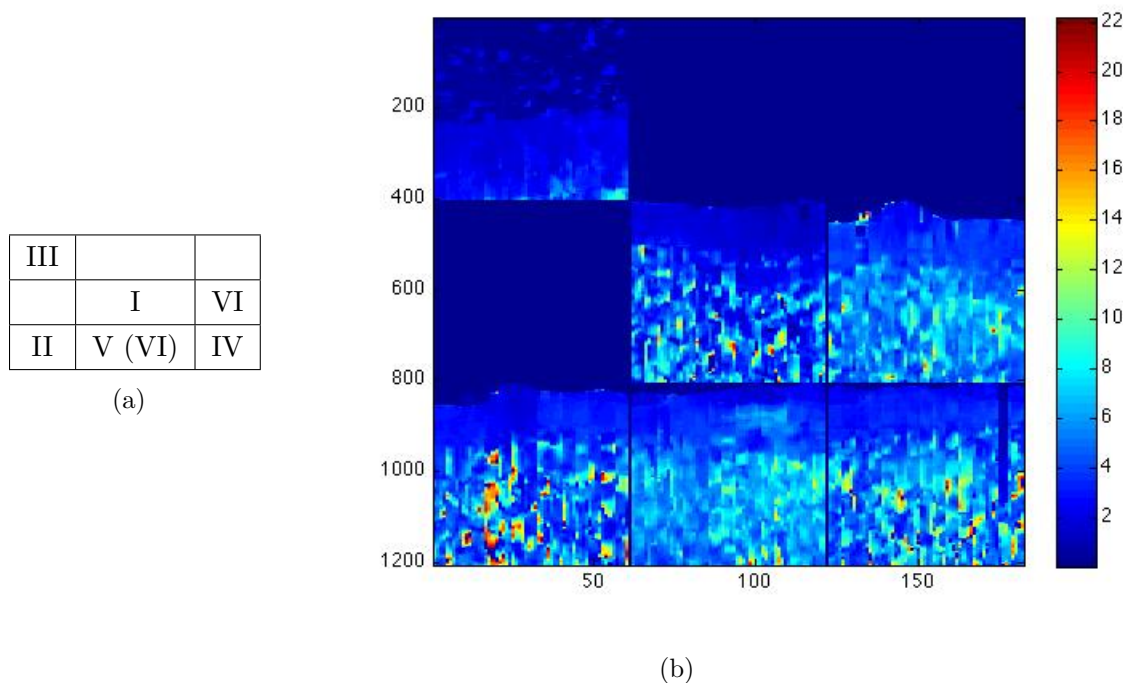


Figure 4.20: The table in (a) shows the final position of each of the original patterns chosen for the training set. The averages of the training data mapped to each neuron after the SOM was trained is shown in (b).

Table 4.8: The results of the validation on the trained SOM shown in Figure 4.20. The numbers in parenthesis after each pattern number indicates the number of images of that pattern type clustered to the neuron.

| | | |
|----------------------------|--|-------------------|
| II. (2), III. (2), X. (10) | VII. (1), VIII. (6) | unused |
| VII. (6) | I. (9) | V. (5) |
| I. (1), II. (9) | IV. (4), V. (2), VI. (5), VIII. (1), IX. (8), X. (1) | III. (3), IV. (1) |

Results for clustering images in the validation set using the trained SOM are shown in Table 4.8. We observe that two previously unused neurons (2 and 4) had data cluster to them during validation. Hence, neuron 3 is the only unused neuron subsequent to training the SOM followed by using the remaining images for validation. The validation images clustering to both (originally unused) neurons 2 and 4 were from patterns that had not been used for training. After clustering images from the validation set some mismatches are apparent (Table 4.8). However, no pattern has its associated images map to more than two clusters. Patterns I, IV, VII, VIII and X have only one mismatched image, while patterns II., III., and V. had two mismatched images. Pattern III has a 3-2 split and pattern V has a 5-2 split. Overall, among patterns that were in the training set, 23 of 43 (53%) mapped to the cluster identified after training. Among patterns in the validation set, 30 of 33 (91%) mapped to the cluster containing the larger number of images for that pattern. Among all patterns, 35 of 43 (81%) mapped to the cluster containing the larger number of images for that pattern.

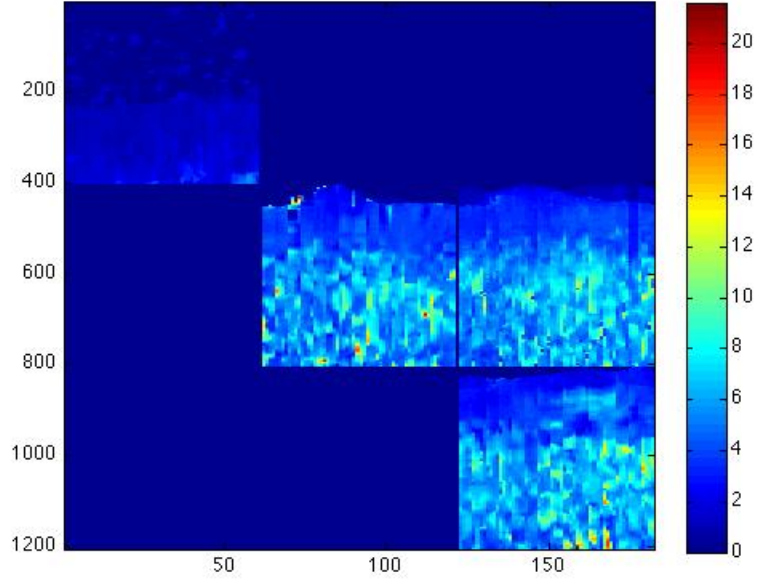
To illustrate a less than optimal case, Figure 4.21 shows the SOM that was trained on the original porcine disease patterns III., IV., V., VI., IX., and X. Figure 4.21(c) demonstrates that there were five unused neurons after the SOM had been trained on data corresponding to six patterns, and Figure 4.21(a) shows where the training data clustered by pattern number. Figure 4.21(b) shows how many training images mapped to each neuron after the SOM was trained. Patterns III., IV., V., and IX. were the training patterns with no mismatches—all the training images based on these particular patterns clustered to the same neuron as the original pattern during training. Unfortunately, neuron 9 is the only neuron that contains images of only one pattern type after training. The validation for this SOM is shown in Table 4.9. After training the SOM and using the validation set, the previously unused neuron 4 is now used, but there are still four unused neurons. There are also many mismatches in the validation set. For example, images based on the original porcine disease pattern I. are clustered to four different neurons in validation.

| | | |
|--------|----|---------------|
| III, X | | |
| | VI | IV, VI, IX, X |
| | | V |

(a)

| | | |
|---|---|----|
| 7 | 0 | 0 |
| 0 | 3 | 10 |
| 0 | 0 | 4 |

(b)



(c)

Figure 4.21: The table in (a) shows the final position of each of the original patterns chosen for the training set. The number of images mapped to each neuron after training is shown in (b). The averages of the training data mapped to each neuron after the SOM was trained is shown in (c).

Table 4.9: Results of the validation set of images for the trained SOM shown in Figure 4.21. The numbers in parenthesis after the pattern type give the number of images of that pattern mapped to that neuron.

| | | |
|--|-----------------|--|
| I. (2), II. (3), III. (4), VIII. (2) | unused | unused |
| III. (1), IV. (5), V. (7), VIII. (3), IX. (4), X. (6) | I. (2), II. (3) | I. (4), II. (9), VI. (5), VII. (7), VIII. (2), X. (1) |
| unused | unused | I. (6) |

4.3.1 Towards Optimizing the Choice of Training Sets

Based on the many realizations of training and validation data subsets considered in the previous section, a natural question is whether a feature or statistic of the training set correlates to success in validation. In a clinical setting, this type of quantitative measure could be used to design or assess the quality of a data set or subsets being considered for training an SOM to be used, later on, for classifying new images. To briefly investigate this question, the following quantitative analysis was performed:

1. For each original pattern D_i in the training set, compute the distance between D_i and all other patterns $D_j (j \neq i)$ in the training set.
2. Record the minimum of these distances, referring to the value as d_k .
3. Formulate a statistic ρ as the sum of all values of d_k , i.e. for each pattern in the training set.

The statistic ρ provides a quantitative measure of the combined distance between vectors (images) in the training space. When this statistic is calculated across varying candidates for training sets, it measures the extent to which training vectors are spread out in the data space. It is hypothesized that training sets with larger values of ρ will exhibit better performance in both training and validation than those with significantly lower values of ρ . Table 4.10 shows three of the 210 training set combinations that were predicted to perform well and three training set combinations that were predicted to not perform well. A strong correlation between the performance of the SOM in both training and validation and the value of ρ is observed.

Table 4.10: The table displays three training sets predicted to train a SOM well (a), and three training sets predicted to not train a SOM well (b) based on their computed values of the statistic ρ for the training set.

| Training Set | Statistic |
|--------------|-----------|
| R1 | 3404.6 |
| R8 | 3212 |
| R18 | 3015.5 |

(a)

| Training Set | Statistic |
|--------------|-----------|
| R188 | 2604.9 |
| R136 | 2860.6 |
| R94 | 2760.8 |

(b)

The training sets designated by R1 (shown in Figure 4.20), R8, and R18 each have three unused neurons after training a SOM and only one unused neuron after validation. The training

sets designated by R188 (shown in Figure 4.21), R136, and R94 all have five unused neurons after training a SOM and four unused neurons after validation. Values of ρ were significantly higher for the SOMs that performed optimally as compared to those that performed poorly (3016-3405 vs 2605-2861). These results suggest that the statistic ρ can be used to compare a variety of choices for data subsets being considered for use in training a SOM.

4.4 Summary and Conclusions

Heart attacks are a leading cause of death worldwide and the development of novel and innovative techniques to detect atherosclerosis early and characterize a patient's risk level have potential impacts on both the quality of life as well as healthcare costs. Atherosclerosis is initiated when the arterial wall begins to change due to a series of events that can lead to thrombosis. ARFI provides a novel technique for non-invasive imaging of arteries by using the scattering and absorbing properties of acoustic waves applied to tissue. The emitted ultrasonic waves are sequenced in a manner that allows a pulse to be focused both spatially and temporally and the resulting tissue displacement to be recorded and analyzed. This information provides peak displacement and recovery time data at each point in space for the region of interest in the tissue. Much work has also been done to mask the noise created by blood flowing through imaged arteries in order to produce useful images [3, 20, 21, 22]. Transformed images based on tissue responses measured using ARFI provide more details about the mechanical properties of tissue regions containing plaque than traditional B-mode ultrasound images. They are therefore useful for diagnosing both the presence and the severity of atherosclerosis and the type of plaque. In this dissertation, specialized clustering algorithms were developed for analyzing ARFI ultrasound imaging data for gel-derived tissue phantoms and for diseased tissue from porcine iliac arteries. The clustering algorithms were based on the use of Self-Organizing Maps (SOM), a type of artificial neural network based on unsupervised learning that employs an underlying network of connected neurons whose weight vectors represent patterns identified in the data set.

In our application to ultrasound imaging data, SOMs achieve clustering by projecting high dimensional data onto a lower dimensional space of neurons arranged on a two dimensional lattice. These neurons are interconnected and the notion of neighboring neurons occurs based on the distances between neurons in the SOM. In Chapter 2, the basic SOM algorithm was presented and factors relating to efficient implementation for high dimensional data, expansion methods for data sets of real images, and approaches for validation were developed. Based on the dimension, size and scope of image patterns considered in this work, the SOM configuration used for all data sets considered was a 3x3 lattice of nine neurons. Once a SOM is trained using a particular data set, the number of used and unused neurons serves as an indicator of the scope

of patterns identified in the data set. Since only a relatively small number of original images were available, SOMs trained on only original images commonly had a large number of unused neurons. As a result, two data expansion techniques were considered to increase the number of image samples as well as the variety of patterns contained in the images used to train the SOMs. The first expansion method perturbed images at each pixel by randomly sampling a normal distribution based on average pixel value in the original image and a specified variance. Alternatively, the second expansion scheme used weighted averages of original images to create an expanded data set with more realistic looking images that exhibited a more continuous transition in plaque features. After training a SOM, two metrics were also proposed in order to determine the quality of the data clustering. The first metric determined the success of a trained SOM by examining the number of unused neurons after both training and validation. The second metric was a statistic calculated from the training data set that measured the extent to which these patterns were spread out in the data space. It was hypothesized that higher values of this measure would correlate with successful SOM performance.

In Chapter 3, data sets of increasing complexity were considered to develop and refine the SOM methods presented in Chapter 2. The first set of data considered was obtained from the Gallippi Ultrasound Lab at the University of North Carolina at Chapel Hill. This data set was comprised of ARFI peak displacement and relaxation time images for gels constructed to mimic arterial tissues with atherosclerotic plaques (phantoms). In the synthesis of these tissue phantoms, two inclusion sizes and two inclusions stiffnesses were manufactured. When the data was used to train a SOM with vectorization, the trained SOM used two neurons to correctly identify the primary distinguishing feature as plaque stiffness (hard or soft), but failed to cluster on the other features reflected in the data set. However, subsequent training of SOMs on each data subgroup (soft and hard), resulted in a clustering that used three neurons and identified additional features such as plaque size and location. The difficulty in identifying secondary features in the full data set occurred because displacements in plaque regions have relatively large magnitudes and affect substantial portion of each image's pixels. In an attempt to utilize more neurons in a trained SOM, two data sets were manufactured to mimic larger data sets with continuous variations in major features. The first manufactured data set contained rectangles of various sizes and stiffnesses to represent plaques, and the second manufactured data set contained ellipses of different sizes and stiffnesses. The SOMs trained on these two manufactured data sets achieved better clusterings, using up to five of the nine available neurons. This led us to search for a realistic data set for ARFI imaging of actual tissues that also contained more continuous variations in features.

In Chapter 4, the SOM techniques developed and evaluated in Chapters 2 and 3, were used to evaluate SOM-based data clustering for ARFI images of porcine tissues obtained from the Gallippi Ultrasound Lab. This time the data came from an *ex vivo* ARFI imaging study

on porcine iliac arteries. Compared to the data sets considered in Chapter 3, there was more variation in the diseased samples as well as more noise since samples were obtained from actual tissue (as opposed to tissue phantoms) and since ARFI imaging was conducted on live pigs. These pigs were either fed a high fat diet to produce hypercholesterolemia and atherosclerosis or fed a low cholesterol diet to maintain good health. Expansion of the data set was considered using both methods summarized above. New patterns obtained by the pixel-by-pixel expansion with noise resembled the original pattern to the extent that the pattern number was used to evaluate clustering by the trained SOM. The SOM trained on this data had no mismatches and used all nine neurons. To ensure that the SOM was not distinguishing patterns based on features in the image away from the plaque region, the data set was cropped. Again, the re-trained SOMs did not have any mismatches within pattern types and used all nine neurons. In order to evaluate SOM interpolation and pattern recognition capabilities for larger, more varied data sets, the expanded data set was partitioned into training (of less than nine patterns) and validation subsets. An extensive analysis of SOM performance was carried out by training SOMs for possible combinations of training sets comprised of either 6 or 7 patterns. In each instance, the number of unused neurons following both the training and validation of each SOM was examined to evaluate success of the method. An optimal sort resulting in only one unused neuron was identified. The second data expansion method was then applied to the original disease patterns by taking weighted averages of the original patterns, creating an expanded data set with more realistic and continuous variations in plaque features relative to the first expansion method. This expanded data set was partitioned into training and validation sets and each possible combination of six original patterns was used to train a SOM. On this more varied data set, optimally trained SOMs now contained mismatches but each pattern in the training or validation set mapped to no more than two neurons in the trained SOM. For clustering of data in both the training and validation sets, roughly 80% of the original disease patterns clustered to the same neuron. In order to identify when a training set is correlated with optimal SOM clustering, the statistic developed in Chapter 2 was evaluated across all training sets. Results demonstrated a strong correlation between larger values of the statistic and the ability of the trained SOM to successfully cluster data in both the training and validation sets.

While methods were developed and capabilities evaluated in the context of ARFI ultrasound imaging data sets, several considerations were beyond the scope of this work. First, the data sets considered in this study were almost exclusively based on values of peak displacement measured using ARFI imaging. Combined analysis of image data for both peak displacement and relaxation time could potentially expand the feature identification capabilities of the SOM. In particular delineation or the correlation of elastic or viscoelastic plaque properties with structural features of the plaque (e.g. collagen or elastic composition) would be possible. The approach used for weight initialization was almost exclusively based on knowledge of the data

used for training. This technique is most relevant to studies in which the SOM is trained on a limited set of patient images and later used to cluster and classify image patterns obtained from new patients. While patterns in the SOM are initially set to specific data vectors, the relationship between distances among patterns in data space and in the space of the lower dimension projection onto the 2-D SOM is non-trivial. Future studies could consider alternate approaches to weight vector initialization such as the use of Principle Components Analysis for initialization prior to SOM training. All results presented in this study were based on the use of a 3x3 map of nine neurons. This choice was well suited to the number of varied patterns and the size of the data sets considered in this work. Studies on larger and more diverse data sets based on imaging of normal and diseased tissue across broader patient populations could explore the use of self-organizing networks that grow in the number of neurons and the dimension of the SOM during training [44]. While the methods developed in this study were in a research setting, in clinical practice, cropping methods and implementing multiple training sets could pose potential problems for real-time in-clinic use. Since the training phase for large data sets can take a substantial amount of time, it would be most useful, in practice, for clinics to obtain a SOM that is already trained appropriately and then used to evaluate new patient data as demonstrated for our validation sets in Chapter 4 in order to classify new images relative to existing, well established disease patterns. Overall, the artificial neural network methods and algorithms, data expansion techniques, and approaches to validation developed in this study demonstrate the strong potential for application of SOM-based clustering to pattern recognition and classification in ultrasound imaging of atherosclerotic tissues.

REFERENCES

- [1] A. Alizad, M. Fatemi, L.E. Wold, and J.F. Greenleaf. Performance of vibroacoustiography in detecting microcalcifications in excised human breast tissue: A study of 74 tissue samples. *IEEE Trans Med Imaging*, 23:307–312, 2004.
- [2] H. Ashrafian, T. Lim, and R. Senior. Carotid ultrasound imaging: An effective technique for detecting early atherosclerosis—vascular screening for cvd has come of age. *J Am Soc Echocardiogr*, 20:1–3, 2007.
- [3] R. Behler, T. Nichols, H. Zhu, E.P. Merricks, and C.M. Gallippi. Arfi imaging for noninvasive material characterization of atherosclerosis part ii: toward in vivo characterization. *Ultrasound in Med Biol*, 35(2):278–295, 2009.
- [4] R.H. Behler, D.M. Dumont, T.C. Nichols, E.P. Merricks, and C.M. Gallippi. Arfi ultrasound for characterizing atherosclerosis. *IEEE Ultrason Symp Proc*, pages 722–727, 2006.
- [5] R.H. Behler, T.C. Nichols, E.P. Merricks, and C.M. Gallippi. A rigid wall approach to physiologic motion rejection in arfi imaging. *IEEE Ultrason Symp Proc*, 2007.
- [6] R.H. Behler, M.R. Scola, and C.M. Gallippi. Reverberation artifact rejection and masking in arterial arfi imaging. *Proc. IEEE Ultrason. Symp.*, pages 2367–2370, 2009.
- [7] J. Bercoff, M. Pernot, M. Tanter, and M. Fink. Monitoring thermally-induced lesions with supersonic shear imaging. *Ultrason Imaging*, 26:71–84, 2004.
- [8] D. Boese, A. Shumermund, P. Margolis, A. Nair, G. Vince, H. Eggebrecht, C. Naber, and R. Erbel. Intravascular ultrasound radiofrequency analysis identifies plaque composition: Virtual histology. *Eur Heart J*, 25, 2004.
- [9] J. Bojunga, N. Dauth, C. Berner, G. Meyer, K. Holzer, L. Voelkl, E. Hermann,

- H. Schroeter, and S. Zeuzem. Acoustic radiation force impulse imaging for differentiation of thyroid nodules. *PLoS One*, 7(8):e42735, 2012.
- [10] D. Bose, C. von Birgelen, and R. Erbel. Intravascular ultrasound for the evaluation of therapies targeting coronary atherosclerosis. *J Am Coll Cardiol*, 49:925–932, 2007.
- [11] D.P. Bradway, S.J. Hsu, B.J. Fahey, J.J. Dahl, T.C. Nichols, and G.E. Trahey. Transthoracic cardiac arfi: A feasibility study. *IEEE International Ultrasonics Symposium*, 2007.
- [12] N. Brodala, E.P. Merricks, D.A. Bellinger, D. Damrongsri, S. Offenbacher, J. Beck, P. Madianos, D. Sotres, Y.L. Chang, G. Koch, and T.C. Nichols. Porphyromonas gingivalis bacteremia induces coronary and aortic atherosclerosis in normocholesterolemic and hypercholesterolemic pigs. *Arterioscler Thromb Vasc Biol*, 25:1451–1456, 2005.
- [13] D.R. Chen, R.F. Chang, and Y.L. Huang. Breast cancer diagnosis using self-organizing map for sonography. *Ultrasound in Med and Biol*, 26(3):405–411, 2000.
- [14] J.J. Dahl, D.M. Dumont, E.M. Miller, E. Schwark, J.D. Allen, and G.E. Trahey. Characterization of in vivo atherosclerosis plaques in the carotid artery with acoustic radiation force impulse imaging. *Proc IEEE Ultrason Symp*, 1:706–709, 2006.
- [15] C. de Korte, M. Sierevogel, F. Mastik, C. Strijder, J.A. Schaar, E. Velma, G. Pasterkamp, P.W. Serruys, and A.F. van der Steen. Identification of atherosclerosis plaque components with intravascular ultrasound elastography in vivo: a yucatan pig study. *Circulation*, 105:1627–1630, 2002.
- [16] D. Dumont, R. Behler, T. Nichols, E.P. Merricks, and C.M. Gallippi. Arfi imaging for noninvasive material characterization of atherosclerosis. *Ultrasound in Med Biol*, 32(11):1703–1711, 2006.
- [17] B.J. Fahey, K.R. Nightingale, R.C. Nelson, M.L. Palmeri, and G.E. Trahey. Acoustic radiation force impulse imaging of the abdomen: Demonstration of feasibility and utility. *Ultrasound Med Biol*, 31:1185–1198, 2005.

- [18] Y.C. Fung. *Biomechanics: Mechanical properties of living tissues*. Springer, New York, 1993.
- [19] C.M. Gallippi. Acoustic radiation force (arf) ultrasound. Available at:<http://www.bme.unc.edu/cmgallip/Site/Research.html>.
- [20] C.M. Gallippi, K.R. Nightingale, and G.E. Trahey. Bss-based filtering of physiological and arfi-induced tissue and blood motion. *Ultrasound in Med and Biol*, 29(11):1583–1592, 2003.
- [21] C.M. Gallippi and G.E. Trahey. Adaptive clutter filtering via blind signal separation for lateral blood velocity measurement. *Ultrasonics Symposium*, 2:1517–1521, 2002.
- [22] C.M. Gallippi and G.E. Trahey. Adaptive clutter filtering via blind source separation for two-dimensional ultrasonic blood velocity measurement. *Ultrasonic Imaging*, 24(4):193–214, 2002.
- [23] C.M. Gallippi and G.E. Trahey. Complex blind source separation for acoustic radiation force impulse imaging in the peripheral vasculature, in vivo. *Ultrason Symp*, 1:596–601, 2004.
- [24] R.G. Gerrity, R. Natarajan, J.L. Nadler, and T. Kimsey. Diabetes-induced accerlerated atherosclerosis in swine. *Diabetes*, 50:1654–1665, 2001.
- [25] R.E. Gerszten and T.J. Wang. The search for new cardiovascular biomarkers. *Nature*, 451:949–952, 2008.
- [26] S. Glagov. Compensatory enlargement of human atherosclerosis coronary arteries. *N Eng J Med*, 316:1371–1375, 1987.
- [27] S. Hanquinet, A.L. Rougemont, D. Courvoisier, L. Rubbia-Brandt, V. McLin, M. Tempia, and M. Anooshiravani. Acoustic radiation force impulse (arfi) elastography for the noninvasive diagnosis of liver fibrosis in children. *Pediatric Radiology*, 43(5):545–551, 2013.

- [28] D. Hebb. *Organization of Behavior*. Wiley, New York, 1949.
- [29] S.J. Hsu, J.L. Hubert, P.D. Wolf, and G.E. Trahey. Acoustic radiation force impulse imaging of mechanical stiffness propagation within myocardial tissue. *IEEE UFFC International Ultrasonics Symposium*, October 2007.
- [30] L.C. Junqueira, W. Cossermelli, and R. Bertani. Picosirius staining plus polarization microscopy, a specific model for collagen detection in tissue sections. *Histochemical J*, 11:447–455, 1979.
- [31] L.C. Junqueira, W. Cossermelli, and R. Bretani. Differential staining of collagens type i, ii, and iii by sirius red and polarization microscopy. *Archiv Histol Jap*, 41:267–274, 1978.
- [32] J.H. Kaas, M.M. Merzenich, and H.P. Killackey. The recognition of somatosensory cortex following peripheral nerve damage in adult and developing mammals. *Ann. Rev. Neurosci.*, 6:325–356, 1983.
- [33] J.H. Kaas, R.J. Nelson, M. Sur, C.S. Lin, and M.M. Merzenich. Multiple representations of the body within the primary somatosensory cortex of primates. *Science*, 204:521–523, 1979.
- [34] T. Kohonen. Analysis of a simple self-organizing process. *Bio. Cybern.*, 44:135–140, 1982.
- [35] T. Kohonen. Self-organized formation of topologically correct feature maps. *Bio. Cybern.*, 43:59–69, 1982.
- [36] T. Kohonen. *Self-Organization and Associative Memory*. Springer, Verlag, 1984.
- [37] P. Libby, P.M. Ridker, and G.K. Hansson. Progress and challenges in translating the biology of atherosclerosis. *Nature*, 473:317–325, May 2011.
- [38] P. Libby and P. Theroux. Pathophysiology of coronary artery disease. *Circulation*, 111:3481–3488, 2005.

- [39] D. Liu and E. Ebbini. Viscoelastic property measurement in thin tissue constructs using ultrasound. *Ultrasound in Med Biol*, 29(12):1715–1723, 2003.
- [40] J. Mackay and G. Mensah. Atlas of heart disease and stroke. *World Health Organization*, 2004.
- [41] E.L. Madsen, M.A. Hobson, H. Shi, T. Varghese, and G.R. Frank. Tissue-mimicking agar/gelatin materials for use in heterogeneous elastography phantoms. *Phys. Med. Biol.*, 50:5597–5597, 2005.
- [42] M.K. Markey, J.Y. Lo, G.D. Toursassi, and C.E. Jr. Floyd. Self-organizing map for cluster analysis of a breast cancer database. *Artificial Intelligence in Medicine*, 27(2):113–127, 2003.
- [43] S. Marsland. *Machine Learning: An Algorithmic Perspective*. Taylor and Francis Group, 2009.
- [44] S. Marsland, J. Shapiro, and U Nehmzow. A self-organizing network that grows when required. *Neural Netw*, 15(8-9):1041–1058, 2002.
- [45] F.R. Mauldin, M.A. Haider, E.G. Loba, and R.H. Behler. Monitored steady-state excitation and recovery (msser) radiation force imaging using viscoelastic models. *Ultrasonics, Ferroelectrics and Frequency Control*, 55(7):1597–1610, 2008.
- [46] F.R. Mauldin, J.H. Levy, R.H. Behler, T.C. Nichols, J.S. Marron, and C.M. Gallippi. P2e-3 blind source separation and k-means clustering for vascular arfi image segmentation, in vivo and ex vivo. *Ultrason Sym*, pages 1666–1671, 2006.
- [47] F.W. Mauldin, H.T. Zhu, R.H. Behler, T.C. Nichols, and C.M. Gallippi. Robust principal component analysis and clustering methods for automated classification of tissue response to arfi excitation. *Ultrasound in Medicine and Biology*, 32(2):309–325, 2008.
- [48] W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5:115–133, 1943.

- [49] P.R. Moreno, K.R. Purushothaman, V. Fuster, and W.N. O'Connor. Intimomedial interface damage is increased beneath disrupted and adventitial inflammation atherosclerosis in the aorta—implications for plaque vulnerability. *Circ*, 105:2504–2511, 2002.
- [50] G. Mozes, R.R. Kinnick, P. Gloviczki, R.E. Bruhnke, M. Carmo, T.L. Hoskin, K.E. Bennett, and J.F. Greenleaf. Noninvasive measurement of aortic aneurysm sac tension with vibrometry. *J Vas Surg*, 42:963–971, 2005.
- [51] A. Nair, B.D. Kuban, E.M. Tuzcu, P. Schoenhagen, S.E. Nissen, and D.G. Vince. Coronary plaque classification with intravascular ultrasound radiofrequency data analysis. *Circulation*, 106:2200–2206, 2002.
- [52] T.C. Nichols, D.A. Bellinger, K.E. Davis, G.G. Koch, R.L. Reddick, M.S. Read, J. Rapacz, J. Hasler-Rapacz, Brinkhouse K.M., and T.R. Griggs. Porcine von willebrand disease and atherosclerosis: Influence of polymorphism in apolipoprotein b100 genotype. *Am J Path*, 140:403–415, 1992.
- [53] T.C. Nichols, D.A. Bellinger, D.A. Tate, R.L. Reddick, M.S. Read, G.G. Koch, K.M. Brinkhous, and T.R. Griggs. von willebrand factor and occlusive arterial thrombosis: A study in normal and von willebrand's disease pigs with diet-induced hypercholesterolemia and atherosclerosis. *Atherosclerosis*, 10:449–461, 1990.
- [54] T.C. Nichols, W.H. Jr. Busby, E. Merricks, J. Sipos, M. Rowland, K. Sitko, and D.R. Clemmons. Protease resistant igfbp-4 inhibits igf-i actions and neointimal expansion in a porcine model of neointimal hyperplasia. *Endocrinol*, 10:5002–5010, 2007.
- [55] T.C. Nichols, T. du Laney, B. Zheng, D.A. Bellinger, G.A. Nickols, W. Engleman, and D.R. Clemmons. Reduction in atherosclerotic lesion size in pigs by alphavbeta3 inhibitors is associated with inhibition of insulin-like growth factor-i-mediated signaling. *Circ Res*, 85:1040–1045, 1999.

- [56] K. Nightingale. Acoustic radiation force impulse (arfi) imaging: a review. *Imaging Review*, 7(4):328–339, 2011.
- [57] K. Nightingale, S. McAleavy, and G. Trahey. Shear-wave generation using acoustic radiation force: in vivo and ex vivo results. *Ultrasound in Med Biol*, 29(12):1715–1723, 2003.
- [58] K. Nightingale, M.S. Soo, R. Nightingale, and G.E. Trahey. Acoustic radiation force impulse imaging: In vivo demonstration of clinical feasibility. *Ultrasound Med Biol*, 28(2):227–235, 2002.
- [59] K.R. Nightingale. Acoustic radiation force impulse (arfi) imaging overview. Available at: <http://kathynightingalelab.pratt.duke.edu/radiationforcebasedimaging>.
- [60] K.R. Nightingale, P.J. Komguth, W.F. Walker, B.A. McDermott, and G.E. Trahey. A novel technique for differentiating cysts from solid lesions: preliminary results in the breast. *Ultrasound Med Biol*, 21(6):745–751, 1995.
- [61] K.R. Nightingale, L. Zhai, J. Dahl, K. Frinkley, and M. Palmeri. Shear wave velocity estimation using acoustic radiation force impulse excitation in liver. *IEEE Ultrason Symp*, pages 1156–1160, 2006.
- [62] M. Palmeri, K. Frinkley, K. Oldenburg, and K. Nightingale. Characterizing acoustic attenuation of homogeneous media using forced impulsive acoustic radiation force. *Ultrasonic Imaging*, 18(2):114–128, 2006.
- [63] M. Palmeri, K.D. Frinkley, L. Zhai, M. Gottfried, R.C. Bentley, K. Ludwig, and K.R. Nightingale. Arfi imaging of the gastrointestinal tract. *Ultrasound Imaging*, 27:75–88, 2005a.
- [64] M. Palmeri, S. McAleavey, G. Trahey, and K. Nightingale. Ultrasonic tracking of acoustic radiation force induced displacements in homogenous media. *IEEE Transaction on Ultrasonics, Ferroelectrics, and Frequency Control*, 53(7):1300–1313, 2006.

- [65] M.F. Prescott, C.H. McBride, J. Hasler-Rapacz, J. Von Linden, and J. Rapacz. Development of complex atherosclerotic lesions in pigs with inherited hepyer-ldl cholesterolemia bearing mutant alleles for apolipoprotein b. *Am J Pathol*, 139:139–147, 1991.
- [66] M. Pulido, D. Angiolillo, and M. Costa. Imaging of the atherosclerotic plaque. *Int J Cardiovasc Imaging*, 10:553–559, 2004.
- [67] D.J. Rader and A. Daugherty. Translating molecular discoveries into new therapies for atherosclerosis. *Nature*, 451:904–913, 2008.
- [68] P. Raggi, A. Taylor, Z. Fayad, D. O’Leary, S. Nissen, D. Rader, and L.J. Shaw. Atherosclerotic plaque imaging—contemporary role in preventive cardiology. *Arch Intern Med*, 165:2345–2353, 2005.
- [69] M.D. Rekhter. Collagen synthesis in atherosclerosis: Too much and not enough. *Cardio-vasc Res*, 41:376–384, 1999.
- [70] H. Ritter. *Neural Computation and Self-Organizing Maps*. Addison-Wesley, 1992.
- [71] W. Rosamond, K. Flegal, K. Furie, A. Go, N. Greenlund, K. an Haase, S.M. Hailpern, M. Ho, V. Howard, B. Kissela, S. Kittner, D. Lloyd-Jones, M. McDermott, J. Meigs, C. Moy, G. Nichol, C. O’Donnell, V. Roger, P. Sorlie, J. Steinberger, T. Thom, M. Wilson, and Yu Hong. American heart association statistics committee and stroke statistics subcommittee. heart disease and stroke statistics–2008 update: A report from the american heart association statistics committee and stroke statistics subcommittee. *Circulation*, 117:e25–e146, 2008.
- [72] T. Rundek, B. Arif, H. Boden-Albala, M.S. Elkind, M.C. Paik, and R.L. Sacco. Carotid plaque, a subclinical precursor of vascular events—the northern manhattan study. *Neurology*, 70:1200–1207, 2008.
- [73] R. Salonen and J. Salonen. Progression of carotid atherosclerosis and its determinants: a population-based ultrasonography study. *Atherosclerosis*, 81(1):33–40, 1990.

- [74] J. Sanz and Z.A. Fayad. Imaging of atherosclerotic cardiovascular disease. *Nature*, 451:953–957, 2008.
- [75] N. Suga and W.E. O’Neil. Neural axis representing target range in the auditory cortex of the mustache bat. *Science*, 206:351–353, 1979.
- [76] G. Trahey, M. Palmeri, R. Bentley, and K. Nightingale. Acoustic radiation force impulse imaging of the mechanical properties of arteries. *Ultrasound Med Biol*, 30:1163–1171, 2004.
- [77] G.E. Trahey, J.J. Dahl, S.A. McAleavey, C.M. Gallippi, and K.R. Nightingale. Arterial stiffness measurements with acoustic radiation force impulse imaging. *Proc. SPIE 5035, Medical Imaging 2003: Ultrasonic Imaging and Signal Processing*, 235:doi:10.1117/12.479966, 2003.
- [78] C. Vijayakumar, G. Damayanti, R. Pant, and C.M. Sreedhar. Segmentation and grading of brain tumors on apparent diffusion coefficient images using self-organizing maps. *Computerized Medical Imaging and Graphics*, 31(7):473–484, 2007.
- [79] F. Viola, M.D. Kramer, M.B. Lawrence, J.P. Oberhauser, and W.F. Walker. Sonorheometry: A noncontact methods for the dynamic assessment of thrombosis. *Ann Biomed Eng*, 32:696–705, 2004.
- [80] J.R. Womersley. Oscillatory motion of a viscous liquid in a thin-walled elastic tube-i: The linear approximation for long waves. *Phil Mag*, 46:199–221, 1955.
- [81] C Wu, M.H. Lin, and J.L. Su. Computer-aided diagnosis system for pancreatic tumor detection in ultrasound images. *Biomedical Engineering, IFMBE Proceedings*, 35:627–630, 2011.
- [82] C. Yuan, W.S. Kerwin, V.L. Yarnykh, J. Cai, T. Saam, B. Chu, N. Takaya, M.S. Ferguson, H. Underhill, D. Xu, F. Liu, and T.S. Hatsukami. Mri of atherosclerosis in clinical trials. *NMR Biomed*, 19:636–654, 2006.

- [83] X.M. Zhang and J.F. Greenleaf. Noninvasive generation and measurement of propagating waves in arterial wall. *J Acoust Soc Am*, 119:1238–1243, 2006.
- [84] X.M. Zhang and J.F. Greenleaf. Estimation of tissue’s elasticity with surface wave speed (1). *J Acoust Soc Am*, 122:2522–2525, 2007.
- [85] Y. Zheng, J.F. Greenleaf, and J.J. Gisvold. Reduction of breast biopsies with a modified self-organizing map. *Neural Networks*, 8(6):1386–1396, 1997.