Ultrasound-based Approaches to Tissue Classification for Breast and Prostate Cancer Diagnosis

by

Nishant Uniyal

B.Sc. Electrical Engineering, Missouri University of Science & Technology, 2012
B.Sc. Computer Engineering, Missouri University of Science & Technology, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Masters of Applied Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Electrical and Computer Engineering)

The University Of British Columbia
(Vancouver)

August 2014

© Nishant Uniyal, 2014
Abstract

Ultrasound-based cancer diagnosis could improve the current breast and prostate cancer diagnosis methods. In this thesis, an ultrasound-based approach is evaluated as a method for breast and prostate cancer diagnosis. Ultrasound RF time series along with a comprehensive machine learning framework is used for accurate classification of tissue samples. The RF time series method requires only a few seconds of raw ultrasound data with no need for additional instrumentation. The developed method produces cancer probability/likelihood maps which show the probability of the tissue under study being cancerous. These probability maps could provide radiologists with a real-time cancer diagnosis tool which could improve cancer yield and significantly reduce the number of negative biopsies.

To prove the utility of ultrasound RF time series as a tissue classification method, an *in vivo* breast lesion classification study of 22 subjects and an *in vivo* prostate biopsy core classification study involving 18 subjects is presented in this thesis. A comprehensive machine learning framework with a new semi-supervised learning technique for tissue classification is also presented in this work. An experimental study to substantiate the ultrasound RF time series hypotheses by studying the effects of ultrasound imaging parameters on animal tissue classification is also presented. Using the ultrasound RF time series method and the developed machine learning framework— we calculated the area under the receiver operating characteristics curve to be 85.6% for breast lesion classification and 91.5% for prostate tissue classification. Increasing the frame rate and the length of the time series, and decreasing the imaging depth we observed consistent improvement in tissue classification results for the animal study.

The results of this thesis suggest the potential of ultrasound RF time series
as a tissue classification method. Ultrasound RF time series along with other ultrasound-based methods could be a valuable and practical addition to the current cancer diagnosis procedures. It has been shown here that a high level of accuracy can be attained using these tools which are non-invasive, inexpensive and readily available to the clinician.
Preface

A short version of Chapter 2 has been published in the proceedings of IEEE Ultrasonics Symposium 2013 as "A new approach to ultrasonic detection of malignant breast tumors" [1]. The article was co-authored by Hani Eskandari, Purang Abolmaesumi, Samira Sojoudi, Paula Gordon, Linda Warren, Robert N. Rohling, Septimiu E. Salcudean, and Mehdi Moradi. Chapter 2 has been submitted for publication in IEEE Transactions on Medical Imaging. A revised version was submitted in May 2014. The author performed all the analysis on the data and developed the required algorithms.

The data collection for the breast study was performed during a study approved by the Clinical Research Ethics Board at the University of British Columbia, Vancouver. The UBC CREB number of this study is H12-00889. Tim S.E. Salcudean is the principal investigator for this study. This study was designed to evaluate ultrasound-based tissue typing and vibro-elastography for breast lesion classification. The vibro-elastography technology was developed by Salcudean et al. Hani Eskandari and Samira Sojoudi collected the data for this study.

Chapter 3 is accepted for publication as a chapter in Lecture Notes in Computer Science series and presentation at Medical Image Computing and Computer Assisted Intervention (MICCAI) 2014 workshop on clinical image-based procedures, CLIP 2014. The report is co-authored by Farhad Imani, Amir Tahmasebi, Harsh Agarwal, Shyam Bharat, Pingkun Yan, Jochen Kruecker, Jin Tae Kwak, Sheng Xu, Bradford Wood, Peter Pinto, Baris Turkbey, Peter Choyke, Purang Abolmaesumi, and Baris Turkbey.

Parvin Mousavi, and Mehdi Moradi. Farhad Imani is also a lead co-author on this paper.

Data for the prostate study was collected at the National Institutes of Health Clinical Center (NIH-CC, Bethesda, MD). This work was a joint effort with NIH, Philips, and Queens University.
Table of Contents

Abstract .........................................................  ii
Preface .........................................................  iv
Table of Contents ............................................... vi
List of Tables ................................................... ix
List of Figures ................................................. x
Glossary ......................................................... xvi
Acknowledgments .............................................. xviii
Dedication ....................................................... xix

1 Introduction ................................................... 1
  1.1 Motivation ................................................ 1
  1.2 Thesis Objectives ........................................ 5
  1.3 Organization of Thesis .................................... 5

2 Ultrasound RF Time Series for Classification of Breast Lesions .... 7
  2.1 Introduction ................................................ 7
  2.2 Methods ..................................................... 9
    2.2.1 Data Collection ...................................... 10
    2.2.2 Features ............................................ 12
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.3</td>
<td>Classification and Estimation of Cancer Likelihood</td>
<td>17</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Evaluation of Classification</td>
<td>20</td>
</tr>
<tr>
<td>2.3</td>
<td>Results</td>
<td>21</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Feature Selection</td>
<td>21</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Classification Results</td>
<td>21</td>
</tr>
<tr>
<td>2.4</td>
<td>Discussion and Conclusion</td>
<td>29</td>
</tr>
<tr>
<td>2.5</td>
<td>Chapter Summary</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>Ultrasound-based Prediction of Prostate Cancer in MRI-guided Biopsy</td>
<td>35</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>35</td>
</tr>
<tr>
<td>3.2</td>
<td>Materials and Methods</td>
<td>37</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Data Acquisition</td>
<td>37</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Feature Extraction</td>
<td>38</td>
</tr>
<tr>
<td>3.2.3</td>
<td>The Proposed Classification Framework</td>
<td>40</td>
</tr>
<tr>
<td>3.3</td>
<td>Results</td>
<td>41</td>
</tr>
<tr>
<td>3.4</td>
<td>Discussion and Conclusion</td>
<td>44</td>
</tr>
<tr>
<td>3.5</td>
<td>Chapter Summary</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>Experiments to Provide Evidence on the Physical Basis of Tissue Classification using RF Time Series</td>
<td>54</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>54</td>
</tr>
<tr>
<td>4.2</td>
<td>Materials and Methods</td>
<td>55</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Types of Tissue</td>
<td>56</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Data Collection</td>
<td>57</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Experiment Design</td>
<td>58</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Features</td>
<td>59</td>
</tr>
<tr>
<td>4.2.5</td>
<td>Classification and Cross-validation</td>
<td>59</td>
</tr>
<tr>
<td>4.3</td>
<td>Results</td>
<td>60</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Effects of Imaging Depth on Tissue Classification</td>
<td>60</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Effects of Frame Rate on Tissue Classification</td>
<td>61</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Effects of Time Series Length on Tissue Classification</td>
<td>63</td>
</tr>
<tr>
<td>4.4</td>
<td>Discussions and Conclusions</td>
<td>63</td>
</tr>
<tr>
<td>4.5</td>
<td>Chapter Summary</td>
<td>65</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1 Ultrasound RF time series, B-mode texture, and attenuation features calculated for classification of malignant and benign breast lesions ........................................... 13
Table 2.2 Sensitivity, specificity at a cut-off $T_{h_c}=0.5$, and area under ROC curve. ..................................................... 26
Table 2.3 Classification results for the 22 subjects. ..................................................... 26
Table 3.1 SVM and Random Forests cancer probabilities. ............................................ 48
Table 4.1 Parameters changed during each experiment. The parameters in **bold** were changed for the corresponding experiment. ........... 56
Table 4.2 1 mm$^2$ ROI size in the RF domain for different imaging depths. 59
Table 4.3 Area under the ROC curve with changing imaging depth. ........ 60
Table 4.4 Area under the ROC curve for different frame rates. ........... 62
Table 4.5 Area under the ROC curve for different time series length. .... 63
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>Overall classification framework. This figure shows a graphical representation of the proposed approach.</td>
<td>10</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Figure showing the ROI selection. Each small rectangle in the larger rectangle, inside the lesion, is an ROI and features were calculated for each of the ROI. This figure does not correctly represent the number of ROIs defined, it is for visual interpretation only.</td>
<td>12</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>The average power spectrum of the reflected RF signal for breast tissue.</td>
<td>15</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Attenuation estimation by calculating the slope of the calibrated spectrum.</td>
<td>16</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Box plots of the nine RF time series features. Table 2.1 provides a description of the features. From the box plots, it is apparent that the best features selected by the RFE algorithm show a difference in the class distribution. As can be seen in the figure, ”Feature07” is the best feature that differentiates the malignant samples from the benign, and it was also selected as the best feature by our feature selection method.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Correlation plot of the 164 texture features extracted from the B-mode image. From the figure, it is observable that most of the features are highly correlated. High positive correlation suggests the need for feature selection prior to classification.</td>
<td>23</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>ROC curve for SVM and Random Forests using three features.</td>
<td>24</td>
</tr>
</tbody>
</table>
Figure 2.8  Classification results for the 22 subjects. As can be seen in the figure, five out of seven malignant subjects were correctly classified and 13 out of 15 benign subjects were correctly classified.

Figure 2.9  The malignancy map created by plotting the value cancer likelihood \( L_c \) of the ROIs overlaid on the B-mode image. The images on the top are from malignant subjects, and the ones on the bottom are benign subjects.

Figure 2.10  The malignancy maps for the subjects that the classifiers did not correctly label. On the top are two incorrectly classified malignant subjects (Subjects 3 and 7), and the ones on the bottom row are the incorrectly classified benign subjects (Subjects 10 and 16).

Figure 2.11  Scatter plot showing the ROI samples from the 22 subjects using features 2 and 3. Red and blue markers represent malignant and benign ROI samples from 22 subjects.

Figure 2.12  SVM decision function plot for the malignant subjects. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the figure, malignant subjects, 3 and 7, lie in the benign region and are therefore misclassified by the classification algorithms. See Figure 2.11 for distribution of the ROI samples.

Figure 2.13  Random Forests decision function plot for the malignant subjects. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the figure, malignant subjects, 3 and 7, lie in the benign region and are therefore misclassified by the classification algorithms. See Figure 2.11 for distribution of the ROI samples.
Figure 2.14 SVM decision function plot for the benign subjects. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the figure, benign subjects, 10 and 16, lie in the malignant region and are therefore misclassified by the classification algorithm. Subject 10 is an outlier in the dataset and significantly effects the classification results. See Figure 2.11 for distribution of the ROI samples.

Figure 2.15 Random Forests decision function plot for the benign subjects. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the figure, benign subjects, 10 and 16, lie in the malignant region and are therefore misclassified by the classification algorithm. Subject 10 is an outlier in the dataset and significantly effects the classification results. See Figure 2.11 for distribution of the ROI samples.

Figure 3.1 Philips UroNav platform (left) and a Philips iU22 ultrasound machine (right).

Figure 3.2 B-mode image showing the ROI selection. ROIs were selected along the biopsy needle trajectory as shown in the figure.

Figure 3.3 An overview of the classification framework employed in this work. Clustering of the data is performed prior to classification to eliminate the affect of outliers and improve the classification accuracy.

Figure 3.4 Boxplots of features.

Figure 3.5 Correlation plots of the RF time series features. As can be seen in the figure, all the time series features are positively correlated, except for features 6 and 9. Due to high correlation among the features, we perform feature selection before classification.

Figure 3.6 Histograms of the features. The best selected features, features 3 and 4, are highlighted in red.
Figure 3.7  Clustering performed on all 360 training samples.

Figure 3.8  ROC curves for SVM and Random Forests (each performed after clustering).

Figure 3.9  Classification results for the 18 biopsy cores. As can be seen in the figure, all biopsy cores were correctly classified.

Figure 3.10  Cancer likelihood colormaps of the 18 biopsy cores from 14 subjects with leave-one-subject-out cross validation using the best two RF time series features. Clustering and SVM classification is used.

Figure 3.11  Scatter plot showing the ROI samples from the 18 biopsy cores using features 3 and 4. Red and blue markers represent malignant and benign ROI samples from 18 cores.

Figure 3.12  SVM decision function plots for the malignant cores. The decision plots shown here are without clustering. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. It is apparent in the above figure, some ROI samples from cores, 3, 8, 9, and 11 are in the benign region and without clustering would have been misclassified. These decision functions (plotted without clustering) show the utility of clustering prior to classification. See Figure 3.11 for distribution of the ROI samples.

Figure 3.13  Random Forests decision function plots for the malignant cores. The decision plots shown here are without clustering. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. It is apparent in the above figure, some ROI samples from cores, 3, 8, 9, and 11 are in the benign region and without clustering would have been misclassified. These decision functions (plotted without clustering) show the utility of clustering prior to classification. See Figure 3.11 for distribution of the ROI samples.
Figure 3.14  SVM decision function plots for the benign cores. The decision plots shown here are without clustering. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the above figure, most of the benign ROI samples lie in the benign region (cyan). Some ROI samples from core 6 are in the malignant region and hence the classification accuracy for core 6 is lower than the other benign cores. See Figure 3.11 for distribution of the ROI samples.

Figure 3.15  Random Forests decision function plots for the benign cores. The decision plots shown here are without clustering. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the above figure, most of the benign ROI samples lie in the benign region (cyan). Some ROI samples from core 6 are in the malignant region and hence the classification accuracy for core 6 is lower than the other benign cores. See Figure 3.11 for distribution of the ROI samples.

Figure 4.1  Experiment setup. The two animal tissues are Chicken breast on the left and bovine on the right surrounded by an ultrasound absorbing pad.

Figure 4.2  B-mode image of the two animal tissues. Bovine on the left and chicken (right).

Figure 4.3  The effect of imaging depth on the classification accuracy of the two tissue types, bovine and chicken. As can be seen in the figure, increasing the imaging depth decreased the classification accuracy. In other words, using the RF time series method, targets that are closer to the ultrasound transducer can be classified more accurately than the targets that are farther away from the transducer. The results are consistent with the previous studies.
Figure 4.4 The effect of increasing frame rate on the classification accuracy of the two tissue types, bovine and chicken. Increasing the frame rate significantly improved the classification results. It is apparent from the above figure, increasing the frame rate from 40 to 71 frames per second improved the area under the ROC curve from 77% to 87%. Previous studies have shown that increasing the frame rate improved the classification performance, therefore our results are consistent with the previous studies.

Figure 4.5 The effect of the length of the time series on the classification accuracy of the two tissue types, bovine and chicken. As can be seen in the above figure, increasing the number of analysed RF frames, for the RF time series feature calculation, improved the classification performance. These results show, the tissue typing information improved as more RF frames are analysed. These observations are consistent with the previous published literature.
Glossary

3D  three-dimensional
2D  two-dimensional
RF  radio-frequency
SVM support vector machines
MRI  magnetic resonance imaging
CT  computer tomography
RFE  recursive feature elimination
DCE  dynamic contrast enhanced
DWI  diffusion weighted imaging
TRUS  transrectal ultrasound
CI  confidence interval
EM  electromagnetic
ROI  region of interest
ROC  receiver operating characteristic
TGC  time gain control
GLCM  gray-level co-occurrence matrix
RBF  radial basis function
DFT  discrete Fourier transform
MCF  mean center frequency
Acknowledgments

I would first like to thank my supervisor, Dr. Mehdi Moradi for the support, guidance, and the encouragement throughout my post graduate studies. Without Mehdi’s motivation this thesis would not have been possible. Dr. Purang Abolmaesumi for his support and finding time to proofread the manuscripts and providing invaluable comments. Dr. Robert Rohling and Dr. Septimiu Salcudean for their help with the manuscripts and sharing their expertise.

A special thanks to Hani Eskandari for providing the data for the breast study, helping with the manuscript, and for the guidance throughout my learning phase. Farhad Imani and Dr. Parvin Mousavi for their help on the prostate study. Bo Zhuang for helping with the data collection for the animal tissue experiments. Amir Tahmasebi and Shyam Bharat for their help while I was at Philips.

Thank you to all my friends and colleagues in the Robotics and Controls Lab at UBC. Thank you for a friendly lab environment and for all the social gatherings. I will miss our Friday lunches and the lab events.

Last but not the least, I would like to thank my mother for her support and encouragement throughout my career. Without her I wouldn’t have been able to achieve this milestone. Thank you to my sister who has always encouraged me to pursue my dreams.

Finally, I would like to thank the funding sources for my project and research: The Natural Sciences and Engineering Research Council of Canada, Canadian Institutes of Health Research, and the Canadian Foundation for Innovation.
Dedication

To my mother
Chapter 1

Introduction

*If I have seen farther it is by standing on the shoulders of Giants.*

— Sir Isaac Newton (1855)

1.1 Motivation

After lung cancer, breast and prostate cancer are the most common causes of cancer deaths in North America [1]. Among females, breast cancer is the leading cause of death and the most frequently diagnosed cancer in the world [2]. Early detection and advanced diagnosis play a vital role in reducing the number of fatalities due to breast and prostate cancer and medical imaging is a promising assessment tool for this purpose.

Imaging modalities including, X-ray or computer tomography (CT), magnetic resonance imaging (MRI), and ultrasound have been used for breast and prostate cancer screening and diagnosis. Ultrasound is the most versatile and widely used imaging modality. Unlike X-ray, ultrasound does not expose the human body to ionizing radiation. Ultrasound is less expensive, more portable, and more widely available than MRI.

Mammography, X-ray of the breast, is the primary screening modality for breast cancer. Subjects with a suspicious lesion in mammography undergo an ultrasound examination before an ultrasound-guided biopsy is performed. However, almost 80% of the biopsies carried out turn out to be negative for cancer [3] [4].
Recently, a 25-year study of 89,835 women in Canada, aged 40-59, found that mammography screening does not reduce mortality from breast cancer and results in over-diagnosis of breast cancer [5]. Mammography alone is limited in its ability to detect cancer in dense breasts [6, 7]. Ultrasound, along with mammography, has been proven to improve the sensitivity in palpable breast masses [8–10] but due to its low sensitivity towards non-palpable and non-cystic breast lesions, it is not used as a primary screening modality. Ultrasound has also been found to be useful in classification of benign and malignant breast lesions [4, 11]. However, low contrast, high speckle brightness mode (B-mode) images, and operator variability make ultrasound based diagnosis more difficult [11]. To address the limitation of ultrasound B-mode breast tissue classification several research groups have proposed alternative methods for breast tissue classification [12–17]. A thorough literature review of ultrasound-based breast lesion classification methods can be found in Section 2.1.

Similarly, MRI has been used for prostate cancer diagnosis to improve high grade prostate cancer yield [18–21]. However, MRI-based prostate cancer diagnosis is difficult, expensive, and time-consuming [22]. MRI in this context refers to multi-parametric MRI (mpMRI) which combines dynamic contrast enhanced (DCE) and diffusion weighted imaging (DWI) [23]. Fusion of ultrasound and MRI has been used to improve prostate cancer detection by enabling targeting of the cancer foci pre-determined in mpMRI during transrectal ultrasound (TRUS)-guided prostate biopsy [18]. Biopsy core locations determined by a radiologist through examination of mpMRI data and images are translated to patient coordinates using pre-procedure three-dimensional (3D) TRUS and its registration to MRI [24]. 3D TRUS to MRI registration requires either sophisticated mechanical systems [25] to guide the biopsy needles or, if performed by software only [26], does not fully account for patient motion or organ deformation occurring during biopsy. Alternatively, other ultrasound-based prostate cancer diagnosis techniques such as radio-frequency (RF) data analysis [27], elastography [28, 29], and Doppler imaging [30] have been reported [31]. However, these technologies, individually, have not entirely succeeded in accurate identification of high grade cancer.

In recent works, ultrasound RF time series, comprising a sequence of ultrasound RF frames captured in time from a stationary tissue location, has been pro-
posed as a new tissue classification method [32]. In this method of analysis, the tissue typing parameters are extracted from the temporal changes of the signal, as opposed to the classical method of spectral analysis on spatial segments of the RF signals [33]. Ultrasound RF time series method for tissue classification is completely non invasive and does not require specialized equipment. As such, it can be a valuable and practical addition to other ultrasound-based methods for tissue classification such as elastography or single frame RF analysis. The next paragraph provides a history and literature review of ultrasound RF time series.

Ultrasound RF time series was first proposed back in 2006 when it was used for detection of prostate cancer [34]. Moradi et al. published a study on fractal analysis of ultrasound RF time series signals for prostate cancer diagnosis. Their work presented the use of a fractal dimension feature and B-mode texture features for prostate tissue classification [34]. After that, Moradi et al. [35] described a study in which they performed discrete Fourier transform (DFT) of ultrasound RF time series signals and calculated six features. A neural network classifier was then employed to classify region of interest (ROI) samples described by the six DFT features. Their work reported 91% mean accuracy and a mean sensitivity and specificity of 92% and 90%. In the same year, Moradi et al. [36] reported the use of high frequency RF time series data for classification of animal tissue. They calculated the fractal dimension features from the high frequency RF time series signal and used a Bayesian classifier to estimate posterior class probabilities for animal tissue classification. Their work reported tissue classification accuracy of upto 98%. In another study, Moradi et al. [32] used ultrasound RF time series for classifying prostate cancer in 35 ex vivo specimens. Using a leave-one-patient-out cross-validation scheme, a support vector machines (SVM) classifier, and RF time series features they reported an area under the receiver operating characteristic (ROC) curve of 0.82. In 2010, Moradi et al. [37] used ultrasound RF time series features to differentiate animal tissue samples. In their report, they also studied the effects of ultrasound imaging parameters, such as transmit power and frame rate, on tissue classification. Their work showed that increased energy delivered through power and frame rate improves the tissue classification results. Their results also reported high classification accuracy of 95% at 6.6 MHz and 98% at 55 MHz. In 2011, Imani et al. [38] extracted novel features, mean center frequency (MCF) val-
ues and slope and intercept fitted to MCF values, from ultrasound RF time series signal to differentiate in vitro animal tissue samples. Using the MCF features and an SVM classifier their study reported 99-100% accuracy in differentiating bovine muscle, bovine liver, and chicken breast. In 2013, various in vivo studies reporting the use of RF time series were published. Moradi et al. [39] published the first in vivo paper employing RF time series features and SVM to classify prostate tissue samples from six subjects. In their work, they reported an average area under the ROC curve of 0.76. Imani et al. [40] in their report on in vivo prostate tissue classification proposed novel wavelet features extracted from RF time series signal. Their work, reported an average area under the ROC curve of 0.83 on data from seven subjects. Imani et al. also showed that their proposed features significantly outperformed the previous proposed six DFT RF time series features. RF time series method was also successfully used by Imani et al. [41] for differentiation of ablated and non-ablated tissue. In this work, ultrasound RF time series has been used for the first time to classify breast lesions. A new semi-supervised machine learning framework is developed to classify MRI-targeted prostate biopsy cores. An animal tissue study is completed to provide experimental evidence on the physical basis of tissue typing using RF time series and prove that is can be used as a tissue typing method.

In the previous paragraph, we discussed various studies that showed that ultrasound RF time series method can be used for tissue classification. While ultrasound RF time series method has yielded consistent results, the source of tissue classification using this method is still mostly unknown. This suggests a need for additional experimental studies to address the physical basis of the tissue typing information extracted from ultrasound RF time series. Several hypotheses describing the physical phenomenon exists and have been explored in different studies. Daoud et al. have shown experimentally and in simulations, that the temperature increase in controlled irradiation of tissue with RF time series can partly explain the phenomenon. Also, Moradi et al. [37] and Imani et al. [41] reported that increasing the energy delivered through transmit power and frame rate improves the tissue classification accuracy in phantom and animal studies. Given these observations, micro-vibrations of the tissue microstructure caused by acoustic radiation force is another likely phenomenon. The acoustic radiation force is related to both the
acoustic energy and to the attenuation and scattering properties of the tissue, which are different for different pathological tissue types. To substantiate the tissue classification power of ultrasound RF time series and build upon previous hypotheses, an animal tissue experimental study is presented in this thesis.

1.2 Thesis Objectives

The overall objective of this thesis was to develop a tissue classification framework and test it on breast and prostate dataset.

1. To develop a breast lesion classification framework based on features extracted from ultrasound RF spectral data, B-mode image texture data, and ultrasound RF time series data: a machine learning package was developed to classify lesions described by the extracted features. The outcome of this study could potentially improve the current breast cancer diagnosis methods and provide radiologists with a readily available diagnosis tool.

2. To augment the current prostate cancer diagnosis by developing an innovative machine learning framework to classify MRI-targeted prostate biopsy samples: in this study, previously proposed ultrasound RF time series features were used in a new machine learning package to classify prostate biopsy samples with high accuracy. This work could potentially lead to the fusion of the proposed tissue classification framework with the Philips UroNav platform to complement MRI-targeted TRUS guided prostate biopsy and provide real-time diagnosis of prostate cancer.

3. To substantiate the RF time series hypotheses by studying the effects of ultrasound imaging parameters such as imaging depth, frame rate, and time series length, on tissue classification through experiments on animal tissue: this part of the work could provide more evidence on the physical basis of tissue classification using ultrasound RF time series method.

1.3 Organization of Thesis

The outline of the thesis is as follows:

Chapter 2 describes a study on breast lesion classification using ultrasound RF time series features and B-mode texture features along with a machine learning framework. This is the first report on using the RF time series analysis in the
context of breast lesions classification.

Chapter 3 is based on prediction of prostate cancer in MRI-targeted biopsy. Ultrasound RF time series features are used within an innovative semi-supervised classification framework. In this chapter, a new semi-supervised machine learning framework is presented that improves prostate tissue classification.

Chapter 4 presents the results of experiments conducted on animal tissue to study the effects of ultrasound imaging parameters, such as imaging depth, frame rate, and time series length, on tissue classification. This chapter addresses the need for experimental evidence on the RF time series hypothesis.

Lastly, Chapter 5 describes the conclusions, thesis contributions, and future work in this field.
Chapter 2

Ultrasound RF Time Series for Classification of Breast Lesions

2.1 Introduction

According to the statistics published by the American Cancer Society in 2013, it was estimated that breast cancer is one of the most common types of cancer in women accounting for 29% (232,340) of all cancer cases [42]. The World Health Organization estimated that almost 1.38 million women worldwide are diagnosed with breast cancer annually. That accounts for 23% of all cancer cases [43]. Early detection and better diagnosis methods play a significant role in reducing the number of fatalities due to breast cancer.

Based on mammography images, the American College of Radiology has developed a method called BI-RADS (Breast Imaging-Reporting and Data System) for deciding whether biopsy of an identified suspicious lesion is indicated. Biopsy is the gold standard for breast cancer diagnosis; however it is an expensive, discomforting, and invasive procedure. Almost 80% of the biopsies carried out based on BI-RADS score turn out to be benign [3]. Hence, there exists a need for reducing the number of unnecessary breast biopsies and augment the current diagnosis methods.

Mammography alone is limited in its ability to detect cancer. Studies report 73% sensitivity for all breast types and only 48% for dense breasts [8]. Almost
50% of women under the age of 50 have dense breasts [44] and could be potentially deemed undiagnosed after mammography screening [7, 45]. Ultrasound, along with mammography, has been proven to improve the sensitivity in palpable breast masses [8–10], but due to its low sensitivity towards non-palpable and non-cystic breast lesions, it is not used as a primary screening modality. Ultrasound has also been found to be useful in classification of benign and malignant breast lesions [4, 11]. However, its drawback lies in low contrast, high speckle B-mode images, and operator variability [11]. To address these limitations, several research groups are working towards ultrasound-based computer-aided diagnosis (CAD) for breast cancer diagnosis.

For example, Zheng et al. [46] used sonographic texture features along with self-organizing maps for classification of breast abnormalities. Chen et al. [47] also used neural networks and texture information extracted from ultrasound images to classify breast nodules. Giger et al. [48] calculated features related to lesion margin, shape, homogeneity (texture), and posterior acoustic attenuation patterns in ultrasound images and used linear discriminant analysis to classify breast lesions. Lefebvre et al. [49] used texture and morphometric parameters with linear discriminant analysis and leave-one-out cross-validation to classify breast lesions. Donohue et al. [50] described a breast lesion classification method using texture and generalized spectrum features along with linear and quadratic discriminant analysis. Shankar et al. investigated the Nakagami distribution [51] and non-Rayleigh statistics [52] of the backscattered envelope for classification of breast lesions. Chang et al. [53, 54] and Huang et al. [55] used the same texture analysis approach but employed support vector machines to classify breast lesions. Drukker et al. [56, 57] performed analysis of posterior acoustic shadowing for breast lesion detection. Joo et al. [58] employed an artificial neural network to detect breast lesions based on five morphological features representing the shape, edge characteristics, and darkness of a lesion. Chen et al. [59] described an approach based on fractal parameter computation from ultrasound images and the k-means classification algorithm. Alam et al. [13] reported the performance of logistic regression in classification of breast lesions based on quantitative acoustic parameters calculated using the spectrum analysis of ultrasound RF echo signals and morphometric features related to the lesion shape. In a more recent work, Gomez et al. [60] inves-
tigated the effectiveness of co-occurrence texture statistics calculated at different quantization levels, as a method for classification of breast images. Yang et al. [61] reported another method for ultrasound image diagnosis using gray-scale invariant features extracted via multi-resolution ranklet transform along with support vector machines. Tan et al. [17] developed a CAD system to detect cancer in 3-D breast ultrasound. A detailed review of the literature in the area of ultrasound-based CAD for breast cancer is reported in [12].

Ultrasound strain imaging [14, 62], Acoustic Radiation Force Impulse (ARFI) imaging [16], and supersonic shear-wave imaging [15, 63–65] have also been used for breast lesion characterization. These methodologies require the use of excitation mechanisms or specialized equipment.

As introduced in Chapter 1, ultrasound RF time series analysis has been proposed as a novel tissue classification method by our research group. The RF time series method is based on a temporal analysis of beamformed RF signals. This method does not require additional equipment and the region of interest selection is as simple as drawing a rectangular box inside the lesion. The task of size selection and placement of the rectangular box is much less subjective compared to manual contouring of the lesions. As such, it can be a valuable and practical addition to the previously reported ultrasound-based methods for breast lesion classification.

This is the first report of using the RF time series analysis in the context of lesion classification in breast images. Preliminary results have been previously presented in [66]. For this study, we use quantitative analysis of spectral and fractal parameters extracted from RF time series. We report the use of both SVM and Random Forests classification methods with RF time series features and demonstrate accurate malignancy maps that can be used for decision support in biopsy recommendation.

2.2 Methods

We describe the use of a machine learning framework to assign labels to each sample and also estimate the cancer likelihood (for SVM) and probability (in case of Random Forests). The estimated cancer likelihood is then used to generate a malignancy map. Figure 2.1 shows a graphical representation of this approach.
Figure 2.1: Overall classification framework. This figure shows a graphical representation of the proposed approach.

2.2.1 Data Collection

An RF time series is formed by the sequence of RF echoes received from one location in the tissue over time. To acquire the RF time series data, the ultrasound probe and the tissue remain fixed and frames of RF signals are acquired. In this method of analysis, the tissue typing parameters are extracted from the temporal changes of the signal, as opposed to the classical method of spectral analysis on spatial segments of the RF signals [33].

The data collection for this work was performed during a study approved by the Clinical Research Ethics Board at the University of British Columbia. This study was designed to evaluate ultrasound-based tissue typing and vibro-elastography for breast lesion classification. The vibro-elastography technology was developed by Salcudean et al., and the preliminary results were reported in [67]. Subjects referred to ultrasound-guided biopsy, based on mammography screening, were con-
sented for collection of vibro-elastography and RF data during biopsy. The study was conducted between September 2012 and January 2013. Data was obtained on a SonixTouch ultrasound machine (Ultrasonix Medical Corp., Richmond, BC, Canada). The research platform provided by the manufacturer enabled acquisition of beamformed RF signals in real time.

For every subject, the sonographer first performed a preliminary ultrasound scan to find the suspected lesion. Once the lesion was located, the sonographer would hold her hand steady for 5 seconds while a computer program stored the RF data into the memory and consequently saved it in a file. Imaging was performed with an L14-5/38 ultrasound transducer at a center frequency of 10 MHz and a depth of 4 cm. Each RF line was sampled at 40 MHz and a total of 128 scanlines were acquired for each RF frame. With these image settings, RF data at a frame-rate of 98 frames per second, for five seconds, was obtained. The data collection of each subject was followed by a routine ultrasound exam and a core needle biopsy of the lesion under ultrasound guidance by the physician. In total, 35 subjects were imaged. However, in five subjects the radiologist decided that biopsy was unnecessary. Also, in eight subjects where biopsies were performed, the exact location of the biopsy sample could not be identified within the ultrasound image. As a result, we used the data corresponding to 22 subjects for whom the biopsy result was available and could be mapped to a specific lesion. Biopsy results for these subjects showed seven malignant lesions, all of the invasive ductal carcinoma type and 15 benign subjects, mostly of fibroadenoma type.

To use the RF time series method, patient motion has to be minimized or compensated. In the current study, we used only the first 128 frames of the data, which accounts for 1.3 seconds of acquisition, to form the RF time series. Our motion estimation using the method described in [68] showed that when the time series is limited to 128 frames, the displacement caused by patient motion is on average smaller than the resolution of the RF samples. This means that we could, on average, assume that series of RF samples obtained at a certain image coordinate were, in fact, from the same physical location in the tissue.
2.2.2 Features

The biopsied lesions were divided into 1 mm$^2$ regions of interest (ROIs). The tissue typing features were extracted from these ROIs. In the RF domain, this ROI size was equivalent to 3×52 samples each forming a time series. In the interpolated B-mode image the ROI size of 1 mm$^2$ was equivalent to 26×26 pixels. The RF frame size was 128×2080 and the B-mode image resolution was 988×1040. Note that from the 22 lesions in the dataset, a total of 863 ROIs were extracted. Among these, 241 were malignant. From each subject, we extracted multiple ROI and the number of ROIs depends on the size of the lesion.

![Image](image.png)

**Figure 2.2:** Figure showing the ROI selection. Each small rectangle in the larger rectangle, inside the lesion, is an ROI and features were calculated for each of the ROI. This figure does not correctly represent the number of ROIs defined, it is for visual interpretation only.

In addition to the RF time series features, we also calculated and used the values of texture features extracted from the first B-mode image and spectral RF
features extracted from the first RF frame. A similar approach to B-mode texture profiling for tissue typing is reported in prior work \cite{47, 50, 60, 61, 69-71}. A complete list of the studied parameters can be found in Table\ 2.1.

**Table 2.1:** Ultrasound RF time series, B-mode texture, and attenuation features calculated for classification of malignant and benign breast lesions

<table>
<thead>
<tr>
<th>#</th>
<th>Feature</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RF time series first quadrant</td>
<td>RF time series</td>
</tr>
<tr>
<td>2</td>
<td>RF time series second quadrant</td>
<td>RF time series</td>
</tr>
<tr>
<td>3</td>
<td>RF time series third quadrant</td>
<td>RF time series</td>
</tr>
<tr>
<td>4</td>
<td>RF time series fourth quadrant</td>
<td>RF time series</td>
</tr>
<tr>
<td>5</td>
<td>Intercept of regression line fitted to norm. spect.</td>
<td>RF time series</td>
</tr>
<tr>
<td>6</td>
<td>Slope of regression line fitted to norm. spect.</td>
<td>RF time series</td>
</tr>
<tr>
<td>7</td>
<td>Higuchi fractal dimension</td>
<td>RF time series</td>
</tr>
<tr>
<td>8</td>
<td>Intercept of line fitted to calibrated spectrum</td>
<td>Single-frame RF</td>
</tr>
<tr>
<td>9</td>
<td>Slope of line fitted to calibrated spectrum</td>
<td>Single-frame RF</td>
</tr>
<tr>
<td>10</td>
<td>Texture: Mean</td>
<td>B-mode</td>
</tr>
<tr>
<td>11</td>
<td>Texture: Standard Deviation</td>
<td>B-mode</td>
</tr>
<tr>
<td>12</td>
<td>Texture: Skewness</td>
<td>B-mode</td>
</tr>
<tr>
<td>13</td>
<td>Texture: Kurtosis</td>
<td>B-mode</td>
</tr>
<tr>
<td>14-17</td>
<td>Correlation: $F_d(\theta) = \sum_{i,j} (i-\mu_i)(j-\mu_j)p(i,j)/\sigma_i\sigma_j$</td>
<td>B-mode</td>
</tr>
<tr>
<td>18-21</td>
<td>Energy: $F_d(\theta) = \sum_{i,j} p(i,j)^2$</td>
<td>B-mode</td>
</tr>
<tr>
<td>22-25</td>
<td>Contrast: $F_d(\theta) = \sum_{i,j}</td>
<td>i-j</td>
</tr>
<tr>
<td>26-29</td>
<td>Homogeneity: $F_d(\theta) = \sum_{i,j} p(i,j)/(1+</td>
<td>i-j</td>
</tr>
</tbody>
</table>

*† Features 14-29 are calculated for each $d \in (1, 2, \ldots, 10)$ resulting in 160 ($16 \times 10$) feature values. $
\theta$ is the orientation angle for GLCM calculation ($\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$). $d$ is the pixel distance for GLCM calculation. $p(i,j)$ is the $(i,j)$th entry in the co-occurrence probability matrix. $\mu_i = \sum_{i,j} i \cdot p(i,j)$ and $\mu_j = \sum_{i,j} j \cdot p(i,j)$ $\sigma_i^2 = \sum_{i,j} (i-\mu_i)^2 \cdot p(i,j)$ and $\sigma_j^2 = \sum_{i,j} (j-\mu_j)^2 \cdot p(i,j)$

**RF time series features:** A variety of methods have been used for tissue classifi-
cation using RF time series. We use the method originally described in [35] which proposes summarizing the power spectrum of the RF time series in six parameters. The frequency spectrum was estimated by calculating the FFT-based periodogram of the Hamming windowed time series. This estimated spectrum was divided into four frequency bands and each averaged to deliver a feature. In other words, the first four features (Features 1-4) were the average of the frequency spectrum in \([0, \pi/4), (\pi/4, \pi/2), (\pi/2, 3\pi/4), (3\pi/4, \pi)\) frequency bands in the discrete frequency domain. Note that the sampling rate here is equivalent to the frame rate of the ultrasound machine. Two other spectral features were the intercept (Feature 5) and the slope (Feature 6) of a regression line fitted to the magnitude of the spectrum versus normalized frequency.

Feature 7 was the average fractal dimension (FD) of RF time series in a region of interest. In this context, fractal dimension is a measure of the non-linear complexity of the signal. For calculation of the FD, we used the algorithm proposed by Higuchi [72] which decomposes the signal into different scales and evaluates the signal complexity. We used the Higuchi algorithm with 16 levels of decomposition for the time series of length 128. The use of this feature in ultrasound-based tissue typing was originally proposed in [34].

**Single-frame RF spectral features:** The tissue attenuation coefficient and RF spectral parameters that are related to attenuation have been used for tissue typing. Similar to texture features, these features were also calculated using the first RF frame. One common method for evaluating the frequency-dependent attenuation of ultrasound in tissue is the spectral analysis of segments of echoed RF signals, after calibration to remove the effects of the imaging system. Researchers at the Lizzi Center for Biomedical Engineering at Riverside Research have shown that the parameters of a linear model fitted to the calibrated power spectrum of an RF segment contain valuable tissue classification information [73] [74]. To estimate these parameters, the averaged tissue power spectrum within an ROI had to be calibrated to account for the ultrasound machine transfer function. The calibration spectrum was acquired from the surface of a flat glass plate in a water bath at the transducer focal zone, with minimum amplifier gain and flat time gain control (TGC). Following [73], we subtracted the logarithm of the averaged spectrum of the glass from the logarithm of the spectrum of the tissue. For our data, we found that the useful
part of the spectrum was at the range of 1-10 MHz and this is apparent in Figure 2.3. Outside of this range, the power drops to less than half of the peak value. Therefore, we fitted the regression line to the data in this range. As can be seen in, Figure 2.4 the calibrated spectrum shows a quasi-linear behaviour. Features 8 and 9 are the intercept and slope of the linear model fitted to the normalized frequency spectrum.

![Power spectrum graph](image)

**Figure 2.3:** The average power spectrum of the reflected RF signal for breast tissue.

*B-mode Texture features:* Features 10–173 are 164 texture parameters calculated using the B-mode image reconstructed from the first RF frame. The RF frame was converted into a B-mode image by taking the Hilbert transform and then resized to 988×1040 to match the physical dimension (3.8×4 cm) of the image. The first four parameters were simply the mean, standard deviation, skewness and kurtosis of the pixel intensities in an ROI and the other 160 parameters were the correlation, energy, contrast, and homogeneity calculated from the gray-level
Figure 2.4: Attenuation estimation by calculating the slope of the calibrated spectrum.

c-co-occurrence matrix (GLCM). The GLCM was created by calculating how often a pixel with a grayscale intensity \( i \) occurs adjacent to a pixel with the value \( j \). For the purpose of this study, we calculated the co-occurrence matrices at a distance of ten pixel \( (d = 1, 2, \ldots, 10) \), for directions of \( \theta = 0^\circ, \theta = 45^\circ, \theta = 90^\circ, \) and \( \theta = 135^\circ \). For every one of the 40 combinations of \( d \) and \( \theta \) values a GLCM was calculated. Four texture statistics are calculated from each GLCM and the definition of these features is provided in Table 2.1.

We also calculated the single-frame RF spectral features and B-mode texture features for ROI sizes of 4 mm\(^2\) and 9 mm\(^2\), for comparison with ROIs of 1 mm\(^2\).
2.2.3 Classification and Estimation of Cancer Likelihood

Support vector machine (SVM): SVM is a widely used maximum margin classifier. The soft margin SVM classifier is a hyperplane of the form $w \phi(x_i) + b$, where $\phi$ is the function that maps the feature vector, $x_i$, to a higher dimensional space and $w$ and $b$ are determined to minimize [75]:

$$\frac{1}{2} \times w^T w + C \sum_{i=1}^{N} \xi_i$$

subject to

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i$$

where $C > 0$ is the regularization or penalty parameter that minimizes the error by controlling the trade-off between the slack variable penalty and the margin, and $\xi_i \geq 0$ are the slack variables that provide flexibility when fitting the data by permitting incorrect classification of noisy and difficult data points [75]. The function $\phi(.)$ maps the data to a higher dimensional space. This new space is defined by its kernel function. It can be shown that in the SVM formulation, one does not need an explicit expression for the kernel function and the SVM optimization and decision hyperplane are defined fully given the dot product format of the kernel $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$.

The classification problem can also be written in the form of its dual representation in terms of the kernel function. In this representation, the class label is determined based on the sign of $y(x)$ in the following equation:

$$y(x) = \sum_{i=1}^{N} a_i t_i K(x, x_i) + b$$

where $y(x)$ is the model prediction of input $x$, $t_i$ is the true label for the support vector $x_i$, $a_i$ is the Lagrange multiplier used to convert the maximum margin SVM optimization problem from 2.1 to its dual representation. The solution to the SVM training is a quadratic programming problem that solves for $a_i$’s.

The soft margin SVM algorithm implemented in the scikit-learn Python package and reported in [76] was used for tissue classification. The radial basis function (RBF) kernel was employed. The values of the RBF exponent ($\gamma$), which de-
terminates the range of support, and the parameter $C$, which controls the trade-off
between the slack penalty and the margin in the soft margin SVM, were chosen
based on a grid search through leave-one-subject-out cross-validation.

Cancer likelihood is estimated using Platt’s algorithm [77] as follows. Assume
that the SVM hyperplane obtained after training is $w\phi(x) + b$ where $x$ is the feature
vector and $\phi$ is the function that maps the feature vector to a higher dimensional
space. Cancer likelihood ($L_c$) is computed by mapping the distance of each test
sample to the decision boundary using a sigmoid function of form:

$$L_c = (c | (w\phi(x_i) + b)) = \frac{1}{1 + \exp(\alpha(w\phi(x_i) + b) + \beta)} \quad (2.4)$$

Where $c$ stands for cancer (class). Maximum likelihood estimation from the obser-
vations for which the true labels are known (training data) was used to calculate the
values of the parameters $\alpha$ and $\beta$. See [75] for details and [78] for implementation.

**Random Forests:** Random Forests is a robust classification algorithm employed
for medical image processing applications [79, 80]. Random Forests algorithm is a
collection of decision trees that exploits the bias variance trade-off. The underlying
concept behind the algorithm is that by taking the prediction of each decision tree
and averaging it across the ensemble/forest the high variance in the prediction of
decision trees can be reduced. The Random Forests algorithm can be described as:

1. For $1$ to $T$ trees:
   (a) Draw a bootstrap (bagging) sample from the training data
   (b) Recursively repeat the steps below at each terminal leaf, until the de-
sired depth is reached, to construct a random tree from the bootstrapped
data.
      i. Randomly select a subset of $n$ features from $f$ features
      ii. Select the split point that results in maximum information gain
          along each $n$ feature
      iii. Split the leaf into two child leaves
2. Output the forest of $T$ trees
The objective function at each leaf of the tree is \( I = I(S_j, \theta) \), where \( S_j \) is the \( j^{th} \) node, \( \theta \) is the leaf parameter or the split point, and \( I \) is the information gain. The information gain is calculated as

\[
I(S, \theta) = H(S) - \sum_{i \in \{L, R\}} \frac{|S_i|}{|S|} H(S_i)
\]  

(2.5)

where, \( H(S) \) is Shannon’s entropy calculated as

\[
H(S) = -\sum_{c \in C} p(c) \log(p(c))
\]  

(2.6)

In the above equation \( C \) is the output/classes. From the \( n \) randomly selected features, the feature that results in the highest information gain is used at the root node and the best split point is also calculated with the objective of maximizing the information gain. During testing, the sample is pushed down each tree simultaneously until it reaches the corresponding leaf.

Random Forests implementation within the scikit-learn Python package was used [76]. Forest class a posteriori probabilities or malignancy probabilities \( p(c|v) \), is calculated as the average of all tree a posteriori probabilities \( p_t(c|v) \) and can be described as:

\[
p(c|v) = \frac{1}{T} \sum_{t=1}^{T} p_t(c|v)
\]  

(2.7)

where \( v \) is the input data, \( c \) stands for cancer (class), and \( T \) is the number of decision trees in the forest. The a posteriori class probability \( p_t(c|v) \) is calculated at each leaf of the decision tree as the fraction of the number of samples with the majority label over the total number of samples. The label assigned to the test data is based on the forest class a posteriori probability [81]. The algorithm parameters were optimized by performing a grid search and the decision trees were constructed shallow to avoid over-fitting. The classification algorithm implemented in scikit-learn package closely follows the method described in [79] and [80].

**Feature and Model Selection:** To use the most informative features and reduce the dimensionality of the data, only a subset of the features were chosen for classification. It is important to note that the Random Forests algorithm provides a built-in method for ranking features in terms of their ability to separate samples
from the two classes. This is not the case for SVM. Therefore, our approach to feature selection was different with the two classifier models.

In case of the SVM classifier, we used recursive feature elimination (RFE). The RFE algorithm, originally reported in [82], recursively eliminates the features with the smallest contribution to an estimated linear maximum margin model and is shown to outperform correlation-based methods in gene selection [82]. In this method, features are eliminated recursively based on their corresponding weight in a linear SVM classifier of the form $y = w^T x + b$. Initially, the model is trained on all the features and weights are calculated for each one of them. Then the feature with the smallest absolute weight value is eliminated. This process is then repeated recursively until the desired number of features is reached [82].

In Random Forests algorithm, feature importance is estimated as follows: each decision tree is created using a bootstrap sample of the original data. Hence, the left out data is labelled as out-of-bag (OOB) samples and can be used as “test data” to estimate feature importance. In this step, first all OOB samples are propagated through the trained model and the accuracy is recorded. Then the values of one feature for OOB samples are randomly permuted and the OOB samples with this permuted feature value are propagated down the trees. The expectation is that the permutation results in decreased accuracy. The permutation of the “important” features will cause a larger reduction in classification accuracy. The difference between average ensemble accuracy with and without permutation provides a raw measure of feature importance. The features were ranked based on this importance measure. The choice of the number of features to select for classification was considered as one of the model parameters. All model parameters, which included the depth of trees, the number of trees, and the number of features were optimized through a grid search. Leave-one-subject-out cross-validation was used for this grid search.

### 2.2.4 Evaluation of Classification

The classification label was compared with the biopsy result, which was used as the ground truth, to compute the classification accuracy. We assess the performance of the classifiers by calculating the AUC. The ROC curves were generated using the
method described in [83] and the AUC was estimated using the trapezoidal rule. We also calculated the standard errors and $p$-values for the AUCs using the Hanley and McNeil method reported in [84].

In order to evaluate the performance of other groups of features as opposed to RF time series, we trained an SVM classifier using parameter selection that included only texture and attenuation parameters. All other aspects of the method were similar to the time series experiment.

### 2.3 Results

#### 2.3.1 Feature Selection

The feature selection algorithm showed that using three features resulted in the lowest cross-validation error in both SVM and Random Forests. Features 2, 3, and 7 were the best performing for classification using SVM and features 2, 5, and 7 yielded the highest importance score for classification using Random Forests. It should be noted, that the fractal dimension feature extracted from the RF time series, feature 7, was the best overall. Note that for both classifiers the final set of features only included RF time series features.

#### 2.3.2 Classification Results

Table 2.3 shows the BI-RADS score, tissue type, classification result, and classification accuracies for all the 22 subjects (IDC stands for invasive ductal carcinoma, FA stands for Fibroadenoma, Infl. stands for Chronic Inflammation, and NS stands for Nodular Sclerosis). The tissue type is not available for subject 12 because the cytology result was unknown due to low cellularity.

Both of the classification algorithms, Random Forests and SVM, correctly classified 18 out of 22 subjects. We define success as correct labelling of at least 60% of the 1 mm$^2$ ROIs from the lesion. Cancer likelihood for each ROI, which is a number between 0 and 1, was translated to a colormap applied to the lesion area on the B-mode image. The result for several cases is illustrated in Figure 2.9 and shows the utility of the method in visualizing the likelihood of malignancy. As can be seen in Table 2.3, Subjects 3, 7, 10, and 16 were incorrectly classified. Further
Figure 2.5: Box plots of the nine RF time series features. Table 2.1 provides a description of the features. From the box plots, it is apparent that the best features selected by the RFE algorithm show a difference in the class distribution. As can be seen in the figure, "Feature07" is the best feature that differentiates the malignant samples from the benign, and it was also selected as the best feature by our feature selection method.

investigation of these subjects showed calcification inside the lesion for Subject 3 and poor lesion boundaries for Subjects 7 and 16. It should also be noted that Subject 7, one of the incorrectly classified malignant cases, was the smallest lesion in our dataset. Figure 2.10 shows the malignancy maps for these subjects.
Figure 2.6: Correlation plot of the 164 texture features extracted from the B-mode image. From the figure, it is observable that most of the features are highly correlated. High positive correlation suggests the need for feature selection prior to classification.

As shown in Figure 2.7, the AUC using SVM was 85.6% (95% confidence interval (CI): 83.9%–89.8%) and using Random Forests was 81.3% (95% CI: 78%–85%). We also calculated the classification accuracies for each test subject and these can be found in Table 2.3. The mean classification accuracy for each subject was calculated as 76% for SVM and 74% for Random Forests. As depicted in Table 2.2, using a cut-off threshold of $T_h = 0.5$, classification using SVM shows sensitivity of 82% and specificity of 80%. However, a drop in sensitivity (73%) and a rise in specificity (88%) was seen when using Random Forests. Our dataset consists of mainly benign lesions and hence the majority of the training data belongs to the benign class. The decrease in sensitivity could be due to this class imbalance problem [85, 86]. Another possible cause of the lower classification accuracy
using the Random Forests algorithm is the small number of selected features. The power of Random Forests is in random sampling of features, which is most useful in high-dimensional feature vectors [80].

Using the statistical tests proposed in [84], the standard errors were calculated as 0.016 and 0.018 for AUC using SVM and Random Forests, respectively. Using the same method, and performing the z test, the $p$-value was calculated to be 0.008, hence, the difference in performance of the two classifiers was statistically significant ($p < 0.05$) [84].
Figure 2.8: Classification results for the 22 subjects. As can be seen in the figure, five out of seven malignant subjects were correctly classified and 13 out of 15 benign subjects were correctly classified.

Figure 2.9: The malignancy map created by plotting the value cancer likelihood ($L_c$) of the ROIs overlaid on the B-mode image. The images on the top are from malignant subjects, and the ones on the bottom are benign subjects.
Table 2.2: Sensitivity, specificity at a cut-off $T_h_c=0.5$, and area under ROC curve.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Sensitivity</th>
<th>specificity</th>
<th>AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>82%</td>
<td>80%</td>
<td>85.6% (83.9%–89.8%)</td>
</tr>
<tr>
<td>Random Forests</td>
<td>73%</td>
<td>88%</td>
<td>81.3% (78.0%–85.0%)</td>
</tr>
</tbody>
</table>

Table 2.3: Classification results for the 22 subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>BI-RADS</th>
<th>Tissue Type</th>
<th>Classification Type</th>
<th>Classification Result</th>
<th>SVM Accuracy</th>
<th>Random Forests Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>IDC</td>
<td>Malignant</td>
<td>100%</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>IDC</td>
<td>Malignant</td>
<td>94%</td>
<td>93%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>IDC</td>
<td>Benign</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4B</td>
<td>IDC</td>
<td>Malignant</td>
<td>96%</td>
<td>71%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4C</td>
<td>IDC</td>
<td>Malignant</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4C</td>
<td>IDC</td>
<td>Malignant</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4B</td>
<td>IDC</td>
<td>Benign</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4A</td>
<td>FA</td>
<td>Benign</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>FA</td>
<td>Benign</td>
<td>86%</td>
<td>73%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>FA</td>
<td>Malignant</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>4B</td>
<td>FA</td>
<td>Benign</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>4A</td>
<td>Unknown</td>
<td>Benign</td>
<td>88%</td>
<td>83%</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4B</td>
<td>FA</td>
<td>Benign</td>
<td>96%</td>
<td>93%</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>4A</td>
<td>FA</td>
<td>Benign</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>4B</td>
<td>Infl.</td>
<td>Benign</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4A</td>
<td>FA</td>
<td>Malignant</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>4A</td>
<td>FA</td>
<td>Benign</td>
<td>69%</td>
<td>94%</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>4A</td>
<td>FA</td>
<td>Benign</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>4A</td>
<td>Cyst</td>
<td>Benign</td>
<td>60%</td>
<td>53%</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>4A</td>
<td>FA</td>
<td>Benign</td>
<td>75%</td>
<td>73%</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>4A</td>
<td>FA</td>
<td>Benign</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>4A</td>
<td>NS</td>
<td>Benign</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

The effectiveness of the RF time series method is evident from the fact that only RF time series features were selected by our systematic ranking method. Classification using the best three single-frame RF spectral and B-mode texture features was also performed for comparison with time series features. The three selected
features in this experiment were the intercept and slope from the attenuation group (Features 8 and 9) and mean intensity from the texture group (Feature 10). This combination resulted in an AUC of 68%. However, these parameters provide a benefit. Specifically, the classifier trained on texture and attenuation parameters was successful in classifying both of the benign subjects that were mis-classified by the time series method. Using the best three texture and attenuation features the classification accuracies for subjects 10 and 16 were calculated as 70% and 91%. This was, however, at the cost of reduced sensitivity and overall accuracy.
We suspected that the poor classification accuracy of the single-frame RF spectral features and the B-mode texture features could be due to the small (1 mm$^2$) ROI size, hence, we increased the ROI size to 4 and 9 mm$^2$ and calculated the AUCs as 69.4% and 70.2%, respectively. The difference in AUCs based on ROI sizes was not significant because the $p$-value was calculated as 0.91. We did not go above 9 mm$^2$ for the ROI size because that would have resulted in loss of data from two subjects due to small lesion size.

Figure 2.11: Scatter plot showing the ROI samples from the 22 subjects using features 2 and 3. Red and blue markers represent malignant and benign ROI samples from 22 subjects.
Figure 2.12: SVM decision function plot for the malignant subjects. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the figure, malignant subjects, 3 and 7, lie in the benign region and are therefore misclassified by the classification algorithms. See Figure 2.11 for distribution of the ROI samples.

2.4 Discussion and Conclusion

There exists a need to reduce the number of false positives based on a BI-RADS score, hence, augmentation of the current breast cancer diagnosis methods is a pressing need. In this work we presented an ultrasound-based method for classification of malignant breast lesions. This method uses RF time series parameters within a machine learning framework. A set of 21 parameters extracted from RF time series analysis, B-mode texture, and single-frame RF spectral analysis were ranked for their importance with two different feature ranking approaches. In both feature-ranking algorithms, all the three best performing features were from the RF
Figure 2.13: Random Forests decision function plot for the malignant subjects. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the figure, malignant subjects, 3 and 7, lie in the benign region and are therefore misclassified by the classification algorithms. See Figure 2.11 for distribution of the ROI samples.

time series group. Using these three features, we obtained an AUC of 85.6% and 81.3% on SVM and Random Forests classifiers, respectively. The best performing subset of features that were not from RF time series resulted in an AUC of 68%.

The performance of B-mode texture features in breast cancer classification was poor in our study. However, it should be noted that the B-mode images used in this work were reconstructed from the RF signals offline. The B-mode images from the commercial ultrasound machines are filtered and optimized in terms of dynamic range and have a higher quality compared to the B-mode images used here. Therefore, the reported performance of the B-mode texture features could be
Figure 2.14: SVM decision function plot for the benign subjects. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the figure, benign subjects, 10 and 16, lie in the malignant region and are therefore misclassified by the classification algorithm. Subject 10 is an outlier in the dataset and significantly affects the classification results. See Figure 2.11 for distribution of the ROI samples.
Figure 2.15: Random Forests decision function plot for the benign subjects.
The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the figure, benign subjects, 10 and 16, lie in the malignant region and are therefore misclassified by the classification algorithm. Subject 10 is an outlier in the dataset and significantly effects the classification results. See Figure 2.11 for distribution of the ROI samples.
potentially improved by using the B-mode images produced by the scanner.

RF time series method is susceptible to patient motion. The time series used in this work were obtained over less than 2 seconds. We found that by asking the subjects to hold their breath, the amount of motion could be minimized. An ultrasound transducer stabilization mechanism could also reduce the operator-introduced motion and increase the length of motion-free time series, which is linked to improved tissue classification in previous work [37]. Our measurement based on [68] showed that the amount of displacement in the RF data had the highest value in the misclassified malignant subjects. Also, our dataset includes mostly benign subjects which introduces a class imbalance problem. This affects the classification performance, specially the sensitivity. We need a larger dataset with improved representation of calcification and malignancy. As the size of the dataset increases, we anticipate improved performance.

The physical basis of the tissue typing capabilities of the RF time series method has been the topic of a few previous works. In [87], a model was developed to relate the variations of the ultrasound backscattering to the variations in tissue temperature and speed of sound that take place during the RF time series scanning procedures. The measurements of the variations in the ultrasound backscattering were then used in a tissue classifier. Other reports show that an increase in the frame rate and the acoustic power of the ultrasound beam result in an improved accuracy of tissue classification [37]. To prove the tissue typing capabilities of ultrasound RF time series we have conducted experiments on animal tissue to study the effects of ultrasound imaging parameters on tissue classification. The results of these experiments can be found in Chapter 4.

Accurate malignancy maps could eventually be used as a diagnosis method which would significantly reduce the negative biopsy outcome. Our proposed method can be part of the overall solution for multiparametric ultrasound analysis of breast cancer. We argue that RF time series can be a practical component of that approach without the need for additional equipment. This method is also independent of the contour shape and the potentially subjective process of segmentation.
2.5 Chapter Summary

This chapter reported the use of ultrasound RF time series analysis as a method for ultrasound-based classification of malignant breast lesions. The RF time series method is versatile and requires only a few seconds of raw ultrasound data with no need for additional instrumentation. Using the RF time series features, and a machine learning framework, cancer maps were generated, from the estimated cancer likelihood, for decision support in biopsy recommendation. These maps depict the likelihood of malignancy for regions of size 1 mm$^2$ within the suspicious lesions. We reported an area under receiver operating characteristics curve of 85.6\% using support vector machines and 81.3\% using Random Forests classification algorithms, on 22 subjects with leave-one-subject-out cross-validation. Changing the classification method had an insignificant effect on the accuracy which indicates the robustness of the tissue typing method. The findings of this chapter suggest that ultrasound RF time series, along with the developed machine learning framework, can help in differentiating malignant from benign breast lesions, subsequently reducing the number of unnecessary biopsies after mammography screening.
Chapter 3

Ultrasound-based Prediction of Prostate Cancer in MRI-guided Biopsy

3.1 Introduction

Prostate cancer is the most common type of solid tumor, and the second leading cause of cancer-related deaths in North American and European men. Early stage prostate cancer, which represents the majority of cases diagnosed today, has many therapy options, including surgery, radiation therapy, brachytherapy, thermal ablation, and active surveillance. Selection of the optimal therapy and therapeutic dosage are chiefly determined by diagnosis and staging. Definitive diagnosis of prostate cancer requires core needle biopsy, typically guided by TRUS. Current biopsy regimens involve systematic sampling of the prostate from eight or more predefined anatomical locations, followed by histopathological evaluation of these samples. The biopsy regimen is scaled to the prostate gland based on its size and using nomograms but otherwise not tailored to the individual. TRUS-guided biopsy has rather poor sensitivity, with positive predictive values between 40-60% [88]. Improved cancer yield can be achieved if patient-specific targeting is combined with systematic sampling. However, this is not feasible using TRUS
alone.

In order to enable patient-specific targeting, other modes of ultrasound imaging [36], such as RF data analysis [27], elastography [28], and Doppler imaging [30] have been explored. These technologies, individually, have not entirely proved successful in accurate identification of high grade cancer.

MRI has been used as an alternative modality to improve high grade prostate cancer yield [18]. Guidelines for structured reporting of prostate cancer assessments based on multi-parametric MRI have been developed, involving simultaneous examination of T2-weighted, DCE T1-weighted, and DWI sequences [23]. MRI-guided biopsy is, however, difficult, costly, time-consuming and not available in many parts of the world [22]. Fusion of ultrasound and MRI has been used to improve prostate cancer detection by enabling targeting of the cancer foci pre-determined in MRI during TRUS-guided biopsy [18]. Biopsy core locations determined in MRI are translated to patient coordinates using pre-procedure 3D TRUS and its registration to MRI [24]. 3D TRUS to MRI registration requires either sophisticated mechanical systems [25] to guide the biopsy needles or, if performed by software only [26], does not fully account for patient motion or organ deformation occurring during biopsy.

In Chapter 1 we introduced ultrasound RF time series as a tissue classification method. An RF time series is formed when ultrasound RF frames are captured in time from a stationary tissue location. In this thesis, we propose to use ultrasound RF time series to complement MR-targeted biopsy procedures by providing cancer likelihood maps around MRI targets during biopsy. We envision that this solution should increase positive cancer yield in both MR-targeted and/or TRUS-guided biopsy procedures. It will also provide an opportunity to correct for misregistrations of MR and TRUS images prior to sampling the tissue.

In the proposed solution, RF time series features have been used within an innovative computational framework that combines unsupervised clustering of the data with supervised classification. We use the histopathology of the biopsy cores for evaluation of cancer detection. Cancer likelihood maps are also shown that highlight the distribution and the likelihood of cancerous tissue within the biopsy cores. In a single centre feasibility trial with data obtained from 14 subjects at 18 biopsy targets, we are able to predict the pathology of MRI-identified targets with
high specificity and sensitivity.

3.2 Materials and Methods

3.2.1 Data Acquisition

Ultrasound RF time-series data is acquired on a Philips iU22 US scanner during MRI-guided targeted TRUS biopsies performed at the National Institutes of Health Clinical Center (NIH-CC, Bethesda, MD) using the Philips UroNav platform. For targeted biopsy, pre-acquired T2-weighted MRI images are automatically fused with real-time TRUS images of the prostate [22]. Initially, the desired targets are delineated on the T2-weighted MRI image by a clinician based on the examination of four multi-parametric MR images: T2-weighted, DWI, DCE, and MR spectroscopy. At the beginning of the biopsy procedure, a series of electromagnetically tracked two-dimensional (2D) TRUS images of the prostate are acquired from base to apex. Next, a 3D US volume is reconstructed based on electromagnetic (EM) tracking data and registered to the MRI scan in the UroNav software. Following the registration of US and MR volumes, the targeted locations for biopsy are transformed to the EM coordinate frame. During the biopsy, the clinician navigates through the prostate volume to reach the desired target location for acquiring a core. Immediately prior to taking the biopsy, the clinician holds the TRUS transducer steady for 4-5 sec to acquire RF time series data. Typically, 100 frames of RF time series data are acquired from each biopsy core. RF data is obtained prior to one, and in some cases, two biopsies of the MR-identified targets.

Ultrasound RF time series data is used from 18 biopsy cores of 14 subjects. Although RF time series data is collected in the axial plane, two biopsies are taken from axial and sagittal planes for each subject from the same location. The recording of the RF data and acquisition of the biopsy core are performed in sequence, not simultaneously, to avoid the appearance of the needle in the images. As a result, hand motion maybe present in some cases, between data and biopsy acquisition as well as during RF data recording. A quality control step is necessary to obtain a dataset with reliable reference label. In this step, we only choose to include subjects for which the histopathology of the axial and sagittal biopsies agree, and
no excessive motion is present during RF time series acquisition. In our data, 10 biopsy cores are cancerous with Gleason scores above 6 and tumor areas >40%. Eight biopsy cores are benign with consistent histopathological information.

3.2.2 Feature Extraction

Regions of Interest (ROIs): For each registered biopsy target, we analyze an area of 2 mm×10 mm in the lateral and axial directions, respectively, along the projected needle path in the RF data, and centred on the target. The width of this area is close to the width of the biopsy core. The length of the biopsy core is typically larger than the 10 mm considered here; however, to account for mis-registration errors and possible hand-motions, we use a conservative estimate in this study. The selected 2×10 mm area is divided into 20 ROIs of size 1×1 mm resulting a total of 360 ROIs from all biopsy cores. For each ROI, we calculate the features described below.

Features: Nine tissue typing parameters are extracted using the spectral, fractal,
and wavelet analysis of the RF time series data. Each RF time series contains 96 sequentially acquired frames of each RF sample of the imaging plane. We compute the spectrum of the zero-mean, hamming windowed, time series of an RF sample and average the values over an ROI. Summation of the spectrum in four equally-spaced frequency bands constitute features 1-4 [32]. The intercept and slope of the fitted line to the spectrum in the entire frequency range are features 5 and 6. Fractal dimension of the time series is computed using Higuchi’s method and averaged over an ROI as feature 7 [32]. We also calculate the central frequency (CF) of the spectrum as the mean of the spectrum bandwidth of the time series of an RF sample. The mean of the CF values (MCF) over an ROI is used as feature 8 [40]. Finally, we apply the discrete wavelet transform to the ultrasound RF time series of each RF sample using Daubechies 4 filter bank, where the signal is decomposed into approximation and detail coefficients at each decomposition level. The first approximation coefficient is computed for each RF sample in the imaging plane at the coarsest level (n=3) of decomposition and averaged over each ROI as feature 9 [40].
3.2.3 The Proposed Classification Framework

*Feature selection:* Feature selection is performed using Recursive Feature Elimination (RFE), prior to classification to identify the optimal combination of the nine features described above for cancer detection. In this method, features are eliminated recursively based on their corresponding weight in a linear SVM classifier. Initially, the model is trained on all the features and their weights are calculated. Then, the feature with the smallest absolute weight value is eliminated. This process is repeated recursively; the number and combination of features resulting in the highest classification accuracy are used as the stopping criteria. In our case, the combination of two features resulted in the highest classification accuracy.

*Figure 3.3:* An overview of the classification framework employed in this work. Clustering of the data is performed prior to classification to eliminate the affect of outliers and improve the classification accuracy.

*Classification:* Even though our biopsy cores are assigned to cancer or benign pathologies, the selected tissue types are heterogeneous within these classes and could potentially be differentiated based on other structural differences. One approach to overcome “within class” differences is to first cluster the ROIs in an unsupervised manner. This could result in identifying the outliers of each class from the main distribution of the class. A cluster-specific classifier can then be used to differentiate cancerous and benign tissue in a supervised manner.

*Experiments:* We follow a leave-one-subject-out cross-validation strategy. Here, we train a classifier using the features extracted from the cancerous and benign
ROIs of biopsy cores from 13 subjects and test on the features extracted from the ROIs of an unseen subject. In the first step of the process, ROIs from all 13 training subjects are clustered into two groups using k-means algorithm. Within each cluster, we train a Support Vector Machine (SVM) classifier to separate cancerous from benign ROIs. The next step constitutes testing, where we first assign the ROIs of the unseen subject to one of the clusters based on their Euclidean distances from the centroids of the clusters. The ROIs of the test subject are then classified using the classifier corresponding to their respective clusters. This process is repeated 14 times where every subject is left out for testing once. If in any of these leave-one-subject-out trials, a resulting cluster after the k-means step is over 90% imbalanced (over 90% benign or cancer), we do not train a classifier for that cluster and the label of test samples are determined based on majority voting in that cluster. In order to ensure that our process is not tailored to one type of classifier, we also use a Random Forests classifier and report our results using the two classification methods.

K-means clustering, SVM and Random Forests algorithms are implemented in the Scikit-learn machine learning package [76]. In addition to the binary class labels, we also estimated the cancer likelihood [77] for the biopsy core of each subject. The hyper-parameters that need to be determined for the classifiers included the Radial Basis Function (RBF) exponent and the soft margin penalty coefficient for SVM, and the number and depth of the trees in the Random Forests. These are tuned using a grid search approach.

3.3 Results

The RFE feature selection process was repeated for every leave-one-subject-out experiment. It consistently isolated features 3 and 4 as the combination of features that result in the highest classification accuracy between cancerous versus benign tissue. These are both spectral parameters of the RF time series. Henceforth, we only use these two features in clustering and classification of the biopsy cores. Figure 3.7 shows the two clusters that are created by k-means for ROIs from all subjects. 190 out of 200 malignant ROIs are assigned to cluster 1 and 125 out of 160 benign ROIs are assigned to cluster 2. In other words, 95% of all cancerous
samples and 78% of all benign samples are grouped in clusters 1 and 2, respectively. Based on this observation, and in order to maximize the number of training data per cluster, we limit the number of clusters to two.

The ROC curves are found in Figure 3.8. The area under the curve is 91.5% and 87.4% (95% CI: 85.7%–92.7%) for SVM and Random Forests methods, respectively. The difference in the area under the curves, of SVM and Random Forests, was statistically significant ($p < 0.05$). Colormaps that depict the cancer likelihood of ROIs in each of the 18 biopsy cores are illustrated in Figure 3.10. The likelihood
Figure 3.5: Correlation plots of the RF time series features. As can be seen in the figure, all the time series features are positively correlated, except for features 6 and 9. Due to high correlation among the features, we perform feature selection before classification.

threshold to label an ROI cancerous in the cancer likelihood maps is chosen to be 60%. It is noteworthy that if we eliminate the clustering step and perform classification with all training samples, we obtain an area under the curve of 86.9% and 87.1% for SVM and Random Forests methods, respectively.

Table [3.1] shows the percentage of the number of ROIs predicted as cancerous in each core, found in test samples in the leave-one-subject-out classification. The two different columns report the outcome for our method using SVM and Random Forests as classifiers. Using the SVM classifier, the percentage of cancer found in all benign cores is 45% or smaller and in five out of eight benign subjects this number is zero. In the positive biopsy cores, we notice that the predicted percentage of cancer is above 60% using the SVM classifier.
3.4 Discussion and Conclusion

We present a machine learning framework, consisting of supervised and unsupervised learning approaches, that uses RF time series analysis for the prediction of the histopathology of MR-guided targeted prostate biopsies. In a leave-one-subject-out study with data obtained from 18 biopsy cores in 14 subjects, we are able to accurately predict the pathology of MRI-identified targets with high specificity and sensitivity. In ROIs as small as 1 mm×1 mm, and using only two spectral features of RF time series, an area under ROC curve of 91.5% is achieved. Using k-means clustering, we show that these two features are able to separate cancerous and benign biopsy cores. Following classification, we calculate similar area under the ROC curve independently with SVM and Random Forests; this points to the stability of the proposed framework for tissue classification. We also present
Figure 3.7: Clustering performed on all 360 training samples.

colormaps that depict the cancer likelihood of ROIs in biopsy cores. These maps closely match the histopathology results of each biopsy core. As Table 3.1 shows, we report low cancer probabilities for all benign cores; specifically we predict zero cancer likelihood for five out of eight benign cores. In other words, 63% of the negative biopsies could have been avoided had we known the cancer likelihood of that area using RF time series during biopsy. In terms of sensitivity, as is observed in Table 3.1 we report at least 60% (mainly 70% and up) cancer likelihood for all positive cores.

Our results demonstrate that RF time series can be used to complement MR-targeted biopsy procedures, by providing cancer likelihood maps around MRI targets during biopsy. Our proposed method could potentially increase positive cancer yield in both MR-targeted and/or TRUS-guided biopsy procedures. It could also be used to compensate for mis-registrations of MR and TRUS images prior to sampling the tissue for MRI guided prostate biopsies.
A limitation of our study is the size of the dataset. This is partly due to our conservative quality control step where we drop data from targets with conflicting pathology results in axial and sagittal planes. In addition, to minimize the impact of registration and targeting error on our analysis, we only choose ROIs in 10 mm length of the RF data centred around the target along the needle trajectory. A typical biopsy core could be as long as 18 mm. Data acquisition for a large clinical study is ongoing; the aim is to also incorporate a detailed histopathology report-
Figure 3.9: Classification results for the 18 biopsy cores. As can be seen in the figure, all biopsy cores were correctly classified.

Figure 3.10: Cancer likelihood colormaps of the 18 biopsy cores from 14 subjects with leave-one-subject-out cross validation using the best two RF time series features. Clustering and SVM classification is used.
Table 3.1: SVM and Random Forests cancer probabilities.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Biopsy Core</th>
<th>Biopsy Result</th>
<th>Gleason Score</th>
<th>Percentage of Cancer SVM</th>
<th>Percentage of Cancer Random Forests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Core 1</td>
<td>Adenocarcinoma</td>
<td>7</td>
<td>100%</td>
<td>75%</td>
</tr>
<tr>
<td>2</td>
<td>Core 2</td>
<td>Adenocarcinoma</td>
<td>8</td>
<td>90%</td>
<td>85%</td>
</tr>
<tr>
<td></td>
<td>Core 3</td>
<td>Adenocarcinoma</td>
<td>8</td>
<td>70%</td>
<td>70%</td>
</tr>
<tr>
<td>3</td>
<td>Core 4</td>
<td>Adenocarcinoma</td>
<td>6</td>
<td>85%</td>
<td>70%</td>
</tr>
<tr>
<td>4</td>
<td>Core 5</td>
<td>Adenocarcinoma</td>
<td>9</td>
<td>95%</td>
<td>95%</td>
</tr>
<tr>
<td></td>
<td>Core 6</td>
<td>Benign</td>
<td>0</td>
<td>45%</td>
<td>25%</td>
</tr>
<tr>
<td>5</td>
<td>Core 7</td>
<td>Adenocarcinoma</td>
<td>8</td>
<td>75%</td>
<td>90%</td>
</tr>
<tr>
<td></td>
<td>Core 8</td>
<td>Adenocarcinoma</td>
<td>8</td>
<td>95%</td>
<td>100%</td>
</tr>
<tr>
<td>6</td>
<td>Core 9</td>
<td>Adenocarcinoma</td>
<td>7</td>
<td>75%</td>
<td>70%</td>
</tr>
<tr>
<td>7</td>
<td>Core 10</td>
<td>Adenocarcinoma</td>
<td>7</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>8</td>
<td>Core 11</td>
<td>Adenocarcinoma</td>
<td>7</td>
<td>60%</td>
<td>70%</td>
</tr>
<tr>
<td>9</td>
<td>Core 12</td>
<td>Benign</td>
<td>0</td>
<td>0%</td>
<td>60%</td>
</tr>
<tr>
<td>10</td>
<td>Core 13</td>
<td>Benign</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>11</td>
<td>Core 14</td>
<td>Benign</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Core 15</td>
<td>Benign</td>
<td>0</td>
<td>30%</td>
<td>10%</td>
</tr>
<tr>
<td>12</td>
<td>Core 16</td>
<td>Benign</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>13</td>
<td>Core 17</td>
<td>Benign</td>
<td>0</td>
<td>15%</td>
<td>10%</td>
</tr>
<tr>
<td>14</td>
<td>Core 18</td>
<td>Benign</td>
<td>0</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

ing scheme where the direction of the cancer in a core is marked and results are reported in quarters along the biopsy core. We expect a larger dataset and more accurate mapping of histopathology to RF time series would further improve the results.

3.5 Chapter Summary

In this chapter, an in vivo clinical feasibility study for ultrasound-based detection of prostate cancer in MRI selected biopsy targets was reported. Spectral analysis of a temporal sequence of ultrasound RF data reflected from a fixed location in the tissue results in features that can be used for separating cancerous from benign biopsies. Data from 18 biopsy cores and their respective histopathology were used in an innovative computational framework, consisting of unsupervised and super-
Figure 3.11: Scatter plot showing the ROI samples from the 18 biopsy cores using features 3 and 4. Red and blue markers represent malignant and benign ROI samples from 18 cores.

Supervised learning, to identify and verify cancer in regions as small as 1 mm × 1 mm. In leave-one-subject-out cross validation experiments, an area under ROC of 91.5% (95% CI: 90.6%–95.7%) was obtained for cancer detection in the biopsy cores. Cancer likelihood maps that highlight the predicted distribution of cancer along the biopsy core, also closely match histopathology. The results of this chapter demonstrate the potential of the RF time series to assist patient-specific targeting during prostate biopsy.
Figure 3.12: SVM decision function plots for the malignant cores. The decision plots shown here are without clustering. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. It is apparent in the above figure, some ROI samples from cores 3, 8, 9, and 11 are in the benign region and without clustering would have been misclassified. These decision functions (plotted without clustering) show the utility of clustering prior to classification. See Figure 3.11 for distribution of the ROI samples.
Figure 3.13: Random Forests decision function plots for the malignant cores. The decision plots shown here are without clustering. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. It is apparent in the above figure, some ROI samples from cores 3, 8, 9, and 11 are in the benign region and without clustering would have been misclassified. These decision functions (plotted without clustering) show the utility of clustering prior to classification. See Figure 3.11 for distribution of the ROI samples.
Figure 3.14: SVM decision function plots for the benign cores. The decision plots shown here are without clustering. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the above figure, most of the benign ROI samples lie in the benign region (cyan). Some ROI samples from core 6 are in the malignant region and hence the classification accuracy for core 6 is lower than the other benign cores. See Figure 3.11 for distribution of the ROI samples.
Figure 3.15: Random Forests decision function plots for the benign cores. The decision plots shown here are without clustering. The red region is the decision boundary for the malignant ROI samples and the cyan region is the decision boundary for the benign ROI samples. As can be seen in the above figure, most of the benign ROI samples lie in the benign region (cyan). Some ROI samples from core 6 are in the malignant region and hence the classification accuracy for core 6 is lower than the other benign cores. See Figure 3.11 for distribution of the ROI samples.
Chapter 4

Experiments to Provide Evidence on the Physical Basis of Tissue Classification using RF Time Series

4.1 Introduction

Ultrasound RF time series method is a data-driven approach that has yielded consistent results in classification of animal tissue [37], prostate cancer ex-vivo [32], prostate cancer in-vivo [39, 40], breast lesions [66], and in monitoring of tissue ablation [41]. While we have not discussed the potential reasons for improved performance obtained from the use of a sequence of frames, several hypotheses describing the physical phenomenon exist and have been explored elsewhere.

It has been shown experimentally and in simulations, that the temperature increase in controlled irradiation of tissue with RF time series can partly explain the phenomenon [87]. Also, increased energy delivered to the tissue through increased frame rate or transmit power result in improved tissue classification in phantom and animal studies [37, 41]. Given these observations, micro-vibrations of the tissue microstructure caused by acoustic radiation force is another likely phenomenon.
The acoustic radiation force is related to both the acoustic energy and to the attenuation and scattering properties of the tissue, which are different for different pathological tissue types.

Despite the increased use of ultrasound RF time series as a method for tissue classification, limited work has been done to provide evidence on the physical basis of tissue typing using this method. In the past, it has been shown that changing the transmit power and the frame rate improves the tissue classification performance of RF time series [32, 37, 41]. However, further experiments exploring the effects of other ultrasound imaging parameters, which can increase the energy delivered to the tissue, could provide sound evidence on the source of the tissue typing information extracted using RF time series.

In this chapter, we build upon previous hypotheses by studying the effects of additional ultrasound imaging parameters such as imaging depth and time series length on tissue classification. We describe experimental studies on animal tissue to provide physical evidence of tissue typing using ultrasound RF time series. The effects of three imaging parameters: imaging depth, frame rate, and the length of the time series, on tissue classification are studied. In this work, RF time series data from two animal tissues, bovine and chicken were collected and the effects of these imaging parameters on tissue classification were studied. 13 experiments were conducted and during each experiment a parameter was changed and the RF time series data was saved. The data for each experiment was analysed by calculating the RF time series features and then using two linear classifiers, linear SVM and logistic regression for tissue classification. For each experimental dataset, the tissue classification performance was evaluated by a 20 fold stratified cross-validation and a calculation of cumulative area under the ROC curve.

4.2 Materials and Methods

This section describes the overall framework of this study. RF time series data was collected for bovine and chicken. After collecting the data, we extracted seven RF time series features for 200 1 mm\(^2\) ROIs centered around the ultrasound beam focal point. This resulted in 100 chicken and 100 beef (10 mm \(\times\) 10 mm window) training samples described by seven RF time series features. For classification of the ROI
samples we used a logistic regression and a linear SVM classifier. The choice of a simple linear approach to classification was made to minimize the effects of machine-learning model selection on the outcome of these comparative studies. The entire process was repeated 13 times with each experiment focusing on one parameter, as described in Table 4.1. Figure 4.1 shows the experimental setup for the study.

Table 4.1: Parameters changed during each experiment. The parameters in bold were changed for the corresponding experiment.

<table>
<thead>
<tr>
<th>Exp. #</th>
<th>Frequency (MHz)</th>
<th>Focal Point (cm)</th>
<th>Imaging Depth (cm)</th>
<th>Frame Rate (fps)</th>
<th>Time Series (frames)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.6</td>
<td>1.5</td>
<td>4</td>
<td>50</td>
<td>1000</td>
</tr>
<tr>
<td>2</td>
<td>6.6</td>
<td>1</td>
<td>4</td>
<td>50</td>
<td>1000</td>
</tr>
<tr>
<td>3</td>
<td>6.6</td>
<td>2</td>
<td>4</td>
<td>50</td>
<td>1000</td>
</tr>
<tr>
<td>4</td>
<td>6.6</td>
<td>2.5</td>
<td>4</td>
<td>50</td>
<td>1000</td>
</tr>
<tr>
<td>5</td>
<td>6.6</td>
<td>3</td>
<td>4</td>
<td>50</td>
<td>1000</td>
</tr>
<tr>
<td>6</td>
<td>6.6</td>
<td>3.5</td>
<td>4</td>
<td>50</td>
<td>1000</td>
</tr>
<tr>
<td>7</td>
<td>6.6</td>
<td>1.5</td>
<td>5.5</td>
<td>40</td>
<td>1000</td>
</tr>
<tr>
<td>8</td>
<td>6.6</td>
<td>1.5</td>
<td>3</td>
<td>60</td>
<td>1000</td>
</tr>
<tr>
<td>9</td>
<td>6.6</td>
<td>1.5</td>
<td>3</td>
<td>71</td>
<td>1000</td>
</tr>
<tr>
<td>10</td>
<td>6.6</td>
<td>1.5</td>
<td>4</td>
<td>50</td>
<td>200</td>
</tr>
<tr>
<td>11</td>
<td>6.6</td>
<td>1.5</td>
<td>4</td>
<td>50</td>
<td>400</td>
</tr>
<tr>
<td>12</td>
<td>6.6</td>
<td>1.5</td>
<td>4</td>
<td>50</td>
<td>600</td>
</tr>
<tr>
<td>13</td>
<td>6.6</td>
<td>1.5</td>
<td>4</td>
<td>50</td>
<td>800</td>
</tr>
</tbody>
</table>

4.2.1 Types of Tissue

Two types of animal tissue were used, bovine tissue and chicken breast. Both tissues have fibrous structures which makes them harder to differentiate based on the B-mode appearance. Shown in Figure 4.2 the B-mode appearance of the two tissue types is very similar. This lack of visually distinct appearance of the tissue types in B-mode images suggests a need for an ultrasound based tissue differentiation method that is independent of B-mode appearance. Ultrasound RF time series is a tissue typing method that does not rely on B-mode appearance but in this method of analysis spectral analysis of ultrasound time series signals is performed. The
animal tissues were bought fresh from the butcher shop 60 minutes before the data collection. This allowed the tissue samples to attain room temperature. The two types of tissue were separated by ultrasound gel and imaged together, as shown in Figure 4.1. The ultrasound B-mode image of the two tissue types can be seen in Figure 4.2.

4.2.2 Data Collection

Data for this study was collected on a SonixMDP ultrasound machine using a L14-3/38 linear ultrasound probe at Ultrasonix Medical Corp., Richmond, BC, Canada. The two pieces of animal tissue were placed together in a tub surrounded by an absorbing pad. Ultrasound gel was put between the two tissues to ensure good contact. To minimize operator hand motion and collect data from the same tissue location for all experiments the ultrasound probe was held in place by a clamp. Be-
Figure 4.2: B-mode image of the two animal tissues. Bovine on the left and chicken (right).

Between each experiment, we waited four minutes to minimize the effect of previous round of ultrasound RF data collection in terms of heating. Figure 4.1 shows the experimental setup for the conducted experiments.

4.2.3 Experiment Design

13 experiments were conducted. For each experiment one ultrasound imaging parameter was changed and the RF data was stored. The parameter changed for each
experiment is shown in Table 4.1. The default imaging parameters were as follows: Center frequency = 6.6 MHz, focal point = 1.5 cm, imaging depth = 4 cm, frame rate = 50, and RF time series length = 1000 frames.

### 4.2.4 Features

100 ROIs of size 1 mm\(^2\) centered at the ultrasound beam focal point were extracted for each tissue type. Six RF time series features and the fractal dimension feature was calculated for each ROI. The RF frame size was 256\(\times\)2080. The number of pixels in 1 mm\(^2\) ROIs was dependent on the imaging depth. The ROI sizes in the RF domain for the each imaging depth, can be found in Table 4.2. Description of the seven RF time series features can be found in Section 2.2.2.

**Table 4.2:** 1 mm\(^2\) ROI size in the RF domain for different imaging depths.

<table>
<thead>
<tr>
<th>Imaging depth (cm)</th>
<th>1 mm(^2) ROI size (RF samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6(\times)69</td>
</tr>
<tr>
<td>4</td>
<td>6(\times)52</td>
</tr>
<tr>
<td>5.5</td>
<td>6(\times)38</td>
</tr>
</tbody>
</table>

### 4.2.5 Classification and Cross-validation

**Classification:** In this study, a logistic regression classifier and a linear SVM (no kernel trick) classifier were used. We chose simple linear classifiers to avoid overfitting and show the robustness of the RF time series method. The results of the report can further be improved by using state-of-the-art classifiers like kernel SVM (kernel machines) and Random Forests. However, in this chapter the goal is to observe trends in response to varying parameters of ultrasound imaging. We use a simple model to minimize the effects of cross-validation on the outcome. Two classifiers were used to prove the stability of the results. The performance of each classifier was measured by computing the AUCs. We also calculated the \(p\)-value for statistically comparing the performance of the two classifiers used in this work.

**Cross-validation:** A stratified k-fold cross-validation technique was employed to evaluate the classification performance. 20 stratified (equal number of test sam-

59
amples from each class) training and test sets were formed and a combined ROC curve was computed for the 20 folds.

### 4.3 Results

Table 4.3 shows the areas under the ROC curve for the 13 experiments performed. As can be seen in the figures, the AUC values for both, linear SVM and logistic regression, are very similar. The classification performance difference between the linear SVM classifier and the logistic regression classifier was statistically not significant ($p > 0.05$). The statistical insignificance ($p > 0.05$) in the performance of the two classification algorithms proves the stability of the results. AUC values using an RBF kernel SVM are also reported. The effects of the imaging depth, frame rate, and time series length on tissue typing are presented in the following sections.

#### 4.3.1 Effects of Imaging Depth on Tissue Classification

After analysis of the AUC values and performing statistical significance tests, it is apparent that increasing the ultrasound imaging depth significantly ($p < 0.001$) decreased the tissue classification performance. This decreasing trend can be seen in Figure 4.3. The difference in AUC values for different imaging depths (experiments 2, 3, 4, 5 and 6) was statistically highly significant ($p$-value $= 0.000005$) and can be found in Table 4.3.

**Table 4.3: Area under the ROC curve with changing imaging depth.**

<table>
<thead>
<tr>
<th>Exp. #</th>
<th>Linear SVM</th>
<th>Logistic Regression</th>
<th>RBF SVM</th>
<th>Imaging Depth (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>71.8%</td>
<td>73.8%</td>
<td>79.9%</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>88.7%</td>
<td>88.4%</td>
<td>86.2%</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>82.9%</td>
<td>82.6%</td>
<td>82.5%</td>
<td>2.5</td>
</tr>
<tr>
<td>5</td>
<td>79.1%</td>
<td>79.0%</td>
<td>77.3%</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>69.6%</td>
<td>69.0%</td>
<td>69.5%</td>
<td>3.5</td>
</tr>
</tbody>
</table>
Figure 4.3: The effect of imaging depth on the classification accuracy of the two tissue types, bovine and chicken. As can be seen in the figure, increasing the imaging depth decreased the classification accuracy. In other words, using the RF time series method, targets that are closer to the ultrasound transducer can be classified more accurately than the targets that are farther away from the transducer. The results are consistent with the previous studies.

4.3.2 Effects of Frame Rate on Tissue Classification

The AUC values for the experiments performed to study the effects of ultrasound frame rate on tissue classification are reported in Table 4.4. From the AUC values, it is apparent that increasing the frame rate resulted in improved tissue classification. This improvement in the classification results as a result of increasing frame rate was statistically significant ($p$-value = 0.027). The increasing AUC trend as a result of increasing frame rate can be seen in Figure 4.4.

![Graph showing the effect of imaging depth on classification accuracy](image)

**Figure 4.3:** The effect of imaging depth on the classification accuracy of the two tissue types, bovine and chicken. As can be seen in the figure, increasing the imaging depth decreased the classification accuracy. In other words, using the RF time series method, targets that are closer to the ultrasound transducer can be classified more accurately than the targets that are farther away from the transducer. The results are consistent with the previous studies.
Table 4.4: Area under the ROC curve for different frame rates.

<table>
<thead>
<tr>
<th>Exp. #</th>
<th>Linear SVM</th>
<th>Logistic Regression</th>
<th>RBF SVM</th>
<th>Frame Rate (fps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>77.7%</td>
<td>78.5%</td>
<td>77.2%</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>79.7%</td>
<td>79.3%</td>
<td>79.5%</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>81.2%</td>
<td>82.6%</td>
<td>84.2%</td>
<td>60</td>
</tr>
<tr>
<td>9</td>
<td>87.7%</td>
<td>87.1%</td>
<td>92.6%</td>
<td>71</td>
</tr>
</tbody>
</table>

Figure 4.4: The effect of increasing frame rate on the classification accuracy of the two tissue types, bovine and chicken. Increasing the frame rate significantly improved the classification results. It is apparent from the above figure, increasing the frame rate from 40 to 71 frames per second improved the area under the ROC curve from 77% to 87%. Previous studies have shown that increasing the frame rate improved the classification performance, therefore our results are consistent with the previous studies.
4.3.3 Effects of Time Series Length on Tissue Classification

Table 4.5 shows the tissue classification performance in terms of the AUC values for different time series lengths. It was found that increasing the time series length significantly \( (p\text{-value} = 0.00014) \) improved the tissue classification results. This observation suggests that using more frames for RF time series feature calculation can result in strong features that distinguish the tissue types more accurately. The increasing trend in the AUC values as the number of RF frames analyzed is increased can be seen in Figure 4.5.

Table 4.5: Area under the ROC curve for different time series length.

<table>
<thead>
<tr>
<th>Exp. #</th>
<th>Linear SVM</th>
<th>Logistic Regression</th>
<th>RBF SVM</th>
<th>Time Series Length (frames)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>49.0%</td>
<td>43.5%</td>
<td>61.5%</td>
<td>200</td>
</tr>
<tr>
<td>11</td>
<td>69.1%</td>
<td>68.8%</td>
<td>70.0%</td>
<td>400</td>
</tr>
<tr>
<td>12</td>
<td>75.3%</td>
<td>75.8%</td>
<td>75.0%</td>
<td>600</td>
</tr>
<tr>
<td>13</td>
<td>79.1%</td>
<td>78.1%</td>
<td>79.2%</td>
<td>800</td>
</tr>
<tr>
<td>1</td>
<td>79.7%</td>
<td>79.3%</td>
<td>79.5%</td>
<td>1000</td>
</tr>
</tbody>
</table>

To further investigate the decreased performance when using 200 frames (4 seconds of RF data), we used the last 200 frames, instead of the first 200, and recorded the AUC as 54% and 56% using Logistic Regressions and SVM classifiers. We also calculated the AUC using 200 RF frames and at an imaging depth of 2 cms (Experiment # 3), as compared to 1.5 cms, and the result was 52% for classification using Logistic Regression and 53% using linear SVM.

4.4 Discussions and Conclusions

This chapter described an experimental study on animal tissue to empirically test RF time series as a method for tissue typing. Previous studies on ultrasound RF time series and Chapter 2 and Chapter 3 have reported consistent tissue classification results using this method. The source of tissue typing information extracted from this type of analysis has not been empirically confirmed. In this chapter we studied the effects of ultrasound imaging depth, frame rate, and the length of the
Figure 4.5: The effect of the length of the time series on the classification accuracy of the two tissue types, bovine and chicken. As can be seen in the above figure, increasing the number of analysed RF frames, for the RF time series feature calculation, improved the classification performance. These results show, the tissue typing information improved as more RF frames are analysed. These observations are consistent with the previous published literature.

time series on the classification of two animal tissue types. It has been demonstrated that increasing the energy delivered to the tissue improved the tissue classification accuracy.

The results of this study suggest that:

1. Using ultrasound RF time series, as the ultrasound imaging depth increases, the animal tissue classification performance significantly decreases ($p < 0.001$).

2. Increasing the frame rate of the ultrasound machine significantly improved the animal tissue classification result using ultrasound RF time series method ($p < 0.05$).
3. In ultrasound RF time series analysis, increasing the number of RF frames analyzed (time series length) significantly enhances the animal tissue classification performance \((p < 0.001)\).

A lower tissue classification accuracy was observed at an imaging depth of 1 cm. This could be attributed to the near field behaviour of the ultrasound beam and to the fact that 1 cm was below the focal range of the transducer. In our experiments, using four seconds of data at 50 frames per second resulted in poor classification results. On the other hand, as more frames were analysed, the tissue classification was significantly better. This could be due to the temperature increase in the tissue as a result of the increased time series length.

Overall, it is clear that increasing the amount of the energy delivered to the tissue improves the performance of the tissue classifiers. The improved classification performance as a result of changing the imaging depth, frame rate, and the length of the time series could be due to the temperature increase in controlled irradiation of tissue with RF time series [87]. Increased energy delivered to the tissue through increased frame rate and shallower imaging depth could also be the cause of the improved tissue classification accuracy [37, 41]. An additional probable phenomenon for the improved tissue classification result could be micro-vibrations of the tissue microstructure caused by acoustic radiation force. The acoustic radiation force is related to both the acoustic energy and to the attenuation and scattering properties of the tissue, which are different for different pathological tissue types. Simulation studies to measure these micro-vibrations are ongoing to further validate these ultrasound RF time series hypotheses.

The results of this study are consistent with the previous reported results which suggest that increased energy delivered to the tissue through increased frame rate or transmit power results in improved tissue classification accuracy in phantom and animal studies [37, 41].

4.5 Chapter Summary

In this chapter the effects of three ultrasound imaging parameters--imaging depth, frame rate, and RF time series length--on tissue classification performance were studied to provide further evidence on the source of tissue typing information ex-
tracted using ultrasound RF time series. The results of this chapter are consistent with the previously reported studies, which suggest that increased energy delivered to the tissue results in improved tissue classification. From the observations reported in this chapter, it can be concluded that decreasing the imaging depth, increasing the frame rate, and increasing the length of the time series improves animal tissue classification using ultrasound RF time series.
Chapter 5

Conclusions

Augmentation of the current breast and prostate cancer diagnosis is a pressing need due to the rate of over-diagnosis and number of unnecessary biopsies performed. In this work, an ultrasound-based tissue classification method for breast and prostate cancer diagnosis was presented.

Ultrasound is a rapidly growing imaging modality that is inexpensive as compared to MRI and unlike X-ray or CT, does not expose the patient to ionizing radiation. Ultrasound-based cancer diagnosis methods are versatile and do not require additional hardware. Ultrasound RF time series is an ultrasound-based cancer diagnosis technique that is non invasive and has been proven successful in classification of animal tissue [37], prostate cancer ex vivo [32] and in vivo [39, 40], and tissue ablation monitoring [41]. Ultrasound RF time series along with other ultrasound-based diagnosis methods could provide radiologists a real-time breast and prostate cancer diagnosis tool that could significantly reduce the number of negative biopsies.

A tissue classification framework, comprising ultrasound RF time series signal analysis, ultrasound RF signal analysis, B-mode texture analysis, and machine learning techniques has been proposed to address the need for ultrasound-based cancer diagnosis. Breast and prostate tissue classification studies were performed to evaluate the performance of the tissue classification framework and prove the validity of RF time series analysis as a tissue typing method. The results of these studies suggest the potential of ultrasound RF time series as a useful cancer di-
agnosis method. Cancer probability maps presented in this work could provide radiologists with key information about the cancer affected region, which could improve biopsy targeting and in some cases eliminating the need for biopsy.

5.1 Summary of Contributions

- We developed a method for classifying breast lesions based on ultrasound RF time series analysis. This was the first RF time series study on breast data. Data from 22 subjects was analyzed and promising results were presented. Ultrasound RF time series, single RF frame, and B-mode texture features were used along with a machine learning framework to diagnose cancer in breast lesions. The clinical significance of this work are the reported cancer probability maps. Cancer probability maps could provide radiologist a real-time cancer diagnosis tool which could significantly improve the cancer yield and reduce the number of unnecessary breast biopsies. Feature calculation scripts were developed to calculate RF time series features, B-mode texture features, and RF spectral features. A machine learning package was also developed for classification of ROIs described by the calculated features. Data analysis and visualization scripts were also developed for exploratory analysis.

- A new semi-supervised machine learning technique was proposed for \textit{in vivo} prostate tissue classification. The new semi-supervised classification method performs clustering of the data prior to classification to eliminate the effect of outliers and improve the differentiation between the malignant and the benign subjects. The effect of outliers was minimized by training classifier models on cluster specific samples. In this prostate tissue classification study, ultrasound RF time series features were used for classification of MRI-targeted biopsy cores. Data from 18 cores (14 subjects) was analysed and highly accurate cancer prediction was reported. The results of this study suggests that clustering of data before classification can improve the classification results. This work could potentially lead to the fusion of the proposed tissue classification framework with the Philips UroNav platform to complement MRI-targeted TRUS guided prostate biopsy and provide real-
time diagnosis of prostate cancer. Feature calculation scripts were developed in MATLAB and the machine learning algorithms were written in Python. The developed algorithms could benefit future studies on ultrasound RF time series.

- An animal tissue study was also completed to substantiate the ultrasound RF time series hypothesis. The effects of ultrasound imaging parameters: imaging depth, frame rate, and the time series length were studied and the results were documented. This work could provide more evidence on the physical basis of tissue typing using ultrasound RF time series method. Data from two animal tissue, steak and chicken, was collected and analysed. Ultrasound RF time series features were calculated and linear classifiers were employed for classification. The results of this study are consistent with the previously reported and simulation studies are ongoing to further validate the RF time series hypotheses. Feature calculation scripts along with machine learning algorithms were developed for this study. A comprehensive analysis of the results was performed to study the effect of each ultrasound imaging parameter on the tissue classification accuracy.

5.2 Future Work

Future work in this field would include researching more robust features for tissue classification. Wavelet analysis of RF time series along with the current features could result in improved tissue classification.

The methods presented in this thesis may be incorporated into the Philips UroNav platform consequently providing radiologists with a real-time cancer diagnosis tool.

Experimental studies further exploring the source of tissue typing information using RF time series would prove beneficial to future developments employing the RF time series technique. Simulations studies to measure the micro-vibrations and temperature changes in the tissue due to energy delivered would help prove the physical basis of tissue typing using RF time series.

Hyper-parameter optimization techniques would significantly reduce the computation time required to select the optimum parameters for a classification algo-
rithm. Currently, a grid search over the parameter space is performed. This approach is time consuming and not practical. It has been shown that random search for hyper-parameter optimization is more efficient than the traditional grid search or manual search for finding the best hyper-parameter of a classification algorithm [89]. For real-time cancer diagnosis the computation time would have to be significantly reduced and this can be achieved by developing or employing advanced hyper-parameter optimization techniques like random search.

Approaches such as outlier detection, Density-based spatial clustering of applications with noise (DBSCAN) [90], and label propagation [91, 92] are some of the techniques that could improve the current classification approach. By using advanced clustering algorithms like DBSCAN, which is more robust to noisy dataset, the novel semi-supervised machine learning approach presented in Chapter 3 could be improved.
Bibliography


74


79
Zanetti, R., Neyroud-Caspar, I., Bouchardy, C.: Importance and
determinants of Gleason score undergrading on biopsy sample of prostate
pages 35


226–231 → pages 70


International Workshop on Artificial Intelligence and Statistics. (2005)
96–103 → pages 70

80