Automated Detection of Breast Cancer Using SAXS Data and Wavelet Features

A thesis submitted to the College of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Master of Science
Division of Biomedical Engineering
University of Saskatchewan
Saskatoon

By
Carissa Erickson

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Division of Biomedical Engineering
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A9

ABSTRACT

The overarching goal of this project was to improve breast cancer screening protocols first by collecting small angle x-ray scattering (SAXS) images from breast biopsy tissue, and second, by applying pattern recognition techniques as a semi-automatic screen. Wavelet based features were generated from the SAXS image data. The features were supplied to a classifier, which sorted the images into distinct groups, such as “normal” and “tumor”.

The main problem in the project was to find a set of features that provided sufficient separation for classification into groups of “normal” and “tumor.” In the original SAXS patterns, information useful for classification was obscured. The wavelet maps allowed new scale-based information to be uncovered from each SAXS pattern. The new information was subsequently used to define features that allowed for classification. Several calculations were tested to extract useful features from the wavelet decomposition maps. The wavelet map average intensity feature was selected as the most promising feature. The wavelet map intensity feature was improved by using pre-processing to remove the high central intensities from the SAXS patterns, and by using different wavelet bases for the wavelet decomposition.

The investigation undertaken for this project showed very promising results. A classification rate of 100% was achieved for distinguishing between normal samples and tumor samples. The system also showed promising results when tested on unrelated MRI data. In the future, the semi-automatic pattern recognition tool

developed for this project could be automated. With a larger set of data for training and testing, the tool could be improved upon and used to assist radiologists in the detection and classification of breast lesions.

ACKNOWLEDGEMENTS

I would like to thank my research supervisor, Dr. Edward Kendall, for giving me the opportunity to work on this project. It has been a very challenging but rewarding experience. Thank you for your knowledge, encouragement, and patience throughout the process.

I would like to thank my external examiner, Dr. Ron Bolton, and the members of my supervisory committee, Dr. Bill Thomlinson and Dr. Mark Eramian. A special thank you to Dr. Bill Thomlinson for allowing me to visit the medical beamline at the ESRF to observe the data collection process for a diffraction enhanced imaging experiment.

I would like to thank Dr. Rob Lewis for providing me with the data used in this project.

Thank you to the Medical Imaging Research Group and the Saskatchewan Synchrotron Institute for providing funding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Bior	Biorthogonal
CAD	Computer Aided Detection
Db	Debauchies
ECM	Extracellular Matrix
FT	Fourier Transform
FFT	Fast Fourier Transform
WFT	Windowed Fourier Transform
MRI	Magnetic Resonance Imaging
ROC	Receiver Operating Characteristic
SAXS	Small Angle X-ray Scattering
SLGD	Spatial Grey Level Dependency

1 INTRODUCTION

1.1 Thesis Outline

This thesis describes the semi-automatic pattern recognition system that was developed for breast cancer detection.

Chapter one introduces the objectives of the paper. Background information on the limitations of current methods of breast cancer detection, as well as possible improvements through computer aided diagnosis (CAD) are discussed. Next, the small angle x-ray scattering (SAXS) data that was used for this project is introduced, including review of diffraction theory. Synchrotron technology, which was used to acquire the SAXS data, is also discussed.

Wavelet analysis was used to parse useful information out of the SAXS data. A detailed introduction to wavelet analysis is presented in chapter two. Wavelets are introduced through a review of wavelet applications. An intuitive description is then followed by details on wavelet theory.

Pattern recognition techniques were employed in conjunction with the wavelet analysis in order to achieve a semi-automated CAD system. Chapter three focuses on pattern recognition. The chapter starts with an overview of some common pattern recognition approaches as well as an introduction to the original pattern recognition approach used

for this project. Finally, a more concrete description of a generic pattern recognition system is given.

Chapter four is a detailed description of the methods used in the analysis of the data for this project. Chapter four shows how the generic pattern recognition system introduced in Chapter three was adapted into an original system designed to use wavelet analysis on SAXS data in order to create classifiable features.

The results of the project are given in chapter five, the discussion is provided in chapter six, and the conclusions are given in chapter seven.

1.2 Objectives

This project had three main objectives:

1. To develop a semi-automatic pattern recognition tool to assist radiologists in breast cancer detection.
2. To investigate specialized SAXS data acquired using synchrotron imaging.
3. To investigate wavelets as a tool for parsing the SAXS data.

1.3 Breast Cancer Detection

Breast cancer is the most frequently diagnosed cancer in Canadian women. In 2004, an estimated 21,200 Canadian women will be diagnosed with breast cancer and 5,200 will die of it. Since 1988, breast cancer incidence rates have risen by 10%, but death rates

have dropped by 19%. The decline in death rates is believed to be due to the benefits of breast cancer screening programs and improved treatments [38].

When breast cancer is detected and treated early, the chances for recovery are better. In March 2002, the World Health Organization's International Agency for Research on Cancer Working Group confirmed early detection and treatment are considered the most promising approach to reduce breast cancer mortality [4] [23]. The current methods for early detection of breast cancer are clinical breast exams and mammography, which is a low-dose x-ray imaging technique. The Canadian Cancer Society recommends that all women between the ages of 50 and 69 have a screening mammogram as well as a clinical breast examination every two years. Although mammography is able to show changes in the breast up to two years before a patient or physician could feel them [41], the process is not foolproof.

Mammography is reported to have a sensitivity of 70% to90% [9]. That means that the false negative rate is between 10% and 30%. In other words, mammograms can miss over one quarter of all tumors [54] [39]. False negatives occur when the mammogram is interpreted as negative when cancer is present. False negatives occur most often with dense breasts that make the masses difficult to distinguish. Cancers are easier to detect in fatty breasts that are less dense [33].

False positives occur when a mammogram is read as abnormal when no cancer is present. Abnormal mammograms are followed up with biopsy procedures to determine

whether the abnormality is cancerous. The main problem with false positive results is that the patient is required to undergo medical procedures that would have been avoided with an accurate screening result. Elmore's study showed that over a time period of ten years, one third of all women screened for breast cancer had an abnormal screening result that resulted in additional evaluation even though no cancer was present [15].

Christiansen reported several factors that lead to false positive screening results. According to the study, false positives were seen more often in women who were pre-menopausal, women who were post-menopausal and taking estrogen, women who had undergone previous biopsies, and women with a family history of breast cancer. The false positive rates were also dependent on the doctor. Some radiologists had higher rates of false positive diagnosis than others did. It was also reported that the false positive rate was higher if the radiologist did not compare the current mammogram to previous mammograms [10].

One method that has been suggested for reducing false positive rates is double reading of mammograms. Double reading requires the same mammogram to be analyzed by two different radiologists. Although double reading has been shown to increase the sensitivity of mammogram results by as much as 15%, it is a very time consuming and costly procedure [43]. CAD is an active area of study because it may provide the benefits of double reading in an efficient and cost-effective way [2]. The CAD system would take the place of one radiologist, saving considerable time and money.

The goal of CAD systems is to reduce errors by drawing radiologists' attention to possible abnormalities [20]. CAD systems could also be applied to biopsy screening. The goal of this project was to design a semi-automatic CAD system for identifying breast cancer from a SAXS image of a biopsy sample. The system could be used to assist radiologists in improving the accuracy of breast cancer diagnosis.

1.4 Synchrotron Technology

The second objective for this project was to investigate the use of specialized data to improve the accuracy of breast cancer screening. The synchrotron is an exciting tool for collecting specialized data, and has been used to investigate problems in many areas including materials science, environmental science, crystallography, biology, and medicine. For this project, the synchrotron was used to collect SAXS images of breast tissue.

1.4.1 Synchrotron radiation properties

Synchrotron radiation exhibits certain properties that make it a valuable research tool. Some of these properties are high intensity, broad spectral range, and collimation. The synchrotron beam is millions of times more intense than a conventional medical x-ray beam. This allows for quick data collection. The spectrum is continuous, and ranges from infrared to x-ray wavelengths. The researcher can select the necessary wavelength for specific applications. Collimation means that the beam of light does not spread out,

as visible light would, allowing the beam to be focused on a precise area smaller than a micron.

1.4.2 Small angle x-ray scattering

Synchrotron radiation is an ideal source for collecting certain types of specialized data. One example is small angle x-ray scattering data. SAXS requires a monochromatic beam that is well collimated. These are two features inherent in synchrotron radiation.

SAXS is a technique that is used to study non-crystalline biological materials such as proteins in solutions or biological fibers. SAXS can be distinguished from wide-angle x-ray scattering by the location of the scattering pattern of interest. SAXS looks at scattering near the primary beam, corresponding to structural features ranging in size from tens to thousands of angstroms [45]. Wide angle scattering typically originates from structural features less than 8 Å.

In this project, breast tissue collagen provided the well-organized structure that produced the SAXS data. The basic unit of collagen is a triple-stranded helical molecule about 300nm long. In fibrous collagens, such as types I, II, III, and V, the basic molecules pack together side by side, forming fibrils with a diameter of 50 – 200 nm. In fibrils, adjacent collagen molecules are displaced from one another by 67 nm [1]. Collagens are the major proteins present in the extracellular matrix. The extracellular matrix, or ECM, is an intricate network of macromolecules that fill up the spaces between cells in tissues. Diffraction patterns of normal breast tissue have shown

the ECM to contain both type I and III collagens [26]. Although the exact role of the ECM in cancer is not well understood, it has been shown that its structure is seriously disturbed in malignant breast lesions [42]. It has been shown by Pucci-Minafra [40], that the invasive tumour growth of breast carcinomas is characterized by drastic changes in the collagen scaffold. These changes are correlated with significant biochemical changes.

Structural changes in the collagen of the ECM caused by breast cancer allow collagen structure to be studied in order to detect the disease. Because of the high degree of order in collagen structures, they can be studied using x-ray diffraction. Changes in the diffraction pattern of a breast tissue sample would indicate damage to the collagen structure in the ECM, which could be linked to cancer.

1.4.3 Diffraction Theory

Diffraction occurs if x-rays interact with well-ordered structures. The x-rays scattered by the structures produce a characteristic diffraction pattern of constructive and destructive interference. The observable value in a diffraction pattern is the intensity described by equation 1.1.

$$I_{Tot}(\vec{S}) = \left|F_{Tot}(\vec{S})\right|^2 = f^2(s)\left(\frac{\sin[(2N+1)\pi\vec{S}\cdot\vec{a}]}{\sin(\pi\vec{S}\cdot\vec{a})}\right)^2 \qquad (1.1)$$

The intensity, I_{Tot}, of the scattering vector, S, is a function of the number of atoms N and their location on a lattice $\vec{a}$.

For most values of $\vec{S} \cdot \vec{a}$, the value of $\sin(\pi \vec{S} \cdot \vec{a})$ lies between 0.1 and 1.0 (or –0.1 and –1.0), and $\sin[(2N+1)\pi \vec{S} \cdot \vec{a}]$ oscillates approximately between 0 and 1. Therefore, the quotient falls in the range of –10 to 10 regardless of the value of *N*.

However, when $\sin(\pi \vec{S} \cdot \vec{a})$ approaches 0, the result is different. To see this, take the limit where $\vec{S} \cdot \vec{a} \to 0$. If the series expansion of sin is used ($\sin(x)=x-x^3/3!+\ldots$) and only the first term is kept, equation 1.1 becomes:

$$F_{Tot}(\vec{S}) = f(S)\frac{(2N+1)\pi\vec{S} \cdot \vec{a}}{\pi\vec{S} \cdot \vec{a}} = (2N+1) f(S) \qquad (1.2)$$

In a crystal, *N* could be greater than10^6, so $F_{Tot}(\vec{S})$ becomes very large when $\sin(\pi\vec{S} \cdot \vec{a}) \to 0$. This occurs each time $\vec{S} \cdot \vec{a}$ approaches an integer value. Compared to this sharp intensity peak, all other values of $\vec{S} \cdot \vec{a}$ are negligible.

Only certain orientations of sample and detector, defined by the von Laue conditions, allow the intensity peaks to be observed. The von Laue conditions are described in equation 1.3:

$$\vec{S} \cdot \vec{a} = \text{h},\ \vec{S} \cdot \vec{b} = \text{k} \text{ and } \vec{S} \cdot \vec{c} = \text{l} \quad \text{where h, k and l} = 0, +/-1, +/-2, \ldots \qquad (1.3)$$

The von Laue conditions define three sets of planes, $\vec{a}$, $\vec{b}$, and $\vec{c}$, in reciprocal space. These planes are generated by successive values of the Miller indices, *h, k,* and *l*. One

set is perpendicular to $\vec{a}$ (spaced $\frac{1}{|\vec{a}|}$ apart), one is perpendicular to $\vec{b}$ (spaced $\frac{1}{|\vec{b}|}$ apart), and the other is perpendicular to $\vec{c}$ (spaced $\frac{1}{|\vec{c}|}$ apart). The intersection points of these three planes form the reciprocal lattice.

The von Laue conditions can be used to derive Bragg's Law for scattering. Three dimensional lattice planes exist that intersect an axis every $\frac{\vec{a}}{h}$, another every $\frac{\vec{b}}{k}$, and another every $\frac{\vec{c}}{l}$. These planes can be specified by the Miller indices *h, k,* and *l*. The spacing, *d,* between the planes is inversely related to the magnitude of the indices and indicates a quantitative relationship with the reciprocal lattice.

In Bragg's law, an x-ray that interacts with a lattice plane at an angle θ is reflected at an equal angle θ. This corresponds to the deflection angle *2*θ.

Bragg's law requires that the path difference between reflected beams from adjacent lattice planes be an integer number of wavelengths:

$$2d\sin(\theta) = n\lambda \qquad (1.4)$$

n is and integer and *d* is the distance between lattice planes.

The scattering vector, $\vec{S}$, is also related to the lattice plane. Consider a lattice plane that intersects the three axes of the unit cell at $\frac{\vec{a}}{h}$, $\frac{\vec{b}}{k}$, and $\frac{\vec{c}}{l}$. The vector $\vec{r}$ goes from the origin to any point on this plane. A scattered wave that satisfies the condition $\vec{S} \cdot \vec{r} = 1$, $\vec{S} \cdot \vec{r} = 1$ defines a plane perpendicular to $\vec{S}$, because it means that the projection of $\vec{r}$ onto $\vec{S}$ is a constant. When the von Laue conditions:

$$\frac{\vec{S} \cdot \vec{a}}{h} = \frac{\vec{S} \cdot \vec{b}}{k} = \frac{\vec{S} \cdot \vec{c}}{l} = 1 \tag{1.5}$$

are satisfied, the scattering may be observed.

The length of the vector $\vec{r}_0$ defines the spacing between two adjacent planes. Since $\vec{r}_0$ and $\vec{S}$ are parallel, θ = 0 and cos (θ) = 1, so the condition $\vec{S} \cdot \vec{r}_0 = 1$ implies that

$$d = |\vec{r}_0| = \frac{1}{|\vec{S}|}. \tag{1.6}$$

From the definition of the scattering vector,

$$|\vec{S}| = \frac{2|\sin(\theta)|}{\lambda} \tag{1.7}$$

where λ= x-ray wavelength and θ= scattering angle

Therefore,

$$|\sin(\theta)| = \frac{\lambda|\vec{S}|}{2}$$

$$\sin(\theta) = \frac{\lambda}{2d}$$

$$2d\sin(\theta) = \lambda \tag{1.8}$$

Equation 1.8 is equivalent to the Bragg condition of Equation 1.4 with n=1.

The data used for this project consisted of diffraction patterns of Bragg peaks created by the well-ordered structure of collagen in breast tissue. The diffraction patterns required special analysis to reduce the complexity of the data and emphasize important features in the data. Wavelet analysis was used to accomplish these tasks.

2 WAVELETS

2.1 Introduction to wavelets

The wavelet transform is one of several types of mathematical transforms. Mathematical transforms are applied to signals to obtain information that is not readily available in the raw signal. The SAXS patterns used for this project contained information useful for classifying the patterns into classes such as “normal” and “tumor.” However, the information was not accessible in the original pattern. The wavelet transform was applied to the SAXS pattern to uncover the information that allowed for classification into the “normal” and “tumor” classes.

One of the most familiar types of transforms is the Fourier transform. Section 2.2 introduces the concept of the wavelet transform by giving an overview of wavelet applications. An intuitive description of the wavelet analysis process is given in Section 2.3. Section 2.4 compares and contrasts the wavelet transform with the Fourier transform. In order to understand the wavelet transform, it is important to be familiar with the concept of multiresolution. Multiresolution will be described in section 2.5. Section 2.6 describes the scaling function and wavelet function. The material up to this point has been presented in terms of the one-dimensional continuous wavelet transform. Section 2.7 introduces the discreet wavelet transform, and section 2.8 introduces the two-dimensional wavelet transform. Section 2.9 describes the different wavelet

families considered in this project. Finally, Section 2.10 discusses wavelet reconstruction.

2.2 Wavelet Applications

A functional objective for this project was to investigate the use of wavelets for parsing the SAXS data set. Wavelets have many advantages in isolating discrete features contained in a dataset. Detailed analysis can be achieved by using a long time interval to capture low frequency information and a shorter interval to retrieve high frequency information. In addition, Hubbard maintains wavelets "make it easier to transmit, compress, and analyze information or to extract information from surrounding 'noise' – even to do faster calculations [22]." Some of the areas where wavelets have been successfully applied include compression, de-noising, and image enhancement.

The FBI has adopted a wavelet-based method for compressing digital fingerprints. The FBI currently has over two million fingerprints in its database. With each image requiring a resolution of 500 pixels per inch with 256 grayscale intensities, a single fingerprint takes up approximately 600 kilobytes [12]. The FBI's wavelet-based compression standard has a ratio of around 20:1, which allows for efficient storage and transmission of digital fingerprint images.

Another interesting application of wavelets is in signal de-noising. Coifman and colleagues at Yale University were able to used wavelets to clean up a recording of Brahms playing his First Hungarian Dance on the piano. The original recording consisted of a re-recording of a radio broadcast of a 78 record copied from a partially melted wax cylinder, and was unrecognizable as music. However, with wavelet

analysis, much of the noise was removed, and Brahms' original playing was revealed [12].

In medical imaging, wavelets have been used for many applications including feature extraction. One example is the extraction of microcalcifications from mammograms. A mammogram can be decomposed with wavelets into high and low frequency components. Microcalcifications appear as small bright spots on a mammogram, and are represented by the high frequency components of the decomposition. By suppressing the low-frequency components when the image is reconstructed, the microcalcifications are enhanced, allowing them to be segmented from the mammograms. This was the technique used by Wang and to enhance microcalcifications [50].

For this project, the goal was to use wavelets to uncover features in the SAXS data that could be used to distinguish normal samples from tumor samples. Two-dimensional wavelet decomposition was applied to the breast tissue SAXS patterns in order to extract features that would be useful in classifying the patterns.

2.3 Wavelet Transform – Intuitive Description

The wavelet transform provides new information about a dataset that is not apparent in the original form of the dataset. For the purposes of this discussion, the original dataset is in the time domain. The wavelet transform reveals frequency information, called

scale information, as well as information about the times at which different frequencies occur, called translation information.

The wavelet transform starts with a mother wavelet. The mother wavelet is an irregular, asymmetric waveform of limited duration. There are many different mother wavelets, the choice of which depends on the application. The mother wavelet can be thought of as a "window" that is shifted along the original signal. At each location, or translation, along the signal the wavelet is correlated with the signal at that particular point. Once the wavelet has been translated to every point along the signal, the process is repeated. This time the wavelet is stretched, or dilated, to a larger scale. The wavelet scale is inversely related to frequency. A large scale corresponds to a low frequency, while a short scale corresponds to a high frequency. The process continues with larger and larger wavelet scales. The final result of the process is a map of correlation values, called wavelet coefficients, corresponding to each translation (time) and scale (frequency).

For example, consider an arbitrary waveform, and an arbitrary mother wavelet as shown on the following page in Figure 2.1. In the Step 1, the wavelet coefficient, C, is calculated by correlating the wavelet with the waveform at each position as it is translated along the waveform. In Step 2, the mother wavelet is dilated to a higher scale, and the process is repeated. The process is repeated for each scale. The result is a map of the wavelet coefficients at each scale and translation, with the light color representing large coefficients and the dark color representing small coefficients. The

coefficients at the higher scales provide information about the coarser, low frequency features of the dataset. The smallest scale coefficients provide information about the finer, high frequency information in the signal.

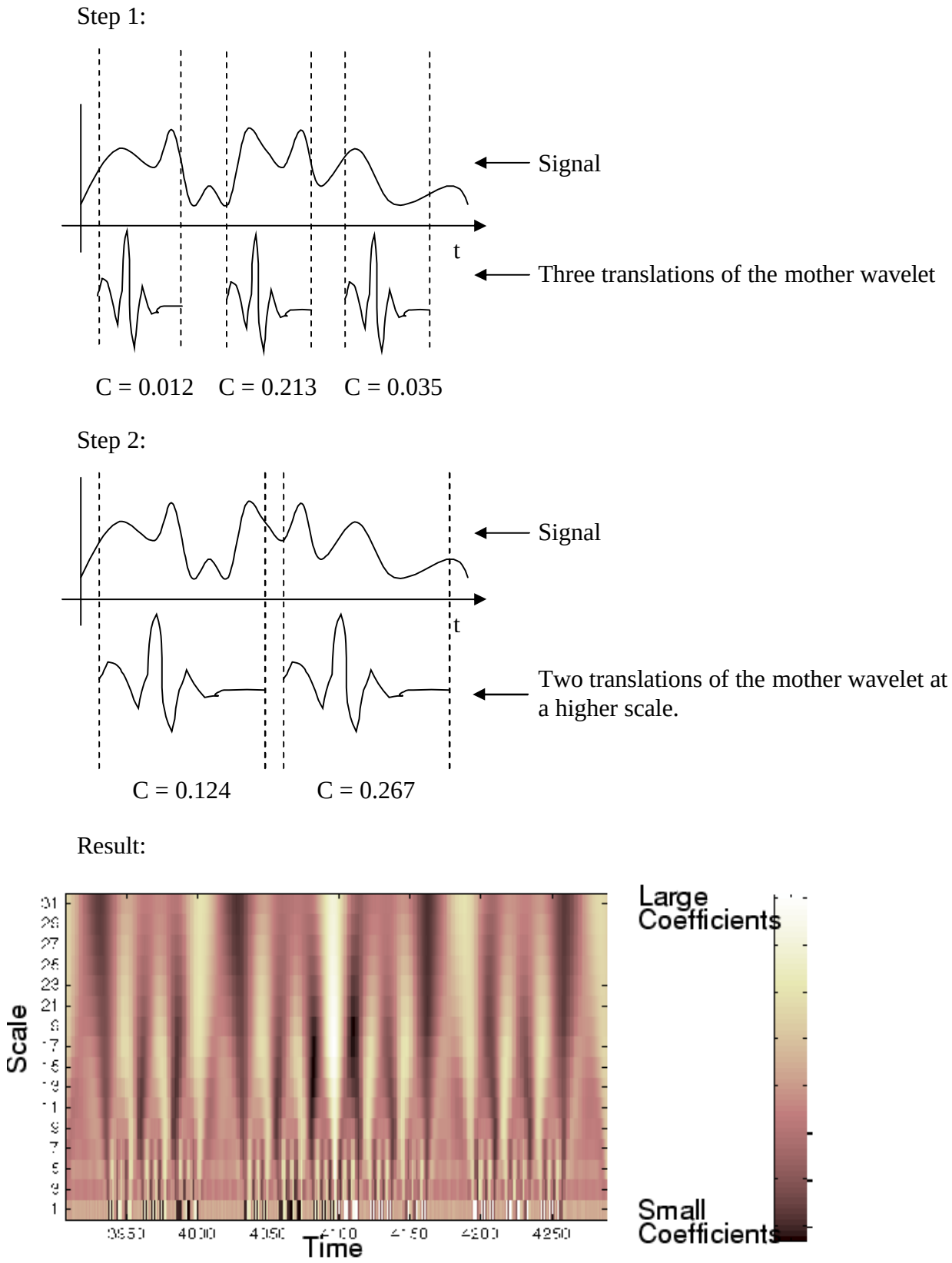


Figure 2.1 Intuitive example of a wavelet transform. Step 1 shows the correlation of the mother wavelet at three translations along the waveform, resulting in three wavelet coefficients. Step 2 shows the same process with a higher scale of the mother wavelet. The result is the Time-Scale plot representing the coefficients by colour. Adapted from [36].

2.4 Wavelet Transforms vs. Fourier Transforms

The Fourier transform is one of the best known and understood mathematical transforms. Therefore, it makes sense to discuss the similarities and differences between the Fourier transform and the wavelet transform.

The continuous 1D Fourier transform can be written as follows:

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \tag{2.1}$$

f is the signal in the time domain, *F* is the signal in the frequency domain, *t* is time, and ω is frequency.

The continuous wavelet transform can be written as follows:

$$C(s,\tau) = \int f(t) \psi^*_{s,\tau}(t) dt \tag{2.2}$$

Again, *f* is the signal in the time domain, *t* is time, C is the wavelet coefficient, *s* is scale, and τ is translation. $\psi^*_{s,\tau}$ is called the mother wavelet.

Both the Fourier transform and wavelet transform allow a temporal signal to be analyzed for its frequency content. The Fourier transform is a linear transform that represents a function with a basis of sine and cosine functions. Similarly, the wavelet transform is a linear transform that represents a function with a basis of wavelet functions. With both the Fourier transform and wavelet transform, an inverse transform returns the original signal.

There are also key differences between the Fourier transform and the wavelet transform. Fourier transforms have a single set of basis functions made up of sine and cosine functions, while wavelet transforms have an infinite number of mother wavelets. In contrast to sinusoids, wavelets are irregular, asymmetric waveforms with limited duration. This makes wavelets better suited to describe sharp changes and local features.

It was mentioned above that both Fourier transforms and wavelet transforms provide frequency information about a signal. The Fourier transform can provide frequency information about a time domain signal, or time information about a frequency spectrum, but never frequency and time information at the same time. The wavelet transform provides both time and frequency information about a signal at the same time.

The windowed Fourier transform was developed as a way to access both time and frequency information with the Fourier transform. The idea of the windowed Fourier transform was to multiply the signal by a window function that effectively isolated one section of the signal to be analyzed for its frequency content separately from the rest of the signal. The windowed Fourier transform can be written as follows:

$$WFT(t',\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[f(t) \cdot W(t-t') \right] \cdot e^{-i\omega t} dt \qquad (2.3)$$

f is the signal in the time domain and *t* is time. *WFT* is the windowed Fourier transform which represents the signal in the frequency domain, and ω is frequency. *W* is the window function, and t' is the location in time of the window function.

The windowed Fourier transform has resolution limitations due to the window size being the same for all frequencies. If a wide window is chosen, frequency resolution is good, but time resolution is poor. If a narrow window is chosen, time resolution is good, but frequency resolution is poor.

On the other hand, wavelet function windows vary in size. With wavelets, detailed analysis of localized areas of large signals can be achieved using a long time interval to acquire low frequency information and a shorter interval to retrieve high frequency information. The main advantage achieved is the ability to "zoom in" on any part of the signal.

The different window sizes in the wavelet transform lead up to the idea of multiresolution, which is discussed in the next section.

2.5 Multiresolution

Multiresolution refers to the simultaneous presence of different resolutions in a signal [51]. For example, a signal can be broken down into a smooth background with fluctuations on top of it.

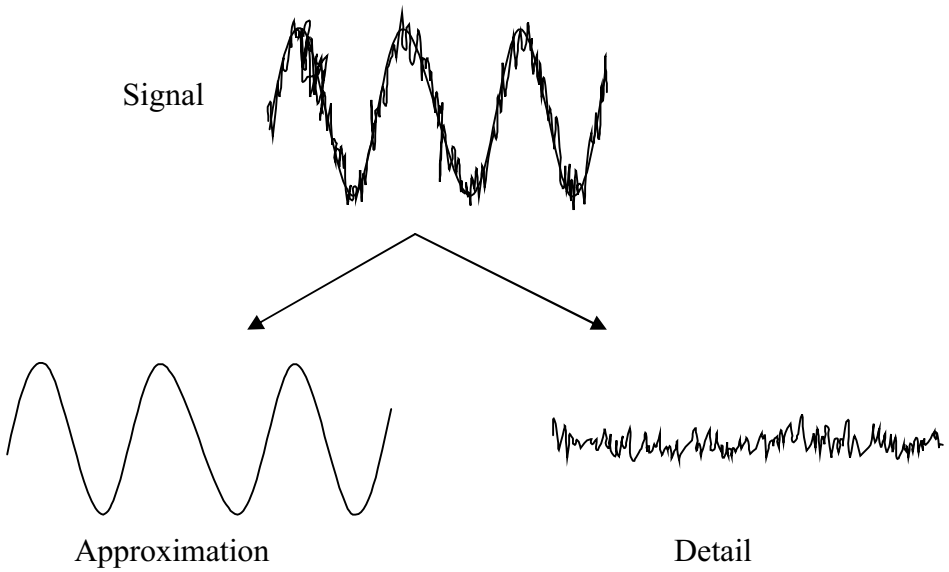


Figure 2.2 Multiresolution of a Signal. An arbitrary signal can be decomposed into approximation and detail components.

Figure 2.2 shows an arbitrary signal. The smooth background (low frequency) is known as the approximation, and the fluctuations (high frequency) are known as the details. Resolution increases as finer and finer details are added to a signal. At a lower resolution, a signal is approximated by the smooth signal, ignoring the detail fluctuations. The smooth, low frequency signal in Figure 2.2 is at a lower resolution than the original signal in Figure 2.2, because the original signal includes the high frequency detail fluctuations.

In a multiresolution analysis, a dataset is broken down into a hierarchy of several levels of approximation and detail maps. The approximation maps contain the image's low frequency information and the detail maps contain the high frequency information. At each hierarchical level, the approximation map is decomposed into descendent detail and approximation maps. In the hierarchical model, the resolution is the highest at the

lowest level. As the levels increase, details are removed, and the approximation maps are at increasingly lower resolutions.

Consider a multiresolution hierarchy with j levels. Each level contains an approximation, A_j, and details D_j. The original data can be thought of as A_0. Approximation A_1 is the low frequency components of A_0, and D_1 is the high frequency components of A_0. The detail can be thought of as the difference between A_1 and A_0 so that $A_0=A_1+D_1=A_2+D_2+D_1$ etc… Figure 2.3 illustrates the hierarchical model.

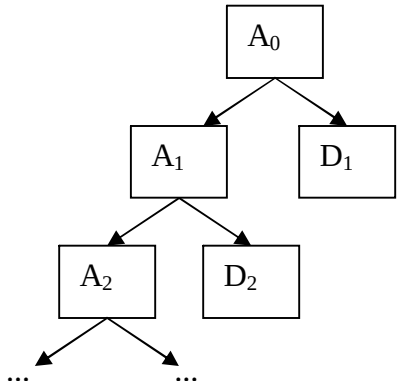


Figure 2.3 Multiresolution Hierarchical Model. A_0 is the original image. A represents the approximation maps, D represents the detail maps, and the digits represent the hierarchy level.

The mathematical description of the multiresolution hierarchy follows [51]:

$$a_0(t) = a_j(t) + \sum_{k=1}^{j} d_k(t) \tag{2.4}$$

$$a_j(t) = a_{j+1}(t) + d_{j+1} \tag{2.5}$$

$a_0(t)$ is the original signal, $a(t)$ is the approximation, $d(t)$ is the detail, and j is the level.

Equation 2.4 says that the original signal can be retrieved from the approximation map at a given level, *j*, by adding the detail maps from all previous levels to the approximation map $a_j(t)$. Equation 2.5 says that the difference between the approximation map at a given level, *j*, and the approximation map at the next level, *j+1*, is the detail map at the next level, *j+1*.

An original dataset is decomposed into a multiresolution hierarchy by exposing the dataset to a filter bank made up of a high pass and low pass filter. The original dataset is at the top of the hierarchy. To construct the next level of the hierarchy, the original dataset is exposed to a high pass filter, which removes the high frequency or detail information, and a low pass filter, which separates the low frequency or approximation information. Iteratively exposing the approximation map at each level to the high pass and low pass filter bank creates the next level of the hierarchy.

Wavelet decomposition is a type of multiresolution decomposition. The wavelet decomposition can be thought of as a filter bank is made up of a high pass component called the wavelet function, and a low pass component called the scaling function.

Figure 2.4 demonstrates what happens in frequency space during the wavelet decomposition. The scaling function at the first level is the original signal. At each subsequent level, the scaling function is split into approximation and detail information with wavelet and scaling filters.

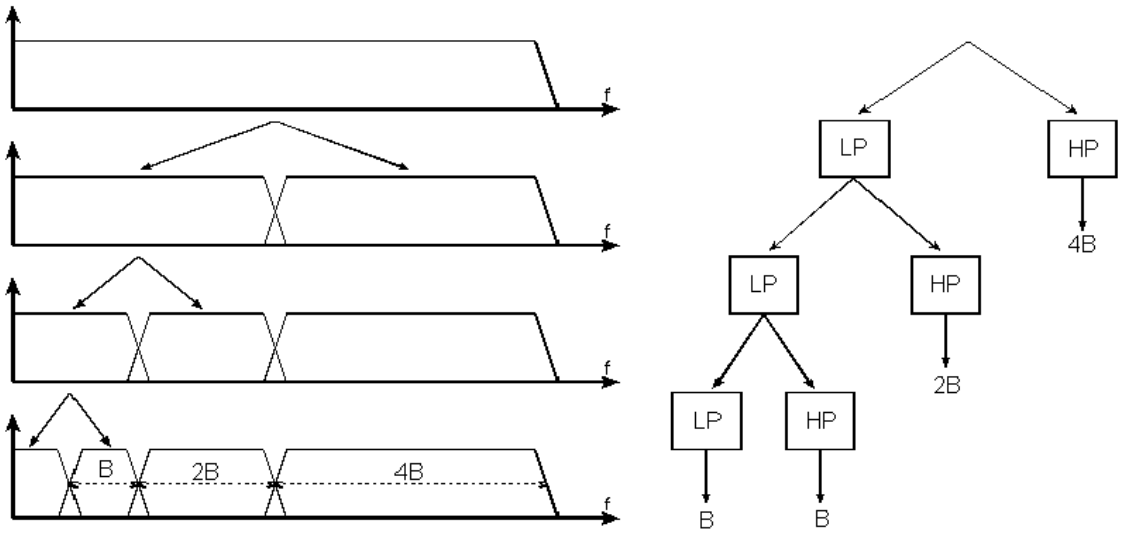


Figure 2.4 Wavelet Decomposition Example. *HP* is the high pass wavelet filter, and *LP* is the low pass scaling filter. *B, 2B,* and *4B* are the coefficients at each level. At each level the low pass approximation information is divided into another level of approximation and detail information by applying the low pass scaling filter and the high pass wavelet filter [49].

The next section describes the wavelet function and scaling function in more detail.

2.6 Wavelet Function and Scaling Function

The scaling function makes up the low pass component of the wavelet decomposition filter bank. The scaling function generates the basis functions for the approximation maps. The scaling function is defined as [51]:

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \tag{2.6}$$

The decomposition level is j, k represents the translation, and t is time. The scaling function coefficients, $\phi_{j,k}$, are the coefficients used to make the approximation maps.

As shown by equations 2.4 and 2.5, there is a relationship between the approximation maps at different levels. The relationship between the scaling functions at two successive scales, *t* and *2t* is called the dilation equation, or two-scale equation [51]:

$$\phi_{0,0}(t) = \sum_k h_k \phi_{1,k}(t) = \sqrt{2} \sum_k h_k \phi(2t - k) \tag{2.7}$$

The coefficients, h_k, are important because they make up the low pass reconstruction filter that is used to construct the approximation map from the approximation coefficients, $\phi_{j,k}$.

The wavelet function is the basis function for the detail maps in the wavelet decomposition. The wavelet function is defined as [51]:

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k) \tag{2.8}$$

Again, the scale is represented by *j*, translation by *k*, and time by *t*, and $\psi_{j,k}$ are the coefficients for the detail maps.

The wavelet function can be expressed in terms of the scaling function [51]:

$$\psi(t) = \sum_k g_k \phi_{1,k}(t) = \sqrt{2} \sum_k g_k \phi(2t - k) \tag{2.9}$$

Writing the wavelet equation in this form shows the relationship between the mother wavelet, $\psi(t)$, and the scaling function at the next scale. The coefficients, g_k, make up the high pass filter that is used with the wavelet coefficients, $\psi_{j,k}$, to reconstruct the detail maps in the wavelet decomposition.

2.7 Continuous and Discrete Wavelet Transforms

The intuitive description of the wavelet transform given in section 2.1 describes a continuous wavelet transform. The mother wavelet is continuously scaled and shifted along the data, potentially generating an infinite number of representations. This makes the continuous wavelet transform highly redundant and impractical to use. In practice, a discrete wavelet transform is used, allowing a predefined number of derivative datasets to be generated.

The discrete wavelet transform allows a signal to be sampled at discrete points, resulting in efficient computation. Discrete wavelets are scaled and translated in discrete steps [13]. This is achieved using scaling and translation integers instead of real numbers. The following is the discrete wavelet transform equation.

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j} \right) \tag{2.10}$$

As in the previous discussion, j and k are integers with j determining the scale, and k the translation. The scale describes the time domain width of the wavelet and the translation identifies the position of the wavelet with respect to the dataset. The rate of scale dilation is s_0, and the translation step magnitude is τ_0. The rate of scale dilation, together with the size of the dataset, governs the number of scales generated.

Dyadic sampling is the usual approach. At each scale the number of data points is reduced by half. Clearly, a minimum of two points is required for a representation of

the data, thus establishing the maximum number of scales. Figure 2.5 gives an example of dyadic sampling in time-scale space.

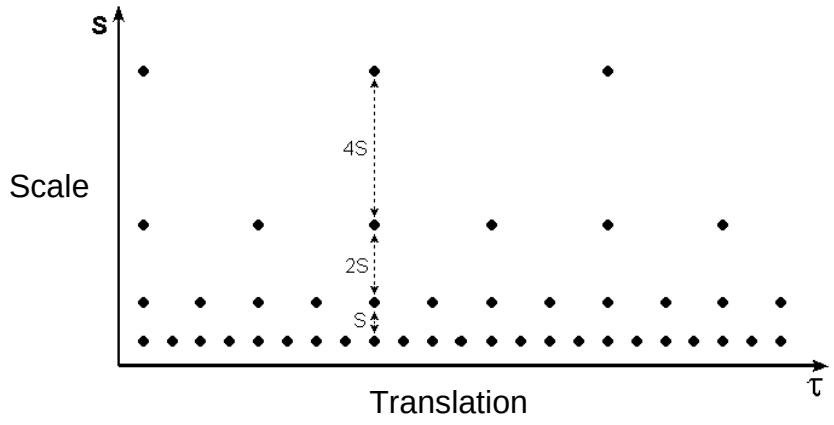


Figure 2.5 Dyadic Sampling Grid. The grid shows the location of discrete wavelets sampled on a dyadic grid [49].

2.8 Two-dimensional Wavelet Analysis

For this project, the dataset was not a one-dimensional signal, but a two-dimensional SAXS pattern. Two-dimensional wavelet analysis is based on one scaling function $\phi(x,y)=\phi(x)\phi(y)$, and three wavelets $\psi_1(x,y)=\phi(x)\psi(y)$, $\psi_2(x,y)=\psi(x)\phi(y)$, $\psi_3(x,y)=\psi(x)\psi(y)$. This section describes how the wavelet transform is actually accomplished on an image.

The wavelet and scaling function coefficients make up a transform matrix, which is applied hierarchically to the original image. The odd rows of the transform matrix

contain low pass filter information, and the even rows contain high pass filter information. Dyadic decimation occurs after each filter application. Each matrix application brings out a higher resolution of the data while smoothing the remaining data

The following algorithm describes how wavelet decomposition works on the image [36]. Figure 2.6 shows a block diagram of the process.

1 Convolution of raw data rows with the transform matrix of scaling filter and wavelet filter coefficients

2 Dyadic decimation on columns of both results

3 Convolution of columns from both results with the scaling filter and the wavelet filter coefficients

4 Dyadic decimation on rows of all four results

5 Final results are: low pass followed by low pass – Approximation
low pass followed by high pass – Horizontal Detail
high pass followed by low pass – Vertical Detail
high pass followed by high pass – Diagonal Detail

6 For the next level decomposition, repeat steps 1 through 5 using the Approximation data for step 1

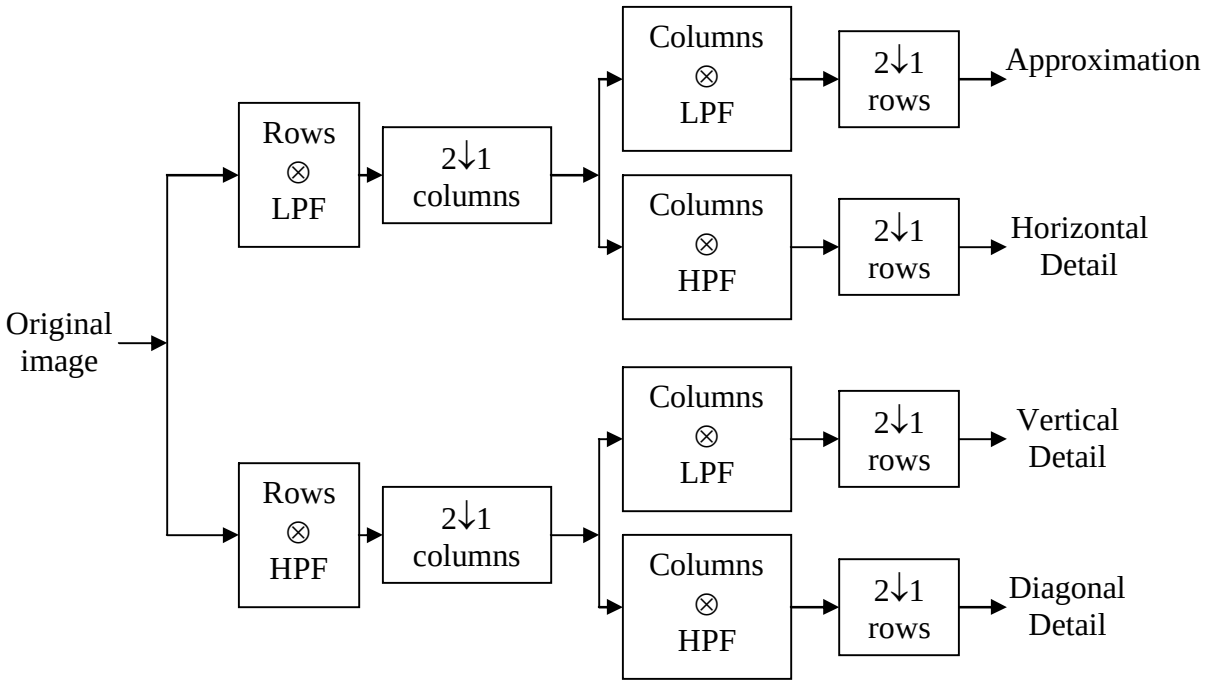


Figure 2.6 Block diagram of Two Dimensional Wavelet Transform. The diagram shows how the transform matrix is applied to an image to achieve a wavelet transform. ⊗ represents convolution and 2↓1 represents dyadic decimation [36].

On the following page, Figure 2.7 shows the result of completing the wavelet decomposition on an image to eight levels. The hierarchy includes an approximation map as well as horizontal, vertical and diagonal detail maps for each decomposition level. The eight-level two-dimensional wavelet decomposition results in thirty-two wavelet maps.

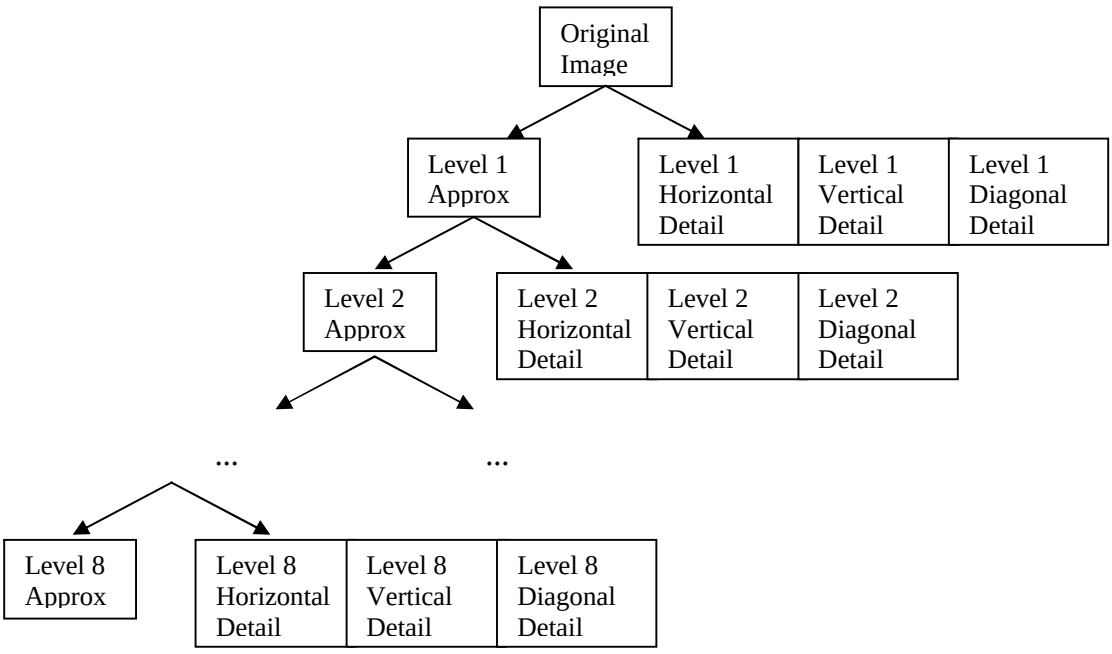


Figure 2.7 Hierarchy of Approximation and Detail Maps. This diagram shows the wavelet feature maps produced in a two dimensional wavelet decomposition. 8 levels x 4 views = 32 feature maps per decomposition.

2.9 Wavelet Families

Different wavelets are designed for different tasks, and are referred to as wavelet families. The choice of wavelet basis is an important topic in wavelet decomposition methods. The choice of an appropriate basis can have a significant effect on the results of the analysis.

There are several well-know families of wavelet bases such as the Daubechies family, and the biorthogonal family. New wavelet bases have also been created for specific applications. For example, Lemaur investigated a new wavelet basis that improved the classification of microcalcifications in mammograms [30].

It is not often obvious which wavelet basis would best suit the application [30]. For this project, a literature search was conducted, but failed to provide any insight into appropriate wavelet bases for the analysis of SAXS data. Therefore, several of the classical wavelet bases were tested. This section reviews some of the commonly used wavelet families.

The simplest wavelet basis is the Haar wavelet. It is discontinuous and resembles a step function. The Haar wavelet is the same as the Db1 wavelet of the Daubechies family. The Daubechies family was invented by Ingrid Daubechies and is made up of compactly supported orthonormal wavelets. Compact support means that the wavelet basis function is non-zero for a finite interval. Compact support allows for the efficient representation of signals with localized features. Orthonormal refers to the way the coefficients are calculated. A basis is orthonormal if the mother wavelet is chosen so that [49]:

$$\int \psi_{j,k}(t)\psi^{*}_{m,n}(t)dt = \begin{cases} 1 & \text{if } j=m \text{ and } k=n \\ 0 & \end{cases} \tag{2.11}$$

$$\int \psi_{j,k}(t)\psi^{*}_{m,n}(t)dt = 0 \quad \text{otherwise} \tag{2.12}$$

The biorthogonal family of wavelets also has the property of compact support. However, they are biorthogonal rather than orthogonal. One wavelet is used for decomposition, and a different wavelet is used for reconstruction. The advantage of biorthogonal wavelets is that they allow symmetry and exact image reconstruction.

For this project, several wavelet bases were tested from each of the Haar, Debauchies, and biorthogonal families. Figure 2.8 shows examples of wavelet bases functions from each of these families.

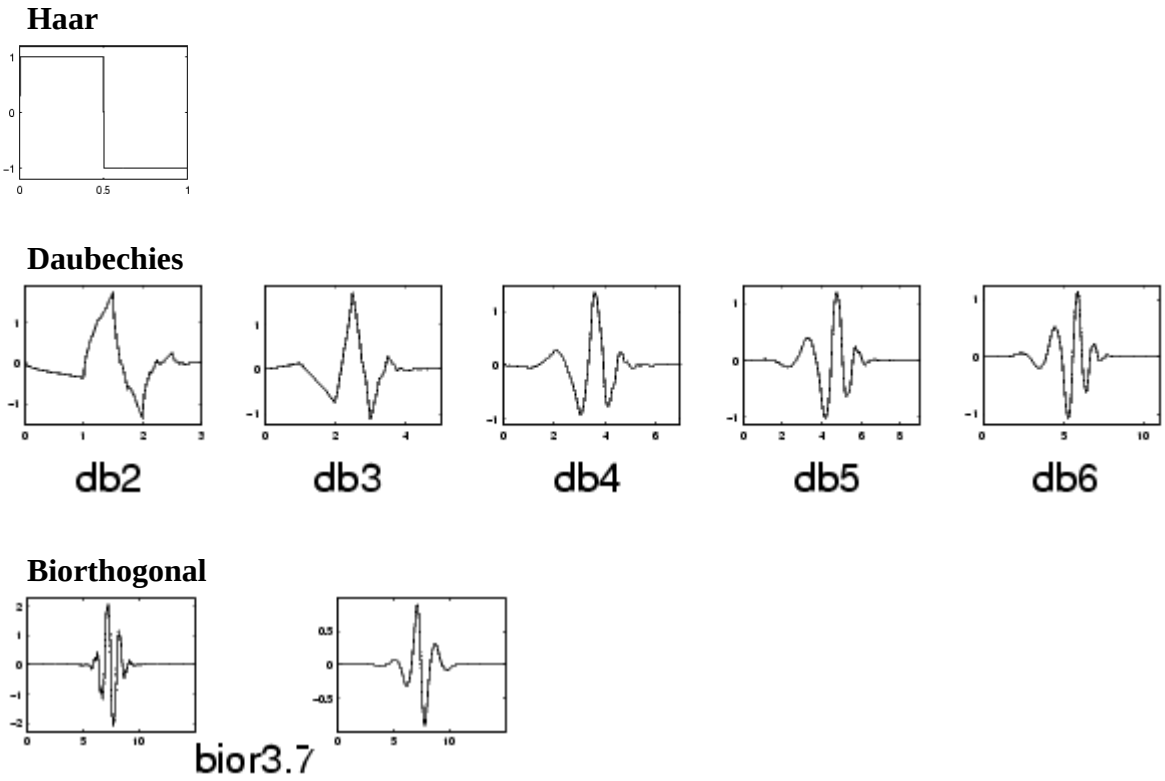


Figure 2.8 Wavelet Bases. These are examples of wavelet bases from the Haar, Daubechies and Biorthogonal wavelet families. The first Biorthogonal wavelet is used for decomposition and the second for reconstruction. (Adapted from [36].)

2.10 Reconstruction

In some applications, such as image de-noising, it is necessary to regenerate the original dataset from a subset of wavelet maps. Reconstruction is possible with the discrete wavelet transform. The necessary and sufficient condition for reconstruction is that the

energy of the wavelet coefficients must lie between two positive bounds, *A* and *B*, called a frame [13].

$$A\|f\|^2 \le \sum_{j,k} \left|\langle f, \psi_{j,k} \rangle\right|^2 \le B\|f\|^2 \tag{2.13}$$

$||f||^2$ is the energy of *f(t)*, $A>0$, $B<\infty$, and *A* and *B* are independent of *f(t)*. When $A=B$, discrete wavelets behave like an orthonormal basis. When $A \neq B$, a dual frame is necessary for reconstruction.

The inverse wavelet transform is [44]:

$$f(t) = \sum_{j,k} \gamma(j,k) \psi_{j,k}(t) \tag{2.14}$$

This means that a signal can be reconstructed by summing the orthonormal wavelet basis functions weighted by the wavelet transform coefficients.

3 PATTERN RECOGNITION

3.1 Common Pattern Recognition Approaches

The goal of pattern recognition is to classify patterns into classes. Classification may be either supervised, where the input pattern is labeled as a member of a known class, or unsupervised, where the input pattern is unlabeled.

Some common approaches to pattern recognition problems are: template matching, syntactic pattern recognition, neural networks, and statistical pattern recognition [35]. In template matching, a prototype of the pattern to be identified is compared to the input pattern and a measure of similarity, such as correlation, is used for classification. While template matching is one of the simplest methods of pattern recognition, it is computationally expensive, and it is not robust in cases where the input patterns have large intra-class variation [14]. Syntactic pattern recognition is a hierarchical approach

to pattern recognition. Each pattern class can be described by a set of primitives, which are combined using rules called a grammar. The syntactic approach is not effective with noisy data, where primitives are difficult to distinguish [14]. Neural networks are parallel processing systems with complex interconnections between elements. One advantage of neural networks is that they are able to learn complex non-linear input-output relationships from the data. Although the implementations are different, most neural network architectures are based on statistical pattern recognition [35]. A statistical pattern recognition approach was chosen for this project. In statistical pattern recognition systems, each pattern is represented with *d* features, and is viewed as a point in a *d*-dimensional feature space. The goal is to find features that give a high level of separation between classes, so that a decision boundary can be constructed between the different classes.

CAD is a popular area of research, and a wide range of approaches applying pattern recognition to breast cancer detection can be found. Some common techniques include morphological operations, statistical texture analysis techniques, fractal theory, and fuzzy logic. According to the current literature, the combination of wavelet-based pattern recognition with SAXS images is a unique approach to automated breast cancer detection. This section describes the common approaches found in the literature.

3.1.1 Morphological Operations

Morphological operations analyze images in terms of shape. The value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image

with its neighbors. Structuring elements are used to choose the size and shape of the neighborhood, making it possible to construct morphological operations that are sensitive to specific shapes in the input image. Morphological operations can be used to perform common image processing tasks, such as contrast enhancement, noise removal, and segmentation. The advantage of morphological approaches is that they make use of geometric features of images, and the shape of these features is not lost in processing.

Nagel used a morphological technique to enhance microcalcifications. Potential microcalcifications were extracted with global thresholding based on an erosion operator and local adaptive thresholding. False positives were then eliminated by texture analysis, and the remaining candidates were classified with a non-linear clustering algorithm. In an independent database of 50 images, at a sensitivity of 83%, the average number of false positive (FP) detections per image reported was 0.8 [38].

The main disadvantage of morphological operations is that they require a priori information about the feature to be enhanced in order to create an effective structuring element [9]. In the case of enhancing microcalcifications in an image, this information would not be difficult to acquire, because microcalcifications have some known properties such as approximate shape, size, and texture. However, with the diffraction data used for this project, the interesting features were not confined to certain regions of the image. Identifying a suitable morphological operation would have been difficult because geometrical shape features suitable for being extracted with morphological operations were not present.

3.1.2 Fuzzy Logic

Classic logical systems are based on Boolean logic, which assumes that every element is either a member or a non-member of a given set. Fuzzy logic extends Boolean logic to handle approximate information and uncertainty in decision-making. To express imprecision, fuzzy logic introduces a set membership function that maps elements to values between zero and one. The value indicates the "degree" to which an element belongs to a set. A membership value of zero indicates that the element is entirely outside the set, and a value of one indicates that the element lies entirely inside a given set. Any value between the two extremes indicates a degree of partial membership to the set. Fuzzy logic provides a simple way to make a definite conclusion based on vague, imprecise, or noisy data. Fuzzy logic handles the uncertainty and imprecision found in mammograms, such as indistinct borders, ill-defined shapes and varying densities.

Cheng demonstrated a system based on fuzzy logic. Images were first fuzzified, using the fuzzy entropy prinicipal and fuzzy set theory to automatically determine the fuzzy membership function. Next the image was enhanced using a homogeneity measurement. The Laplacian of a Gaussian filter was used to find local maxima in the image that represented the possible microcalcifications. Thresholds for selecting the actual microcalcifications out of the candidates were determined using a neural network. The receiver operating characteristic curve showed that the method achieved

an accuracy of greater than 97% true positive rate with the false positive rate of three clusters per image [8].

The major advantage of the fuzzy logic method described above was the ability to detect microcalcifications in breasts of various densities. One disadvantage was that the determination of a fuzzy membership function was complex [9]. Another disadvantage was that the method relied on many parameters that would require a priori information such as the threshold value for classification, the enhancement parameters, and the image normalization parameters. For mammograms, the fuzzy logic method was advantageous because it allowed regions to be identified even with inherent uncertainty in the data. The fuzzy logic approach was not chosen for this project because the diffraction data eliminated the problems of density variation and uncertainty due to indistinct borders and ill-defined shapes that would be present in a mammogram.

3.1.3 Fractal Theory

Images can be modeled by deterministic fractals, which are attractors of sets of two-dimensional affine transformations [32]. Fractals are ideal for modeling texture in images that show a high degree of self-similarity. Background breast structure in a mammogram has a high degree of local self-similarity, so it is an excellent candidate to be modeled by a fractal approach.

Li used a deterministic fractal approach to enhance microcalcifications in mammograms. Deterministic fractals were used to model the breast background

structures. The difference of the model image and the actual image was calculated in order to enhance microcalcifications present in the mammogram. The approach was qualitatively compared to wavelet and morphological approaches, and it was determined that the fractal approach removed the background structure more effectively than the other two approaches, but did not preserve the overall shape of the microcalcifications as well as the wavelet approach [32].

The advantage of this approach was that it was very effective for removing the breast background structure from an image. The effectiveness of the fractal approach was due to the fact that microcalcifications do not demonstrate the same type of self-similar structure that background breast tissue does, allowing microcalcifications to be segmented from the image. For this project, the fractal approach was not implemented because segmentation was not the goal. Another reason that this approach was not considered further is that it was very computationally intensive [9].

3.1.4 Statistical or Texture Analysis Methods

Statistical or texture analysis approaches look at local texture in mammograms and measure features such as correlation, contrast, and entropy in order to distinguish normal tissue from cancerous structures. Statistical features based on texture introduce new information in addition to intensity information in an image. Texture features are suited to the analysis of mammograms because the different tissue structures in the mammogram display different statistical texture features.

Mendez used bilateral subtraction to identify asymmetries between left and right breast images. A threshold was applied to obtain a binary image of suspicious areas. A region growing algorithm was used to define the suspicious regions, and size and eccentricity tests were used to eliminate false-positives. When tested on 70 mammograms, the approach achieved a true-positive rate of 71% with an average number of 0.67 false positives per image. Using ROC analysis, the value of Az=0.667 was achieved [34].

Chan used texture features derived from spatial Grey Level Dependency (SGLD) matrices to classify mammograms containing microcalcifications as normal or benign. Several texture features were extracted from the SGLD matrices: correlation, entropy, angle of second moment, inertia, inverse difference moment, sum average, sum entropy, difference entropy. A feed-forward back propagating neural network was used as the classifier. The classifier achieved an area under the ROC curve of 0.88. Additionally, 11 of the 28 benign cases were correctly identified (39% specificity) without missing any malignant cases (100% sensitivity) [7].

The advantage to using statistical texture features is that they provide additional information about a region of an image. They can be used to characterize a whole image, not just a small region within an image. Because of these advantages, statistical texture features were used for this project, although wavelet decomposition maps were used rather than SGLD matrices.

3.1.5 Wavelet Approaches

Wavelet approaches have become popular choices for the analysis of mammograms. One of the main advantages of wavelet decomposition is the scale space localization. Images can be decomposed into well-localized components that allow for "zooming in" on interesting features. The image is decomposed into coefficients that describe the image in terms of scale and orientation. Different scales and orientations of these details often reveal information that is obscured in the original image. Important details can be enhanced, and uninteresting or noisy details can be eliminated during reconstruction. Another advantage of wavelet decomposition is that the multiple decomposition scales allow for flexibility with respect to image resolution. The input image does not require a specific resolution for the method to be effective. A detailed description of wavelet decomposition can be found in Chapter 2.

Many of the wavelet-based approaches in the literature focus on the enhancement or identification of microcalcifications in mammograms, such as Wang [50], and Yu [53]. A few methods also attempted to identify masses, such as Laine [29].

Wang used wavelet decomposition to enhance microcalcifications in mammograms. The mammograms were decomposed using the Debauchies 4 and Debauchies 20 wavelets. Since microcalcifications corresponded to high-frequency components of the decomposition, the low-frequency components were compressed and the mammogram was reconstructed using only the components containing high frequencies. Intensity thresholding was then used to segment the microcalcifications from the image [50].

Yu used a wavelet system to detect microcalcifications in mammograms. Wavelet features combined with gray level statistical features were used to segment potential microcalcifications. A neural network classified a set of 31 features extracted from the potential individual microcalcification objects to reduce false positives. The method was applied to a database of 40 mammograms containing 105 clusters of microcalcifications. Results showed that a 90% mean true positive detection rate was achieved at the cost of 0.5 false positive per image [53].

Laine used wavelet transforms to enhance microcalcifications as well as tumor masses in mammograms. The wavelet transform provided a hierarchy of multiscale images, which localized important image information at different spatial frequencies. At each level of resolution multiscale edges were used to emphasize the desired features. The results were qualitatively compared with traditional methods used for image enhancement such as unsharp masking and adaptive histogram equalization and showed that the wavelet-based processing algorithms were superior [29].

Zhen used a wavelet transform in combination with several other artificial intelligent techniques for detection of masses in mammograms. First, fractal dimension analysis was used to determine the approximate locations of the suspicious regions in the mammogram. In the second step, a discrete wavelet transform based segmentation algorithm was used to remove image noise caused by veins and fibers. In the classification step, features were generated from the segmentation step and a binary

decision tree was used to label the suspicious areas. The technique was tested with 322 mammograms from the Mammographic Image Analysis Society Database, and resulted in a sensitivity of 97.3% with 3.92 false positives per image [54].

The wavelet methods described above demonstrate the effectiveness of wavelet analysis on mammograms. In each of the cases, wavelets were used to enhance and segment microcalcifications or tumor masses. For this project, SAXS data was used instead of conventional mammograms. Unlike mammograms, the important features in the diffraction patterns were not geometric shape features. Therefore, the task of segmentation was not the goal.

3.1.6 Original Approach Used in this Project

For this project, a unique approach was designed to use features arrived at through wavelet analysis in conjunction with a naïve Bayesian classifier to detect abnormalities in the SAXS patterns.

A statistical measure based on wavelet decomposition was used as the feature for classification. Each image was decomposed into wavelet maps representing eight levels with horizontal detail, vertical detail, diagonal detail, and image approximation maps. The average intensity of each map was calculated and used as the feature. Different combinations of the average wavelet map intensity features were used for classification

with a naïve Bayesian classifier. A complete description of the approach can be found in the methods and results sections.

3.2 Pattern Recognition System Tasks

The three main tasks of a pattern recognition system are: preprocessing, feature extraction, and classification. Preprocessing is used to define the pattern of interest, and could include segmentation of the pattern from the background, noise removal, normalization, or any other operation that would help to define the pattern. The feature extractor measures different attributes of the pattern that can then be used for classification. The classification step uses the measured features to assign the pattern to one of the classes based on a statistical decision rule. A generic pattern recognition system is depicted in Figure 3.1.

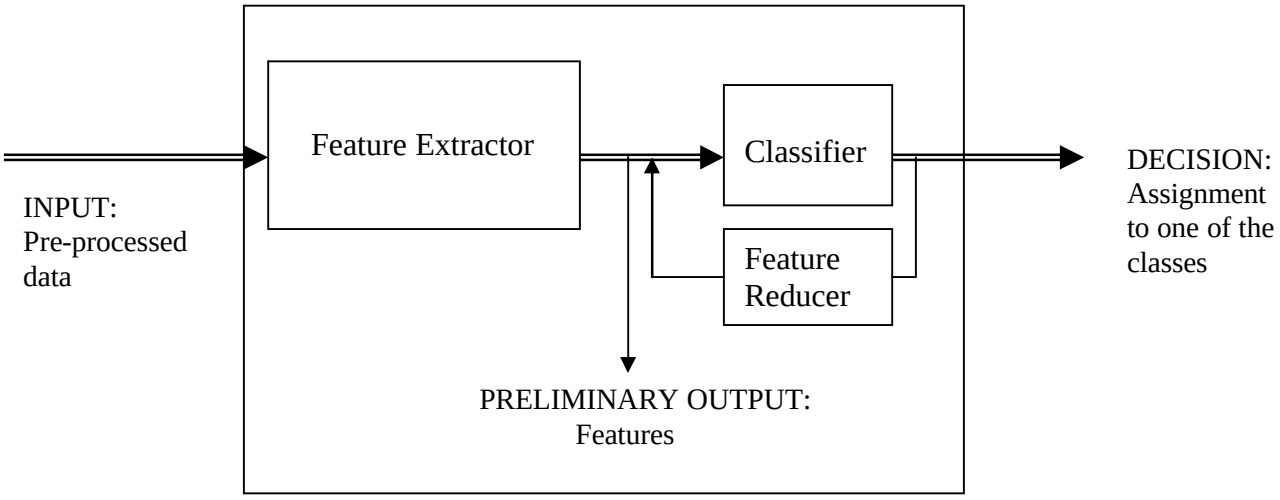


Figure 3.1 Generic pattern recognition system. This figure shows the components required for a pattern recognition system, as well as the flow of the data through the system, and the output of the system.

3.3 Pattern Recognition System Performance

The performance of a pattern recognition system depends on the number of features used, the number of classes present, and the number of training samples available. The "curse of dimensionality" is a well-known problem that refers to the performance of the classifier being degraded if a small number of training samples are used with a large number of features. Although there is no exact rule to define the relationship between the performance of the classifier, the sample size, and the number features, a general guideline has been established [24].

$$\frac{n}{d} > 10 \tag{3.1}$$

n is the number of training samples per class, and d is the number of features used. In other words, there should be at least ten times as many training samples per class as there are features.

4 METHODS

4.1 Research Methods Used

This chapter introduces the research methods that were used for this project. The project involved the implementation of a semi-automatic pattern recognition system. The system is summarized using the block diagram depicted in Figure 4.1. Section 4.2 describes the feature extractor block, Section 4.3 describes the feature reducer, and Section 4.4 describes the classifier. Section 4.5 describes the data set that was used, as well as the pre-processing steps involved.

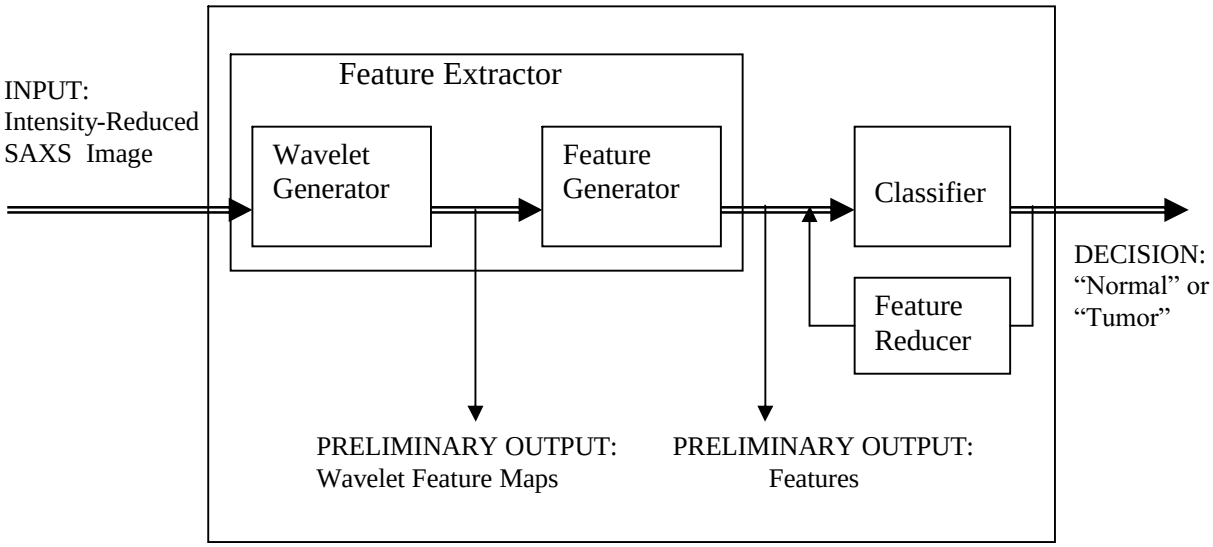


Figure 4.1 Automated Pattern Recognition Tool Designed for this Project. This figure shows the specific pattern recognition tool designed for this project including the wavelet generator, with preliminary output of wavelet feature maps, as well as the feature generator with preliminary output of features derived from the maps. The classifier includes a feature reducer. The final output is a label of either "normal" or "tumor".

4.2 Feature Extractor

The feature extractor consists of two components: the wavelet generator, which decomposes the input image into wavelet maps, and the feature generator, which measures certain aspects of the wavelet map to be used as features.

4.2.1 Wavelet Generator

The first step in the feature extraction procedure is to decompose the input image using wavelet analysis. Wavelet analysis decomposes an image into a hierarchical set of approximation and detail wavelet maps. The approximation map contains the image's low frequency information, while the detail maps contain the high frequency information. At each level, the wavelet transform is applied to the approximation map, breaking it down into further approximation and detail maps.

Each image was decomposed to eight levels, with each level resulting in four views: horizontal detail, vertical detail, diagonal detail, and the approximated image. For a detailed description of the wavelet decomposition process, refer to section 2.8. Figure 4.2 shows the four views of the wavelet decomposition at scale level 6 for a normal sample. To keep track of the decomposition maps, they were labeled as follows: a letter represented the view (*A*=approximation, *H*=horizontal detail, *V*=vertical detail, and *D*=diagonal detail); a digit from 1 to 8 represented the level. For example, the map at level 6 in the vertical view was labeled *V6*.

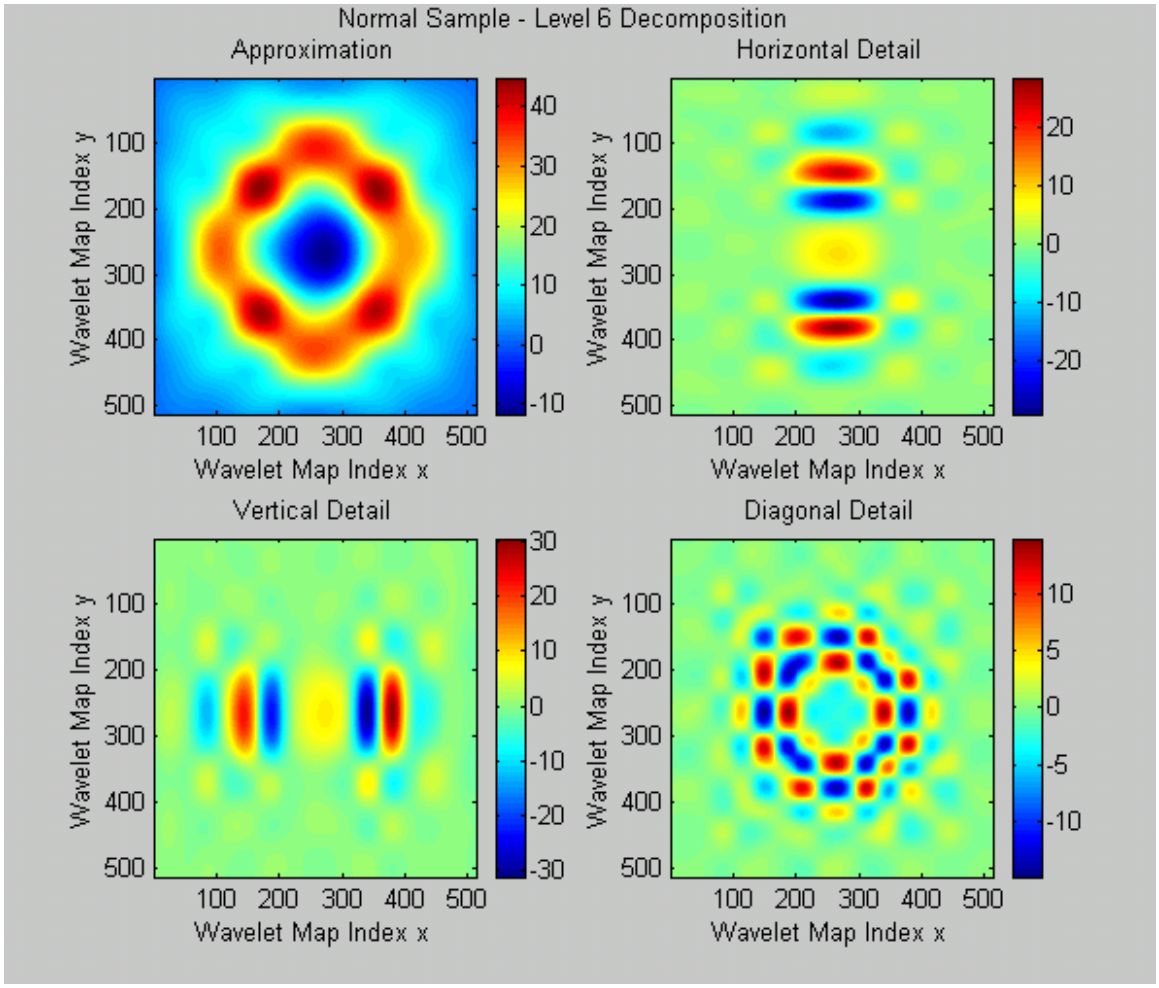


Figure 4.2 Normal Sample Wavelet Decomposition Maps. This figure shows the wavelet decomposition of a normal sample at decomposition Level 6 in all 4 views: approximation, horizontal detail, vertical detail, and diagonal detail. a digit from 1 to 8 represented the level. The color scale represents strong or weak correlation with the wavelet basis at each position on the map.

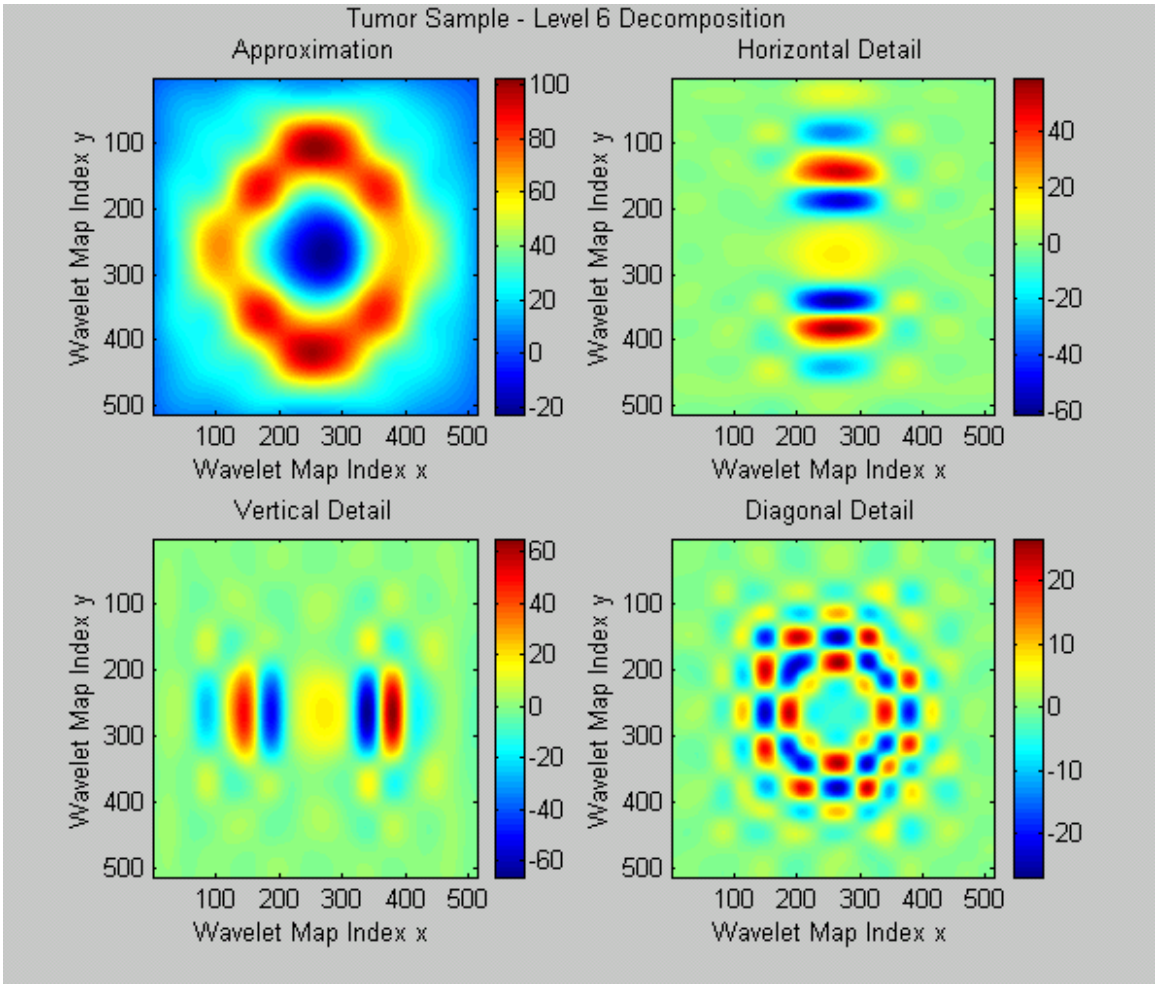


Figure 4.3 Tumor Sample Wavelet Decomposition Maps. This figure shows the wavelet decomposition of a tumor sample at decomposition Level 6 in all 4 views: approximation, horizontal detail, vertical detail, and diagonal detail. a digit from 1 to 8 represented the level. The color scale represents strong or weak correlation with the wavelet basis at each position on the map.

4.2.1.1 Wavelet Basis Selection

As discussed in section 2.9, the wavelet maps could be created using one of many different wavelet bases. Wavelet bases respond differently to different data sets. No literature was found describing the use of wavelets with a dataset similar to the one used in this project. Therefore, the best wavelet basis was determined by iteratively testing an arbitrary set of bases.

The wavelet bases tested for this project were the Haar, Debauchies 2, Debauchies 4, Debauchies 8, Biorthogonal 2.2, Biorthogonal 3.7, and Biorthogonal 6.8 wavelet bases. Descriptions of these bases can be found in section 2.7. The Haar wavelet was chosen because it is the most simple wavelet basis. The Debauchies family bases were chosen because of their properties of compact support and orthonormality. The Debauchies family of wavelets is represented by the abbreviation "Db". The Biorthogonal wavelets were chosen for their property of exact reconstruction. The Biorthogonal wavelets are represented by the abbreviation, "Bior".

Figures 4.4 through 4.0 show the level 6 wavelet decomposition maps that resulted for the same normal sample shown in Figure 4.2 when the different wavelet bases were introduced.

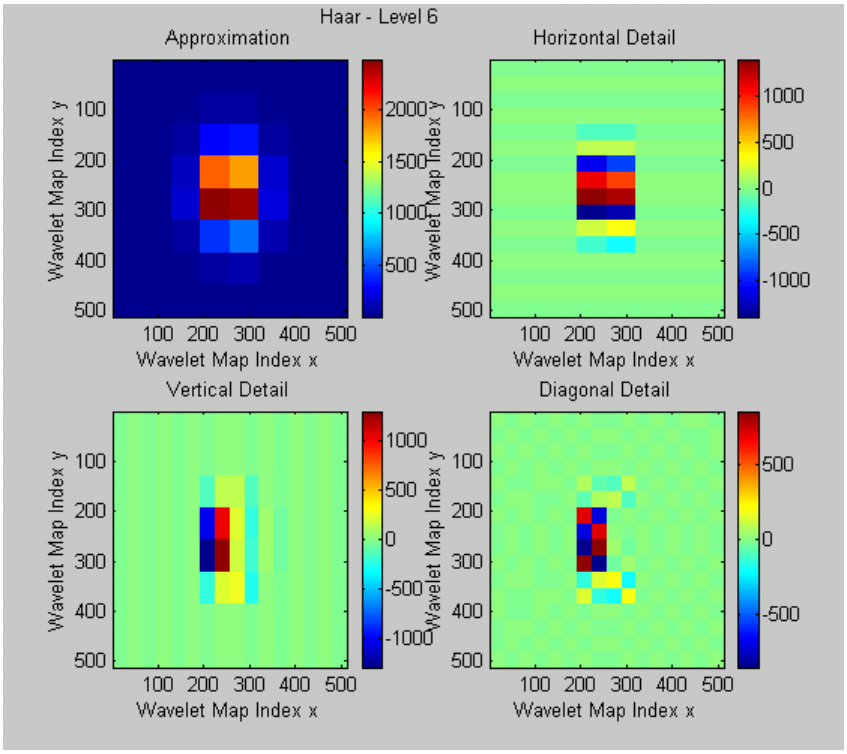


Figure 4.4 Haar Basis Decomposition Maps. This figure shows the wavelet decomposition of a normal sample at decomposition Level 6 in all 4 views using the Haar basis.

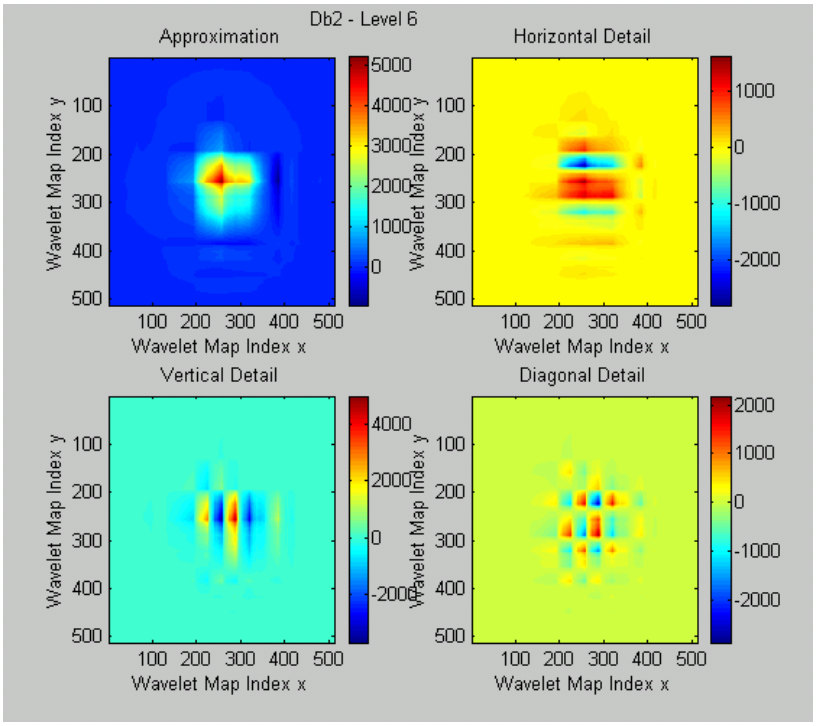


Figure 4.5 Db2 Basis Decomposition Maps. This figure shows the wavelet decomposition of a normal sample at decomposition Level 6 in all 4 views using the Db2 basis.

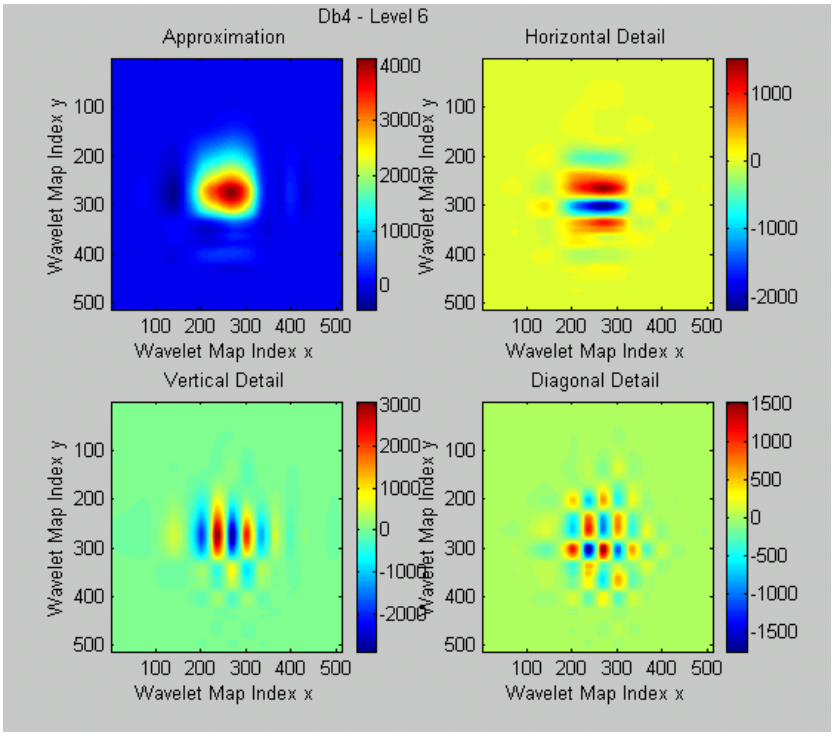


Figure 4.6 Db4 Basis Decomposition Maps. This figure shows the wavelet decomposition of a normal sample at decomposition Level 6 in all 4 views using the Db4 basis.

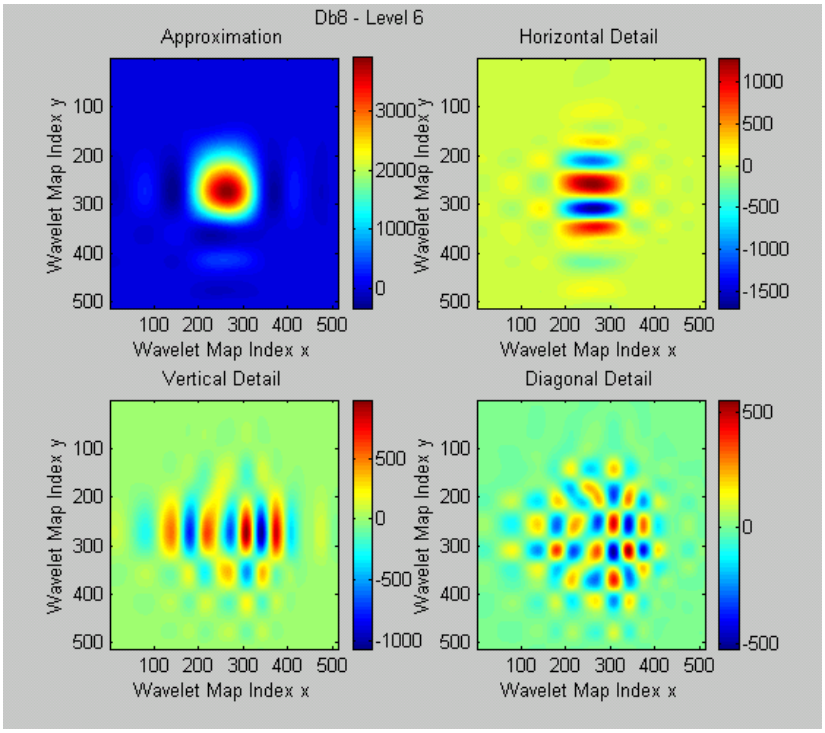


Figure 4.7 Db8 Basis Decomposition Maps. This figure shows the wavelet decomposition of a normal sample at decomposition Level 6 in all 4 views using the Db8 basis.

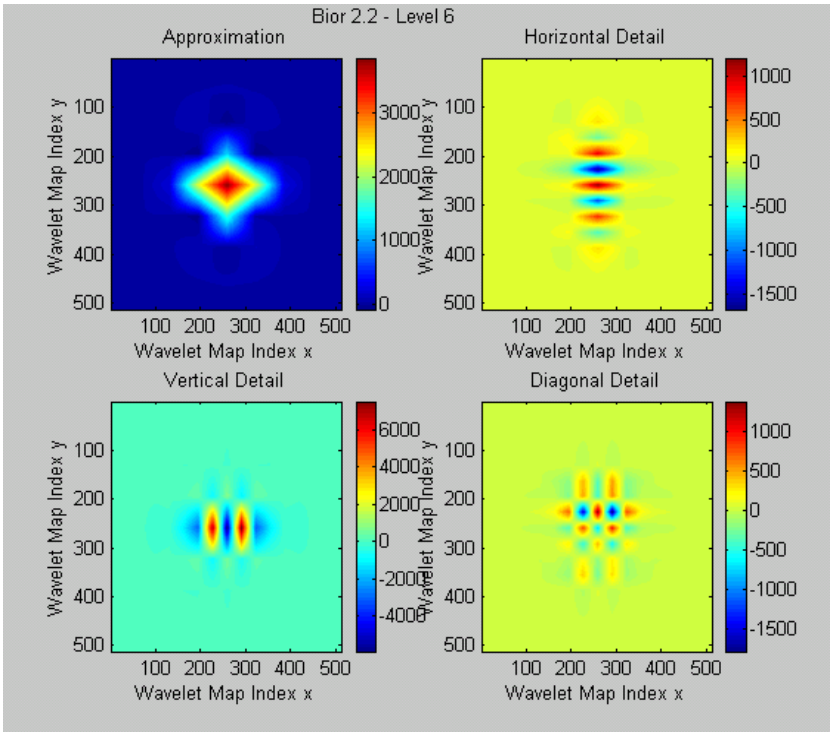


Figure 4.8 Bior2.2 Basis Decomposition Maps. This figure shows the wavelet decomposition of a normal sample at decomposition Level 6 in all 4 views using the Bior2.2 basis.

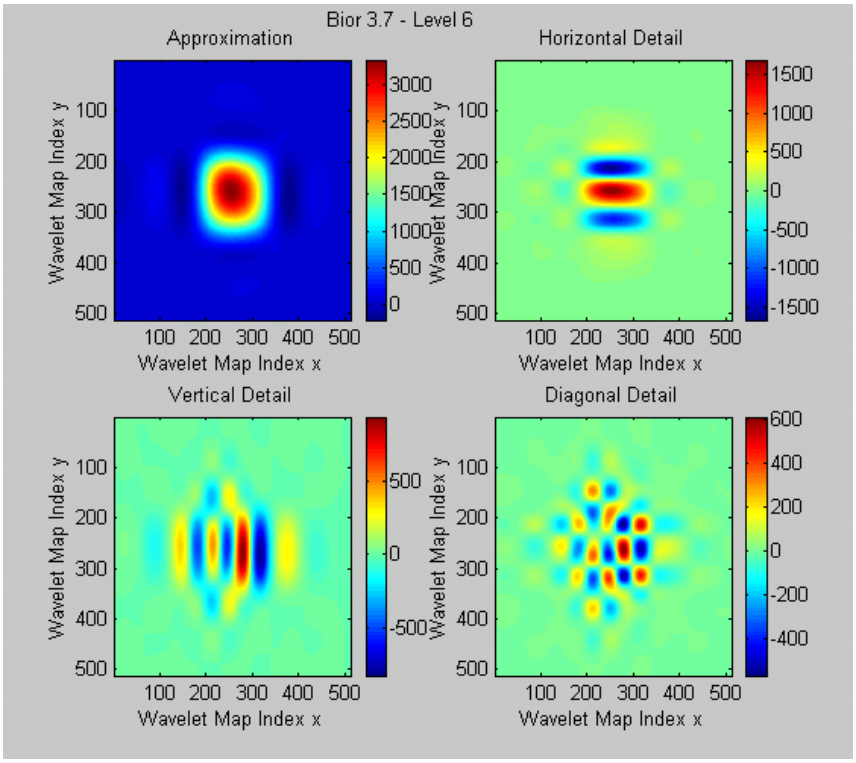


Figure 4.9 Bior3.7 Basis Decomposition Maps. This figure shows the wavelet decomposition of a normal sample at decomposition Level 6 in all 4 views using the Bior3.7 basis.

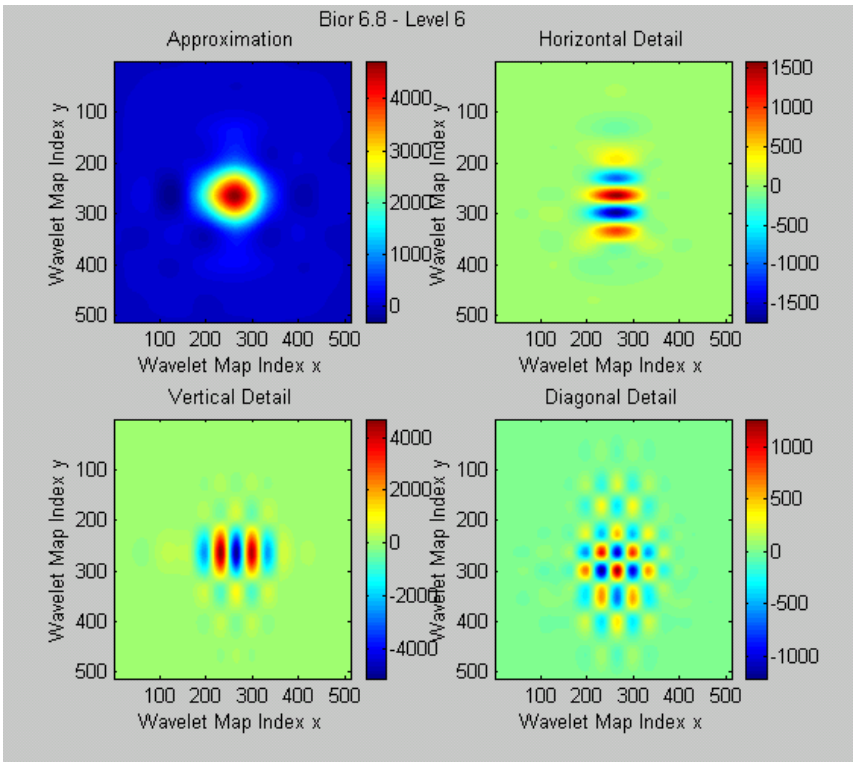


Figure 4.10 Bior6.8 Basis Decomposition Maps. This figure shows the wavelet decomposition of a normal sample at decomposition Level 6 in all 4 views using the Bior6.8 basis.

4.2.2 Feature Generator

After the wavelet maps were generated, features useful for classification were extracted from the maps. Three measurements were taken: wavelet map intensity, wavelet map standard deviation, and Fourier transform of wavelet map. These features were calculated for every wavelet map generated in the wavelet decomposition. One additional feature was considered that did not depend on wavelet analysis. The diffraction image intensity profile feature was measured directly from the original image. A detailed description of each of these features follows.

4.2.2.1 Wavelet Map Mean Intensity

The mean intensity of each wavelet map provided a simple way to represent the texture of each wavelet map. Therefore, the mean intensity of the map, $\langle I \rangle$, was then calculated as shown in equation 4.1, and used as a feature.

$$\langle I \rangle = \frac{\sum_{x=1}^{m} \sum_{y=1}^{n} I_{x,y}}{mn} \tag{4.1}$$

where m and n are the image dimensions.

4.2.2.2 Wavelet Map Standard Deviation

The standard deviation could be used to measure the surface energy of the map. Therefore, the standard deviation of each wavelet map was calculated according to equation 4.2 and used as a feature.

$$\sigma = \sqrt{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(I_{mn} - \langle I \rangle\right)^2} \tag{4.2}$$

4.2.2.3 Wavelet Map Fourier Transform

The Fourier transform of the wavelet map was tested as a way to access frequency information inherent in the wavelet map. Several steps were involved in calculating a feature with the Fourier transform of the wavelet map. The two-dimensional fast Fourier transform was computed for each wavelet map, as shown in equation 4.3.

$$F(u,v) = \frac{1}{mn}\sum_{x=1}^{m}\sum_{y=1}^{n} f(x,y)e^{-2\pi i(xu+yv)/mn} \tag{4.3}$$

m and *n* are the image dimensions, *u* and *v* are the image coordinates in the frequency domain, *x* and *y* are the image coordinates in the time domain.

Next, the power spectrum was calculated.

$$P_{m,n} = \sqrt{r_{mn}^{\ 2} + i_{mn}^{\ 2}} \tag{4.4}$$

Where *m* and *n* are the image dimensions, *r* is the real part of the Fourier transform, and *i* is the imaginary part.

An image was formed from the logarithm of the power spectrum $(r^2 + i^2)^{1/2}$. Figure 4.11 shows this image for the level six, horizontal view wavelet map from figure 4.2.

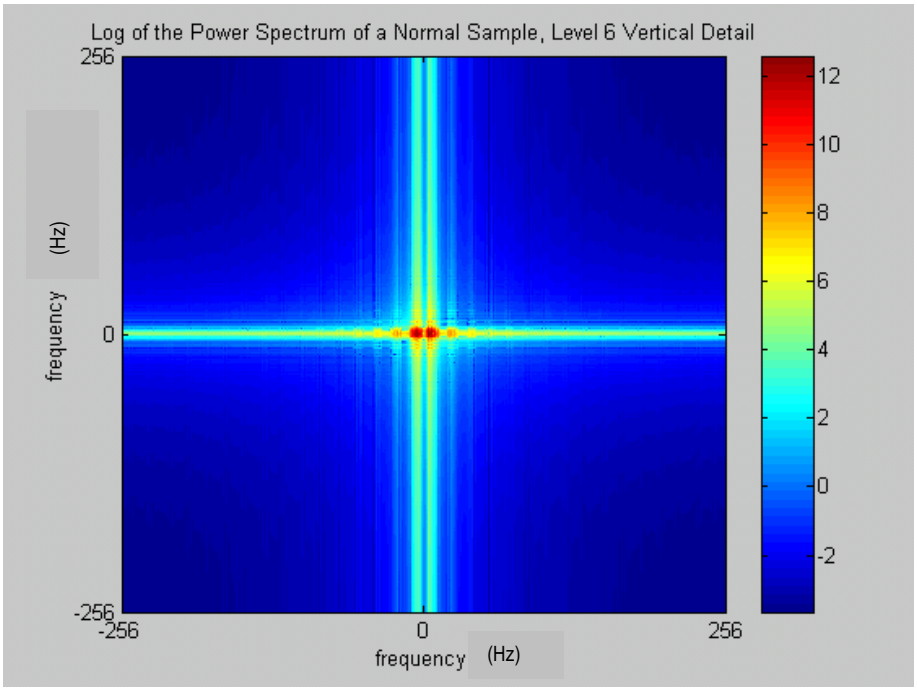


Figure 4.11 Fourier Transform of a Wavelet Decomposition Map. This figure shows Fourier transform of the Level 6, Horizontal view wavelet map depicted in Figure 4.2.

Figure 4.12 is a plot of the intensity of the first 50 points along this line. In order to use this as a feature, the intensities were normalized as a percentage of the maximum intensity, and the area under the Fourier transform intensity profile curve was calculated.

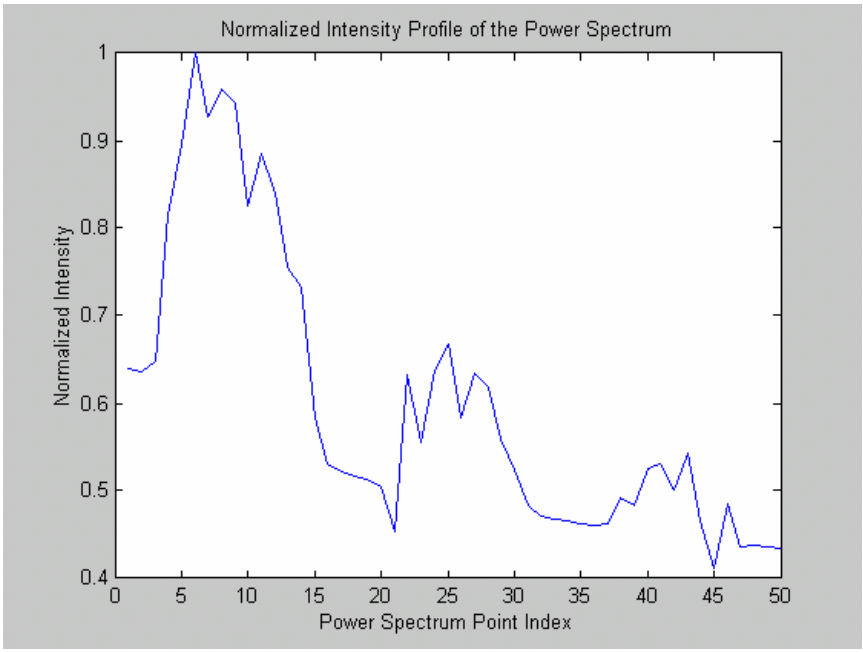


Figure 4.12 Intensity Profile of the Fourier Transform of the Wavelet Decomposition Map. This shows the normalized intensity profile of the first 50 points along the centerline between the 1st and 4th quadrants of the Fourier transform from Figure 4.11.

4.2.2.4 Diffraction Image Intensity Profile

The diffraction image intensity profile feature was the only one that did not involve a wavelet transform. This feature was used as a way to represent the original data set. The feature was obtained by calculating a one-dimensional curve to represent the intensities of the 1st quadrant of the original image. Because of the symmetry of the image, the intensities in the first quadrant were sufficient to represent the image. The first quadrant intensities were approximated by running a diagonal line from the center of the image to the top right corner, and summing the intensities along the x and y directions for each point on the line. Figure 4.13 illustrates how the intensity at each point along the diagonal line was calculated.

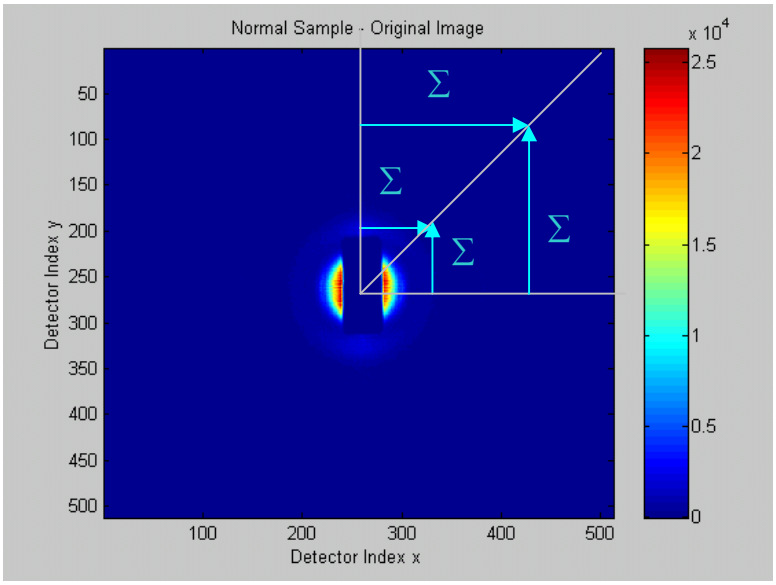


Figure 4.13 Diffraction Image Intensity Profile Illustration. The intensity at each point along the diagonal line was determined by summing all of the intensities along the x and y directions at that point. The symmetry of the image allowed for the intensity to be represented by the first quadrant intensity profile.

Figure 4.14 shows an example of the normalized intensity profile. Normalization was necessary for comparing sample to sample. The profile contains peaks at the 0^{th}, 1^{st}, 3^{rd} … orders. In the low angle region (pixels 250 to 350) information on collagen fiber spacing would be featured. In the wide angle region (beyond 400 pixels) information smaller than 15 angstroms would be featured.

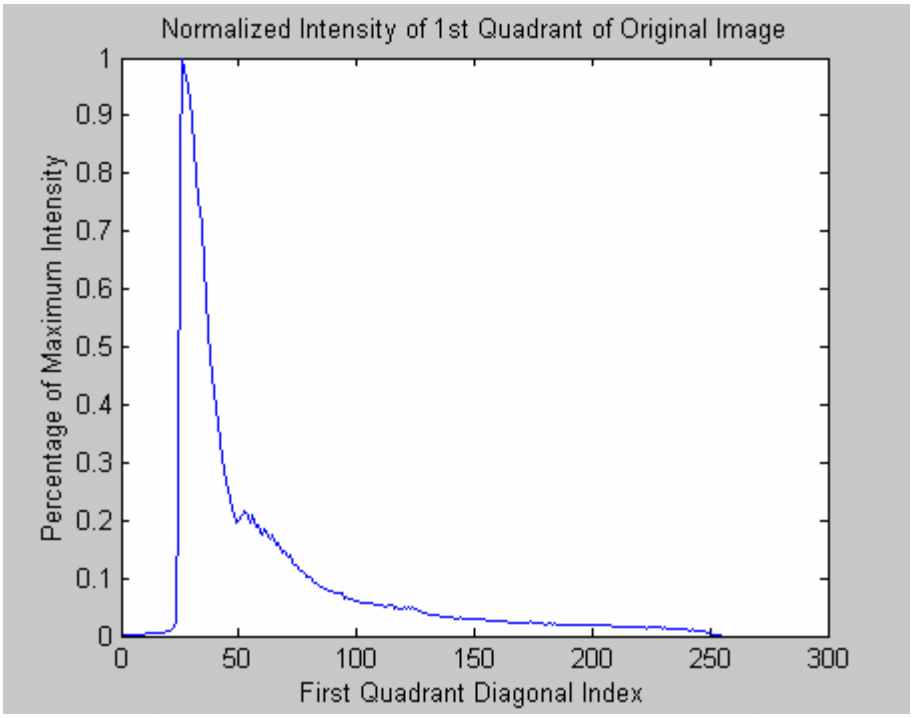


Figure 4.14 Normalized Diffraction Image Intensity Profile. This figure shows the intensity profile of the first quadrant of the diffraction image normalized as a percentage of the maximum intensity.

The diffraction image intensity profile was calculated according to equation 4.5 and used to measure the diffraction image intensity feature.

$$A = \sum_{x=1}^{m} f(x) \tag{4.5}$$

Where *f(x)* is the curve, and *m* is the number of points along the curve.

The diffraction image intensity profile feature was calculated as a benchmark to compare the wavelet features to. It was necessary in determining whether the wavelet features provided any advantage over features arrived at from the original image.

4.3 Feature Reducer

For every input image, thirty-two wavelet features were generated. As discussed in Section 3.3, the ratio of samples per class to features should be at least ten. For this project, the normal class consisted of twenty samples, and the tumor class consisted of twenty-two samples, allowing for a maximum of two features.

In order to reduce the number of features, a variation of an exhaustive search was used. In an exhaustive search, given a set of features, every possible combination of features is tested, and the subset that leads to the lowest classification error is selected. For this project, every combination of the eight levels for each of the four views, and every combination of four views for the eight levels was tested. The combinations with the highest classification rates and fewest features were kept.

4.4 Classifier

4.4.1 Naïve Bayesian Classification

A naïve Bayesian classifier was implemented for the classification step of the pattern recognition tool. The feature vector, $\boldsymbol{x}$, is an element of a d-dimensional Euclidean space $\boldsymbol{R}^d$ called the feature space. $\{c_1, \ldots, c_j\}$ is a finite set of the c possible classes. For this classifier, there are three classes: normal, tumor, and fibro adenoma.

$P(\boldsymbol{x}| c_j)$ is the state conditional probability density function for $\boldsymbol{x}$ conditioned on c_j being the true state of nature. $P(c_j)$ describes the prior probability of the state of nature being

c_j. $P(c_j|\boldsymbol{x})$ is the posterior probability of the state of nature being c_j given the particular feature vector, $\boldsymbol{x}$.

In the naïve Bayesian classifier, the posterior probability can be computed using Bayes rule:

$$\mathrm{P}(c_j \mid \mathbf{x}) = \frac{P(\mathbf{x} \mid c_j)P(c_j)}{P(\mathbf{x})} \tag{4.6}$$

where

$$\mathrm{P}(\mathbf{x}) = \sum_{j=1}^{c} P(\mathbf{x} \mid c_j)P(c_j) \tag{4.7}$$

The unknown sample is assigned to the class that maximizes the posterior probability, $P(c_j|\boldsymbol{x})$.

4.4.2 Determining the Probabilities

The main challenge with Bayesian classifiers is determining the prior probabilities $P(c_j)$, and the state conditional probability densities $P(\boldsymbol{x}|\ c_j)$. In most pattern recognition problems, these probabilities are unknown. General knowledge about the problem must be combined with training data in order to design the classifier.

If information about the problem is known, the prior probabilities, $P(c_j)$, are determined based on that information. For this classifier, information on the probabilities of each of the three classes was not available. Therefore, the prior probabilities for the normal, tumor, and fibro adenoma classes were assumed to be equal

The state conditional probability densities, $P(\mathbf{x}| c_j)$, are more difficult to estimate. Approaches such as maximum-likelihood estimation and Bayesian learning make the assumption that the form of the density is known, and then attempt to estimate the density parameters. For example, it could be assumed that the density is a multivariate normal distribution, and then the mean and covariance matrix would be estimated based on the training data. The problem with these methods is that the probability densities found in most problems rarely fit the common parametric forms. For example, all of the classical parametric densities have a single local maximum, and many practical problems involve densities with more than one local maximum.

The naïve Bayesian classifier addresses the problem of determining the state conditional probability densities by assuming that they can be represented by a naïve Bayesian network. The structure of the network is shown in Figure 4.15.

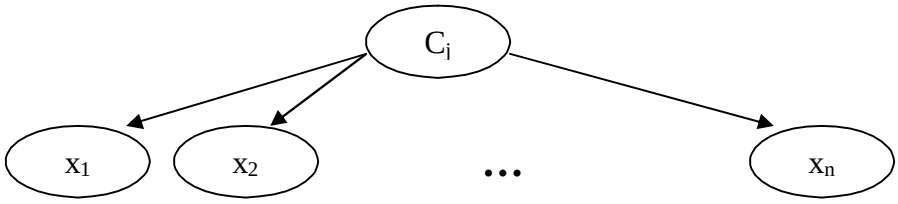


Figure 4.15 Naïve Bayesian Network Structure. C_j is the class label. x_1 through x_n represent the feature vector.

The naïve Bayes network imposes the assumption of independence between the features, given a particular class, c_j. The conditional probabilities for each feature x_i, given each of the three classes, c_i, are learned from the training data.

As shown in equation 4.6, Bayes rule is then used to compute the probability of each class c_j given the particular instance of the feature vector $x_1 \ldots x_n$. The sample is assigned to the class with the highest posterior probability [17].

It is noted frequently in the literature that the performance of naïve Bayesian classifiers is surprising, because of the strong independence assumption that must be made. In most cases, the assumption of independence between the features is unrealistic, as there are usually correlations between features. However, in spite of the unrealistic independence assumption, the naïve Bayesian classifier is one of the most effective classifiers [17].

4.4.3 Training Scheme

When dealing with large amounts of data, the data are usually divided into two equal groups, one for training and one for testing. For this project, the leave one out training scheme was used because the data set was relatively small. The leave one out scheme sets aside one sample for testing, and all other samples are used for training. The procedure is repeated for each of the samples in the data set to determine the percentage of correctly classified samples, or classification rate.

4.5 Data

SAXS data from breast biopsy tissues were collected at the Synchrotron Radiation Source at Daresbury, England, under the direction of Robert Lewis [31]. The literature suggests that invasive tumors in the breast tissue alter collagen structure. While the details of these changes have not been completely resolved, it appears that fundamental changes in collagen fiber spacing occur [31].

4.5.1 Experimental Protocol

The details of the experimental protocol have been previously published [31]. In brief, breast biopsy samples (~ 1 mm diameter, and 3mm to 20mm in length) were placed in glass capillary tubes. The tubes were oscillated vertically through the beam over the length of the specimen to average tissue inhomogeneities. The beam size was ~0.5x0.5mm at the sample and had an energy of 8 keV. The sample to detector distance was 6.25m, and a 200x200 mm^2 imaging multiwire proportional counter operated at

512x512 pixels was used. The configuration allowed length scales of 70 to 1390nm to be sampled. Exposure times were 300 seconds [31].

4.5.2 Pre-processing

In diffraction data most of the intensity is concentrated at low diffraction angles. This area is expected to contain information on structures with length scales 100 – 200 Å. As described by Lewis in [31], the axial D spacing of the molecules within the collagen fibrils (649 Å) resulted in intensity maxima in the diffraction patterns. The 3rd and 5th orders were the most apparent. In order to preferentially sample this region, the low angle intensities were cropped from the dataset.

As a simple intensity reduction technique, circular masks with different sized radii were applied to the center of the diffraction pattern. The mask was used to set all intensities within a given radius to the background level of the image. The background level of each image was obtained by taking the average intensity of a 50 x 50 pixel square in the upper left corner of each image.

Figures 4.16 and 4.17 show examples of an normal and tumor SAXS images before the high intensities concentrated at the center of the image were removed.

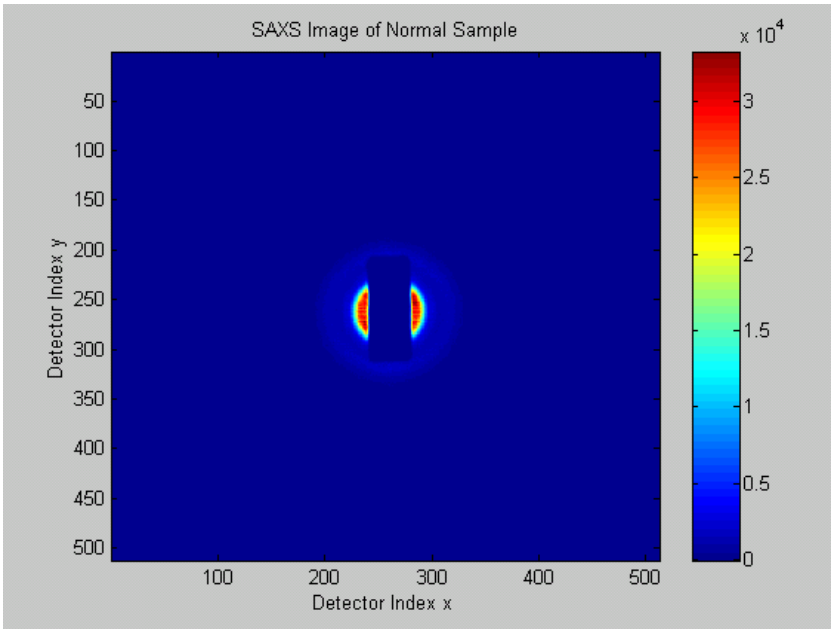


Figure 4.16 SAXS Pattern of a Normal Sample. The rectangular central shape is caused by the beam stop that absorbs non-diffracted x-rays. The vast majority of diffracted intensity is located in the low-angle region.

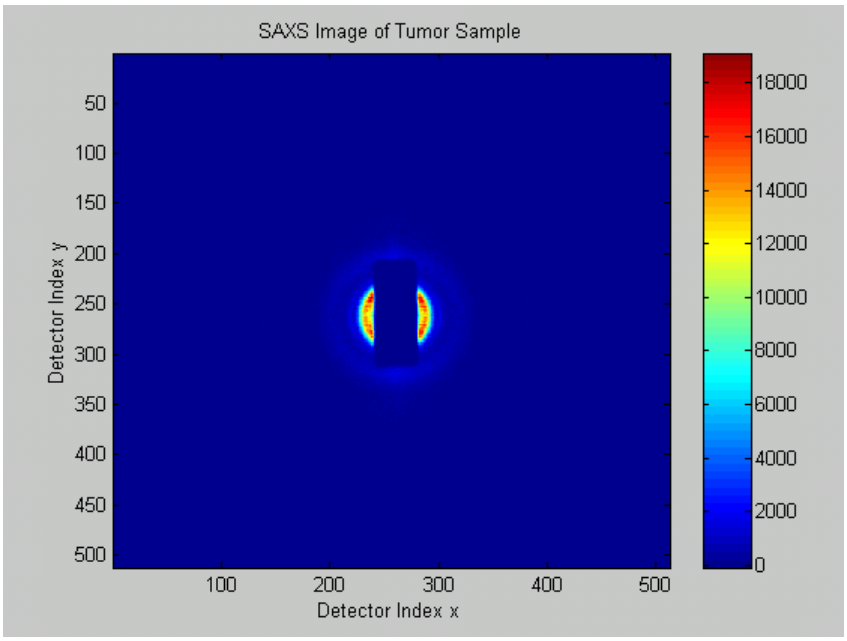


Figure 4.17 SAXS Pattern of a Tumor Sample. The rectangular central shape is caused by the beam stop that absorbs non-diffracted x-rays. The vast majority of diffracted intensity is located in the low-angle region.

Figure 4.18 shows the same sample as in Figure 4.16 with the high intensities at the center "cut out" with a mask radius of 115 pixels. Because the high central intensities are cut out in Figure 4.18, detail is revealed that was obscured in Figure 4.16.

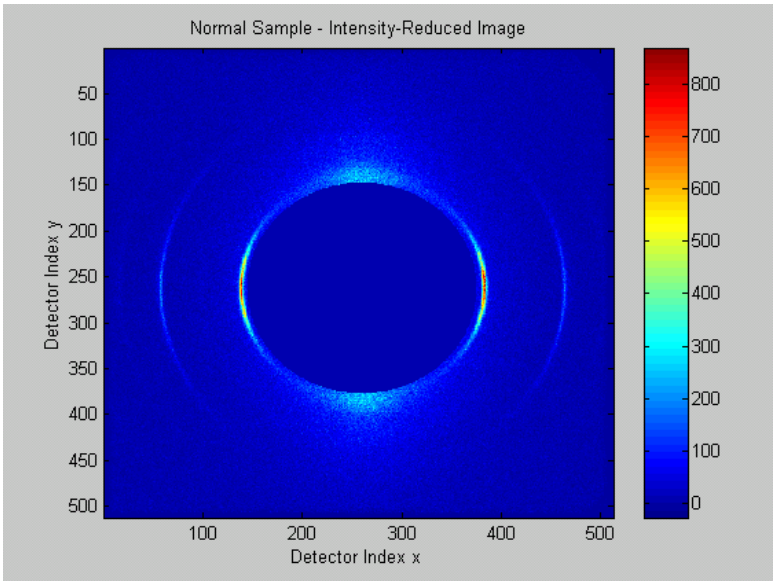


Figure 4.18 Intensity Reduction on Normal SAXS Image. The normal SAXS image from Figure 4.16 with the high center intensities cut out with a mask radius of 115 pixels.

Several different sized radii were tested. However, masks with radii sizes between 50 and 190 pixels were the most effective. These radii emphasized the third and fifth order intensity maxima in the diffraction pattern. Radii of 50 and 100 pixels included both the 3rd and 5th order maxima, while the radius of 190 pixels isolated the 5th order maxima.

Figures 4.19 through 4.21 show average intensity profiles of a SAXS image. The intensity profile shows the average intensity of a strip 10 pixels high along the center line of the SAXS pattern, starting at the center of the pattern and ranging to the right

hand edge of the pattern. Figure 4.19 shows the intensity profile when a 50 pixel radius is cut out of the pattern. The 3rd and 5th orders are visible, but the intensity is still much higher closer to the center of the pattern (or the left hand side of the intensity profile chart.)

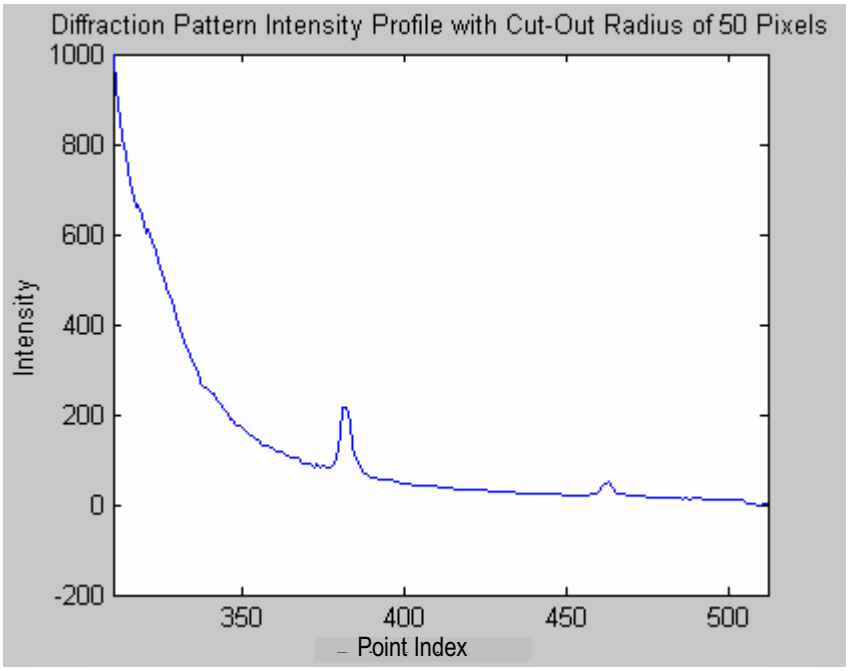


Figure 4.19 Diffraction Pattern Intensity Profile with 50 Pixel Mask Radius.

Figure 4.20 shows the intensity profile achieved with a mask radius of 100 pixels. The 3rd and 5th orders are much more dominant in this figure than in figure 4.19.

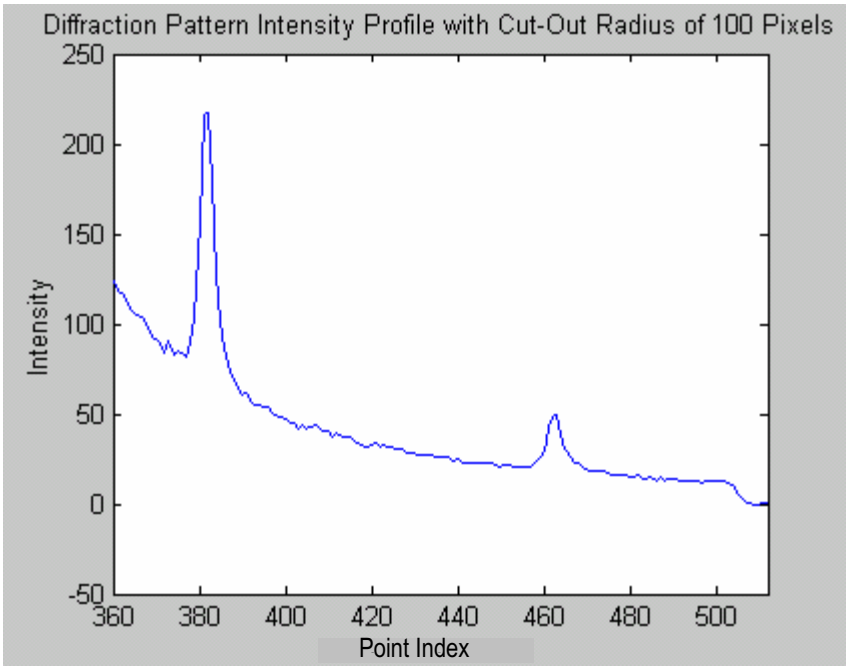


Figure 4.20 Diffraction Pattern Intensity Profile with 100 Pixel Mask Radius.

Figure 4.21 shows the intensity profile with a mask radius of 190 pixels. In this case only the 5th order was included.

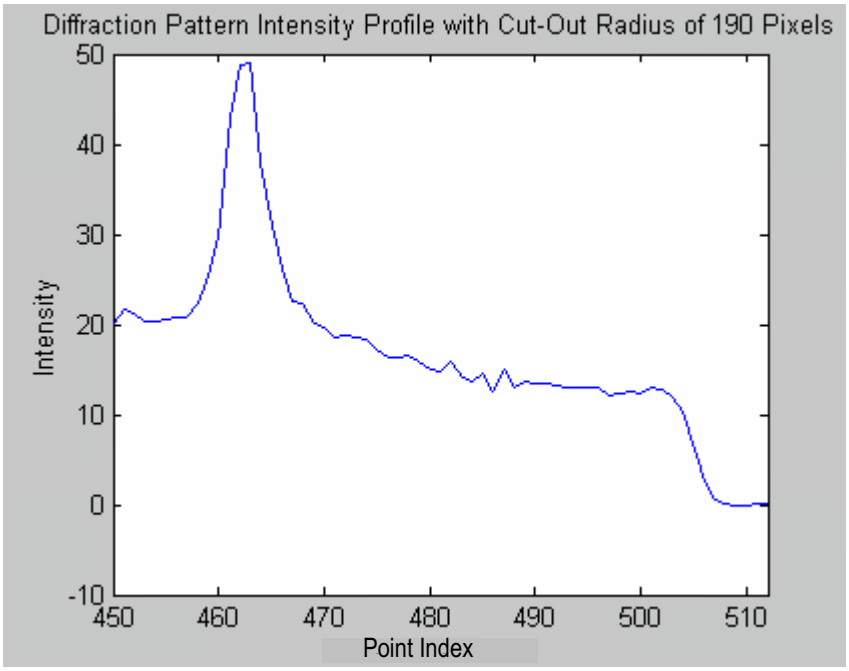


Figure 4.21 Diffraction Pattern Intensity Profile with 190 Pixel Mask Radius.

5 RESULTS

5.1 Results Overview

The purpose of this project was to classify unknown samples into classes of normal, tumor, and fibro adenoma. The main problem was to find a set of features that provided sufficient separation for classification. The analysis consisted of three parts, each described by a section in this chapter. Section 5.1 describes the initial test that was used to narrow down the feature possibilities. Section 5.2 describes the full analysis that was completed for the most promising feature. Section 5.3 gives the results of the classifier when tested using different data sets. Section 5.4 gives the results summary.

5.2 Feature Type Selection

The initial test was used to narrow down the choices of feature type that were described in Section 4.1.2: Wavelet Map Mean Intensity, Wavelet Map Standard Deviation, Wavelet Map Fourier Transform, and Diffraction Image Intensity Profile.

In order to get a top-level comparison of these features types, four constraints were used. First, a single wavelet basis (Bior 3.7) was arbitrarily selected. Since the

literature provided no direction on which wavelet basis would be most suitable for this data set, this initial test used only one arbitrarily selected wavelet basis. The wavelet basis refinement process was left for the full analysis. The second constraint was that no pre-processing was used. The pre-processing of the images also required refinement, which was left for the full analysis. The third constraint was that only the normal vs. tumor samples were tested. It was expected that the Normal vs. Tumor samples would show the largest differences, and would provide the best separation for classification purposes no matter what feature was chosen. Additionally, the most important contrasts for classification are the ones involving tumor samples. The implications are more severe for misclassifying a tumor sample than they would be for a fibro adenoma sample. Therefore, the normal vs. tumor samples were given priority in choosing the best feature type. The fourth constraint was that the combinations of wavelet level and view were not analyzed or optimized. In the training process, specific levels and views of the wavelet features are chosen. However, the goal for this initial experiment was not to determine *specific* wavelet features, but to determine the most promising *type* of features. The specific wavelet features were analyzed and refined in the full analysis.

These four constraints allowed for an initial evaluation of the features. Table 5.1 shows the highest classification rates achieved for each feature type in classifying the Normal vs. Tumor samples.

Table 5.1 Normal vs. Tumor Classification Rates for Each Feature Type. The Classification rates are given for each feature type when classifying the Normal vs. Tumor contrast.

		Classification Rate (%)
Feature Type	Wavelet Map Mean Intensity	98
	Wavelet Map Standard Deviation	81
	Wavelet Map Fourier Transform	93
	Diffraction Image Intensity Profile	88

As seen in Table 5.1, the lowest classification rate achieved was 81% with the Wavelet Map Standard Deviation feature. The wavelet map standard deviation did not provide sufficient information to differentiate the normal samples from the tumor samples. If regions of interest with different textures were being selected from an image, the standard deviation may have been an appropriate measure. However, as expected, the standard deviations of the diffraction patterns for normal samples and tumor samples were very similar.

The Diffraction Image Intensity Profile resulted in a classification rate of 88%. This result reflects the fact that the diffraction images could not be easily sorted by simply looking at the intensity pattern.

The wavelet map fourier transform feature was tested to try utilize frequency information contained in the wavelet decomposition maps. The classification rate achieved for this feature type was 93%, which was a promising result. However, it was not the highest classification rate achieved.

The highest classification rate achieved was 98%, with the Wavelet Map Mean Intensity feature. As a result, the Wavelet Map Mean Intensity feature was selected as the most promising feature for full analysis. The wavelet map average intensity feature provided a simple measure to represent the texture of each wavelet map. The texture of the wavelet map was due to the different values of the decomposition coefficients, which represented the correlation between the wavelet basis at a particular scale and orientation and a particular point on the original image. The texture of the map revealed details at different scales and orientations. These details provided information that was not visible in the original image, but was useful in distinguishing normal samples from tumor samples.

The main advantage of the wavelet map intensity feature is the ease of computation. It required only one step. The disadvantage of calculating the mean of the whole wavelet map was that additional native information contained in the map was not used. For example, the locations of the high intensities are not taken into consideration.

5.3 Full Feature Analysis

Once the Wavelet Map Mean Intensity feature type was chosen as the most promising feature type, the full feature analysis was completed. The purpose of the full analysis was to refine the classification process to improve the classification rates achieved. Several steps were involved in the refinement process. Pre-processing was completed on the data, and different wavelet bases were tested. The full analysis also looked at not

only the normal vs. tumor contrast, but also the contrasts involving the fibro adenoma samples. Finally, the combinations of wavelet level and view were analyzed to determine the best possible feature set. This step could be referred to as feature reduction.

5.3.1 Pre-Processing and Wavelet Basis Selection

The pre-processing step and the wavelet basis selection were tested in combination with each other. The goal was to determine the pre-processing radius that would be used in combination with one of the wavelet bases to produce the best classification results.

For the pre-processing step, masks with radii of 50, 100, 150, 190, and 215 pixels were used to cut out the high center intensities from the original images. The radii were chosen to emphasize different orders of maxima in the diffraction pattern. The pre-processing step is described in detail in section 4.5.2.

Table 5.2 shows the classification rates resulting from the different combinations of mask radius and wavelet basis for the Normal vs. Tumor contrast. The highlighted cells indicate the highest classification rate achieved for each contrast.

Table 5.2 Normal vs Tumor Classification Rates (%) for each Basis-Mask Combination.

		Mask Radii (Pixels)				
		50	**100**	**150**	**190**	**215**
Wavelet Basis	**haar**	76	81	90	93	93
	Db2	95	100	98	95	95
	Db4	100	95	98	98	95
	Db8	93	100	98	100	95
	Bior2.2	98	98	98	98	95
	Bior3.7	98	100	98	100	95
	Bior6.8	93	95	98	98	95

Table 5.2 shows that for the normal vs. tumor contrast, several combinations of mask radius and wavelet basis produced a classification rate of 100%. In order to choose the best combination for classification, the wavelet basis and mask radii combinations for normal vs. fibro adenoma samples and tumor vs. fibro adenoma samples were considered.

Tables 5.3 and 5.4 show the same analysis as Table 5.2, but with the normal vs. fibro adenoma contrast, and the tumor vs. fibro adenoma contrast. The highlighted cells reflect the highest classification rate achieved for the contrast.

Table 5.3 Normal vs Fibro Adenoma Classification Rates (%) for each Basis-Mask Combination.

		Mask Radii (Pixels)				
		50	**100**	**150**	**190**	**215**
Wavelet Basis	**haar**	81	81	85	85	81
	Db2	89	89	89	89	89
	Db4	89	89	89	89	89
	Db8	93	93	96	89	93
	Bior2.2	89	89	89	89	89
	Bior3.7	85	89	89	89	85
	Bior6.8	85	85	85	85	85

Table 5.4 Tumor vs Fibro Adenoma Classification Rates (%) for each Basis-Mask Combination.

		Mask Radii (Pixels)				
		50	**100**	**150**	**190**	**215**
Wavelet Basis	**haar**	86	86	90	90	90
	Db2	86	90	90	90	86
	Db4	97	90	90	90	86
	Db8	90	93	90	93	86
	Bior2.2	90	90	90	90	90
	Bior3.7	90	97	90	97	90
	Bior6.8	90	86	93	86	90

Tables 5.3 and 5.4 show that there were several wavelet basis and mask radius combinations that achieved classification rates greater than 95% for these two contrasts.

In order for a combination of wavelet basis and mask radius to be selected for further analysis, it must achieve a classification of 100% in the normal vs. tumor contrast, and over 95% in one of the other contrasts. In other words, the best wavelet basis and mask radius combinations would produce the highest classification rates for the normal vs. tumor contrast as well as at least one of the fibro adenoma contrasts.

Table 5.5 shows the six wavelet basis and mask radius combinations that achieved 100% classification rates for the normal vs. tumor contrast. It shows the classification rates achieved for the other two contrasts as well. The highlighted rows indicate which wavelet basis and mask radius combinations achieved classification rates of 100% in the normal vs. tumor contrast and over 95% in at least on of the other two contrasts.

Table 5.5 Highest Classification Rates (%). This table summarizes the highest classification rates for each combination of wavelet basis and mask radii for each of the three contrasts. The highlighted combinations were selected for further testing.

			Contrast		
Wavelet Basis and Mask Radius Combination	**Wavelet Basis**	**Mask Radii (pixels)**	**Normal vs. Tumor**	**Normal vs. Fibro Adenoma**	**Tumor vs. Fibro Adenoma**
	Db2	100	100	89	90
	Db4	50	100	89	97
	Db8	100	100	93	93
	Db8	190	100	89	93
	Bior3.7	100	100	89	97
	Bior3.7	190	100	89	97

The three highlighted combinations in Table 5.5 were selected for further testing: Db 4 with a radius of 50 pixels, Bior3.7 with radius of 100, and Bior3.7 with radius of 190 pixels. These three combinations produced the maximum classification rates for both the normal vs. tumor and normal vs. fibro adenoma contrasts.

The next step was to classify the samples when all three classes were presented at the same time. The ideal result would be to find an optimal feature set that would continue to classify the normal and tumor samples at a classification rate of 100% when the fibro adenoma samples were also included in the test data.

Each of the three wavelet-radii combinations that were selected from Table 5.5 was tested with normal, tumor and fibro adenoma samples all included. Table 5.6 summarizes the results of the test.

Table 5.6 Best Rates Achieved for Simultaneous Classification of all Three Classes. The table shows the best rate for each wavelet–radii combination.

Wavelet	Mask Radius (Pixels)	Classification Rate (%)
Db4	50	90
Bior3.7	100	88
Bior3.7	190	86

Table 5.6 shows that the classification rates were significantly lower when all three classes were considered at the same time. However, examination of the misclassified points showed that in all three cases, none of the normal or tumor samples were misclassified.

Table 5.7 shows the details of the misclassified points. For each of the wavelet basis and mask radii combinations, the chart shows the number of times each sample type was classified as normal, tumor, or fibro adenoma. The ideal result is when each sample type is classified 0 times as another sample type. For example, the first line of the table shows the normal samples with the Db4 basis and 50 pixel radius. In this case, all 20 normal samples were classified as normal, with 0 being classified as tumor, and 0 being classified as fibro adenoma.

Table 5.7 Summary of Classification Results. The table shows the number of times the data samples for each of the wavelet-radii combinations were classified into each of the three classes. The table is shows how many samples from each class were misclassified, and what they were classified as.

			Number of Times Classified As:		
Wavelet Basis	Mask Radius (pixels)	Sample	Normal	Tumor	Fibro Adenoma
Db4	**50**	Normal	20	0	0
		Tumor	0	22	0
		Fibro Adenoma	4	1	2
Bior3.7	**100**	Normal	20	0	0
		Tumor	0	22	0
		Fibro Adenoma	4	2	1
Bior3.7	**190**	Normal	20	0	0
		Tumor	0	22	0
		Fibro Adenoma	3	4	0

Table 5.7 shows that the normal and tumor samples were correctly classified with each of the three combinations. Only fibro adenoma samples were misclassified. The Db4 wavelet – 50 pixel radius combination performed the best, with two fibro adenoma samples being classified correctly, four being classified as normal, and only one being classified as tumor. The Bior3.7 wavelet – 190 pixel radius combination performed the worst, misclassifying all of the fibro adenoma samples, and classifying four of them as tumors.

5.3.2 Feature Reduction

The last point to consider before choosing the final feature parameters was the feature reduction step, as discussed in section 4.3. It should be noted that for each of the classification results achieved above, different combinations of the average wavelet map intensity features were used. The wavelet decomposition process results in 32 average intensity features, so the features needed to be narrowed down for classification purposes. Testing all combinations of the features, and selecting the combinations with the fewest features resulting and highest classification rates accomplished this task. Details of the feature reduction process can be found in section 4.3.

Table 5.8 shows the specific wavelet map average intensity features that were used with each of the wavelet basis – mask radius combinations to achieve the classification results shown in Tables 5.6 and 5.7. Digits from 1 through 8 represent the wavelet decomposition level, and letters a, h, v, and d represent the approximation, horizontal, vertical, and diagonal views, respectively.

Table 5.8 Specific Intensity Features Used to Achieve the Highest Classification Rates. The wavelet decomposition level is represented by a digit from 1 through 8, and the wavelet view is represented by the letters a (approximation), h (horizontal), v (vertical) or d (diagonal).

Wavelet Basis	Mask Radius (pixels)	Intensity Features Used
Db4	50	1 2 6 7 v
Bior3.7	100	3 4 6 7 8 v
Bior3.7	190	7 d

According to Table 5.8, the Bior3.7 wavelet – 190 pixel mask radius combination performs the best. It requires only one intensity feature, the level 7 diagonal view, in order to classify all of the normal and tumor samples correctly. This reduces the computation required for feature extraction. As discussed in the methods chapter in section 1.3, the ratio of samples per class to features should be at least 10, allowing for a maximum of two features for this project. The Bior3.7 7D wavelet – 190 pixel mask radius combination meets this goal.

In order to choose the final feature set, it needs to be decided which factor is more important: number of misclassified fibro adenoma samples, or number of features used. If number of misclassified fibro adenoma samples was more important, the Db4 wavelet basis and 50 pixel radius performs the best. If it is more important to use the smallest number of features, the Bior3.7 wavelet basis and 190 pixel radius performs the best.

5.4 New Data Sets

Two experiments were performed to test the performance of the classifier in different situations. The first experiment looked at four samples from one breast, the first sample being located at a tumor, and each of the other three at increasing distances from the tumor. This was a way to test the sensitivity of the classifier to the sampling accuracy. The second experiment applied the pattern recognition software to an unrelated data set consisting of MRI images of rat brains. This showed the ability of the classifier to detect abnormalities in different types of images.

5.4.1 Experiment 1: Increasing Distances from the Tumor

The purpose of this experiment was to classify four samples that were biopsied from the same breast. The samples were taken at increasing distances from a tumor. The first sample was located at the tumor, and the other three were located 20mm, 40mm and 80mm from the tumor.

The classifier was trained using the 20 normal samples and 22 tumor samples. The classifier was tested with each of the three combinations of features and mask radius – wavelet basis combinations discussed in section 5.1.4. The classification results for each of the three combinations are summarized in Table 5.9.

Table 5.9 Classification of Samples at Increasing Distances from the Tumour. Results are shown for three feature parameter combinations.

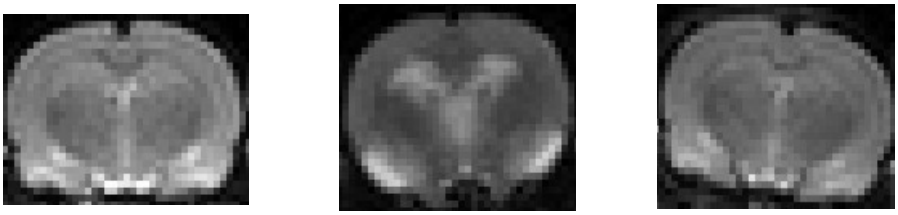
	Wavelet Basis - Mask Radius Combination And Specific Features Used		
Distance from tumor (mm)	**Bior3.7-190 d7**	**Bior3.7-100 v34678**	**Db4-50 v1267**
0	Tumor	Tumor	Tumor
20	Tumor	Tumor	Tumor
40	Normal	Tumor	Tumor
80	Normal	Normal	Normal

Table 5.9 shows that the Bior3.7 wavelet basis, 190 pixel mask radius, and 7D feature, classified the sample as a tumor for samples within 20mm of the tumor. At distances of 40mm and greater, the sample was classified as normal. The other two feature combinations included more intensity features, and detected the tumor at distances up to 40mm from the tumor.

5.4.2 Experiment 2: MRI Data

The classifier was also tested on a set of diffusion weighted MRI images of rat brains. The images showed time-related ischemic injury due to a nerve agent. The data set consisted of 29 control images with no brain damage, seven images taken 12 hours after the nerve agent administration, seven images taken 24 hours after the nerve agent administration, twelve images taken 48 hours after the nerve agent administration, and 21 images taken seven days after the nerve agent administration. The images taken at 12 and 24 hours after the nerve agent injection showed the greatest injury. With increasing time after 24 hours, the images show more resemblance to the normal images.

Figure 5.1 shows examples of the diffusion weighted MRI images of the rat brains.



12 hour **24 hour** **Control**

Figure 5.1 Diffusion Weighted MRI Images. The bright spots on the 12 hour and 24 hour images show areas of acute ischemic injury due to the nerve agent.

Figure 5.2 shows the mean intensities of the control, 12-hour, 24-hour, 48-hour, and 7-day images. All of the intensities are concentrated in the same band, and no separation between the classes is visible.

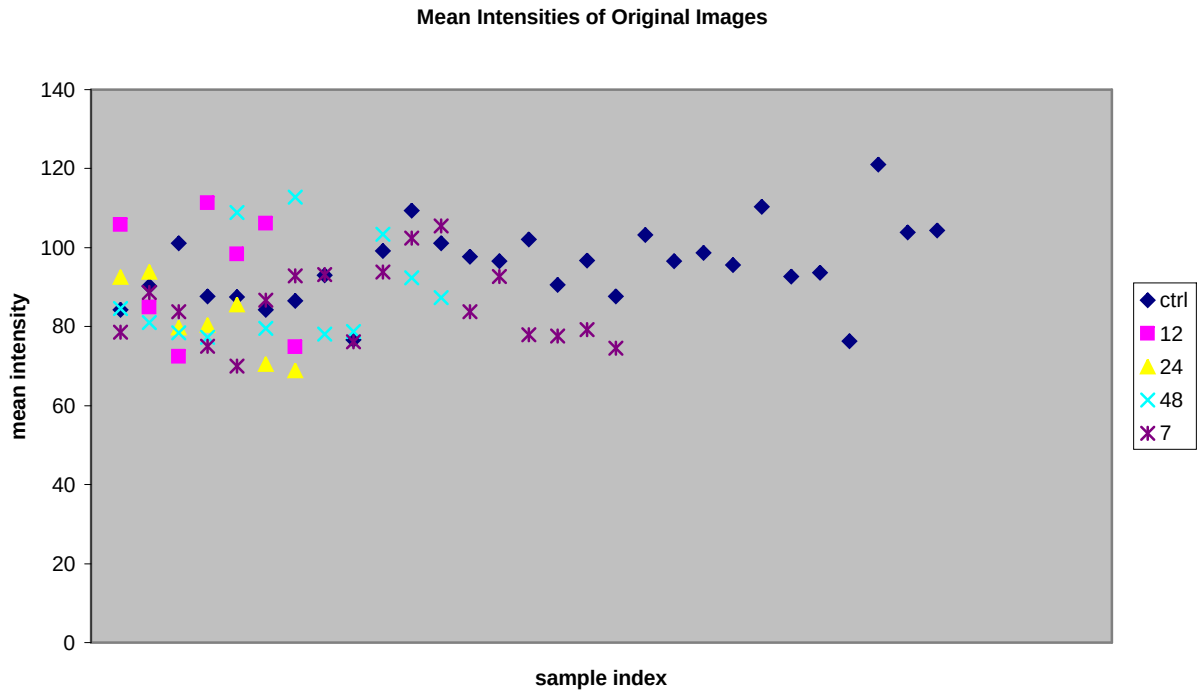


Figure 5.2 Mean Intensities of Rat Brain Images. This figure shows the overlap of the control, 12-hour, 24-hour, 48-hour, and 7-day MRI rat brain image intensities.

In order to classify the images automatically, features were needed that would provide separation between the classes. The wavelet bases selected in section 5.1.2 as the most promising, the Bior3.7 and the Db4, were tested for this experiment. For the MRI data, the high center intensities were not a factor, therefore the pre-processing step was not used.

Table 5.10 shows the best classification results achieved for the MRI rat brain images with each of the wavelet bases. The 12 hour and 24 hour samples were contrasted with

the control samples because they were the ones that were expected to show the largest difference from the control samples. For each of the wavelet bases tested, only one intensity feature was required for 100% classification. H5 means the horizontal view at level 5 of the wavelet decomposition. H8 means the horizontal view at level 8 of the wavelet decomposition.

Table 5.10 Classification Rates (%) for Rat Brain Contrasts. Results are shown for each contrast with each feature.

		Contrast	
		Control vs 12 hour	Control vs 24 hour
Feature	Bior3.7 H5	100	71
	Db4 H8	100	100

As shown in Table 5.10, classification rates of 100% were achieved for both contrasts using the Db4 Horizontal view Level 8 feature. This feature was then used to classify all of the contrasts against each other at the same time.

Figure 5.3 shows the performance of the Db4 H8 feature in classifying all of the contrasts against each other at the same time. There is overlap between the 12 hour and 24 hour classes, as well as the 48 hour, 7 day, and Control classes. However, there is significant separation between the 24 hour and 48 hour classes and the other three classes combined. A classifier based on this feature would be successful in separating either of the 12 hour or 24 hour samples from the other three classes.

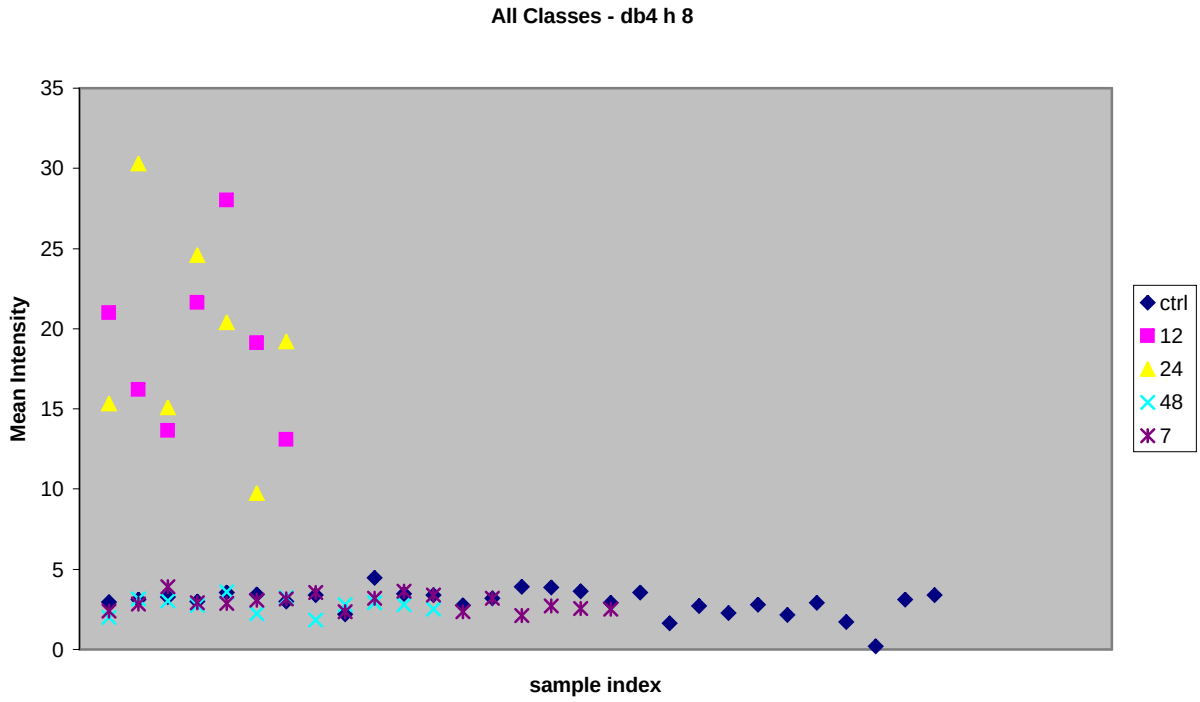


Figure 5.3 Mean Intensities of the Db4 H8 Feature of Rat Brain Images. This figure shows the separation achieved between the 12-hour and 24-hour, and the control, 48-hour, and 7-day MRI rat brain images when the Db4 H8 intensity feature was used.

5.5 Results Summary

Four initial features types were evaluated including wavelet map intensity, wavelet map standard deviation, wavelet map Fourier transform, and diffraction image intensity. It was determined that the most promising feature type was the wavelet map intensity feature.

The wavelet map intensity feature was improved by using pre-processing to remove the high central intensities from the original images, and by using different wavelet bases for the wavelet decomposition. The combination of the Bior3.7 wavelet with a mask radius of 190 pixels and an intensity feature of the diagonal view at level 7 was selected

as the optimal feature parameter set. It used the fewest intensity features and still classified the normal and tumor samples with a classification rate of 100%.

The classifier was tested with samples at increasing distances from the tumor. It was found that with the Bior3.7 wavelet with a mask radius of 190 pixels and the level 7 diagonal view intensity feature, the classifier detected the tumor up to 20mm from the tumor. However, with the Bior3.7 wavelet basis with a 100 pixel mask radius and V34678 feature combination and the Db 4 wavelet basis with a 50 pixel mask radius and V1267 feature combination, the tumor was detected up to 40mm from the tumor. This suggests that the additional features could increase the sensitivity of the classifier to tumor data.

The final test evaluated the classifier on a set of MRI images of rat brains showing time-related injury due to a nerve agent. The Db4 H8 feature provided sufficient separation to distinguish either the 12 hour or 24 hour images from the 48 hour, 7 day, and Control images.

6 DISCUSSION

6.1 Implications of Results

The classification results based on the normal, fibro adenoma, and tumor samples were very encouraging. Classification rates of 100% for normal vs. tumor samples, 96% for normal vs. fibro adenoma samples, and 97% for tumor vs. fibro adenoma samples were achieved. These results compared with the highest sensitivities found in the literature [7] [53] [54].

When fibro adenoma samples were included in the classification process, the classification rates decreased. The normal and tumor samples were still correctly classified 100% of the time. However, because of the misclassified fibro adenoma samples the overall classification rate was reduced to 90%. As only seven fibro adenoma samples were available, further testing with more samples would be required to determine more accurate classification rates when fibro adenoma samples are included in the data.

Common weaknesses of the breast cancer detection systems reviewed in the literature were that the feature extraction processes required many complicated steps, or were computationally expensive. The system designed for this project does not pose those problems. The wavelet map average intensity feature was straightforward to calculate, and required only two steps: wavelet decomposition, and average intensity calculation.

Another strength of the system designed for this project was that the feature extraction process did not require a priori information, as many systems do. The only a priori information required was for the naïve Bayesian classifier. The a priori probabilities for the different classes needed to be specified. However, since the wavelet features were so strong, the choice of classifiers may be flexible. An interesting area for further research would be to test a classifier based on a clustering method that would require no a-priori parameter estimation.

6.2 Implications of the Distance Test

The automated classification system was tested on data sampled at increasing distances from the tumor. There are differences in the diffraction patterns of the tissue samples at increasing distances from a tumor [31]. The results of the classifier implemented in this project agreed with those findings. The tissue was classified as tumor at distances from 20mm to 40mm away from the tumor depending on the feature set used. At distances of 80mm from the tumor all feature combinations classified the tissue as normal.

6.3 Implications of the MRI Results

The classifier was also tested on a completely unrelated data set consisting of MRI images of rat brains that contained time-related pathology. The classifier was able to sort the MRI images into two distinct groups. One group contained the control images plus the images taken at time points of forty-eight hours and seven days. The other group contained the images at time points of 12 hours and 24 hours. These results were encouraging because the damage that was observed by the changes in the apparent diffusion coefficient of the MRIs was most pronounced for the 12 to 24 hour time points. The apparent diffusion coefficients gradually returned to levels close to normal by 7 days. These results agree with those of the classifier.

Although the classifier was designed for classifying diffraction patterns of breast tissue, it worked surprisingly well on a completely unrelated data set of MRI images of rat brain damage. This means that the method may be used to solve other classification problems. Another application currently being investigated is the use of this classification system on digitized mammograms. Preliminary results indicate classification rates of 94% for distinguishing between cancerous calcifications and normal tissue, and 90% for distinguishing between cancerous masses and normal tissue [11].

7 CONCLUSION

Breast cancer is the most commonly diagnosed type of cancer in Canadian women. Although the death rate is the second highest among Canadian women with cancer, early detection of the disease greatly improves the chance of survival. Therefore, it is important to develop new and improved methods for breast cancer screening.

The overarching goal of this project was to improve breast cancer screening protocols by developing a semi-automatic pattern recognition tool to assist radiologists in the classification of breast lesions.

Another of the objectives of the project was to use specialized data that was acquired using synchrotron technology. The synchrotron allowed SAXS images to be collected from breast biopsy tissue.

The main task of the project was to find features in the data that would distinguish normal samples from those containing tumors. That is where wavelet analysis came in.

The third objective for this project was to use wavelets to analyze the data. Wavelet based features were generated from the SAXS image data. The features were supplied to the pattern recognition tool, and were successful in sorting the images into distinct groups of "normal" and "tumor".

Meeting the three objectives for this project was a semi-automatic breast cancer screening tool that showed very promising results in the investigation undertaken for this project. A classification rate of 100% was achieved for distinguishing between normal samples and tumor samples. Classification rates of 96% were achieved for distinguishing between fibro adenoma samples and normal samples as well as fibro adenoma samples and tumor samples. The system also showed promising results when tested on unrelated MRI data, and preliminary research showed that the system may be effective for classifying digital mammograms.

REFERENCES

[1] Alberts, Bruce et al., *Molecular Biology of the Cell. 3rd ed.* New York and London: Garland Publishing, 1994.

[2] Burhenne, Linda J. Warren et al. "Potential Contribution of Computer-aided Detection to the Sensitivity of Screening Mammography." *Radiology.* May 2000: 554-562.

[3] Burrus, C. Sidney, Ramesh A. Gopinath, and Haitao Guo. *Introduction to Wavelets and Wavelet Transforms, A Primer.* Upper Saddle River: Prentice Hall, 1998.

[4] Canadian Cancer Society: **The Canadian Cancer Society's perspective on early detection and screening of breast cancer.** July 28, 2003. http://www.cancer.ca/ccs/internet/standard/0,3182,3172_83256703_275174_langId-en,00.html

[5] Canadian Light Source Inc: **What is a synchrotron?** March 5, 2004. http://www.cls.usask.ca/education/whatis.php

[6] Chan, Heang-Ping et al. "Computerised Classification of Malignant and Benign Microcalcifications on Mammograms: Texture Analysis Using and Artificial Neural Network." *Physics in Medicine and Biology.* Mar. 1997: 549 - 567.

[7] Chan, Heang-Ping et al. "Improvement of Radiologists' Characterization of Mammographic Masses by Using Computer-aided Diagnosis: An ROC Study." *Radiology.* Sept. 1999: 817-827.

[8] Cheng, Heng-Da, Jing li Wang, and Xiangjun Shi. "Microcalcification detection using fuzzy logic and scale space approaches." *Pattern Recognition.* Mar. 2004: 363 – 375.

[9] Cheng, Heng-Da et al. "Computer-aided detection and classification of microcalcifications in mammograms: a survey." *Pattern Recognition.* Dec. 2003: 2967 – 2991.

[10] Christiansen, Cindy L. "Predicting the cumulative risk of false-positive mammograms." *Journal of the National Cancer Institute.* 18 Oct. 2000: 1657-1666.

[11] Chytyk, Krista, Edward Kendall, Carissa Erickson. Semi-automatic classification of breast images. Poster.

[12] Cipra, Barry. "Wavelet Applications Come to the Fore," *SIAM (Society for Industrail and Applied Mathematics) News.* November 1993. <http://www.siam.org/siamnews/mtc/mtc1193.htm>

[13] Daubechies, Ingrid. *Ten Lectures on Wavelets. 2nd ed.* Philadelphia: SIAM, 1992.

[14] Duda, Richard O., Peter E. Hart and David G. Stork. *Pattern Classification 2nd Edition*. New York: Wiley-Interscience, 2001.

[15] Elmore, Joanne G. et al. "Ten-year risk of false positive screening mammograms and clinical breast examinations." *New England Journal of Medicine*. 16 Apr. 1998: 1089-96.

[16] Freer, Timothy W. and Michale J. Ulissey. "Screening Mammography with Computer-Aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center." *Radiology*. Sept. 2001: 781-786.

[17] Friedman, Nir, Dan Geiger and Moises Goldszmidt. "Bayesian Network Classifiers." *Machine Learning.* Kluwer Academic Publishers, 1997.

[18] Garg, Ashutosh, and Dan Roth. Understanding Probabilistic Classifiers. *ECML*, 2001.

[19] Hall, Chris. **Collagen.** July 29, 2003. http://detserv1.dl.ac.uk/Herald/xray_diff_collagen.htm

[20] Hartswood Mark, et al., "Prompting in practice: How can we ensure radiologists make best use of computer-aided detection systems in screening mammography?" *Proceedings of the Fourth International Workshop on Digital Mammography*. Nijmegen, June 1998.

[21] Heckerman, David. "A Tutorial on Learning With Bayesian Networks." *Microsoft Research Technical Report*, 1996.

[22] Hubbard, Barbara Burke. "The World According to Wavelets. The Story of a Mathematical Technique in the Making. 2nd Ed." AK Peters, Ltd. Natick, MA. 1998.

[23] IARC Working Group: **Mammography Screening Can Reduce Deaths From Breast Cancer.** News Release. March 19, 2002.

[24] Jain, Anil K., and B. Chandrasekaran, "Dimensionality and Sample Size Considerations in Pattern Recognition Practice." Handbook of Statistics, P.R. Krishnaiah and L.N. Kanal, eds. Vol.2, p. 835-855, Amsterdam: North-Holland, 1982.

[25] Jain, Anil K., Robert P.W. Duin and Jianchang Mao, "Statistical Pattern Recognition: A Review." *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Jan. 2000: 4 - 37.

[26] James, Veronica J, J. Fraser McConnell and Malcom Capel. "The d-spacing of collagen from mitral heart valves changes with ageing, but not with collagen type III content." *Biochimica et Biophysica Acta.* May 1991:19-22.

[27] Jiang, Yulei et al. "Potential of Computer-aided Diagnosis to Reduce Variability in Radiologists' Interpretations of Mammograms Depicting Microcalcifications." *Radiology*. Sept. 2001: 787-794.

[28] Kreyszig, Erwin. *Advanced Engineering Mathematics, Seventh Edition,* NY: John Wiley & Sons Inc., 1993: 613.

[29] Laine, Andrew et al. "Mammographic Feature Enhancement by Multiscale Analysis." *IEEE Transactions on Medical Imaging.* Dec. 1994: 725 – 740.

[30] Lemaur G., K. Drouiche, and J. DeConinck. "Highly Regular Wavelets for the Detection of Clustered Microcalcifications in Mammograms." *IEEE Transactions on Medical Imaging*. Mar. 2003: 393 - 401.

[31] Lewis, Robert A. et al. "Breast Cancer Diagnosis Using Scattered X-Rays?" *Physics of Medical Imaging, Proceedings of SPIE*. 2001.

[32] Li, Huai, K.J. Ray Liu, and Shih-Chung B. Lo. "Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms." *IEEE Transactions on Medical Imaging*. Dec. 1997: 785 – 798.

[33] Love, Susan. *Dr. Susan Love's Breast Book, 3rd ed.*, MA: Perseus Publishing, 2000: 125.

[34] Mendez, Arturo J. et al. "Computer-aided Diagnosis: Automatic Detection of Malignant Masses in Digitised Mammograms." *Medical Physics.* June 1998: 957-964.

[35] Meisel, William S. *Computer-Oriented Approaches to Pattern Recognition*. New York: Academic Press, 1972.

[36] Misiti, Michel et al. "Advanced Concepts." *MATLAB Wavelet Toolbox.* The MathWorks, 1999.

[37] Nagel, Rufus H. et al. "Analysis of methods for reducing false positives in automated detection of clustered microcalcifications in mammograms." *Medical Physics*. Aug. 1998: 1502 - 1506.

[38] National Cancer Institute of Canada: **Canadian Cancer Statistics 2004**, Toronto, Canada, 2004.

[39] Poplack, Steven P. et al. "Mammography in 53,803 Women from the New Hampshire Mammography Network," *Radiology*. Dec. 2000: 832-840.

[40] Pucci-Minafra, Ida et al. "Absence of regular alpha2(I) collagen chains in colon carcinoma biopsy fragments." *Carcinogenesis*. Apr. 1998: 575-84.

[41] *RadiologyInfo: What is Mammography?* Radiological Society of North America, Inc. (RSNA). July 29, 2003. http://www.radiologyinfo.org/content/mammogram.htm

[42] Raymond Wendy A. and Anthony S.-Y. Leong. "Assessment of invasion of in breast lesions using antibodies to basement membrane components and myoepithelial cells." *Pathology*. Oct. 1991: 291-297.

[43] Shile, Peter E. and Jessica A. Guingrich. "Detecting "missed" breast cancers: A comparison of CAD systems", *Supplement to Applied Radiology*. Oct. 2002.

[44] Sheng, Yunlong. "Wavelet Transform." *The transforms and applications handbook*. Boca Raton, Fl: CRC Press, 1996.

[45] Singh, Marsha. "A Brief Overview of SAXS." *Small Angle X-Ray Scattering* 2002. Queen's University SAXS Laboratory. 28 March 2003. <http://www.physics.queensu.ca/wwwhome/singh/SAXSoverview.html>

[46] Stanford Linear Accelerator Center Virtual Visitor Center. "Accelerator Form and Function." 12 March 2004. http://www2.slac.stanford.edu/vvc/accelerator.html

[47] Stanford Synchrotron Radiation Laboratory. "Biological Small Angle Scattering/Diffraction (B-SAXS/D) on BL 4-2 at SSRL." 12 March 2004. http://www-ssrl.slac.stanford.edu/~saxs

[48] Strang, Gilbert and Truong Nguyen. *Wavelets and Filter Banks*. Wellesley: Wellesley-Cambridge Press, 1996.

[49] Valens, Clemens. *A Really Friendly Guide to Wavelets*. May 20,2003. http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html

[50] Wang, Ted C. and Nicolaos B. Karayiannis. "Detection of Microcalcifications in Digital Mammograms Using Wavelets." *IEEE Transactions on Medical Imaging*. Aug. 1998: 498 – 509.

[51] *Wavelet Explorer Documentation*. Wolfram Research, Inc. August 7, 2004. http://documents.wolfram.com/applications/wavelet

[52] Yankaskas, Bonnie C. et al. "Association of recall rates with sensitivity and positive predictive values of screening mammography," *American Journal of Roentgen Ray Society*. Sep. 2001: 543 - 549.

[53] Yu, Songyang and Ling Guan. "A CAD System for the Automatic Detection of ClusteredMicrocalcifications in Digitized Mammogram Films'', *IEEE Transactions on MedicalImaging*. Feb. 2000: 115-126.

[54] Zhen, Lei and Andrew K. Chan. "An Artificial Intelligent Algorithm for Tumor Detection in Screening Mammogram." *IEEE Transactions on Medical Imaging*. July 2001: 559 - 567.